Philosophical Studies, 159 (2012): 21-39.

Emergent Causation

Simon Prosser
University of St Andrews

1. Introduction

The notion of an *emergent property* has aroused much recent discussion among both scientists and philosophers. Many quite different kinds of properties have been described as 'emergent', the only clear common factor being the broad idea that when a physical system of sufficient complexity is in a suitable configuration new properties 'emerge' in a way that could not have been predicted from the physical laws governing less complex or differently configured systems. Examples of putative emergent properties in the actual world can be found in biology, chemistry, in branches of physics dealing with complex systems, in computer science, and in the philosophy of mind.[1] Many of the examples given by scientists belong to an 'epistemological' variety of emergence, where the emphasis is on the impossibility of predicting the presence of emergent features of macroscopic phenomena based purely on complete knowledge of the microscopic phenomena from which the macroscopic phenomena emerge. This kind of emergence is to be distinguished from an 'ontological' variety of emergence according to which the emergent phenomena are not merely unpredictable from base-level phenomena but are in some sense ontologically distinct from them. This is typically understood in terms of a failure of strong whole-part supervenience; the emergent property of the whole is not logically entailed by the properties of the parts and their mode of combination (I give a more precise formulation below). This latter kind of emergence has tended to attract more interest from philosophers. Perhaps the most common examples of putative

[1] For details see the papers collected in Clayton and Davies 2006 and in Bedau and Humphreys 2008.

ontologically emergent properties, and the historical source of much of the debate, are properties associated with phenomenal consciousness (for a historical survey see McLaughlin 1992). Recently, however, the possibility of ontologically emergent properties has also attracted the attention of philosophers interested in other core issues in metaphysics (see for example Schaffer 2007, 2010, and McDaniel 2007, 2008). For these latter purposes it does not usually matter whether the emergent properties are mental properties; what matters is usually only that ontological emergence is logically possible.

In this paper I discuss ontological emergence in relation to two problems raised by Jaegwon Kim (1999, 2006a, 2006b) under the general heading of the *problem of downward causation*. These are generally regarded as presenting some of the most serious obstacles to the notion of ontological emergence. Following Kim, I call these problems the *causal exclusion problem* and the *causal closure problem*. The first threatens the very possibility of emergent properties; the second threatens the possibility of emergent properties in worlds in which the base level (the 'physical' world) is causally closed, which is often assumed to be the case in the actual world.

Due to its historical origins much of the debate over ontological emergence has focussed on mental properties as putative examples. I suspect that this narrow focus may have drawn attention away from some of the issues of interest to general metaphysics. In any case, in what follows I proceed by first describing two examples of emergent properties, neither of which is mental. I then describe the problems of causal exclusion and causal closure, discussing each problem in relation to the two examples. I conclude that the causal exclusion problem fails to show that emergence is logically impossible, though it does place constraints on what kinds of properties could be emergent. What one says about the causal closure problem depends on just how one defines 'causal closure', but most definitions do rule out ontological emergence in worlds with a causally closed base level. I suggest, however, that what many people have found problematic about ontological emergence is not the incompatibility with causal closure *per se* but rather a supposed incompatibility with the predictions given by the base-level laws. Using the examples I show, however, that these constraints should not be conflated; causal closure can fail, allowing ontological emergence, without any violation of the predictions given by the base-level laws. Moreover, one of the examples suggests that there cannot be empirical evidence sufficient to show that the actual physical world is causally closed in the sense required to rule out all kinds of ontological emergence. It is important to stress, however, that this paper is not an attempt to defend actual emergent properties, mental or otherwise. Although some light might be shed on that issue, the main purpose here is to investigate a more general aspect of the relation between the properties of a whole and the properties its parts.

2. What is an emergent property?

There are many definitions of 'emergence'; and several authors have distinguished various different strengths or kinds of emergence (see for example Bedau 1997, 2003, Van Gulick 2001, Chalmers 2006, McDaniel 2007). The kind of emergence to be discussed here is ontological and is best defined in terms of 'strong' synchronic supervenience (Kim 1984, 1987).[2] In particular I shall borrow the following definition from Brian McLaughlin (1997: 39), which is a modification of an earlier definition by James Van Cleve (1990):

> If P is a property of w, then P is emergent if and only if (1) P supervenes with nomological necessity, but not with logical necessity, on properties the parts of w have taken separately or in other combinations; and (2) some of the supervenience principles linking properties of the parts of w with w's having P are fundamental laws

Here a 'fundamental law' is defined such that a law is fundamental 'if and only if it is not metaphysically necessitated by any other laws, even together with initial conditions' (1997: 39). Thus the definition aims to capture the idea that there is a set of fundamental laws describing the non-emergent, 'base-level' phenomena yet when matter is combined in a certain way a new property is instantiated according to a law

[2] The emphasis on *synchronic* supervenience is relevant because some defenders of emergence argue that the problems of downward causation can be avoided by theories that appeal to *diachronic* supervenience or other, *non*-supervening relations (see for example Humphreys 1997, O'Connor 2000a, 2000b, O'Connor and Wong 2005). I am not challenging these claims; but I think synchronic emergence is a more interesting phenomenon in relation to the metaphysics of the whole-part relation.

whose existence is logically independent of the facts about the base level; the base-level laws and configuration could logically have been exactly the same without the emergent property being instantiated. The clause concerning ‘properties the parts of w have taken separately or in other combinations’ (which comes from Van Cleve’s definition) is intended to exclude putative base-level properties such as *being a part of an object that has property P*. Clearly if such properties could be included in the supervenience base then it would follow somewhat trivially that all properties logically supervened on the base-level properties. The second clause, concerning fundamental laws, is to rule out law-like correlations such as that between mass and weight from allowing a property such as weight to count as emergent. This kind of definition is generally held to best capture what the ‘British Emergentists’ (such as Mill 1843, Alexander 1920, Morgan 1923 and Broad 1925) originally meant by the word ‘emergent’.

Supervenience is often defined as a relation between classes of properties (mental properties and physical properties, for example). For present purposes, however, the relation should be understood as holding between token instantiations of properties that need not belong to different classes. This allows for claims about the supervenience of properties of a given kind instantiated at one ontological level on properties of the same kind instantiated at a lower ontological level. For example, one might claim that the mass of a composite object supervenes on the masses of its atomic parts. This is relevant because the first of our examples of emergent properties allows for a case in which the colour of a composite object does not logically supervene on the colours of its atomic parts.

3. Two kinds of emergent properties

3.1 Zeno objects

A *Zeno object* is composed from an infinite number of atomic parts in such a way that some of its surface properties, including its colour, do not logically supervene on the properties of its atomic parts. The example I shall describe is made of atoms that are *transparent* (light passes right through them) yet the object that they compose is not transparent; it is coloured red and could, logically, have been any colour whatsoever. Many other permutations are possible; the colours of the Zeno object and its atomic parts are logically independent of one another (with the one exception that a Zeno object cannot be transparent without its atomic parts being transparent).[3]

Since Zeno objects exist in far-off worlds it is necessary to clarify what is meant by 'light' and 'colour'. Light, in the worlds in question, need not be what we call 'light' in the actual world; it will suffice for our purposes that light is anything that can reflect from some objects and pass through, or be absorbed by, others. We might also allow that light can have different properties corresponding to 'wavelengths' that allow objects to reflect different wavelengths differentially; but light need not be a wave, and the word 'wavelength' is just a place-holder for whatever property light has that allows differential reflection. Because it will make the argument simpler I shall discuss a world in which light consists of point-sized atoms ('photons'), but this is inessential to the general form of the argument. I shall also speak of the 'colour' of an object. By this, I shall mean the *reflectance* of the object – the percentage of light reflected from the object at each wavelength. It is controversial whether actual colours can be equated with reflectances but for our purposes this does not matter.[4] I take it that whatever the metaphysical status of actual colour, reflectance is a property in good standing; so if necessary the argument below can be understood as an argument for emergent reflectance properties rather than emergent colours. The same applies to the use of the words 'light', 'wavelength' and so on; it does not really matter whether the properties discussed here are the actual properties referred to by those words, provided they are possible properties. Zeno objects will still provide examples of emergent properties, which is all that matters.

The Zeno object that we shall consider is as follows. The object is spherical and has an onion-like internal structure consisting of a solid central spherical part surrounded by an infinite number of concentric spherical layers. Let us suppose for simplicity that each 'layer' is an indivisible atom. Each layer is half the thickness of the layer it

[3] I describe Zeno objects in more detail in Prosser 2009. Here I give a reformulated argument for their properties.
[4] For arguments in favour of the view that actual colours are reflectances, as well as a survey of the alternatives, see Byrne and Hilbert 2003.

surrounds; and, although not essential, I shall stipulate that there are gaps between layers such that each gap is the same thickness as the layer it surrounds. The thicknesses of the layers from the centre outwards thus form an infinite ‘Zeno’ series proportional to the series 1/2, 1/4, 1/8, 1/16, … This is depicted in cross-section in figure 1. Consequently although the Zeno object is of finite diameter it has an infinite number of layers and no outermost layer.

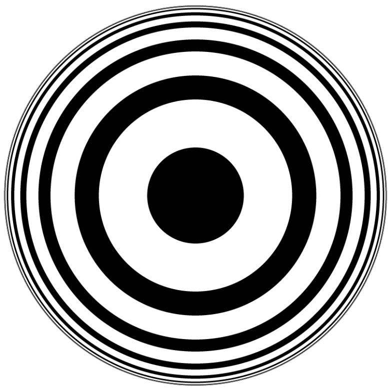


Fig. 1 A Zeno Object

The crucial property of the Zeno object is its lack of an outermost layer; any object with this property will have some surface properties that do not logically supervene on the properties of its atomic parts. To make it clear why, I shall first need to introduce some terminology. An object with a *topologically closed* surface is one that occupies an outermost layer of points. An object with a *topologically open* surface is one that occupies no outermost layer of points. For simplicity I shall stipulate that the atoms all have topologically closed surfaces; whereas it follows from the structure of the Zeno object that it has a topologically open surface.[5] I shall use the terms *overlap* and *boundary* as follows:

[5] See Prosser 2009, however, for discussion of Zeno objects composed from atoms with topologically open surfaces, or with non-atomic parts (or even parts made of infinitely divisible ‘gunk’) in place of the atoms.

Overlap: *x overlaps with y* if and only if there is a space-time point occupied by both *x* and *y*

Boundary: The *boundary*, *B*, of an object, *O*, is defined such that: (i) if *O* has a topologically closed surface then *B* is the outermost layer of points of *O*, or else (ii) if *O* has a topologically open surface then *B* is the layer of points that perfectly encloses *O* (such that there is no empty space between *O* and *B*).

We can now stipulate some of the fundamental laws that hold in the world of the Zeno object. The first law concerns what happens when light strikes an atom (which occurs when a photon reaches the boundary of the atom):

Law #1: If a photon overlaps with the boundary of an atom the photon continues in its trajectory just as if the atom were not there

The atoms are thus *transparent*. Variations on Law #1 could easily be given such as to make an atom absorb light that overlaps its boundary (such that the atom is *black*) or reflect light differentially depending on wavelength, such that the atom has a specific reflectance and therefore a specific colour.

We can also specify a fundamental law concerning what happens when light strikes (reaches the boundary of) a Zeno object:

Law #2: If a photon overlaps with the boundary of a Zeno object the photon is reflected according to reflectance property *R*

Reflectance property *R* could be, for example, the property of reflecting all *red* photons and absorbing all others (perhaps their energy is then distributed through the Zeno object; the details don't matter). Alternatively *R* could be the property of reflecting all photons regardless of wavelength, such that the Zeno object is *white*. Whatever the details, it is sufficient for present purposes that R is *some* kind of reflectance; the Zeno object, unlike its atomic parts, is not transparent and instead reflects light (it has a colour).

Law #1 and Law #2 are logically independent of each other. Moreover, either law could be swapped, without any resulting inconsistency, for various alternative laws that say slightly different things about what happens when a photon reaches a boundary (such that, for example, the atoms also have a colour, but one that may be different from that of the Zeno object). The reason is straightforward: because a Zeno object has no outermost atomic part, a photon can overlap with the boundary of the Zeno object without overlapping with the boundary of any atomic part; and likewise a photon can overlap with the boundary of any atom without thereby overlapping with the boundary of any Zeno object. In effect, laws such as Law #1 and Law #2 determine the colours (or transparency) of atoms and Zeno objects respectively; so it follows from the logical independence of these laws that the colour of a Zeno object is not logically entailed by the colour (or transparency) of its atomic parts. This can be made vivid by imagining a Zeno object with alternating red and blue layers; since there is no outermost layer, the colours of the layers fail, on their own, to determine whether the Zeno object is red, blue, or some other colour. Rather, the colour of the Zeno object nomologically (but not logically) supervenes on the properties of its atomic parts (including their spatial configuration) according to Law #2, which is a fundamental law. Thus, by the definition given above, the colour of a Zeno object is an emergent property.

I have a described a case in which the emergent property is novel; none of the atoms, or finite composites thereof, have reflectances, but when the atoms are suitably composed the resulting Zeno object has a reflectance property (colour) not possessed by any of the atoms. But it is easy to see how the laws could be altered such that the atoms could be coloured, the Zeno object also coloured, and yet the colour of the Zeno object still be an emergent property (that is why we need the broadened notion of supervenience that describes relations between token property instantiations rather than classes of properties, as mentioned above). Note that some variants of Law #2 would include reference to the colours of the atoms – there could be a law-like relation between the colours of the atoms (as well as their spatial configuration) and the colour of the Zeno object. But even when there is a law saying that the colour of the Zeno object is identical to the colour of its constituent atoms the colour of the Zeno object would still count as an emergent property because of its logical independence from the colours of its atomic parts.[6]

3.2 Entanglement

My second example of an emergent property is inspired by some well-known properties of certain quantum mechanical systems. I prefer not to discuss quantum mechanical systems themselves, however, for they bring with them many controversies of interpretation that might cloud the issues.[7]

Consider, then, a world with the following characteristics. Firstly, the world contains atoms that can be in any of three states, which I shall call *black*, *red* and *blue*. These are merely labels; for the purposes of the example it does not matter very much what kinds of properties the atoms can have. Secondly, the world contains a *threshold*, a spatial plane (which may or may not be flat) through which atoms sometimes pass. Again the precise nature of this does not matter, but for example there could be an object surrounded by some kind of field such that crossing the threshold consists in passing within a certain distance of the object.

The base-level physics of this world includes the following fundamental laws:

Law #1: All atoms that have not yet crossed the threshold are *black*

[6] The emergent properties of Zeno objects are distant relatives of the cases of asymptotic 'emergence' described by Batterman 2001; but very distant because, as Batterman notes, his cases have nothing to do with the whole-part relation and require a modified notion of 'emergence' to reflect this, whereas the cases described above depend crucially on features of the whole-part relation (specifically the fact that a Zeno object has no outermost atomic part, allowing light to reflect from the object without reflecting from any atomic part. The emergence is not explained by the fact that the Zeno object has a topologically open surface; no colour property would *emerge* from an atom with an open surface).

[7] Some authors have suggested that entangled quantum mechanical systems do possess emergent 'holistic' properties; however I take no view on this here. For details and further references see Schaffer 2007, 2010. Paul Humphreys (1997) argues that certain entangled quantum mechanical systems exemplify his 'fusion' account of emergence, but this is very different to the kind of emergence discussed here. David Chalmers (2006) suggests that interpretations of quantum mechanics according to which the wave function 'collapses' may involve a kind of downward causation of the kind relevant to emergence, but again this is unrelated to the 'entanglement' property discussed here.

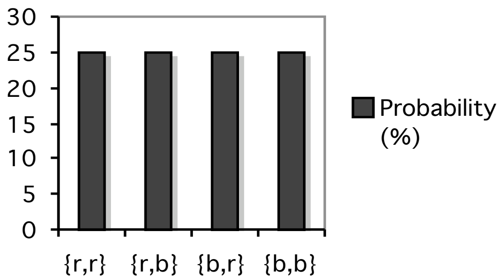
Law #2: When a *black* atom crosses the threshold it turns either *red* or *blue*. This is determined by nothing other than probability, with a probability of 50% for each property.

We need not worry about other laws concerning the behaviour of individual atoms; there are many ways to fill in the details of the general mechanics of the world consistent with what matters here. If one were to follow the career of an individual atom, these laws along with Law #1 and Law #2 would give the most complete (albeit indeterministic) account that can be given of its individual behaviour. In addition to Law #1 and Law #2, however, a further fundamental law holds:

Law #3: When two *black* atoms pass within distance *d* of each other they become *entangled*. The effect of this is that if the two atoms cross the threshold simultaneously one of the pair will turn *red* while the other will turn *blue*; they will not turn the same colour. This is determined by nothing other than probability, and the probability of a {*blue*, *red*} outcome and the probability of a {*red*, *blue*} outcome are each 50%. Once an atom is entangled it does not entangle with any further atoms. If more than two unentangled *black* atoms pass within distance *d* of each other simultaneously they do not entangle.

Law #3 is consistent with Law #1 and Law #2 but is not entailed by them. Even if an atom is entangled when it crosses the threshold its *red*-*blue* probability distribution remains flat, with a probability distribution of 50% for each outcome, as specified by Law #2.

For pairs of atoms there are four possible outcomes, which we can symbolise by the ordered pairs {r,r}, {r,b}, {b,r} and {b,b}. In a world in which Law #3 does not apply (or if we restrict attention to unentangled atoms) the probability distribution across these pairs would be flat, with a probability of 25% for each outcome, as depicted in figure 2. But the probability distribution across pairs of entangled atoms is quite different; the outcomes {r,b} and {b,r} each have a probability of 50% while the outcomes {r,r} and {b,b} each have a probability of 0, as depicted in figure 3. Where there is a shift in a probability distribution, there is a causal power at work; the property of entanglement causes a difference, albeit only probabilistically, to the subsequent distribution of properties at the base level. One can of course imagine far more complex varieties of entanglement involving larger numbers of atoms and with far more complex effects on probability distributions over n-tuples of outcomes while making no difference to probability distributions for individual atoms.



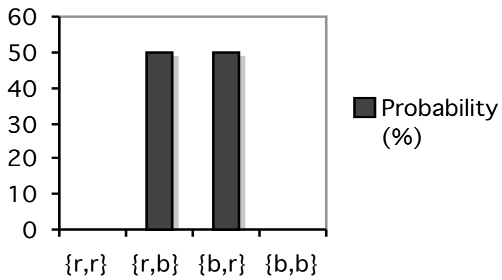


Fig. 2 Pair-wise probabilities for unentangled pairs **Fig. 3** Pair-wise probabilities for entangled pairs

If we think of Law #1 and Law #2 (plus the general mechanics of individual atoms) as constituting the base-level laws then it is clear that *entanglement* is an emergent property. For, since Law #1 and Law #2 do not entail Law #3, Law #3 is a fundamental law; and instantiations of the property of entanglement thus nomologically (but not logically) supervene on the base-level facts in the appropriate way. (We can suppose that the supervenience is synchronic just in the sense that when two atoms pass within distance *d* of each other they are entangled from that moment on; though of course at later times the supervenience base will have to incorporate facts about the history of the atoms.)

4. Two problems of downward causation

Much of the interest in emergent properties lies in the idea that new properties emerge from a complex system in such a way as to affect the subsequent behaviour of the system itself or of adjacent entities. In that case emergent properties must have causal powers to affect the base level – *downward causation*, as it has come to be known.[8] It has been claimed that downward causation raises at least two serious problems for emergent properties. In this section, however, I shall suggest that these problems are not as troublesome for the general notion of emergence as is often supposed (though they might well raise problems for some specific candidates for emergence).

4.1 The causal exclusion problem

Suppose an instantiation of an emergent property, *M*, causes a base-level effect *P**. The first problem, sometimes known as the *causal exclusion problem*, is an objection to the logical possibility of emergent properties (i.e. it is not merely an objection to actual emergence). It is described by Kim as follows:

> Now we face a critical question: if an emergent, *M*, emerges from basal condition *P*, why cannot *P* displace *M* as a cause of any putative effect of *M*? Why cannot *P* do all the work in explaining why any alleged effect of *M* occurred? 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**. It follows that *P* is nomologically sufficient for *P** and hence qualifies as its cause (2006b: 558).

The problem can be illustrated by the following example. Suppose that there is a world containing cube-shaped atoms that are normally transparent. Sometimes groups of these atoms compose larger cube-shaped objects (which I shall call *composite cubes*) just by becoming spatially configured as a cube, with no gaps between the atoms. Suppose the following law holds:

Law #C: Whenever a photon strikes a composite cube (i.e. overlaps with its boundary) the photon is reflected according to reflectance property *R*.

---

[8] One might add that it is hard to make sense of a property with no associated causal powers (indeed Shoemaker 1980 argues that properties are to be *identified* with causal powers, though we shall not need this assumption). But, as Kim (1999, 2006a, 2006b) points out, instantiations of emergent properties cannot have effects only at the emergent level because this would come into conflict with the nomological supervenience of the caused instantiations of emergent properties on the base level.

Suppose that *R* is such that the composite cube will reflect red light; so the composite cube is red (given the equation of 'colours' with reflectances, as described above). This is true despite the fact that the atoms from which the composite cube is composed are transparent when they are not part of a composite cube. Now, there are two quite different ways in which we might describe the metaphysics of that world:

Description #1: When atoms compose a composite cube the composite cube has an emergent property (redness) in accordance with fundamental Law #C; the causal power of reflecting red light belongs to the composite cube as a whole; the causal powers of the individual atoms that compose it remain unchanged.

Description #2: When atoms compose a composite cube the individual atoms exhibit different causal powers from those that they exhibit individually or in other combinations; instead of being transparent they gain reflectance property *R* (i.e. they change colour). Law #C is thus part of the base-level physics; the reflectance property *R* is not an emergent property.

In the case described I think it is clear that we should favour description #2, but it is instructive to consider why. Necessarily, when a photon overlaps with the boundary of a composite cube it thereby overlaps with the boundary of at least one of the atoms. Thus, intuitively, light reflects from the composite cube by reflecting from one (or more) of the atoms that compose it. It therefore seems hard to accept, or even make sense of, the claim that the atoms retain their individual causal powers (i.e. they remain transparent) despite the composite cube gaining a reflectance property. Far better just to say that the atoms change colour when put together in a certain way (and that the colour of the composite cube *logically* supervenes on the properties, including colours, of its atomic parts).

This line of thought is captured by Kim's objection. Rather than hold that a property emerges in a law-like way from the composition of the atoms, and then produces its own effects in a law-like way, we can instead complicate the base-level laws to say that the atoms produce certain effects in a law-like way when composed thus-and-so. In many cases, like the case of the composite cube, this is correct.

4.11 Zeno objects and the causal exclusion problem

The equivalent of description #2 would be far less compelling for Zeno objects, however. There is, at least intuitively, a significant difference between composite cubes and Zeno objects in this respect. The reason for the difference is clear: when light strikes a Zeno object it does not thereby strike any atomic part of the Zeno object. This allows for the logical independence of the laws describing the reflectance properties of the Zeno object and the reflectance properties of its atomic parts, as described above. It seems clear that all of the atoms composing the Zeno object can remain transparent even though the Zeno object is not. This is illustrated by the fact that any photons in the gaps between the atomic layers of the Zeno object pass straight through the atoms in accordance with Law #1 of the Zeno-world, even though photons striking the Zeno object from outside are reflected. Thus, in the case of a Zeno object, it is clear how the atoms can retain exactly the same individual causal powers that they possess when not part of a Zeno object, while the Zeno object itself has a quite different, emergent causal power.

I am not suggesting that it is *impossible* to construe the Zeno object along analogous lines to description #2. I acknowledge that Kim's argument shows that it is always possible to come up with a way of re-describing emergent causal powers in terms of complex, discontinuous causal powers acquired by the atoms just when they compose an object in a certain way. For the Zeno object the atoms would have to work in tandem to repel photons that reached the boundary of the Zeno object while retaining different causal powers with respect to photons at their own boundaries. But just because it is *possible* to describe the situation in this way, it does not follow that it is *correct* to do so. Zeno objects remove one major motivation for keeping everything at the base level by showing how something can causally interact with a composite object without thereby causally interacting with any of its atomic parts.

More generally, the question of emergence (as in the Zeno object) versus changes of base-level properties (as in the composite cube) should be decided on the grounds of theoretical elegance, by achieving the best combination of simplicity and strength.[9] Different theories, positing different properties, are needed to give the best account of each case. To insist that the purely base-level description must always be employed whenever available seems to presuppose a principle of parsimony to the effect that one must not posit emergent properties if they could be avoided, even if doing so would result in an overall theory that is more complex and discontinuous. But it seems to me, by contrast, that in the case of Zeno objects the correct overall principles of theory choice clearly weigh in favour of the description in terms of emergent properties.

This is further illustrated by worlds in which both the atoms and the Zeno objects can have the same range of colours. There are, for example, worlds containing red atoms as well as red Zeno objects that are composed from blue atoms. A satisfactory theoretical description of this world must already ascribe the property of redness to some of the atoms; so why not ascribe that very same property to the red Zeno object? (Indeed on some views of properties the fact that the red Zeno object and the red atom share a causal power in the same world would be sufficient reason to ascribe the same property to both). The alternative description of a Zeno object, avoiding emergence by ascribing complex causal powers to its blue atoms – a description that would presumably still ascribe blueness to the atoms but would not ascribe redness to anything – would be cumbersome by comparison. But once we acknowledge that redness can be a property of a Zeno object in *some* worlds then there seems little obstacle to ascribing it to Zeno objects even in worlds in which it is never a property of an atom.

4.12 Entanglement and the causal exclusion problem

---

[9] Timothy O'Connor (1994, 2000a) has made similar claims about the relevance of principles of theory choice in relation to competing base-level and emergence descriptions.

As with the case of Zeno objects, there is bound to be a way to reconstrue entanglement in terms of base-level properties as per Kim's objection. The obvious way to do this would be to reinterpret the notion of entanglement in terms of relational properties holding at the base level. Thus, instead of saying that the system comprising atoms *A* and *B* is entangled we would say something along the lines that *A* has the property of being opposite-coloured from *B* and *B* has the property of being opposite-coloured from *A* (where 'being opposite-coloured from' means that if the two atoms cross the threshold simultaneously they will turn different colours from one another).

For simple scenarios in which only two atoms can become entangled this does not seem entirely implausible. But matters look different if we complicate the example to allow arbitrarily large numbers of atoms to become entangled. Suppose for example that we replace Law #3, above, with the following:

Law #3*: When a group of *n black* atoms becomes located such that every atom in the group is at a distance *d* from at least one other atom in the group, the group becomes *entangled*. The effect of this is that if the atoms all cross the threshold simultaneously half will turn *red* while the other half will turn *blue* (if *n* is an odd number then there will be one extra *red* or *blue* atom with a probability of 50%). The post-threshold colours of the individual atoms are determined by nothing other than probability, and the *blue-red* probability for each atom remains 50%. Once an atom is entangled it does not entangle with any further atoms.

To make the property of entanglement do a little more theoretical work, suppose further that entangled groups can interact with one another:

Contact: A group of atoms *G1* is in contact with a group of atoms *G2* if and only if at least one atom from *G1* is within a distance *d* of at least one atom from *G2*

Law #4: When entangled groups make contact they disentangle.

Law #5: Once disentangled, an atom is immune from further entanglement.

Notice how simply Law #4 can be stated, and that although groups of any size can be entangled their interactions are still described by Law #4. Now try writing down the laws that would be needed in order to describe the same world in terms of base-level relations between individual atoms rather than in terms of the emergent property of entanglement. This would require positing arbitrarily many *n*-adic relational properties holding between atoms, and the interactions between them would be extremely complex. I conclude that in such a case positing an emergent property gives by far the best combination of theoretical simplicity and strength. The causal exclusion argument is thus overridden.

Some philosophers, faced with the causal exclusion problem, have accepted that emergence as defined above is not possible and have chosen instead to modify the definition of emergence. One way to do this, for example, is to define emergence such that 'wholes (systems) exhibit features, patterns or regularities that cannot be represented (or understood) using the theoretical and representational resources adequate for describing and understanding the features and regularities of their parts' (Van Gulick 2001: 20). This does capture a *kind* of emergence that avoids the causal exclusion problem because the 'emergence' consists in the existence of base-level laws (or clauses in base-level laws) that only have relevance to certain kinds of complex object. But, as Van Gulick notes, this is an epistemological variety of emergence; it is a very different phenomenon from the ontological emergence defended above. In particular it does not involve a breakdown of whole-part supervenience, and is thus of far less interest in relation to metaphysics.

4.2 The causal closure problem

The causal closure problem (e.g. Kim 2006a) is closely related to well-known difficulties in accounting for mental causation within interactionist-dualist or non-reductive physicalist theories of the mind. Put simply, the claim is that if the base level (typically taken to be the physical world) is causally closed then there is no room for the causation of base-level events by anything except other base-level events, and therefore base-level events cannot be caused by instantiations of properties that are emergent relative to that base level. One response would be to accept that emergent properties are impossible in worlds in which the base level is causally closed but to hold that they are nonetheless possible in other worlds. On the assumption that actual physics is causally closed this would rule out emergent properties from the actual world, but their metaphysical possibility might nonetheless still be of interest to debates in metaphysics.

Although my main concern in this paper has been to defend the logical possibility of emergence, I do think that something slightly more interesting can be said about the relation between emergence and causal closure. Much depends on exactly what one means by 'causal closure' and the reasons one might have for believing that the actual world is causally closed. For example, one sometimes sees causal closure defined as the claim that every base-level event has only base-level events among its causes.[10] Thus defined it is indeed *trivially* true that if the base level is causally closed then there is no downward causation, because instantiations of emergent properties are not base-level events. So, of course, if that is what is meant by 'causal closure' the emergentist must immediately surrender. Similarly, causal closure is sometimes defined such that the probability of any base-level event is determined only by the previous base-level events and the base-level laws. This has much the same consequence as the previous definition. The advocate of actual emergence must therefore reject causal closure if it is defined in either of these ways. But there seems to be no good reason to insist that actual physics is committed to causal closure thus defined; to do so would in effect build into the commitments of physics the claim that there are no emergent properties, which seems somewhat question-begging.

Another version of the causal closure principle is as follows: 'If a physical event has a cause at *t*, then it has a physical cause at *t*' (Kim 2005: 15). Kim also gives an

[10] Just to avert a possible terminological confusion: one frequently sees the causal closure of the *physical* world defined in a corresponding way. But if we were to equate the physical with the base level in the actual world, and if quantum entanglement is an emergent property, then we should have to say that entanglement is not a physical property. But this sounds odd, given that quantum entanglement is part of a theory that belongs to what we normally call 'physics'. To help avoid confusion I shall tend to refer to the base level rather than the physical in this section.

explanatory analogue: 'If a physical event has a causal explanation (in terms of an event occurring at *t*), it has a physical causal explanation (in terms of a physical event at *t*)' (Kim 2005: 16). These principles suggest that there are no explanatory gaps at the base-level; nothing happens that stands in need of further explanation through the positing of additional causes from outside the base level. Once again if the base level is causally closed in either of the above senses then it is hard to see what room there could be for downward causation by an emergent property (I am assuming that causal over-determination is not an appealing option). And these principles seem less question-begging than those stated above; they do not directly rule out emergent causes, but instead render them redundant.

I shall suggest in what follows that although many people have found closure principles of this latter kind plausible they go beyond what physicists strictly have evidence for. Insofar as one formulates, and gathers evidence for, laws describing the behaviour of individual atoms, there are certain ways in which one can fail to notice failures of causal closure because these failures do not result in any violations of the predictions made using the base-level laws. Indeed I suspect that the idea that downward emergent causation would inevitably result in observable violations of base-level laws is what has really been behind suspicion of emergent properties. Using my two examples of emergent properties, however, I shall explain that this need not be the case.

Contrary to first appearances there are in fact many ways for an emergent property to cause base-level events without those events conflicting with the predictions of the base-level laws. In order to see why, we must distinguish two ways in which a set of laws can be indeterministic. By 'initial conditions' in what follows I mean the state of the world (excluding the laws) at some particular time (not necessarily the first moment in time):

Deterministic laws:	A set of laws *L* is deterministic if and only if, when conjoined with the initial conditions and the assumption that *L* includes all of the laws, *L* entails every future state of affairs

Weakly indeterministic laws: A set of laws *L* is weakly indeterministic if and only if, when conjoined with the initial conditions (but *not* with the assumption that *L* includes all of the laws), *L* entails a probability for every possible future state of affairs (where some of those probabilities are strictly between 0 and 1)

Strongly indeterministic laws: A set of laws *L* is strongly indeterministic if and only if, when conjoined with the initial conditions (but *not* with the assumption that *L* includes all of the laws), there are possible future states of affairs consistent with *L* but for which *L* does not entail a probability

It is crucial to notice that these definitions concern sets of laws, not worlds – it would be a mistake to talk about weakly or strongly indeterministic worlds, for example, for a subset of the laws that hold in a given world might be strongly indeterministic even though the totality of the laws in that world might be weakly indeterministic. Related to this, it is also crucial to notice that for the two varieties of indeterministic laws no assumption is made that *L* includes *all* the laws of the world in question. Once the latter assumption is added, if *L* fails to determine a probability distribution over some possible future states of affairs then probability theory may be used to assign the remaining probabilities. In other words, if the probability of some future event is not determined by *L* plus the initial conditions, and if *L* includes all of the laws, then whatever remains undetermined may be assumed to be random and may therefore be assigned a probability on that basis. But, without supplementation by an assumption of randomness, a set of laws conjoined with the initial conditions may in some cases fail to determine a probability for some possible future state of affairs. Note, for example, the difference between determining a 'flat' probability distribution (where all outcomes are equally probable) and failing to determine a probability distribution at all. Where no probability distribution is determined by the laws, no frequency distribution can be regarded as violating the laws. This kind of difference is captured by the distinction between weakly and strongly indeterministic sets of laws.

Thus if a set of laws is strongly indeterministic then there is room for a further law to be added, determining a probability distribution over otherwise undetermined states of affairs without thereby conflicting with any predictions made using the existing laws. Strongly indeterministic base-level laws thus allow emergent properties to fill the nomological gaps. In a perfectly deterministic world downward causation would not be possible because there are no such gaps; but elsewhere strongly indeterministic base-level laws are commonplace, as I shall illustrate using my two examples of emergent properties.

4.22 Zeno objects and the base-level laws

In a world containing Zeno objects, let *L* be a set of laws concerning the behaviour of all non-Zeno objects; so *L* includes Law #1 (concerning the reflection of light from atoms) but not Law #2 (concerning the emergent reflection of light from Zeno objects). We can suppose that if scientists in the Zeno-world were to study the behaviour of all matter *other* than Zeno objects – and even the individual atoms that compose Zeno objects - they would find that *L* gave a fully satisfactory base-level physics. *L* could even be fully deterministic with respect to all non-Zeno matter. But *L* does not thereby tell us what happens when light strikes a Zeno object; and *L* is thus strongly indeterministic. This leaves a nomological gap, which is filled by Law #2. Hence there can be downward causation without any violation of the predictions given by the conjunction of the initial conditions with *L*; the emergent reflectance property of the Zeno object makes a difference at the base-level just where the base-level laws fail to make a prediction. If an atom is struck by light reflected from a Zeno object, and this makes a difference to the subsequent behaviour of the atom (e.g. in a world in which atoms are not transparent), *L* will not have made any prediction that could be violated by this.

One notable feature of the Zeno-world is that if one follows the career of an individual photon there will be a point (when it strikes a Zeno object) at which the base-level laws, on their own, fail to make a prediction. If there were no further laws the subsequent behaviour of the photon would be random, at least within the range consistent with base-level laws. In some such worlds Zeno objects might for example turn out to be grey because light would be reflected at random wavelengths. At any rate, a kind of 'decision point' is reached in the career of the photon that strikes a Zeno object, wherein the subsequent behaviour of the photon depends on which further fundamental law, if any, obtains. Consequently if the physicists in the Zeno-world thought carefully enough about the ways in which the atoms could become arranged they would notice that there was a case in which the base-level laws failed to make a prediction and therefore were, in that sense, incomplete. This incompleteness is associated with a violation of causal closure by cases of downward emergent causation in such worlds. As we shall now see, however, the entanglement-world does not share this feature.

4.23 Entanglement and the base-level laws

In the entanglement-world, let *L* be the set of laws concerning the behaviour of individual atoms; so *L* includes Law #1 and Law #2 (concerning what happens to individual atoms crossing the threshold) but not Law #3 (concerning the entanglement of pairs of atoms). *L* thus comprises the set of base-level laws. It is clear that *L* is strongly indeterministic because without the additional assumption that there are no further laws it makes no non-trivial predictions about the behaviour of multiplicities of atoms. That is why Law #3 can be added to such a world; it fills one of the nomological gaps left by *L* (other gaps include aspects of the behaviour of larger groups of atoms). It is clear that if we follow the career of any individual entangled atom there will be no violation of the predictions made by *L*; the *red-blue* probability distribution on crossing the threshold remains 50%-50%. Moreover there are no situations in which *L* fails to make a prediction about the individual atoms. Yet it is also clear that entanglement shifts a probability distribution concerning the behaviour of atoms – the overall configuration of the coloured atoms is likely to be different when there is entanglement – and its effects therefore constitute downward causation. Hence there is no conflict between downward causation by the emergent property of entanglement and the base-level laws.

The kind of downward causation just described violates causal closure in all of the senses defined above (assuming, at any rate, that the base-level events to be explained include distributions of properties across multiplicities of atoms). Yet it involves no violation of the base-level laws. Consequently a failure of causal closure does not entail a violation of base-level laws. A question thus arises as to which principle we have reason to accept: causal closure, or the non-violation of base-level laws. Now, empirical evidence may lead us to accept a set of base-level laws. Further empirical evidence might lead us to discover that pairs of atoms can become entangled. But since there is in principle no limit to the number of atoms that could jointly instantiate an entanglement-style property it is impossible in principle to ever establish through any finite amount of empirical investigation that there are no entanglement properties whatsoever (provided the base-level laws are not deterministic). Consequently, unlike the base-level laws, causal closure can never be empirically established in full generality by finite means. Of course, all finite investigation underdetermines even the base-level laws. But let *n*-entanglement be any entanglement-style property involving *n* atoms being entangled. Then for finite beings there will always be values of *n* for which no investigation whatsoever has been carried out into the existence or non-existence of *n*-entanglement properties. At best, if no entanglement-style properties were discovered at lower values of *n* then one might infer, by a kind of induction, that there were no such properties for higher values of *n*. But it would remain questionable whether one would be justified in saying that causal closure had been established by empirical means, and that emergence of all kinds could therefore be ruled out (which is not to say that one would thereby have any reason to *accept* actual emergence).

5. Concluding remarks

Although I have defended the possibility of emergent properties under certain conditions I have not claimed that *every* putative emergent property is possible. So it is worth briefly summarising the constraints that are placed on emergent properties by the above considerations. Firstly, by considering the causal exclusion problem, I have suggested that in deciding whether to regard a system as instantiating an emergent property, or instead merely complicating the laws at the base level, we should appeal to what makes for the best theoretical account by balancing simplicity and strength. The causal exclusion argument, by contrast, seems to implicitly assume that emergent properties should be eliminated wherever possible, regardless of how much this complicates the base-level laws. It was shown how these considerations favour emergent properties for both the Zeno-world and the entanglement-world; in the former case by showing how it is possible to causally interact with a composite object without thereby interacting with any of its atomic parts, thus allowing the atomic parts to retain their individual causal powers.

Secondly, in relation to the causal closure problem, I suggested that we need to distinguish violations of causal closure from violations of the base-level laws; the former does not entail the latter. Emergent causation will always violate causal closure, but can nonetheless be consistent with the predictions of the base-level laws if those base-level laws are strongly indeterministic. Consequently it would be possible for physicists in a given world to constantly confirm the base-level laws while remaining oblivious to the presence of emergent causation and the associated failure of causal closure. Emergence can always be detected empirically, but one has to look at the right states of affairs (e.g. those involving multiplicities of atoms); if one looks elsewhere (e.g. at individual atoms) one can fail to see it.

Finally, although it has not been the focus here, what are the prospects for emergent properties in the actual world, on the assumption that the predictions of current physics are always correct? I do not claim that the two kinds of example that have been discussed exhaust the possibilities. But with regard to those two kinds of example we can say the following: on the face of it, current physics makes it hard to see how there could be actual Zeno objects. Emergence of the general kind exemplified by entanglement therefore seems the more promising candidate. The issues of interpretation surrounding current theories of physics are of course highly contentious, so perhaps some careful attention to the details is needed before claiming that current physics is strongly, rather than weakly indeterministic. But there does seem to be at least some prima facie plausibility to this claim; if we do not assume that current physics includes all of the actual laws then on the face of it there are numerous ways to add laws concerning multiplicities of particles without thereby coming into conflict with current physics.[11] Whether we have reason to believe that there are actually such properties is another matter; and whether mental properties are among them is another matter again. But, at any rate, if one were determined to look for actual emergent properties then entanglement-style properties might be the best place to look.

Acknowledgements I would like to thank Ralf Bader, Derek Ball, Katherine Hawley, David Landsberg and other members of the St Andrews metaphysics reading group for very helpful discussions on a draft of this paper. Thanks also to Serazer Pekerman for help with figure 1.

References

Alexander, S. (1920). *Space, time, and deity*. 2 vols. London: Macmillan.

Batterman, R. W. (2001). *The devil in the details*. Oxford: Oxford University Press.

Bedau, M. (1997). Weak emergence. *Philosophical Perspectives*, 11, 375-99.

Bedau, M. (2003). Downward causation and autonomy in weak emergence. *Principia Revista Internacional de Epistemologica*, 6, 5-50. Reprinted in Bedau and Humphreys 2008.

Bedau, M. A. and Humphreys, P. (eds.) (2008). *Emergence: Contemporary readings in science and philosophy*. Cambridge, MA: MIT Press.

Broad, C.D. (1925). *The mind and its place in nature*. London: Routledge & Kegan Paul.

Byrne, A. and Hilbert, D. (2003). Color realism and color science. *Behavioural and Brain Sciences*, 26, 3-21.

Chalmers, D. J. (2006). Strong and weak emergence. In Clayton and Davies 2006, 244-254.

Clayton, P. and Davies, P. (eds.) (2006). *The re-emergence of emergence*. Oxford: Oxford University Press.

Humphreys, P. (1997). How properties emerge. *Philosophy of Science*, 64, 1-17.

[11] To give just one possible example: Current physics says that the time at which an unstable atomic nucleus decays is purely a matter of probability. But perhaps a multiplicity of atomic nuclei could become ‘entangled’ such that they will decay simultaneously, without thereby affecting the probability that any individual nucleus will decay during a given interval of time.

Kim, J. (1984). Concepts of supervenience. Reprinted in Kim's *Supervenience and mind: Selected philosophical essays*. Cambridge: Cambridge University Press. 53-78.

Kim, J. (1987). 'Strong' and 'global' supervenience revisited. Reprinted in Kim's *Supervenience and mind: Selected philosophical essays*. Cambridge: Cambridge University Press. 79-91.

Kim, J. (1999). Making sense of emergence. *Philosophical Studies*, 95, 3-36.

Kim, J. (2005). *Physicalism, or something near enough*. Princeton, NJ: Princeton University Press.

Kim, J. (2006a). Being realistic about emergence. In Clayton and Davies 2006, 189-202.

Kim, J. (2006b). Emergence: Core ideas and issues. *Synthese*, 151, 547–559.

McDaniel, K. (2007). Brutal simples. *Oxford Studies in Metaphysics, vol. 3*. Oxford: Oxford University Press.

McDaniel, K. (2008). Against composition as identity. *Analysis* 68, 128–133.

McLaughlin, B. P. (1992). The rise and fall of British Emergentism. In A. Beckerman, H. Flohr and J. Kim (eds.) *Emergence or reduction? Essays on the prospects of nonreductive physicalism*. Berlin: Walter de Gruyter, 49-93. Reprinted in Bedau and Humphreys 2008.

McLaughlin, B. P. (1997). Emergence and supervenience. *Intellectica*, 25, 25-43. Reprinted in Bedau and Humphreys 2008.

Mill, J. S. (1843). *A system of logic*. London: Longmans, Green, Reader and Dyer.

O'Connor, T. (1994). Emergent properties. *American Philosophical Quarterly*, 31, 91-104.

O'Connor, T. (2000a). *Persons and causes*. Oxford: Oxford University Press. Chapter 6.

O'Connor, T. (2000b). Causality, mind and free will. *Philosophical Perspectives*, 14, 105-117.

O'Connor, T., and Wong, H. Y., (2005). The metaphysics of emergence. *Noûs*, 39, 658-678.

Prosser, S. (2009). Zeno objects and supervenience. *Analysis*, 69, 18-26.

Morgan, C. L. (1923). *Emergent evolution*. London: Williams and Norgate.

Schaffer, J. (2007). From nihilism to monism. *Australasian Journal of Philosophy*, 85, 175-191.

Schaffer, J. (2010). Monism: The priority of the whole. *Philosophical Review*, 119, 31-76.

Shoemaker, S. (1980). Causality and properties. In P. V. Inwagen (ed.), *Time and cause*. Dordrecht, Holland: Reidel Publishing Co.

Van Cleve, J. (1990). Mind-dust or magic? Panpsychism versus emergentism. In J. E. Tomberlin (ed.) *Philosophical Perspectives volume 4*. Atascadero, CA : Ridgeview. 215-226.

Van Gulick, R. (2001). Reduction, emergence and other recent options on the mind-body problem: a philosophic overview. *Journal of Consciousness Studies, 8*, 1–34.