

EMERGENCE AND REFLEXIVE DOWNWARD CAUSATION

JOHN SYMONS

University of Texas at El Paso

Abstract

This paper responds to Jaegwon Kim's powerful objection to the very possibility of genuinely novel emergent properties. Kim argues that the incoherence of reflexive downward causation means that the causal power of an emergent phenomenon is ultimately reducible to the causal powers of its constituents. I offer a simple argument showing how to characterize emergent properties in terms of the effects of structural relations on the causal powers of their constituents.

I

Despite undergoing something of a revival in recent years, 'emergence' is a concept that continues to elicit a strongly critical reaction from certain philosophical quarters. Though it has a variety of respectable uses in the natural sciences, reference to emergence in the philosophy of mind evokes visions of wild speculative metaphysics and an abandonment of ontological parsimony. Given its checkered history, this response is understandable.¹ And yet, a suitably-crafted concept of emergence could play a useful role in contemporary metaphysics with respect to the problem of epiphenomenalism for higher-level properties. Unlike traditional functionalist approaches to this issue, emergence promises to provide the conceptual framework necessary for understanding the related notions of causality, explanation and individuation such that we can avoid rendering everything above the level of our microphysics epiphenomenal.

Emergentists have traditionally sought ways of reconciling commonsense materialism with our ordinary assumptions about the kinds

© *Principia*, 6(1) (2002), pp. 183-202. Published by NEL — Epistemology and Logic Research Group, Federal University of Santa Catarina (UFSC), Brazil.

of things that exist. By commonsense materialism I mean the scientifically-informed belief that the things we encounter in our daily lives are constituted of matter and energy. This aspect of the scientific worldview is often seen as clashing with our belief in the reality of the objects of everyday life. Are there really mountains, teacups, trees and people or is there nothing but "atoms in the void"? Virtually all naturalistically-inclined thinkers (emergentists included) will accept that everything in the natural world is constituted of matter and energy that is derived from the original moment of the formation of the universe.

However, even given these physicalist assumptions, there is a difference, as Lynne Rudder-Baker and others have noted, between what something is made of and what it is: (Rudder-Baker 1993) This distinction between constitution and identity is especially significant in the case of people. Ordinarily, a person's name is taken to refer to something that is irreducible to the state of the physical stuff that constitutes a particular body at a particular time. The problem of determining a meaningful sense in which you can be said to differ from or are more than your constituents has a long history. In the past thirty years, considerable effort has been devoted to understanding the difference between constitution and identity in functional terms, and yet the functionalist half-way house between physicalism and dualism faces a set of well-known problems.² Emergence, by contrast is a relatively unexplored way to understand the difference between constitution and identity from within a naturalistic framework.

Understanding what it means to say that minds are real and that reference to concepts like belief and desire has real explanatory value, requires that we develop a proper understanding of how to make room for the causal powers of mental entities in a physical world. Given the causal completeness of physics and the non-reducibility of mental properties, the power of minds to make a causal difference in the world is a tricky metaphysical problem indeed. Many philosophers view functionalist or supervenient characterizations of mental properties as the proper modern response to the problem of mental causation and to the kinds of concerns that motivated the old British Emergentist tradition. Functionalists legitimize statements involving mental entities by reference to the explanatory indispensability of

those entities in the special sciences. However, an argument for explanatory indispensability is not the same as an argument for the reality of an entity. As Dretske (1988) and others have noted, given the usual physicalist ontological assumptions, it is difficult to justify more than an epiphenomenal role to things we individuate functionally. Hence, functions have little ontological weight and hard-nosed reductionists can easily argue that while they might be good for certain explanatory or pragmatic purposes, they will not be counted when we catalog the ontological furniture of the natural world.

The first obstacle to arguing for the reality of higher-level phenomena is the suspicion that the threat of epiphenomenalism is a fake problem. This suspicion motivates Rudder-Baker's (1993) pragmatic response to the problem of mental causation. She advises dropping our metaphysical commitment to the causal closure of the physical world in order to save talk of persons (2000). Her argument derives from pragmatic considerations and as such, it is obviously sound. It would be unimaginably difficult to abandon our belief in people for the sake of our metaphysical scruples. But perhaps rather than simply dismissing the problem for pragmatic purposes, we can understand the instrumental value of our belief in higher-level phenomena as pointing to an alternative metaphysical picture. In contrast with the strategy of pragmatic avoidance, I take the concept of emergence as providing a way to recast our basic metaphysical assumptions so as to account for the usefulness of higher-level phenomena.

The problem of understanding the relationship between the instrumental success of our belief in higher-level phenomena and the truthfulness of those beliefs is especially important with respect to the status of psychological explanation. It is generally assumed that the legitimacy of psychological explanation depends, at least in part, on its ability to uncover distinctively psychological kinds of causes. The trouble for psychological explanation is that standard, non-reductive materialist views of mental life seem unable to differentiate between the causal power of mental events qua mental events and the causal power of the microphysical phenomena that realize or embody them. Our alleged inability to differentiate between what minds do and what their physical constituents do becomes a problem for non-reductive materialists given orthodox views of what it is that makes

something real. If neurons, rather than minds are doing all the work, can we really say that minds exist? For the generalizations of psychology to have more than a merely epiphenomenal grip on reality, proponents of non-reductive materialism must be able to point to a causally relevant role for mental properties.

While there are some philosophers who happily advocate eliminativism with respect to mental life, the idea that the mental properties of an agent will not figure in a genuine explanation of her behavior is unacceptable to most non-reductive materialists. Nevertheless, the crucial problem for all non-reductive materialists is to understand what it means to say that higher-level phenomena (like people) act on their constituents in such a way as to distinguish the causal power of the higher-level phenomena from the causal power of its constituents. As previously mentioned, one prominent response to the problem is to treat higher-level phenomena as supervenient. In essence, advocates of supervenience claim that the properties characterizing the identity of a supervening phenomenon exist only because of the underlying, or 'subjacent' properties of its constituents. On this view, the causal properties of the constituents can account entirely for the causal properties of the supervening phenomenon. For something to count as real, it must have a unique causal role to play in the natural world. A property that supervenes on the properties of its constituents cannot, by definition, have such a role. Hence, supervenience is basically a denial of the reality of higher-level phenomena. This is simply because, in these debates, reality is connected to the notion of causal power.

At first glance, contrary to advocates of supervenience, it seems obvious that complex systems like hurricanes, organisms and epidemics have their own distinctive set of powers. As it digests its dinner, for example, an organism seems to exert a distinctively biological kind of causal power over its newly absorbed constituents. And yet, the commonsense view that new things can exhibit new powers faces a metaphysical objection reminiscent of Ecclesiastes' pessimistic pronouncement that 'there is nothing new under the sun. Is there a thing of which it is said, "See this is new"? It has already been in the ages before us' (1.9-10). The fate of the food as it undergoes digestion, according to the modern Ecclesiastes, can be explained

entirely in terms of a set of chemical reactions, and these in turn can, at least in principle, be explained entirely in terms of the behavior of the atoms in the molecules of our food and our digestive juices at the quantum level. Ultimately, there is no need to look for any ontological content to our talk of organisms and digestion, let alone hunger, kitchens and the pleasure of a good meal. For emergentists, by contrast an organism would be a good example of the kind of thing which manages to act on its own constituents in a way that can be distinguished from the behavior of those constituents.

However, while it certainly seems natural to assume that higher-level phenomena like organisms exhibit their own novel kind of power in the world, on reflection, the kind of downward causation implicit in examples of emergence can seem like a bizarre metaphysical bootstrapping exercise. Can something like an organism really act on its constituents? Wouldn't this require that in acting on its constituents, the emergent property is changing the very things that make it what it is? If so, then wouldn't the identity of the organism be changing in such a way as to make it impossible to say that it is acting on itself? Taken in its strictest sense, it looks like the idea of a system acting on its own constituents reduces to absurdity. The apparent contradiction that seems implicit in such cases leads Jaegwon Kim to conclude that the putative causal powers of higher-level properties are always causally preempted by the properties of their underlying physical constituents. Kim and others argue that, while we can certainly identify new patterns and phenomena for instrumental or other reasons, these can only be shown to be 'real' or, to constitute a 'natural kind', given the identification of a unique set of causal powers. Consequently, a non-trivial model of downward causation can make sense only if we give it a conceptual interpretation: "That is, we interpret the hierarchical levels as levels of concepts and descriptions, or levels within our representational apparatus, rather than levels of properties and phenomena in the world." (Kim 1999, 33). The present paper responds to Kim's criticism of reflexive downward causation and in so doing attempts to make the case that higher-level phenomena are more than merely an artifact of our representational apparatus.

II

To begin with, the case for a realistic form of emergence is blocked by metaphysical principles governing recent debates in philosophy, especially in the philosophy of mind. The problem of determining the causal powers of higher-level phenomena is generated by two well-entrenched metaphysical assumptions shared by virtually all materialists. The first is the claim that all non-basic properties strongly supervene on their physical constituents, the second is that the physical world is causally closed. Given the causal closure principle, it is impossible to imagine causal powers that are not already exhaustively captured by the basic physical constituents of the natural world. In part, this is because such powers (higher-level or otherwise) would enter into intolerable causal competition with their physical constituents.

So, how should emergentists proceed in the face of the preemption argument? At present there are two principal strategies in the literature, each of which has an important place in the emergentist arsenal: First, an argument against preemption from pragmatic considerations and second a criticism of the physicalist's unrealistic physics. While both lines of criticism are reasonable, they fail to respond to the central problem. A third more basic conceptual argument is needed to seal the case for a believable form of emergentism. This third argument must make clear sense of the most paradoxical commitment of traditional emergentism, namely the belief in what Kim has called "synchronic reflexive downward causation".³

Beginning with the pragmatic argument, in the first place critics like Rudder-Baker and others correctly note that the preemption argument, along with the minimalist ontology it imposes, draws its force from the idea that causal power talk makes sense as a way of individuating properties and objects. Since, there is no clear agreement on the nature of causation and given the difficulty of getting clear on the notion of cause, many philosophers have chosen to simply avoid the murky question entirely. For example, in his recent discussions of emergence Robert Batterman has simply denied that natural kinds talk makes sense (Batterman, 2000). While I am sympathetic to Batterman's frustration with the metaphysical argument

concerning causal powers and natural kinds, his way of dismissing the preemption argument, by sidestepping the core conceptual issue, does not satisfy the traditional emergentist demand for a way of legitimizing emergent phenomena as *real*, as opposed to merely *instrumentally convenient*.

There are a variety of pragmatic strategies for continuing to talk about higher-level phenomena, each of which is unsatisfying to emergentists in the long run for similar reasons. For instance, one could admit that causation and scientific explanation are intimately connected but reverse the traditional order of their relationship. Along these lines, Rudder-Baker has argued that, in practice, our explanations are not legitimized by their relationship to some underlying causal structure. Instead, she argues, our notion of cause derives its legitimacy from our explanatory practices. For Rudder-Baker, good explanations lead us to assert the presence of causes and the reality of certain kinds of objects, not vice versa (2000).

Rather than taking the pragmatic way out of the metaphysical problem, emergentists may undertake a second more general critique of the premises that force us into the impasse in the first place. This second strategy for avoiding preemption argues that an inappropriately fundamentalist picture of physics supports the preemption argument. Most scientists would agree, for example, that the picture of physics implicit in the work of most physicalist philosophers of mind is hopelessly unrealistic. In terms of ontology, for example, the pre-emptor's metaphysics is supported by the idea of a mechanical universe where smaller things combine in mechanical relations to produce bigger things. It is well known that the reductive relationship between parts and whole runs up against a number of problematic phenomena in physics. Take, for example, cases where new entities emerge via fusion. For Paul Humphreys, the physical phenomenon of fusion provides an obvious example of emergence. Here, the bearers of two distinct properties combine by fusion to generate a third entity whose properties cannot be said to supervene on the properties of its constituents.

In the case of fusion, the crucial point for the discussion of emergence is whether the emergent entity can be said to supervene over its constituents in the way that functionalists believe functional con-

cepts supervene on physical structures. As Humphreys argues, fusion is clearly not a matter of a functional or a supervenient relationship between the constituents and their fused product. Emergence via fusion differs from a functional relationship between levels of explanation in the following way. In the case of a supervenience relation, properties can be said to supervene on their structural constituents without eliminating or competing with the underlying material structure. A classic example might be for example, the aesthetic properties of a work of art. By contrast, the relationship between the emergent phenomenon and its constituents in the case of fusion takes a different form. There, the emergent product appears only after the union and disappearance of the constituents.⁴ The notion that the bearers of properties can fuse, thereby giving rise to a new product with new properties has a variety of applications in physics. Nuclear fusion and the like would be the obvious example.⁵ Though nuclear fusion is a very specific case, one could generalize the idea of a fusion of properties to other contexts.

In crafting the preemption argument, philosophers of mind clearly assume a great deal about the ontology of physics. Their basic assumption is that the account of elementary particles and fields provided by the standard model supports reductionism. As it happens, physicists are deeply ambivalent about the status of reductionism in their work. We should note for example that the reductionist intuition that bigger things are nothing more than the sum of their parts all the way down is violated consistently in modern formulations of quantum field theory. From the inability of the standard model to successfully reduce hadrons to quarks and gluons to the pluralist ontology of effective field theory, modern physics is not a comfortable place for reductionists.

Arguably, the view of physics supporting the anti-emergentist position is based on a questionable conception of physical law, explanation and ontology.⁶ Once the physicalist's picture of physics is revised, then the emergence debate will look somewhat different. However, even given an updated picture of physics, there remains a major conceptual problem blocking talk of higher-level phenomena. This problem will serve as the focus for the remainder of this paper.

It is generally acknowledged that in order to defend the genuine causal power and thereby the reality of emergent phenomena requires the articulation of a meaningful sense in which they can be said to exhibit downward causation. Downward causation is a slippery notion, since there is a sense in which downward causation follows as a corollary to any ascription of a causal relation above the level of our basic physics. As we shall see below, the real difficulty is not so much in demonstrating that our causal claims commit us to downward causation — that turns out to be trivial, especially given a model of properties which strongly supervene on one their constituents. Rather, the trouble starts when we begin to consider what Jaegwon Kim has called *reflexive downward causation* (Kim 1999, 25). Kim has claimed that this kind of causal power is conceptually incoherent and if he's right, then emergence can be dismissed. While it may be possible to find a variety of apparent examples of reflexive downward causation in nature, this is not my purpose here, nor would it settle the metaphysical question entirely. Instead, my goal is to overcome the conceptual or metaphysical obstacles that Kim articulates on, more or less, their own terms.

The contemporary debate concerning the metaphysical status of higher-level phenomena treats causation and causal powers in terms of the relationship between properties. These properties are situated within a world of distinct layers or levels. It is common, for instance, to speak of higher- and lower-level properties. Ordinarily, these levels or layers are stacked in terms of their putative fundamentality. Our basic physics forms the ontological ground floor, with chemistry, biology, psychology etc., each in turn providing the higher layers. While there may be reasons for doubting its legitimacy, the layered picture of the natural world is widely accepted by both reductionists and non-reductive materialists.

Within this layered world, properties can cause instantiations of other properties in three ways. They can cause instantiations of properties at the same level, at higher levels or at lower levels. Examples of causal relations between properties at the same level are easy to imagine. Likewise for upward causation. In the case of upward causal relations, for example, we can imagine a property at a certain level, say the molecular bonds in a crystal, causing the crystal

to be transparent. Transparency is not a property of molecules, but can instead be thought of as a higher-level property of ensembles of molecules. Upward causation is central to traditional reductionist arguments concerning the status of higher-level phenomena. The heat of a gas, a reductionist will say, is caused by the kinetic action of its constituents.

Downward causation is a little trickier within the layered picture of properties. Kim (1999) describes three approaches to downward causation, supervenient (or conceptual) downward causation, reflexive downward causation, (the kind which he regards as both essential to emergentism and absurd) and a third kind, which he regards as unproblematic, stemming simply from the additive properties of ensembles. Of this third kind he writes for example:

[C]lasses in which higher-level entities and their properties prima facie causally influence lower-level entities and their properties seem legion. The celadon vase on my desk has a mass of 1 kilogram. If it is dropped out the window of my second floor office, it will crash on the paved sidewalk, causing myriads of molecules of all sorts to violently fly away in every which direction. Even before it hits the ground, it will cut a rapid downward swath, causing all sorts of disturbance among the local air molecules. . . There is no question that the vase, in virtue of having this mass, has a set of causal powers that none of its micro-constituents have. (Kim 1999, 25–6)

The possession of this new set of powers by the vase does not make the vase real, in the sense that emergentists have traditionally sought, because being 1 kilogram is not a uniquely vase-like property. It is not 1 kilogram by virtue of being a vase as opposed to something else. Additive properties are the kind of unsurprising consequences of property changes that are equivalent to simply adding constituents together or taking them away. I know that if I keep piling firewood in the driveway it will eventually add up to a ton.

While additive properties like mass and volume are relatively consistent with a reductionist framework, most of our higher-level generalizations are *prima facie* irreducible. In recent years, talk of economies, ecosystems, minds and the like has been reconciled with materialism via some form of functionalism. Supervenient conceptions

of downward causation, in particular are an attempt to make sense of the importance of higher-level causal statements without violating the causal closure principle. Advocates of supervenience handle the problem of downward causation along the following lines: If one accepts that intralevel causal relations occur at a level above that of our most basic physics, then given strong supervenience, all such descriptions, involve the basal conditions from which the caused property emerges or results. So, to accept the truth of causal statements other than those we find in our subatomic physics is to accept downward causation. To say that *A* causes *B*, then I'm saying that *A* causes the basal conditions of *B* call them *b(B)* to be realized. Of course *A* only accomplishes this via its basal conditions, *a(A)*.

The thrust of Kim's discussion of this kind of downward causation, is that all higher-level properties are what they are by virtue of their supervenience on certain lower-level properties. Only in so far as this is true can we make sense of the claim that the higher-level property *A* can be said to cause the instantiation of the lower-level property *b(B)*. Nevertheless, from an emergentist perspective, such claims of downward causal power are empty since the real agent in the event that gave rise to *b(B)* was the lower-level basal condition of *A* namely *a(A)*. Hence, on Kim's view, "higher-level properties can serve as causes in downward causal relations only if they are reducible to lower-level properties" (1999, 33).

Kim's treatment of downward causation via the supervenience relation implies that when push comes to shove, we shall deny that there really are any causal relations above the 'very bottom level'. Such fundamentalism comes at a high price, since it denies legitimacy to scientific generalizations above the level of our basic physics. On this view, not only are minds causally impotent, but biological properties have no influence on chemical events, chemical events cannot influence physical events etc. Insofar as the higher-level sciences succeed at all, then genuine scientific predictions are parasitic on the causal activity of the basic physical constituents.

In order to block any robust form of emergence, Kim tackles reflexive downward causation. This is a third kind of downward causation which Kim rules out as patently absurd. For emergentists, this paradoxical form of downward causation goes to the heart of what

distinguishes the emergent phenomenon from its constituents. Here, the higher-level phenomenon acts on its own constituents. Kim argues that this is an impossibly circular phenomenon since, if causation is transitive, it seems to lead to a kind of self-causation. Kim argues that the kind of self-causation or self-determination that is required for emergence to make sense is "an apparent absurdity" (1999, 28). His argument takes two cases where a whole and its parts are in a causal relation. He considers first, the case where the part and whole are in an instantaneous causal relationship:

Case 1:

At a certain time t , a whole, W , has emergent property M where M emerges from the following configuration of conditions: W has a complete decomposition into parts $a_1 \dots a_n$; each has Property P_i ; and relation R holds for the sequence $a_1 \dots a_n$. For some a_j , W 's having M at t causes a_j to have P_j at t . (Kim 1999, 29)

He calls this a case of "synchronic reflexive downward causation" and sees it as unacceptably circular by virtue of the assumption that for an entity to be responsible for an act, it must have had the power to perform the act prior to performing it. He writes:

we said that the whole, W causes one of its proper parts, a_j , to "have" P . If there is real downward causation, from W 's having M to a_j 's having P , this "having" must be understood as "acquiring". For if a_j already has P_j at t , what role can W 's having M at t play in causing it to have P_j at t ? Obviously none. (Kim 1999, 29)

The synchronic version of reflexive downward causation as characterized by Kim looks like an impossible kind of bootstrapping effort. Given the framework presented here, causation takes place over time and involves property changes that make "self-causing" unacceptably paradoxical. However, since most ordinary cases of downward causation do not leave the whole unchanged as it apparently acts on its parts, emergentists might consider the synchronic case excessively artificial. The synchronic case is, perhaps tailor-made for generating the kind of paradox that serves his argument. In order to block this

line of objection, he moves on to consider the diachronic case, where the relationship between parts and whole takes place over time.

Case 2:

As before W has emergent property M at t , and a_j has P_j at t . We now consider the causal effect of W 's having M at t on a_j at a later time $t + \Delta t$. Suppose, then, that W 's having M at t causes a_j to have Q at $t + \Delta t$. (Kim 1999, 29)

In the diachronic case, the vicious circularity has been removed, but at the expense of the reflexive aspect of the relationship that emergentists hope to retain. Cases of this type end up reducing easily to the kind of supervenient downward causal relationship described above. W 's having M at t causes a_j to have Q at $t + \Delta t$. But since a_j 's having Q at $t + \Delta t$ is not part of the conditions that give rise to the Property M , (since those conditions obtained at another time t) we find no problem of bootstrapping or self-causation. Since these are different basal conditions, the situation reduces to the supervenient downward causation case discussed above.

If Kim is correct, then there seems to be little chance for the kind of synchronic emergence that emergentists require. While it remains possible to demonstrate that when something emerges, it has an identifiable causal effect on the lower-level stuff from which it emerged and that this effect is more than merely an additive consequence (to use the old emergentist phrase) of the interaction of lower level constituents, this diachronic form of emergence is reducible to a significantly less exciting kind of supervenient downward causation. In the diachronic case since a_j 's having Q at $t + \Delta t$ is not part of the conditions that give rise to the Property M , and therefore it does not exhibit the problem of self-causation or self-reference. While the diachronic case will meet with few objections, the real prize is the synchronic case. Arguing for a robust form of emergentism requires that we demonstrate the significance of relations to the causal powers of a system. This, in turn must be done in such a way as to show that emergent properties have causal powers that are due to the structural relations between parts and not to the parts themselves. As Kim's analysis makes clear, the only way to do so in a way

