

Jaegwon KIM*

Explanation, Prediction, and Reduction in Emergentism

The core doctrine of emergentism is that among the properties of a complex system are those that are neither *predictable* nor *explainable* in terms of the properties of the lower-level constituents or subsystems from which they emerge. It is also claimed that these emergent properties are nor *reducible* to these more basic properties. This paper presents a new model of reduction, i.e the functional model (opposed to the classic Nagel model), which gives a unified account of these emergentist claims.

Key words: Emergence, Reduction, Explanation, Prediction, the Mind-Body Problem.

Explication, prévision et réduction dans la doctrine émergentiste :
Selon la doctrine centrale de l'émergentisme, certaines propriétés d'un système complexe ne sont ni *prévisibles* ni *explicables* en termes de propriétés des composants d'un niveau inférieur. Toujours selon cette doctrine, on considère également que des propriétés émergentes ne sont pas *réductibles* aux propriétés plus élémentaires du niveau inférieur à partir duquel l'émergence a lieu. Cet article présente un nouveau modèle de la réduction, à savoir un modèle fonctionnel qui s'oppose au modèle classique de Nagel, et qui permet une formulation unifiée de cette doctrine émergentiste.

Mots clés : Emergence, Réduction, Explication, Prévision, Problème Corps-Esprit.

The concepts of explanation, prediction, and reduction figure prominently at several crucial junctures in the development of the doctrine of emergence. Most importantly, the concept of explanation is invoked in the claim that emergent phenomena or properties, unlike those that are merely "resultant", are *not explainable*, or *reductively*

* Department of Philosophy, Brown University, Providence RI 02912 - USA
E-mail : Jaegwon_Kim@brown.edu

explainable, on the basis of their "basal conditions", the lower-level conditions out of which they emerge. This is frequently coupled with the claim that emergent phenomena are *not predictable* from their emergence base. Some emergentists seem to have taken the two claims to be equivalent, or at least as forming a single package.

Further, emergent properties are claimed to be causally, and hence explanatorily, efficacious properties. This is a fundamental tenet of emergentism, both in the classic emergentism of Samuel Alexander, Lloyd Morgan, and others but also in its modern version defended by Roger W. Sperry and others. Emergentists often contrast their position with epiphenomenalism, dismissing the latter with open scorn. On their view, emergents have causal/explanatory powers in their own right, introducing novel, and hitherto unknown, causal structures into the world. In particular, most emergentists believed in what has now come to be called "downward causation" — causal influence from "higher" to "lower" levels. Thus, emergent phenomena can cause, and hence explain, not only other emergent phenomena and those that may emerge from them, but also phenomena at their basal level, namely events at the level from which they emerge.

In this paper I want to describe a model of reduction in terms of which we can make a clear sense of the emergentist claims about explainability and predictability as regards emergent properties. The ideas that motivate this model of reduction are not new; however, it will become clear, I hope, that an explicit articulation of them as an alternative model of reduction — an alternative to the long-reigning Nagel model of intertheoretic reduction — can illuminate a number of issues involving emergentism, reductionism, and the interlevel relationship of properties.

I

Let us assume that any object has a unique complete microstructural description: that is, any physical system can be exhaustively described in terms of (i) the basic particles that constitute it (this assumes atomism, which most classic emergentists accepted); (ii) all the intrinsic physical properties of these particles; and (iii) the relations that configure these particles into a structure (with "substantial unity", as some emergentists would have said). Such a description will give us the total "relatedness" of basal constituents that the emergentists often spoke of; it also gives us what we may call *the total micro-based property* (or total microstructural property) of the system — a macro-property (macro since it belongs to the system as a whole) that is

constituted by the system's basic micro-constituents, their properties, and the structuring relations that configure them into a system with unity and stability as a substance.

I would expect most emergentists to accept the following thesis of mereological supervenience:

[Mereological supervenience] Systems that have an identical total micro-based property have all other properties in common. Equivalently, all properties of a physical system supervene on, or are determined by, its total micro-based property.

It is a central tenet of classic emergentism that among these properties supervenient on, or determined by, a system's total micro-based property, some have the special characteristic of being "emergent", while the rest are only "resultant". What is the basis for this distinction? Lloyd Morgan says this:

"The concept of emergence was dealt with (to go no further back) by J.S. Mill ... The word 'emergent', as contrasted with 'resultant', was suggested by G.H. Lewes ... Both adduce examples from chemistry and from physiology; both deal with properties; both distinguish those properties (a) which are additive and subtractive only, and predictable, from those (b) which are new and unpredictable."¹

I don't think we need to focus on "additivity" and "subtractability"; I believe these terms were used to indicate the simplicity of logical/mathematical operations involved in the predictive calculation of the properties involved. Predictability is the key idea. In any case, resultant properties are predictable from a system's total micro-based property, but emergent properties are not. Morgan's (b) above introduces the idea of "newness", or "novelty", an idea often invoked by the emergentists. Is he using "new" and "unpredictable" here as expressing more or less the same idea, or is he implying, or at least hinting, that emergent properties are unpredictable because they are new and novel properties? I think these are interesting questions, but let us set aside this issue for now.

In speaking of predictability, it is important to distinguish between *inductive predictability* and *theoretical predictability*, a distinction that the emergentists were clearly aware of. Even emergent properties are inductively predictable: Having observed that an emergent property, E, emerges whenever a microstructural property M is present in a system, we may predict that another system will instantiate E on the

¹ C. Lloyd Morgan, *Emergent Evolution* (London: Williams and Norgate, 1923), pp. 2-3.

basis of the knowledge or belief that it instantiates, or will instantiate, M.² What is being denied is the theoretical predictability of E on the basis of M: we may know all that can be known about M and the laws that govern the properties and relations constitutive of M, and yet this knowledge does not suffice for the prediction of E. This unpredictability may be the result of our not having even the concept of E, this concept lying entirely outside the concepts in which our theory of M is couched. In cases where E is a phenomenal property of experiences (a "quale"), we may have no idea what E is like before we experience it.³ But this isn't the only kind of barrier to predictability. It may well be that we know what E is like — we have experienced E before — but we may be powerless to predict whether or not E — or whether E rather than another emergent E* — will emerge when a complex is formed with a novel microstructure M* that is similar to M in certain significant respects. This means that the emergence law "Whenever M is instantiated by a system, it instantiates E" must be taken as a primitive, stating a brute correlation between M and E. Inductive predictions of emergents become possible when we are allowed to assume such brute emergence laws as part of our evidence base.

The occurrences of conscious states can be inductively predicted in the sense explained, but, if the emergentists were right about anything, they were probably right about the phenomenal properties of conscious experience: these properties are not theoretically predictable on the basis of a complete knowledge of the neurophysiology of the brain. This is reflected in the following interesting difference between phenomenal properties and other mental properties (including cognitive/intentional properties): We can imagine designing and constructing novel physical devices that will instantiate certain cognitive capacities and functions (e.g., perception, information processing, information storage, inference and reasoning, the use of information to guide behavior) — arguably, we have already designed and manufactured such devices in certain robots and other computer-driven devices. But it is difficult to imagine our designing devices and structures that will have phenomenal experiences, for example, structures that will feel pain and itch. I don't think we have any idea where to begin. The only way we can manufacture a conscious device is to make a physical replica of a system that is

² Cf. Morgan: "Lewes says that the nature of emergent characters can only be learnt by experience of their occurrence; hence they are unpredictable before the event", *Emergent Evolution*, p. 5.

³ See, e.g., what Michael Tye calls "perspectival subjectivity", in his *Ten Problems of Consciousness* (Cambridge: MIT Press, 1995).

known (or believed) to be conscious. Notice that this involves inductive prediction, whereas theoretical prediction is what is needed to design new physical devices with consciousness. The emergentists were wrong in thinking that sundry chemical and biological properties were emergent; but this was an understandable mistake given the state of the sciences before the advent of quantum mechanics and molecular biology. The interest of the ideas underlying their distinction between the two kinds of properties need not be diminished by the choice of wrong examples to illustrate it.

II

As was noted at the start of our discussion, another idea that is closely related to the claimed unpredictability of emergents is expressed in the doctrine that the emergence of emergent properties cannot be *explained* on the basis of the underlying processes, and that emergent properties are not *reducible* to the basal conditions from which they emerge. These two claims can be combined into one: Emergent properties are not *reductively explainable* in terms of the underlying processes. Some may wish to distinguish the issue of reduction from that of reductive explanation;⁴ we will address this issue later. I will now present a model of reduction that connects and makes sense of these three ideas, namely that emergent properties are not predictable from their basal conditions, that they are not explainable in terms of them, and that they are irreducible to them.

Let me begin with an example. To reduce the gene to the DNA molecule, we must first interpret the property of being a gene functionally — that is, in terms of the causal work it does. Briefly, the property of being a gene is the property of having some property (or being a mechanism) that performs a certain causal function, namely that of transmitting phenotypic characteristics from parents to offsprings. As it turns out, it is the DNA molecule that fills this causal specification ("causal role"), and we have a theory that explains, at least in broad but persuasive outlines, how the DNA molecule is able to perform this causal task. When all of this is in, we are entitled to the claim that the gene has been reduced to the DNA molecule.

We may now formulate a general model to accommodate reductions of this form. Let **B** be the domain of properties (also phenomena, facts, etc., if you wish) serving as the reduction base — for us; these contain the basal conditions for our emergent properties. The reduction of property E to **B** involves three steps:⁵

⁴ For example, David Chalmers, *The Conscious Mind* (New York: Oxford University Press, 1996), p. 43.

⁵ The fundamental ideas for this view of reduction are present in David Armstrong's *A Materialist Theory of Mind* (New York: Humanities Press, 1964), and David Lewis's "An Argument for the Identity Theory", *Journal of Philosophy* 67 (1970): 203-211. However, neither Armstrong nor Lewis, to my knowledge, think of these ideas in connection with models of reduction. The idea of functional analysis of mental terms or properties is of course essential to the functionalist approach to mentality; however, most functionalists have regarded their approach to be essentially nonreductionist. For similar views on reduction see Robert Van Gulick, "Nonreductive Materialism and the Nature of Intertheoretic Constraint", in *Emergence or Reduction?*, ed. A. Beckermann, H. Flohr, and

Step 1: E must be *functionalized* — that is, E must be construed, or reconstrued, as a property defined by its causal/nomic relations to other properties, specifically properties in the reduction base **B**.

We can think of a functional definition of E in domain **B** as typically taking the following (simplified) form:

Having E = _{def} Having some property P in **B** such that (i) P is typically caused to be instantiated by C₁, ..., C_n, and (ii) P typically causes⁶ F₁, ..., F_m to be instantiated (where the Cs and Fs are in B).

We should allow either (i) or (ii) to be empty. The main point to notice is that the functionalization of E makes E nonintrinsic and relational — relational with respect to other properties. E's being instantiated is for a certain property P to be instantiated, with this instantiation bearing causal/nomic relations to the instantiations of a specified set of other properties. We may call any property P in **B** that satisfies the causal specification (i) and (ii) a "realizer" or "implementer" of E. Clearly, multiple realizers for E are allowed on this account; so multiply realizable properties fall within the scope of the present model of reduction.

Step 2: Find realizers of E in **B**. If the reduction, or reductive explanation, of a particular instance of E in a given system is wanted, find the particular realizing property P in virtue of which E is instantiated on this occasion in this system; similarly, for systems of the same species or structure types.

This of course is a scientifically significant part of the reductive procedure; it took many years of arduous and inspired scientific research to identify the DNA as a realizer of the gene.

Step 3: Find a theory (at the level of **B**) to explain how a given realizer of E performs the causal task as specified in Step 1 (and perhaps other causal/nomic relations in which the given instance of E is implicated).

J. Kim; and, especially, Joseph Levine, "On Leaving Out What It Is Like", in *Consciousness*, ed. Martin Davies and Glyn W. Humphreys (Oxford: Blackwell, 1993). See also Chalmers' discussion of "reductive explanation", in *The Conscious Mind*, ch. 2. I discuss these issues in greater detail in *Mind in a Physical World* (Cambridge: MIT Press, 1998).

⁶ For brevity we will speak somewhat loosely by saying that a property causes another property -- what is meant of course is that an instantiation of a property causes another property to be instantiated.

We presumably have a story at the molecular microbiological level about how DNA molecules manage to code and transmit genetic information. When temperature, for gases, is reduced to mean translational kinetic energy of molecules, we have a theory that explains the myriad causal/nomic relations in which temperature plays a role. Steps 2 and 3 are likely to be part of the same scientific research: Ascertaining the realizers of E will almost certainly involve theories about causal/nomic interrelations of lower-level properties in the base domain.

Notice how this functional conception of reduction differs from the classic Nagel model of intertheoretic reduction⁷ — in particular, there is no talk of "bridge laws" or "derivation of laws". The question whether appropriate bridge laws are available connecting the domain to be reduced with the base domain — for example, whether or not there are bridge laws providing for each mental property a physical property nomologically coextensive with it — has been at center stage in debates over reduction and reductionism. However, from the emergentist point of view, the existence of bridge laws connecting emergent properties with their basal conditions is completely irrelevant to the issues of reducing or explaining the emergents. For it is these bridge laws, laws that state that when specified basal conditions are present a certain novel property is manifested, that are exactly what need explaining. Why is it that pain, not itch or tickle, emerges when a certain neural condition (e.g., C-fiber stimulation) occurs? Why doesn't pain accompany conditions of a different neural type? Why does any phenomenal consciousness emerge when these neural conditions are present? These are the explanatory/reductive questions the emergentists wanted to have answered. And I think they were right. The "mystery" of consciousness cannot be explained away by any reductive procedure that takes these bridge laws as brute unexplained primitives.

The philosophical emptiness of Nagel reduction (at least in contexts like the mind-body reduction) can be seen from the following fact: the Nagel reducibility of psychology to physical theory is compatible with, and sometimes even entailed by, many dualist mind-body theories, such as the double-aspect theory, the theory of preestablished harmony, occasionalism, and epiphenomenalism. The reason is simple: many of these theories actually guarantee an abundance of mind-body bridge laws correlating each mental property with a uniquely corresponding

⁷ See Ernest Nagel, *The Structure of Science* (New York: Harcourt, Brace, and World, 1961).

physical property. Nagel reduction is not even excluded by substantial dualism of mental and physical substances (although Descartes' own interactionist version probably excludes it).

III

We can now see how the functional model of reduction can meet the explanatory/predictive/ontological demands on reductions of genuine philosophical interest. Let E be the property targeted for reduction, and assume that E has been functionalized as the property of having some property P meeting causal specification C.

1. The explanatory question.

Why does this system exhibit E at t? Because having E is having a property with causal role C, and the system, at t, has property Q which fills causal role C. Moreover, we have a theory that explains exactly how Q manages to fill C.

Why do systems of kind K exhibit E when they instantiate Q? Because E is a functional property defined by causal role C, and Q is a realizer of E (in systems of kind K). Again, there is a theory that explains how Q realizes E (in these systems).

Suppose that pain can be given a functional definition — something like this: being in pain is being in some state (or instantiating some property) that is caused by tissue damage and that in turn cause winces and groans. Why are you experiencing pain? Because being in pain is being in a state caused by tissue damage and causing winces and groans, and you are in neural state N, which is one of those states that are caused by tissue damage and that cause winces and groans. Why do people experience pain when they are in neural state N? Because N has these causal/nomic relations, and being in pain is being in some state with just these causal/nomic relations. It is clear that these meet all our explanatory demands. There is nothing further to be explained about why pain occurs, or why pain occurs when neural condition N is present.

But is this a *reductive* explanation? This question is connected with the question whether, and in what sense, the proposed model is a model of reduction, a question that will be considered below.

The remaining question of course is whether pain can be functionalized. We will briefly return to this complex issue later, but at least we have isolated the problem.

2. The predictive question.

Will the system exhibit E at time t? Can we predict this from knowledge of what goes on in the base domain? Yes, because we can know that Q, a property in the base domain, is a realizer of E, solely on the basis of the functional definition of E and the causal/nomic relations of Q in the base domain. Given this knowledge, we can predict that the system will, at t, instantiate property E on the basis of our knowledge that it will instantiate Q at t.

It is clear that what enables the ascent from the reduction base to higher properties is the conceptual connections provided by the functionalization of higher properties. This is in sharp contrast to Nagelian reduction with bridge laws taken as additional auxiliary premises. These laws are standardly conceived as empirical and contingent, and for that reason are incapable of helping us make predictions of the instantiations of higher properties solely on the basis of our knowledge of what goes on in the base domain. These bridge laws must be viewed as net additions to our predictive base, and this means that predictions no longer depend solely on knowledge of the base domain. This is why bridge laws only enable inductive predictions, whereas functionalization makes theoretical predictions possible.

And the reason we cannot design novel physical/biological devices that will exhibit consciousness is that brute bridge laws may be all we can get to connect phenomenal properties with physical/biological properties. What we need is an ability to make theoretical predictions of qualia on the basis of knowledge of the base domain alone, namely physics, biology, and the like. The functionalization of qualia would give us such an ability.

3. The ontological question.

In what sense is the functional model a model of reduction? What does it reduce, and how does it do it? Central to the concept of reduction, I believe, is the idea that what has been reduced need not be countenanced as an independent existent beyond the entities in the reduction base — that if X has been reduced to Y, X is not something "over and above" Y. From an ontological point of view, reduction must mean *reduction* — a simplified ontology. Reduction is not necessarily elimination: reduction of X to Y need not do away with X, for X may be conserved, or retained, as Y. Thus, we can speak of "conservative" reduction (some call this "retentive" reduction), reduction that conserves the reduced entities, as distinguished from "eliminative"

reduction, which rids our ontology of reduced entities. Either way we end up with a simpler, leaner ontology.

Our question, then, is in what ways the model of reduction being recommended here serves the cause of ontological simplification. Two cases may be distinguished: the first case concerns instances of property E; the second concerns property E itself.

(i) System *s* has E, in virtue of *s*'s instantiating one of its realizers, say Q. Now, *s*'s having E is just its having some property meeting causal specification C, and in this particular instance, *s* has Q, where Q meets specification C. Thus, *s*'s having E on this occasion is identical with its having Q on this occasion. There is no fact of the matter about *s*'s having E on this occasion over and above *s*'s having Q.

Each instance of E is an instance of one of E's realizers, and all instances of E can be partitioned into Q₁-instances, Q₂-instances, ..., where the Qs are E's realizers.

Suppose someone were to object as follows: There is no good reason to identify this instance of E with the instance of Q in virtue of which E is realized on this occasion. Rather, *s*'s having E should be identified with *s*'s having some property or other meeting causal specification C, and this latter instance is not identical with *s*'s having Q. For having some property or other meeting C is not the same property as having Q; that is, property E ≠ property Q. How should we counter this line of argument? I think it will be helpful to consider the causal picture, and ask: What are the *causal powers* of this instance of E, namely *s*'s having E on this occasion? If *s* has E in virtue of E's realizer Q, it is difficult to see how we could avoid saying this: the causal powers of this instance of E are exactly the causal powers of this instance of Q. This is what I have elsewhere called the "causal inheritance principle":

If a functional property P is instantiated on a given occasion in virtue of one of its realizers, Q, being instantiated, then the causal powers of this instance of P are identical with the causal powers of this instance of Q.

If this principle is accepted, the E-instance and the Q-instance have identical causal properties, and this exerts powerful pressure to identify them. What good would it do to count them as different? If they were different, the difference would not even be detectable.

Let us now turn to the reduction of E, the property itself. Here we need to come to terms with E's having multiple realizers, Q_1, Q_2, \dots . There are three possible approaches here.

First, one may choose to defend E as a legitimate higher-level property irreducible to the Qs. This is the position taken by most functionalists: Psychological properties are functional properties defined in terms of input/output correlations, with internal physical/biological properties as realizers, and yet they are irreducible to them, constituting an autonomous domain for the special science of psychology (cognitive science, or whatever).

Second, one may choose to identify E with the disjunction of its realizers, $Q_1 \vee Q_2 \vee \dots$. Notice, though, that this identity is not necessary — it does not hold in every possible world — since whether or not a property realizes E depends on the laws that prevail at a given world. The reason is that E is defined in terms of causal/nomic conditions, and whether something satisfies them is clearly dependent on the laws that are in force at a given world. This means that in another world with different laws, E may have a wholly distinct set of realizers, and in others E may have no realizers at all. So the identity, $E = Q_1 \vee Q_2 \vee \dots$ is metaphysically contingent, although nomologically necessary, and "E" becomes nonrigid, though nomologically rigid (as we may call it). For example, in a world with laws quite different from those prevailing in this world, molecules of another kind, not the DNA molecules, may perform the causal task of coding and transmitting genetic information.⁸

Third, we may give up E as a genuine property, only recognizing the expression "E" or the concept E. As it turns out, many different properties are picked out by E, depending on the circumstances — the kind of structures involved and the nomic/causal nature of the world under consideration. One could argue that by forming "second-order" functional expressions by existentially quantifying over "first-order" properties, we cannot be generating new properties, only new ways of indifferently picking out first-order properties in the base domain, in terms of causal specifications.⁹ As noted, the concept is only nomologically rigid, it picks out the same properties only across worlds that are similar in causal/nomological respects.

⁸ This point is valid whether or not E has single or multiple realizers in the actual world. A property may have a single realizer here but multiple realizers in other worlds, and vice versa.

⁹ For more details on this approach see my "The Mind-Body Problem: Taking Stock After 40 Years", forthcoming in *Philosophical Perspectives*, 1997, and *Mind in a Physical World*.

Here I will not argue my points in detail. It seems clear, however, that the second and third approach effectively reduce the target property E: the second is a conservative reduction, retaining E as a disjunction of properties in the base domain. In contrast, the third is eliminative: it recommends the elimination of E as a property, retaining only the concept E (which may play a practically essential role in our discourse, both ordinary and scientific). The first approach, as I said, is one that is most widely accepted: many philosophers, on the basis of multiple realization, want to argue that E is an irreducible property that nonetheless can be a property playing an important role in a special, "higher-level", science. I believe that this position cannot be sustained. For if the "multiplicity" or "diversity" of realizers means anything, it must mean that these realizers are causally and nomologically diverse. Unless two realizers of E show significant causal/nomological diversity, there is no reason to count them as two, not one. This means that multiply realizable properties are ipso facto causally and nomologically heterogeneous. This is especially obvious when one reflects on the causal inheritance principle. All this points to the inescapable conclusion that E, because of its causal/nomic heterogeneity, is unfit to figure in laws, and is thereby disqualified as a useful scientific property. On this approach, then, one could protect E but not as a property with a role in scientific laws and explanations. You could insist on the genuine propertyhood of E as much as you like, but the victory would be empty.¹⁰ The conclusion, therefore, has to be this: As a significant scientific property, E has been reduced — eliminatively.

IV

What I hope I have established is this: Functionalization of a property is both necessary and sufficient for reduction (sufficient at least as a first conceptual step, the rest being scientific research). This accords well with the classic doctrines of emergentism: as I argued, it nicely explains why reducible properties are predictable and explainable, and correlatively it makes it plausible why irreducible properties are not predictable or explainable on the basis of the underlying processes. I believe this makes good sense of the central tenets of emergentism.

However, emergentism may yet be an empty doctrine. For there may not be any emergent properties, all properties being functionalizable

¹⁰ For more details see my *"Multiple Realization and the Metaphysics of Reduction"*, reprinted in *Supervenience and Mind*.

and therefore reducible to physical properties, whether these are basic physical properties of microparticles or microstructural properties of larger complexes of basic particles. So are there emergent properties? Many scientists have argued that certain "self-organizational" properties of organic, living systems are emergent. But it is not clear that these are emergent in our sense of nonfunctionalizability.¹¹ And, as I said earlier, the classic emergentists were mostly wrong in putting forward examples of chemical and biological properties as emergent.¹² It seems to me that if anything is going to be emergent, it is consciousness — or the qualitative properties of our experiences (or "qualia"). Here I don't want to rehearse the standard arguments pro and con, but merely affirm, for what it's worth, my own bias toward the pro side: qualia are intrinsic properties if anything is, and to functionalize them is to eliminate them as intrinsic properties.¹³

In this paper I have not touched on various issues on causation that arise in connection with emergentism, in particular the emergentist doctrine of "downward causation". In classical emergentism early in this century, it was the unpredictability and irreducibility of the emergents that was the primary focus. Lately, the emergentists, in particular Roger Sperry, have emphasized the capacity of emergents to exercise their causal powers "downward" — that is, to causally influence and modify the processes at the lower level. Views of this kind give rise to many perplexing issues. Their discussion must await another occasion.¹⁴

¹¹ This point is made by David Chalmers; see his *The Conscious Mind*, p. 129.

¹² An observation made by Brian McLaughlin in "The Rise and Fall of British Emergentism", in *Emergence or Reduction?*.

¹³ More details and overview of the philosophical terrain involved, see Chalmers, *The Conscious Mind*, ch. 3. Two early papers arguing this point are Joseph Levine, "Materialism and Qualia: the Explanatory Gap", *Pacific Philosophical Quarterly* 64 (1983): 354-61, and Frank Jackson, "Epiphenomenal Qualia", *Philosophical Quarterly* 32 (1982): 127-36. But the debate on this issue is still very much alive.

¹⁴ For some further details see my "'Downward Causation' in Emergentism and Nonreductive Physicalism", in *Emergence or Reduction?*, and "Making Sense of Emergence", forthcoming in *Philosophical Studies*.