

WILEY

The Metaphysics of Emergence

Author(s): Timothy O'Connor and Hong Yu Wong

Source: *Noûs*, Vol. 39, No. 4 (Dec., 2005), pp. 658-678

Published by: Wiley

Stable URL: <http://www.jstor.org/stable/3506115>

Accessed: 05-12-2017 03:22 UTC

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <http://about.jstor.org/terms>



Wiley is collaborating with JSTOR to digitize, preserve and extend access to *Noûs*

The Metaphysics of Emergence

TIMOTHY O'CONNOR
Indiana University

HONG YU WONG
University College London and Budapest University of
Technology and Economics

I. Mental Causation: The Current State of Play

The following framework of theses, roughly hewn, shapes contemporary discussion of the problem of mental causation:

- (1) **Non-Identity of the Mental and the Physical**
Mental properties and states cannot be identified with specific physical properties and states.
- (2) **Causal Closure (Completeness) of the Physical**
The objective probability of every physical event is fixed by prior physical events and laws alone. (This thesis is sometimes expressed in terms of explanation: In tracing the causal history of any physical event, one need not advert to any non-physical events or laws. To the extent that there is any explanation available for a physical event, there is a complete explanation available couched entirely in physical vocabulary. We prefer the probability formulation, as it should be acceptable to any physicalist, though some reject the explanation formulation.)
- (3) **Causal Exclusion**
There is at most one complete and wholly independent explanation for any given event or sequence of events.
- (4) **Causal Relevance of the Mental**

Mental events are partly determinative of, and so have causal-explanatory relevance to, some physical events.

(5) Homogeneity of Mental and Physical Causation.

The mode of causation between mental events and physical or mental events is, in general, metaphysical terms, the same as that holding among purely physical events.

These theses are jointly inconsistent, and contemporary views may be identified by which thesis they reject. Traditional Mind-Body Dualism rejects (2), accepting that physical events and laws alone do not determine the occurrence (or in probabilistic scenarios, the objective chance) of every physical event. Instead, some of these events are partly determined by processes that include events within primitively mental substances. Reductive Physicalism rejects (1), supposing there to be type-identities between mental and physical event kinds (though possibly only highly specific kinds of both sorts). And Nonreductive Physicalism comes in two varieties; one rejects (3) and accepts systematic causal overdetermination of some physical events, and the other rejects (5) and claims ambiguity in our causal ascriptions.

We believe that none of these three options are satisfactory. After briefly indicating why, we will sketch an alternative metaphysic that appears to offer a more promising basis for understanding the mental-physical relationship.¹ Like traditional dualism, it questions thesis (2), though its rejection of (2) is more qualified. We call the doctrine we espouse “the ontological emergence of the mental.” Our favored concept of emergence bears some important similarities to that figuring in accounts of the early twentieth-century British emergentists. Yet it clarifies and likely improves the emergentist model of mental-physical dynamics, and as a result one can see that recent conceptual critiques of emergence on this score are easily turned aside.

II. Problems for Three Central Views

We now set the stage for the presentation of our own proposal with brief and largely familiar criticisms of the dominant alternatives. Our aim is not to persuade reflective adherents but to make clear the sorts of considerations that motivate our sort of emergentist. This is not gratuitous, as the term “emergence” is nowadays used to cover a multitude of sympathies. Besides, reminding ourselves of the intuitive costs of popular views is a hedge against creeping dogmatism.

We begin with two ontologies which clearly allow for mental causation: mind-body dualism and reductive (type-identity) physicalism.

II.1. Mind-Body Dualism

Dualism supposes that minds are basic substances wholly distinct from any physical substances and have primitive causal capacities to influence

physical events within associated brains. This view is casually dismissed by very many theorists of mind, though the reasons cited are often weak or poorly developed. We suggest that the true difficulties with the position are these:

Causal pairing problem

It is exceedingly odd that *particular* minds and brains form a lifelong ‘monogamy’ despite the absence of any apparent relational framework. For it is only within the terms of such a framework that we could explain the persistent individual pairings as a consequence of a contingent, external relationship between them, which relations structure mental-physical causality in a general fashion.² (Compare the role of spatial relations in determining which physical systems interact at a given time. In physical causation, interactions are a general function of intrinsic characteristics and external relations, rather than involving *haecceitistic* affinities between particulars.) This difficulty might be overcome by positing the emergence of the mental substance, so that the asymmetrical dependency of mind on brain grounds their monogamous interaction.

Developmental problem

But even an emergentist version of substance dualism requires what is empirically implausible, viz., that a composite physical system gives rise, all in one go, to a whole, self-contained, organized system of properties bound up with a distinct individual. For we cannot say, as we should want to do, that as the underlying physical structure *develops*, the emergent self does likewise. This would require us to posit changing mereological complexity within the self, which would give rise all over again to problems of endurance that substance dualism is supposed to avoid, and which would run counter to intuitions of primitive unity that substance dualists have regarding persons. No, the emergent dualist view will have to say, instead, that at an early stage of physical development, a self emerges having all the capacities of an adult human self, but most of which lie dormant owing to immaturity in the physical system from which it emerges.³

II.2. Reductive Physicalism

Reductionism nowadays is much disparaged. Yet by our lights, the most plausible variety of physicalism is reductionist, as it does not require one to make dubious moves in the underlying metaphysics of physical properties.

Most plausible variety of physicalism, but still implausible. The claim that ‘token’ mental states—i.e., particular, concrete mental occurrences—just are complex electrochemical events (which just are complex microphysical events) implausibly denies that there is anything distinctive about mental activity in the world. The suggestion that the mental may still be *conceptually* irreducible to the physical is but a sop: from an ontological

point of view, it amounts to nothing more than noting that there can be complex physical structure types which conform to high-level patterns of instantiation. The ontological basis of these high-level patterns are *the nature of fundamental physical capacities and certain contingent, physical boundary conditions obtaining in our neck of the cosmic woods*. (Such conditions ensure the relative stability of complex physical structures, including biological ones.) The flipside of the reductionist view that mental *activity* is not fundamentally distinctive is that the qualitative and intentional character of mental states from which such activity flows is likewise unremarkable. Here, too, the general reductionist line flies in the face of deep intuitive judgments derived from introspection.

II.3. Nonreductive Physicalism (NRP)

Given problems such as these with the more extreme materialist and dualist views, some sort of middle ground position is bound to attract. We ourselves shall propose a *via media* in what follows. But the banner of 'sensible middle ground' has been carried by an apparently different view, 'non-reductive physicalism' (NRP). Central to this view is a distinction between tokens and types of mental states. On its most usual variety, NRP holds that *token* mental states (=the particular instances of a mental property) are identical in each case to a token physical state, but the mental and physical state *types* which are thereby instanced are non-identical. (This is often motivated through arguments for the multiple realizability of functional properties generally, and so mental properties in particular.)⁴ The mental state, they say, is in each case *realized* by, but distinct from, some physical state. Allegedly, the view allows one to consistently adhere to the causal closure of physical transactions while preserving the autonomy of mental explanation: mental states are type-distinct from physical states and explain psychological patterns of instantiation unexplainable in terms of basic physics. In our judgment, maintaining consistency here depends on one's not thinking very deeply about properties and individuals.

As the individual theorist looks to make the ontology rigorous and explicit, NRP will transmute into (less charitably, but more accurately, it will *take shape as*) either a form of reductionism or a mysterious, semi-emergentist form of property dualism.

- (1) If one emphasizes the abstract, non-immanent character of mental property types, as against mental property tokens, the irreducibility appears to be purely conceptual. Indeed, the view would be a mere linguistic variant on a view that declares itself to be reductionist about physical ontology (immanent and token-level, where the real causal action is) but not about mental concepts. And one would be hard-pressed to say how *this* sort of 'non-reductionism' about mental properties is anything more robust—from an ontological

point of view—than a property platonist’s non-reductionism about any number of applicable but explanatorily useless concepts we might specify from our armchairs.

- (2) On the other hand, one can follow Jerry Fodor in placing heavy emphasis on the role of mental concepts in useful and lawlike *ceteris paribus* generalizations, and insist that explanatory success with such concepts is a mark of genuine properties immanent to the nature of things in the world. We’d thereby be led to eschew the identity of physical and mental *tokens*. Such a scenario will indeed yield a robust property dualism, but at a heavy cost. Physical properties will cause and explain other physical properties. Mental properties will cause and explain other mental properties. Finally, mental properties will also cause physical properties, apparently as overdetermining causes. We might soften this by the usual appeal to an asymmetrical supervenience of the mental on the physical, but supervenience here will be inexplicable: it’s not grounded in *causal* relations, as with emergentism (see below), and the token identity of mental and physical events has now been abandoned. (If one continues to speak of mental-physical ‘realization’ in the absence of token-identity, this notion’s content will be obscure, rendering it useless for explanatory purposes.) So we get a weird ontology indeed. There are ubiquitous, causally-interacting fundamental physical properties, some organized clusters of which are associated in regular but unexplained fashion with high-level properties which then interact amongst themselves.

III. Ontological Emergence and its Metaphysical Underpinnings

III.1. *Emergence as a Species of Nonstructurality*

One moral from our previous remarks is that if a notion of emergence is to improve upon the unhappy alternatives rejected above, it had better be robustly ontological. This is contrary to the tendency among recent philosophical commentators to conflate or blur ontological and epistemological issues when applying emergentist ideas to nonlinear phenomena, artificial life, and human mentality. Discussions of senses of ‘unpredictability’ dominate these other expositions. While seeing how an emergent property would be unpredictable from a certain, limited empirical standpoint *is* a useful way of getting a fix on the concept, this is but a consequence of its core metaphysical features. There are two such features, both critical to the usefulness of an emergentist approach to understanding the mind-body relationship.

Older discussions of emergence sometimes spoke of the ‘novelty’ of such properties in relation to more fundamental physical properties. This term is not felicitous, however. Novelty cannot simply mean “not having been

instanced previously,” as this has been true of ever so many non-emergent features at various junctures in the world’s history (e.g., the first occasion on which a composite had determinate mass M , for some arbitrary, large value of M). Nor can it mean “not had by any of the object’s proper parts,” as this is true, e.g., of the mass of *any* composite. What such theorists had in mind, we think, was something like irreducibility in an ontological, rather than semantic, sense, within the context of a property theory that can do some principled differentiation. Platonism, with its abundant properties, is too indiscriminate. We shall want, instead, a sparse theory, on which the properties a thing has are far fewer than the concepts it falls under, such that properties make a difference to how the objects act in at least some circumstances. (Properties in this sense answer to what is alternately termed the “Eleatic Principle,” or “Alexander’s dictum,” on which all existents possess causal power.) Either of the theories of immanent universals or tropes will serve our purposes here. On these theories, there are a quite limited number of types of basic properties, instanced in most cases by fundamental particulars, whatever these might be (quarks and electrons, or what have you). Each type is non-redundant with respect to the others in determining the behavior of such particulars. Now consider an unremarkable composite object, C , comprising the atomic individuals a and b .⁵ It is plausible on empirical grounds that many, if not all, of C ’s properties are complex, being ‘built up’ from the properties of its parts. (David Armstrong 1978) attempts to define this idea through the notion of a ‘structural universal’. A modified version of it is this:

A property, S , is structural if and only if proper parts of particulars having S have properties not identical with S and jointly stand in relation R , and this state of affairs *is* the particular’s having S .

That is to say, there is *nothing more* to having the structural property than being composed by parts having certain other properties and bearing certain relations to one another—it is ontologically reducible.⁶ Consider Armstrong’s favorite example of *being a methane molecule*, or CH_4 . Let us pretend for example’s sake that carbon and hydrogen atoms are mereological atoms and the *properties* of *being a carbon atom* and *being a hydrogen atom* are likewise basic. It will then be plausible to suppose that the property of being a methane molecule just consists in having as parts a carbon atom and four hydrogen atoms bound in the relationship characteristic of methane.

The notion of an emergent property can then be understood in part by way of contrast with structural properties. An emergent property is a property of a composite system that is *wholly nonstructural*. If we allow simple conjunctive properties, the conjunction $F\&G$ of emergent properties F and G will count as nonstructural in this sense, though we should not

deem it basic. The basic properties and relations of our world will be those properties whose instantiation does not even partly consist in the instantiation of distinct properties by the entity or its parts. *It is the thesis of emergentism that some basic properties are had by composite individuals.*

III.2. The Generation of Emergent States

Emergent properties are nonstructural properties of composite individuals. We further presume that they arise from and are sustained by underlying microstructures. How shall we conceive this dependency relationship, and what nonredundant difference can they make to the future distribution of such microstructural properties? We answer these questions from within a traditional causal realist framework, on which properties confer propensities to act. One could adapt some of what we say within a Lewis-style Humean picture, and rather more within the currently fashionable Armstrong-Tooley account of causality as a higher-order relation among universals. (As we see things, the Humean ontology championed by Lewis is deeply counterintuitive and has radically skeptical consequences, and the Armstrong-Tooley picture is simply second-order Humeanism, adding a bit more structure to the inert Humean picture of the world at no explanatory gain whatsoever. But our conviction on these scores is independent of our attraction to and basic understanding of emergence.)

Of central importance is to recognize that the relationship of micro-level structures and macro-level emergent properties is dynamic and causal, not static and formal (in a quasi-logical sense). Contemporary discussions of emergence by (Kim 1999), (McLaughlin 1997, though apparently not 1992), and (Shoemaker 2002) all tend, to varying degrees, to assimilate the concept of emergence to the nonreductive physicalist's picture. Insofar as this leads them to assume that the emergent property synchronically supervenes on the microphysical property which is its 'base', the assimilation generates confusion. Emergent properties are basic properties, *token*-distinct in character and propensity from any microphysically structured properties of their bearers. If their appearance in certain systems is to be explained at all, they must be explained in terms of a *causal*, not purely formal, relationship to underlying, immediately *preceding* structures. And the whole question of whether there is any sense in which they supervene on lower-level features—which we discuss in §IV—is subtle, and should not be built in definitionally from the outset.⁷

Here is how we think of the matter.⁸ An emergent property of type E will appear only in physical systems achieving some specific threshold of organized complexity. From an empirical point of view, this threshold will be arbitrary, one that would not be anticipated by a theorist whose understanding of the world was derived from theories developed entirely from observations of physical systems below the requisite complexity. In optimal circumstances, such a theorist would come to recognize the locally

determinative interactive dispositions of basic physical entities. Hidden from his view, however, would be the tendency (had by each of the basic entities) to generate an emergent state. This tendency is not discernible in contexts lacking the requisite macro-complexity, as it is a tendency towards a joint effect of an organized system of the right kind.⁹ We further suppose that the continuing instantiation of the emergent property depends on the continuing presence of the structural universal that generated it.¹⁰ Clearly, the way in which microdispositions must combine to generate an emergent feature—even the form of the summation function capturing the relevant notion of organized complexity—is a matter for empirical theory, not a priori analysis. But the notion at play here seems in the neighborhood of the familiar concepts of scalar and vector addition in physics.

IV. A Dynamical Emergence Model

Our picture of emergence becomes more complicated once we consider not just the generation of an initial emergent state, but the dynamics of an object's having one or more emergent features for a period of time. Think of what we've just described as a baseline case, involving the onset of an emergent state. Then consider that, as a fundamentally new kind of feature, it will confer causal capacities on the object that go beyond the summation of capacities directly conferred by the object's microstructure. Its effects might include directly determining aspects of the microphysical structure of the object as well as generating other emergent states. In setting forth a general account of how this might go, we are guided not by abstract intuition about how it must go in any possible emergent scenario, but about how it is plausible to suppose it goes with respect to our own mental life, on the supposition that qualitative and intentional features of our mental states are emergent. (The diagram below, then, is not intended to capture a minimally sufficient schema for emergence, but a variant that plausibly applies to the mental, if any emergentist scenario does.)

It is plausible that there are enduring baseline mental states that partially underwrite more specific and often momentary mental states. (Underlying one's visual awareness of a computer screen, e.g., is a more general state of conscious awareness that persists when one looks in another direction. We might plausibly conjecture that underlying our entire mental lives are certain highly general states, themselves mental in character, disposing us towards having specific sorts of mental experiences and cognitive states in suitable circumstances.) Suppose, then, that when a neurophysiological system *S* comes to have a certain kind of complex physical configuration *P** at time *t*₀, the baseline emergent state *E* is the direct result at *t*₁. (*P**, of course, will have to be of a sufficiently general type as to persist through constant and over time dramatic change. Therefore, we cannot equate it with the total physical state of the system at any given time.) *P** will also

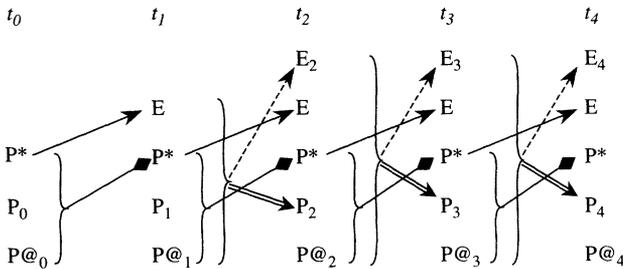
partly determine the underlying physical state of S at time t_1 . Let P_0 be the remaining aspect of S's intrinsic physical state at t_0 , and $P@_0$ be the summation of those physical factors in S's immediate environment that will bear upon the physical state of S at t_1 . Letting " \Rightarrow " represent (minimally sufficient) causation, we have

$$P^* \text{ at } t_0 \Rightarrow E \text{ at } t_1$$

and

$$P^* + P_0 + P@_0 \text{ at } t_0 \Rightarrow P^* + P_1 \text{ at } t_1$$

(The conjunction $P^* + P_1$ is the total intrinsic physical state of S at time t_1 .) Now E at t_1 will help to determine the physical state of S at the subsequent moment, t_2 , but presumably not its continuing to exhibit P^* .¹¹ E, we may suppose, will also help to determine the occurrence at t_2 of another emergent state, E_2 . Diagrammatically, the overall picture is this:



Dynamical evolution of system S over time

- upward causation of baseline emergent state E
- upward causation of super-emergent states E_n
- ◆— maintenance causation of emergence-sustaining configuration P^*
- ⇒ wide horizontal causation (including downward causation)

It would be absurd to suppose that this simple diagram could adequately represent any actual mental episode, with its complex array of mental elements. Nonetheless, a toy example may help one to see the thrust of the view we are promoting. Suppose, then, that E is a very general state of being disposed to visual awareness. At t_0 , P^* obtains, giving rise to the disposition to visual awareness at t_1 . This dispositional state, in conjunction with the total physical state of S at t_1 , and in the presence of certain physical stimuli, cause the visual awareness of a red apple in front of S, E_2 , and an aspect (P_2) of the subsequent physical state. (The physical state of S and its environment at t_1 suffice, apart from E, to maintain the critical structural

feature, P*.) As, let us suppose, the apple is moved and this information is encoded in S's physical state, E₂ accordingly gives way to E₃ as well as influencing the perceptual information encoded in P₃.

For many contemporary theorists, the all-important questions to ask of any seemingly adventurous view of the dynamics of mental properties are whether and in what sense the mental supervenes on the physical. We place less stock than they do in the significance of these questions.¹² But let us approach the matter by first asking whether the emergent properties of S as characterized above *strongly supervene* on its physical properties. Strong supervenience is a relation between families of properties, and is usually thought of as synchronic. The family of emergent properties would strongly supervene on the family of physical properties just in case having an emergent property E at time t implies, of causal necessity, (1) that an object has some physical property P at time t and (2) necessarily, if it has P at t, it has E at t.

The first condition—having some physical properties—is evidently satisfied. The slogan used to capture the second condition is: 'No mental difference without a physical difference'. Consider first the status of our baseline emergent feature E, with reference to times t₀ and t₁ in the diagram. E is absent at t₀ and present immediately thereafter. The underlying physical properties are different, too, but that is not the reason for the difference in emergent properties. For the differentiating factors (P₀, P₁, and the variable P@₀) are, by hypothesis, not directly relevant to the occurrence of E. P* alone is so relevant. Yet E is absent at the first time, since P*'s obtaining at t₀ causally determines not what will occur at that very time, but immediately thereafter. So at the first instant of P*'s instantiation in S, S will not bear E. This indicates that there might be two objects having identical intrinsic physical properties (including P*) and existing in the same external circumstance, yet one has E and the other lacks it.

But this is only a slight departure, restricted to the first instant at which the 'base' property P* occurs. More interesting divergence between emergent properties in the face of physical similarity can be seen when we turn from the baseline emergent property E to the more specific features E₂, E₃, and E₄. Consider E₂ at time t₂. You might have the underlying physical properties P* and P₂ without having had E₂. For E₂ is a causal product of the immediately prior state of S at t₁ (comprised of P*, P₁, and E). This prior state presumably could have been different in such a way that E₂ does not occur at t₂, although the actual physical state at t₂ does. We need only suppose a suitably fortuitous change in the environmental circumstance P@ at the prior time t₁.

However, these two ways for strong supervenience to fail within our framework are consistent with *global* supervenience, where the base of underlying properties fixing the supervening properties is the entire state of the physical universe. (We again restrict the base synchronically. In a moment, we shall consider the effect of lifting this restriction.) For global

supervenience to fail, it is necessary for the causal connections to be probabilistic only. In the case of such a stochastic process, we can hold fixed the immediately prior state of S and its environment and suppose a scenario in which the physical state of S at t_1 were to cause the actual underlying physical state at t_2 , but, owing to a different chancy outcome, cause the occurrence not of E_2 , but of some distinct property, E_2' , at the emergent level.

Finally, we consider whether fixing the physical state of the universe at all times suffices to fix the distribution of emergent properties. Again, we believe the answer is No. For it is possible to imagine a case in which an indeterministic physical state P_1 has two possible emergent outcomes, and these emergent states, in turn, have a possible physical effect in common. (Perhaps they differ in their other possible effects, or perhaps they differ solely in the strength of their propensity for each of the same range of outcomes.) In such a case, it is possible for there to be two physically and nomically indiscernible worlds which nonetheless differ with respect to their emergent properties. Schematically:

$$\begin{aligned} W_1: & P_a \rightarrow P_b \ \& \ E \rightarrow P_c \\ W_2: & P_a \rightarrow P_b \ \& \ E' \rightarrow P_c^{13} \end{aligned}$$

All of this contrasts with what we should suppose (at least on standard physicalist assumptions) regarding ordinary, purely structural properties of composites. If an object's having a certain biological property B just consists in its parts having certain properties and being arranged in a certain fashion, then we cannot suppose a scenario, consistent with natural law, wherein those lower-order properties and relations obtain but B does not. Fix the physical, and the garden varieties of chemical and biological properties are thereby fixed, too. But not so the emergent properties, if such there be.

V. Metaphysical Objections

Emergent properties are epiphenomenal

Are emergent properties inevitably epiphenomenal, at least with respect to the purely physical states of S and its immediate environment? Is the emergent system in its purely physical aspect—is physics more generally, on the present picture—causally closed? Clearly, the answer is No. P_3 's obtaining at t_3 is in part a product of E and E_2 's obtaining at t_2 . Had one or both of these failed to obtain at that previous time, something other than P_3 would have occurred subsequently. Consistent with this, it is true in an emergentist scenario that everything that occurs rests on the complete dispositional profile of the physical properties prior to the onset of emergent features. For the later occurrence of any emergent properties are contained (to some probabilistic measure) within that profile, and so the effects of the emergent features are indirectly a consequence of the physical properties,

too. (We might, then, speak of a *dispositional* supervenience, taking care to note that this does not imply the closure thesis stated at the outset of this paper.) The difference that emergence makes is that what happens transcends the immediate, or local, interactions of the microphysics.¹⁴

Emergence involves either causal circularity or systematic overdetermination
In “Making Sense of Emergence” (1999), Jaegwon Kim argues that unless ontological emergence is given a deflationary, epistemological interpretation, it is implausible because its component notion of downward causation is incoherent. He considers two varieties, synchronic and diachronic. With the synchronic variety, the whole’s having emergent property M_1 at t_1 both supervenes on its physical state P_1 at t_1 and is partly causally determinative of that very physical state. With the diachronic variety, the whole’s having emergent property M at t_1 supervenes on its physical state P_1 at t_1 and is partly causally determinative of its physical state P_2 and mental state M_2 at the subsequent time t_2 (pp. 26–31).

Kim judges the synchronic variety to be absurd on the grounds that it involves causal circularity. (We let this pass, since our view of emergence is not of this type. But note that if one resolves the whole’s physical state into two sub-states, as we do above—a state that generates the emergent property and a distinct state that is causally affected by it—Kim’s charge of causal circularity would be unfounded.) The diachronic variety, he allows, escapes the circularity worry, but it is prone to his causal exclusion argument:

... I earlier argued that any upward causation or same-level causation of effect M^* by cause M presupposes M ’s causation of M^* ’s lower level base, P^* (it is supposed that M^* is a higher-level property with a lower-level base; M^* may or may not be an emergent property). But if this is a case of downward emergent causation, M is a higher-level property and as such it must have an emergent base, P . Now we are faced with P ’s threat to preempt M ’s status as a cause of P^* (and hence of M^*). For if causation is understood as nomological (law-based) sufficiency, P , as M ’s emergence base, is nomologically sufficient for it, and M , as P^* ’s cause, is nomologically sufficient for P^* . Hence P is nomologically sufficient for P^* and hence qualifies as its cause. The same conclusion follows if causation is understood in terms of counterfactuals—roughly, as a condition without which the effect would not have occurred. Moreover, it is not possible to view the situation as involving a causal chain from P to P^* with M as an intermediate causal link. The reason is that the emergence relation from P to M cannot properly be viewed as causal. This appears to make the emergent property M otiose and dispensable as a cause of P^* ; it seems that we can explain the occurrence of P^* simply in terms of P , without invoking M at all. If M is to be retained as a cause of P^* , or of M^* , a positive argument has to be provided, and we have yet to see one. In my opinion, this simple argument has not so far been overcome by an effective counter-argument (p. 32).

Kim's argument against diachronic downward causation boils down to this: The diachronic activity of an emergent property will inevitably be redundant, since its effects are directly (and not just indirectly) attributable to the conditions which sustain it; thus, emergent properties could not, as emergentism demands, confer "genuinely novel causal powers"—powers that "must be capable of making novel causal contributions that *go beyond* the causal powers of the lower-level basal conditions from which they emerge" (p. 25).

We confess to be puzzled by Kim's argument, even given its assumptions about the dynamics of emergence.¹⁵ But Kim's argument clearly cannot get off the ground against the dynamical model of emergence set forth above, which characterizes the relation between emergents and their 'base' conditions as diachronic and causal, rather than as a *sui generis* variety of synchronic supervenience. As noted in discussing the epiphenomenalism charge, the distinctive potentialities of emergent properties do stem indirectly from the total potentialities of the basic physical properties. But they do not determine the emergent effects (or fix the emergent probabilities) *independently* of the causal activity of those emergents.

VI. Epistemological Objections

If the foregoing is sound, emergence is a coherent metaphysical concept capable of grounding an interesting solution to the mind-body problem. However, some have objected to the use of the concept of emergence in a different way. As they see it, emergentist hypotheses are gratuitous or necessarily more objectionable than alternative strategies for explaining the kind of possible phenomena for which emergentist hypotheses are supposed to be ideally suited.

The potential significance of these objections is more limited in the present context than might appear at first glance. For the appeal to emergence within the philosophy of mind is not epistemically on a par with other emergentist hypotheses within theoretical science. In those other cases, one might form emergentist conjectures solely in response to observed failures of ordinary reductionist hypotheses—as occurred, for example, in early twentieth-century biology prior to the discovery of DNA. There, the question of whether other types of hypothesis might inevitably fill such explanatory gaps equally well at less ontological cost is crucial. But we are attracted to an emergentist conception of mental states for more direct reasons: we already reasonably believe on broadly Cartesian, introspective grounds that mental states are states of complex systems that involve distinctive qualities and dynamics *vis-à-vis* the systems' underlying physical states. So even if the differential behavior resulting from emergent mental features could, from a third-person standpoint, be adequately captured in formal dynamical terms by non-emergentist models of the sort described below, we

have independent reasons for rejecting such models. Having said that, we will now argue that, given the right sort of evidence, better alternatives would *not* in principle be available.

Emergent laws need no emergent entities

From the standpoint of empirical theory, the notion of an emergent property is tailor-made for situations in which elegant dynamical laws that are fully adequate to characterize particle behavior below a certain threshold level of systematic complexity prove inadequate, even with ideal technology, to describe situations above the threshold. However, assuming the variation in underlying dynamics beyond the threshold is lawlike, we could simply complicate our original laws by recasting them as nonlinear functions that yield the same predictive outcomes below the threshold value and yield in a compact fashion the newly observed outcomes at and above it. Nothing is gained in explanatory value by a counterpart theory that instead posited an emergent property. (By contrast, we *must* quantify over some *non-emergent* properties if we are to give any lawlike account of phenomena at all.)¹⁶

In effect, the objector proposes that, in the worst-case scenario for the reductionist, we should build holistic causal patterns into our fundamental laws, but deny that we need to correspondingly complicate our basic ontology. In a slogan, emergent laws without emergent entities. In reply, we suggest that here one's basic metaphysical convictions matter. From the standpoint of a causal realist, discontinuous microscopic behavior associated with precisely specifiable macroscopic parameters needs to be explained on both sides of, *and across*, the threshold in causal/dispositional terms. New basic properties are naturally posited to explain systematic differences at a basic level of activity—unless we can get an equally elegant resulting theory by complicating the dispositional structure of the already accepted inventory of basic properties. This leads us to the second strategy. . . .

Latent microphysical dispositions alone

Emergentist hypotheses assert an unnecessary middle term. They suppose that complex microphysical configurations jointly cause the occurrence of an emergent property, which in turn has a causal effect on the evolution of the physical state of the system. Yet suppose we were actually to observe discontinuous patterns of effects within complex systems. It would be simpler to conclude that the microphysical entities have otherwise latent dispositions (triggered only within specified, macroscopically complex contexts) *directly towards large-scale, microphysically-complex effects*, alongside those dispositions which are continuously manifested in (nearly) all contexts, and the observed discontinuity is a result of the manifestation of these latent dispositions. All of what happens, we could continue to suppose, is *directly* a product of the elementary constituents of matter, though the

range of microphysical dispositions would be richer than one might have supposed through observation of relatively simple particle interactions alone.¹⁷

The micro-dispositional approach, we argue, would provide a less plausible interpretation of the appropriate dynamical discontinuity than would a holistic emergentist approach. Consider the complexity of the effects that will be directly ascribed to the microphysical particles (acting in tandem): in quantum-mechanical terms, a large-scale distribution of qualities unevenly distributed across the region in question (and perhaps not at all in some sub-regions). We are thus required to suppose that each of the basic particles is disposed towards a highly complex result (by contrast, in our scenario, the disposition is towards an ontologically simple, or nonstructural, quality). The posited dispositional complexity will greatly increase, furthermore, when we contemplate more complicated scenarios which the emergentist would interpret as the outcome in part of nested structures of emergent properties. A variation on the latent-micro-dispositions proposal will suppose instead subtle differences in the specific dispositions being exercised by the different particles, depending on their location within the appropriate type of complex. Since the particles are of the same type, each has all of these dispositions, but they exercise only those that their different circumstances permit. While perhaps more realistic than the original idea, the variation needs to posit a wide range of hidden dispositions (the number depending on the empirical details) for each specifiable dynamical discontinuity. Our view is simpler and so to be preferred: for each discontinuity, just one basic disposition—to give rise to an emergent property under suitable circumstances—that is had by every fundamental constituent.

In conversation, we find that many philosophers are skeptical about our contention that theoretical simplicity alone would favor the holistic emergence approach over the microdispositional approach as an interpretation of appropriate dynamical discontinuity. (The implications for theoretical simplicity of differences in the complexity of dispositions, or in the number of dispositions of a given type, is not clear cut.) Here is an alternative route to the same destination.¹⁸ Suppose two scenarios in which systems in identical physical states, *P*, and local environments gives rise to two very different effects, *R1* and *R2*. Suppose further that the effects cannot be attributed to multiple probabilistic propensities of the system's physical state. Finally, suppose that the cause of *P* is different in the two scenarios (*H1* and *H2*, respectively). Since the micro-dispositionalist doesn't recognize holistic emergent properties, he cannot account for the different effects of *P*, apart from an objectionable action-at-a-distance from *H1* and *H2*. The emergentist, by contrast, can readily account for these diverging results as follows:



The prior causes, H1 and H2, manifest their difference solely at the emergent level, and the different emergent features, in the presence of the common state P, account for the difference in R1 and R2, respectively.

Taking emergence seriously leads to inductive skepticism

The emergentist supposes that it is an ‘open question’ whether there are any emergent properties and that the required complexity for the appearance of an emergent property is not knowable a priori—it will look ‘arbitrary’ from the point of view of basic physics. It follows that we should be agnostic about *any* generalizations of going theory to situations of a complexity level different from that which we’ve previously observed and verified as consistent with what a nonemergentist picture would lead us to expect. But this is absurd.¹⁹

Reply: the objection misrepresents the emergentist’s view of the epistemological situation. The emergentist can and should allow that there is an epistemological presumption against emergentist hypotheses for systems of currently-untested complexity levels *absent special reason to suspect that they are different from run of the mill cases*. In general, it is enough to forestall unreasonable skepticism towards modest extensions of going theories that one judge the epistemic probability for non-special cases to be low, though non-negligible. Furthermore: theories are necessarily put to the test for a finite number of case variations. Suppose we verify a theory for systems with complexity levels *i* and *i + k*, but not for systems whose complexity falls between these magnitudes. We may confidently disregard emergentist hypotheses concerning systems within the interval as well, since any emergent effects for such systems is apt to be manifested at the higher *i + k* level.

VII. ‘Not a Scintilla of Evidence?’

Brian McLaughlin concludes his lengthy, illuminating account of British Emergentism with a note of ostensible scientific sobriety: while emergence is a coherent concept, it is enormously implausible that there *are* any such features (1992, p. 91). We contend, to the contrary, that it is sweeping judgments such as McLaughlin’s, based on partial evidence, that fall afoul of sober scientific practice—practice that includes measured skepticism regarding broad extrapolation of confirmed results. We allowed above that we reasonably suppose theories to have broader application than to just those types of cases actually tested. However, the further removed a scenario is from the well-confirmed range along a scale of increased

structured complexity, the less confident one should be that the theory is fully adequate to the situation.²⁰

There are strong *methodological* reasons for taking a reductionist approach to poorly understood phenomena. If there are any emergent features involved, we will understand their distinctive contribution only if we have a fairly thorough grasp of the microstructural dynamics with which they interact. But we should not mistake methodological stance for evidential confirmation. In the opposite direction, the stance of typical emergentists of the early twentieth century was flawed in two basic ways: they advanced an oversimplified and insular view of ‘levels’ of nature on the basis of what was essentially an argument from ignorance. Theorists of the early twenty-first century, by contrast, are confronted with a picture of enormous, nested complexity in the biological and psychological realms. Clearly there are no sharp dynamical boundaries between such levels of organization. But for all that, we should not be astonished to discover emergent phenomena within the interconnected whole of nature—as a subtle interplay of microphysical and holistic factors, instead of dramatic supersession. Biological life, so poorly understood in the early twentieth century, was the favorite target of earlier emergentists. Now, of course, the epistemic situation is dramatically different. With the chemical basis of life being further charted with each passing year, there is no positive reason for us to suppose that emergent factors are essentially involved. Even so, absent a fully worked out picture of microbiology and a reasonably good set of mappings from complex physical structures to basic microbiological features, emergence cannot be entirely ruled out, either.

But matters are different, we contend, with respect to mind, and different in ways that make emergentist conjectures more than arguments from ignorance. A person’s experiences and other conscious mental states—states which manifestly influence our behavior—exhibit features quite unlike those of physical objects, whether as revealed in ordinary sense perception or as uncovered in the physical and biological sciences. And the maximally direct nature of our first-person awareness of these conscious states precludes the a posteriori ascription to them of underlying physical micro-structure hidden to introspection. There is, then, positive evidence that strongly favors a dualist direction at least as strong as the emergentist property dualism envisioned here. All evidence is defeasible, of course, and some will contend that the whole conception of mind on which this is based, deeply entrenched as it is, is but an illusion which must give way to reductive, third-person theories. To that we say: believe it if you can. And do not neglect to develop an epistemology—anti-skeptical, lest you saw off the limb on which you stand!—that shows how our ordinary empirical knowledge may comfortably rest on a radical and pervasive cognitive illusion at its very source.

Notes

¹ Our emergentist framework also shares the advantages of traditional dualism over physicalism with respect to the phenomenal character of consciousness, although we will not develop this point here. (Doing so would require us to consider the relationship of the dynamical to the qualitative in the theory of properties, itself a thorny issue.)

² For discussion, see (Foster 1991, pp. 163–172), (Kim 2001) and in reply (O'Connor 2001). The modern source of the pairing problem is (Foster 1968).

³ The developmental problem for dualism is presented in (Armstrong 1968). (Hasker 1999) defends an emergentist form of substance dualism.

⁴ The locus classicus of nonreductive physicalism and multiple realizability is (Fodor 1974). For further discussion see (LePore and Loewer 1987), (Kim 1992), (Fodor 1997) and (Block 1997).

⁵ The question of what composite individuals there are is a non-trivial one within sparse-property ontologies. We shall here blithely assume the existence of composites having no emergent properties, solely for the sake of expository simplicity. We briefly consider below the relevance of emergent properties to 'sparse-*individual*' ontologies. For much further discussion, see (O'Connor and Jacobs 2003).

⁶ Does Armstrong accept this 'nothing buttery' understanding of a structural property? This is not entirely clear. There are some indications that he construes it as something subtly 'extra'—distinct from the underlying states of affairs but strongly supervening upon them (Armstrong 1997, p. 37). But this is hard to square with his adamant contention that they are an ontological free lunch. (See the discussion on pp. 34–45.)

⁷ We recognize that at least some of the British emergentists were committed to something like strong supervenience. But there is nothing in the core concept that requires this, and it is in tension with their commitment to emergents' conferring irreducible causal powers, as we will show below.

⁸ See also (O'Connor 2000, Ch.6, and 2001), which repudiates (O'Connor 1994) on supervenience.

⁹ Compare (Armstrong 1997): "suppose that atomic particular *a* has atomic F, and stands in the atomic relation R to the atomic particular *b*, which has atomic G. It may be that the non-atomic particular *a + b* has the further *atomic* property H, although neither *a* nor *b* (nor, indeed, any sum of proper parts of *a* and *b*) has property H. It might further be that, for *a + b* to instantiate atomic H, one of the two particulars nominally had to have atomic property F while the other had to have atomic property G. In this situation, perhaps, the molecular state of affairs *a's being F + B's being G + a's having R to G* nominally ensures [*a + b*]'s having H. We might call H an 'ontologically *Gestalt*' property. Other, more complex scenarios could be devised. But it will be seen that Logical Atomism could in this way be made compatible with *emergent* laws, as the above law might well be." (p. 153)

¹⁰ In several articles (1996, 1997a, 1997b), Paul Humphreys defends a picture of ontological emergence that in certain respects resembles our own. While we cannot discuss his picture in detail here, we want to make a few points concerning it. As with our own view, Humphreys' account involves nonstructural elements having "novel causal powers," though it's unclear, even in general terms, how the dynamics are supposed to go (1997a, p. 8). What is distinctive in his picture is a notion of property fusion. When properties at the generating level *i* are 'fused', the individual properties are destroyed and are non-individuable within the emergent fusate existing at the *i + 1* level. (Humphreys emphasizes that fusion is an ontological, not logical, operation.)

We do not deny that fusion characterizes a possible variety of emergent properties. But we do think it is very implausible on general empirical grounds that it might apply to the distinctive features of any actual macroscopic systems:

1) Consider such a system S having an emergent property E. The subvening properties giving rise to E also constitute myriad nonemergent, structural properties of S. If these lower

level properties literally ceased to be in fusing into E, then so, it seems, would those structural properties, including those which may be crucial to other functions, e.g. life preservation.

2) Humphreys will presumably allow that fused properties are naturally decomposable into their original constituent properties (the “arguments” of the fusion “operator”—to employ a logical vocabulary for this metaphysical operation). But then it is not so clear that they ever lose their identity in the first place. Since decomposition is a disposition (power), within an ontology where properties are closely associated with their possessors’ causal powers, the fused entities have not fully lost their identity, since they retain the power to de-couple and differentially interact with other properties in specifiable conditions.

3) The question of how decomposition of fusates works is nontrivial. If fusates are truly nonstructural, and given Humphreys’ belief that fusion likely occurs at many levels of complexity, then it should not in general be possible to anticipate *a priori* how this will go. The decomposition of a fairly high-level emergent at level *n* might result in features at any level $< n$. Yet the decomposition of actual complex systems, including living systems, do not yield any surprises. By way of contrast, consider the familiar process of ‘decomposing’ cheese and broccoli soup. We might decide to decompose two bowls of soup in two different ways. The first way is to put it into a food grinder, with ground bits of broccoli, cheese and other constituents as the result. The second way is to use lytic enzymes which break down the soup into sugars and various other chemicals. In the latter case, decomposition occurs a couple levels down (the level of macromolecules as opposed to the level of ground broccoli). We can anticipate the sort of result stemming from each of these processes based on our knowledge of the structural features of both the catalytic factors and the receptive system. But if an emergent system *loses* (through fusion) all or most of its underlying, lower-level features, decomposition should be just as unpredictable, *a priori*, as fusion is.

¹¹ Recall that P* is a specific, albeit partial, configurational property, one that suffices for generating E. It would be odd—we do not say impossible—that the event of P* at t_0 causes E at t_1 , yet E at a later moment t_n is a contributing cause to P*’s subsequent persistence.

¹² And we will argue below that the present picture points to at least one distinct variety than has yet to be conceived even in the mind of Jaegwon Kim!

¹³ A referee for *Nous* wondered whether our multi-layered picture of emergence was a necessary component of our argument for the non-supervenience of dynamical emergence. Our simplified scenario here in comparing W1 and W2 demonstrates that it is not. We emphasize that the multi-layered picture is not required for a satisfactory reply to any of the objections we consider in this paper. We claim simply that it is intuitive to divide conscious mental phenomena into enduring and momentary states, and this division has to be reflected in the dynamical evolution of mental states.

¹⁴ (Loewer 2001) gives a counterfactuals-based argument for the epiphenomenality of emergence. His argument *begins* from the premise that physics is causally complete and deterministic. He contends that it would go through even assuming indeterminism, but he must mean a kind of indeterminism on which the physical facts at *t* plus the non-emergence-involving laws fix the chances at subsequent times *t*. None of this is accepted by the position defended here. (Also, his Lewisian way of evaluating counterfactuals seems to presuppose Humeanism about laws, since it assumes that a non-actual world W* containing an event P that in context is contrary to an actual-world law is necessarily a world where the laws are different. From our non-Humean perspective, one can readily imagine cases of miraculous counterintervention of law where the laws are still the same as those in the otherwise highly similar actual world. But this contention is not necessary for our rejection of Loewer’s argument.) The proper lesson of Loewer’s argument is that certain assumptions which are natural within a strongly reductionist picture are disastrous for working out the concept of emergence, so don’t make them.

¹⁵ It seems at first blush to be driven by the use of inconsistent criteria for causal sufficiency to the relation between P and P*, on the one hand, and P and M (or P* and M*),

on the other. If it is taken as given that the relation between P and M is not one of causal sufficiency (in accordance with Kim's interpretation of the British emergentists), why does this not also preclude causal sufficiency in the larger link between P and P*? Perhaps the thought is that while a *sui generis*, non-causal relation may enter into (as a partial constituent of) a sufficiency relation, it cannot of itself constitute one.

¹⁶ This objection was first proposed to one of us by Anil Gupta. It is discussed in (O'Connor 1994).

¹⁷ A relative of this view was first proposed to one of us by Sydney Shoemaker and is discussed in (O'Connor 1994 and 2001). (Shoemaker 2002) sets out this picture as a way of *explicating* a concept of emergence. A variation on it was suggested in discussion and correspondence by Dan Ryder.

¹⁸ We owe this argument to Michael Tooley, which he proposed in conversation.

¹⁹ Kelsey Rinella suggested this objection to us.

²⁰ Note that we refer not merely to the complexity of aggregation, but of nested hierarchy. This simple distinction suffices to separate us from the stronger type of skepticism regarding extrapolation that is expressed, as best we can tell, by (Cartwright 2000).

References

- Armstrong, David. (1968) *A Materialist Theory of Mind*, London: Routledge & Kegan Paul.
———(1997) *A World of States of Affairs*, Cambridge: Cambridge University Press.
- Block, Ned. (1997) "Anti-Reductionism Slaps Back," *Philosophical Perspectives, 11: Mind, Causation, and World*, Cambridge, MA: Blackwell, pp. 107–133.
- Cartwright, Nancy. (2000) *The Dappled World: A Study of the Boundaries of Science*, Cambridge: Cambridge University Press.
- Fodor, Jerry. (1974) "Special Sciences (or: The Disunity of Science as a Working Hypothesis)," *Synthese* 28, pp. 97–115.
———(1997) "Special Sciences: Still Autonomous After All These Years," *Philosophical Perspectives, 11: Mind, Causation, and World*, Cambridge, MA: Blackwell, pp. 149–63.
- Foster, John. (1968) "Psychophysical Causal Relations," *American Philosophical Quarterly* 5, pp. 64–70.
———(1991) *The Immaterial Self*, London: Routledge.
- Hasker, William. (1999) *The Emergent Self*, Ithaca: Cornell University Press.
- Humphreys, Paul. (1996) "Aspects of Emergence," *Philosophical Topics*, 24 (1), pp. 53–70.
———(1997a) "How Properties Emerge," *Philosophy of Science*, 64, pp. 1–17.
———(1997b) "Emergence, Not Supervenience," *Philosophy of Science*, 64, pp. S337–S345.
- Kim, Jaegwon. (1992) "Multiple Realization and the Metaphysics of Reduction," *Philosophy and Phenomenological Research* 52, pp. 1–26.
———(1999) "Making Sense of Emergence," *Philosophical Studies*, 95, pp. 3–36.
———(2001) "Lonely Souls: Causality and Substance Dualism," in K. Corcoran, ed., *Soul, Body, and Survival*, Ithaca: Cornell University Press, pp. 30–43.
- Loewer, Barry and LePore, Ernest. (1987) "Mind Matters," *Journal of Philosophy* 84, pp. 630–642.
- Loewer, Barry. (2001) "From Physics to Physicalism," in Loewer and Gillett, eds., *Physicalism and Its Discontents*, Cambridge: Cambridge University Press, pp. 37–56.
- McLaughlin, Brian. (1992) "The Rise and Fall of British Emergentism," in A. Beckermann, H. Flohr, and J. Kim, eds., *Emergence or Reduction?*, Berlin: De Gruyter, pp. 49–93.
———(1997) "Emergence and Supervenience," *Intellectica* 25, pp. 25–43.
- O'Connor, Timothy. (1994) "Emergent Properties," *American Philosophical Quarterly* 31, pp. 91–104.
———(2000) *Persons and Causes: The Metaphysics of Free Will*, New York: Oxford University Press.

- (2001) "Causality, Mind and Free Will," in K. Corcoran, ed., *Soul, Body, and Survival*, Cornell University Press, pp. 44–58. (Originally published in *Philosophical Perspectives 14: Action and Freedom*, 2000, Cambridge, MA: Blackwell, pp. 105–117.)
- and Jacobs, Jonathan. (2003) "Emergent Individuals," *Philosophical Quarterly* 53 (Oct.), pp. 540–555.
- Shoemaker, Sydney. (2002) "Kim on Emergence," *Philosophical Studies* 108 (March), pp. 53–63.