

The significance of emergence

Tim Crane¹

This paper is an attempt to understand the content of, and motivation for, a popular form of physicalism, which I call ‘non-reductive physicalism’. Non-reductive physicalism claims although the mind is physical (in some sense), mental properties are nonetheless not identical to (or reducible to) physical properties. This suggests that mental properties are, in earlier terminology, ‘emergent properties’ of physical entities. Yet many non-reductive physicalists have denied this. In what follows, I examine their denial, and I argue that on a plausible understanding of what ‘emergent’ means, the denial is indefensible: non-reductive physicalism is committed to mental properties being emergent properties. It follows that the problems for emergentism—especially the problems of mental causation—are also problems for non-reductive physicalism, and they are problems for the same reason.

The structure of the paper is as follows. In section 1, I outline what I take to be essential to non-reductive physicalism. In the second section I attempt to clarify what is meant by ‘emergent’, and I argue that the notion of emergence is best understood in terms of the idea of emergent properties having causal powers which are independent of the causal powers of the objects from which they emerge. This idea—‘downward causation’—is examined in section 3. In the final section I draw the lessons of this discussion for the contemporary debate on the mind-body problem.

1. Non-reductive physicalism

In his discussion of the mind-body problem in *The View from Nowhere*, Thomas Nagel claims that

¹ Department of Philosophy, University College London, Gower Street, London WC1E 6BT, UK.

what is needed is something we do not have: a theory of conscious organisms as physical systems composed of chemical elements and occupying space, which also have an individual perspective on the world, and in some cases a capacity for self-awareness as well. In some way that we do not now understand our minds as well as our bodies come into being when these materials are suitably combined and organised. The strange truth seems to be that certain complex, biologically generated physical systems, of which each of us is an example, have rich nonphysical properties.²

It seems to me that the position Nagel describes, and the conception of the problem this position generates, are quite widely held in current philosophy of mind.³ According to this position, mental properties are not physical properties, yet cartesian dualism is false: minds are ‘biologically generated physical systems’, yet mental properties are ‘non-physical’ and ‘come into being’ when the elements of thinking organisms are ‘suitably combined’. Furthermore, according to Nagel, the mind-body problem consists in the fact that we do not understand how these claims can all be true. For Nagel, and for many others, the mind-body problem is not solved by physicalism—rather, physicalism is what gives rise to the problem.⁴

I shall call the position described by Nagel, ‘non-reductive physicalism’. Non-reductive physicalism is characterised by (at least) the following two theses:

Distinctness: Mental properties are distinct from physical properties.

Dependence: Mental properties are properties of physical objects.

Non-reductive physicalism can therefore be distinguished from the property or ‘type’ identity theory on the one hand, and cartesian dualism on the other. The identity theory asserts that mental properties are identical with physical properties,

²*The View From Nowhere* (Oxford: Clarendon Press 1985) p.51.

³Similar ideas are found in David Chalmers, *The Conscious Mind* (Oxford: Oxford University Press 1996); Joseph Levine, ‘On leaving out what it’s like’ in Martin Davies and Glyn Humphreys (eds.) *Consciousness* (Oxford: Blackwell 1993), Colin McGinn, ‘Consciousness and content’ and ‘Can we solve the mind-body problem?’ in *The Problem of Consciousness* (Oxford: Blackwell 1991).

⁴See, for example, the discussions by S. Shoemaker and J. Fodor in T. Szubka and R. Warner (eds.) *The Mind-Body Problem* (Oxford: Blackwell 1994).

which is the denial of *Distinctness*; and cartesian dualism asserts that mental properties are properties of mental substances, which is one way of denying *Dependence*. The commitment to *Dependence* is what makes non-reductive physicalism *physicalism*, and the commitment to *Distinctness* is what makes it non-reductive. ('Non-reductive' and 'physicalism' have meant many things in recent philosophy. Here I intend these terms simply to be labels for the characteristic positions expressed by the principles of *Distinctness* and *Dependence*.)

In recent years, non-reductive physicalists have felt the need to distinguish their own doctrine from a doctrine which, like theirs, upholds *Distinctness* and *Dependence*: this is the doctrine that the mind is *emergent*.⁵ According to those who hold this doctrine—'emergentists'—mental properties are distinct from physical properties, cartesian dualism is false, and mental properties 'emerge' from complex arrangements of matter in a way that is inexplicable from the perspective of the sciences of matter. Emergentism was held to be the truth about many non-physical (or 'special') properties by a number of British philosophers of the late 19th and early 20th century.⁶ More must be said to clarify this doctrine, but the essential idea of emergentism is that special properties 'emerge' from their underlying physical substrates, in a way that cannot be predicted or explained from the perspective of the sciences of these physical substrates.

So why do non-reductive physicalists distinguish themselves from emergentists? Terence Horgan claims that according to emergentism, the laws

⁵For discussion, see Thomas Nagel, 'Panpsychism' in *Mortal Questions* (Cambridge: Cambridge University Press 1979), Colin McGinn, 'Consciousness and content', James Van Cleve, 'Emergence or pansychism: magic or mind-dust?' in J. Tomberlin (ed.) *Philosophical Perspectives Vol. 4* (Atascadero: Ridgeview 1990), Brian McLaughlin, 'The rise and fall of British Emergentism' in Beckerman, Flohr and Kim (eds.) *Emergence or Reduction?* (Berlin: de Gruyter 1992), John R. Searle, *The Rediscovery of the Mind* (Cambridge Mass.: MIT Press 1992), Terence Horgan, 'From supervenience to superdupervenience: meeting the demands of a material world' *Mind* 102 (1993).

⁶See the valuable discussion in Brian McLaughlin, 'The rise and fall of British Emergentism'. Despite my disagreement with some of the claims made in McLaughlin's paper, I am greatly indebted to it here. The central emergentist texts McLaughlin discusses are: J.S. Mill, *A System of Logic* (London: Longmans 1875); S. Alexander, *Space, Time and Deity* (London: Macmillan 1920); C. Lloyd Morgan, *Emergent Evolution* (London: Williams and Norgate 1923); and C.D. Broad, *The Mind and its Place in Nature* (London: Routledge and Kegan Paul 1923).

which determine the emergence of higher-level properties are 'metaphysically and scientifically basic, in much the same way that fundamental laws of physics are basic; they are unexplained explainers'.⁷ And this, Horgan claims, is what must be denied by any form of physicalism or materialism:

A materialist position should surely assert, contrary to emergentism, (i) that physics is causally complete (i.e. all fundamental causal forces are physical forces, and the laws of physics are never violated); and (ii) that any metaphysically basic facts or laws—any unexplained explainers, so to speak—are facts or laws within physics itself.⁸

Horgan's view is that emergentism denies that non-physical facts and laws (e.g. psychophysical laws) must be ultimately explicable in terms of physical facts and laws. And since it is this claim which, according to Horgan, is essential to a genuine form of materialism (or physicalism),⁹ it is the emergentists' denial of the claim which distinguishes their doctrine from non-reductive physicalism.

Looked at in this way, the difference between non-reductive physicalism and emergentism appears quite significant. Emergentism holds, for instance, that mental properties have causal powers which are not explicable in terms of the causal powers of their physical substrates (see below, §3). But non-reductive physicalism apparently holds that the causal powers of the mental are explicable in terms of underlying physical properties and laws. For this reason, it appears important for non-reductive physicalists to deny that mental properties are emergent properties.

The aim of this paper is to argue that this appearance is an illusion. For when the notion of an emergent property is examined, it turns out that the distinction between emergentism and non-reductive physicalism cannot adequately be drawn on metaphysical grounds. The only satisfactory way of drawing the distinction is on

⁷From supervenience to superdupervenience: meeting the demands of a material world', 557-558.

⁸From supervenience to superdupervenience', 560.

⁹A materialistic metaphysical position should assert that all supervenience facts are explainable—indeed, explainable in some materialistically acceptable way.' 'From supervenience to superdupervenience' 560.

epistemological grounds: in terms of the limits of our *a priori* expectations of what must be explicable. While this distinction is important, it is not the one by means of which non-reductive physicalists hope to distinguish their view from emergentism. For this distinction was supposed to be a metaphysical one: some properties were claimed to *be* emergent by the emergentists, and this is what the non-reductive physicalist is supposed to deny. But it turns out that on the non-reductionist's view, mental properties are emergent properties too, in the only interesting and plausible sense that can be given to the term.

Once this is recognised, then non-reductive physicalists must accept that whatever problems attach to emergentism attach themselves to non-reductive physicalism too. In essence, emergentism is generally thought to encounter two problems: first, it is committed to mental properties having their own causal powers, and second, it is committed to the inexplicability of the mind-body relation. But if I am right, then non-reductive physicalists must share these commitments. This is problematic for non-reductive physicalists, since their physicalism ought to commit them to the causal closure of the physical world, and to the physical explicability (in principle) of the mind-body relation.¹⁰ These issues have been greatly discussed in recent philosophy of mind: the standard physicalist response is to develop distinctive accounts of mental causation, and to declare the inexplicability of mind (especially consciousness) to be the 'hard problem' of closing the 'explanatory gap'.¹¹

Emergentists, by contrast, are quite happy to accept these consequences of their position. The British emergentists of the early 20th century, for instance, were happy to accept that mental properties have independent causal powers, and happy

¹⁰For these physicalist commitments, see (for example) David Lewis, 'An argument for the identity theory' in his *Philosophical Papers* Vol.1 (Oxford: Oxford University Press 1983), David Papineau, *Philosophical Naturalism* (Oxford: Blackwell 1993) chapter 1—on the closure of the physical—and Horgan, 'From supervenience to superdupervenience'—on physical explicability.

¹¹The phrase, 'the hard problem' is David Chalmers': see *The Conscious Mind* xi-xii; for the 'explanatory gap' see Joseph Levine, 'Consciousness and the explanatory gap' *Pacific Philosophical Quarterly* 64, 1983.

to accept that the mind-body relation is (in some sense) inexplicable. What for the non-reductive physicalist is the 'hard problem' about mind, is for the emergentists a brute fact which must be swallowed with what Broad called the 'philosophic jam' of 'natural piety'.¹² This is an aspect of what we can think of as the *epistemological attitude* of emergentism, not its *metaphysical content*. If emergentism is coherent, which even its critics admit it is, then surely this epistemological attitude is coherent too.

However, my aim here is not primarily to defend the truth of British emergentism, but to examine the *significance* of the doctrine in helping us to understand current debates in philosophy of mind. For it seems to me that there is more to be said for the attitude of 'natural piety' than is often recognised in recent philosophy of mind. I will suggest at the end of this paper that if we must accept mental properties as non-physical properties, then we would do well to consider favourably the epistemological attitude of emergentism as one of the available options for 'bridging the explanatory gap'. But first I must clarify the doctrine of emergence.

2. Emergent properties

Emergentism, like non-reductive physicalism, is committed to the truth of *Distinctness* and *Dependence*. Like non-reductive physicalism, it is also a 'naturalistic' view, in the (admittedly vague) sense that it regards the natural sciences as employing the correct method for investigating mental and other macroscopic and higher level phenomena. The apparent difference between the two views is that emergentism accepts, and non-reductive physicalism officially denies, that there are *emergent properties*. But how should we understand the idea of an emergent property? The intuitive idea of an emergent property is the idea of a 'novel'

¹²See Broad, *The Mind and its Place in Nature* 55. The phrase 'natural piety' derives from Samuel Alexander: see *Space, Time and Deity* 47.

property of a whole or complex which ‘emerges’ from the parts of the whole and the way the parts are put together. But how should we make this idea more precise?

There are many ways the properties of things are related to one another. Some properties of wholes are straightforward combinations of properties of their parts. Suppose an object weighing ten kilos has ten parts, each weighing one kilo. Then the property of the whole is determined by adding the properties of its parts. Weighing ten kilos is a distinct property from weighing one kilo, of course, but nonetheless it is not a different *kind* of property: weighing one kilo and weighing ten kilos are determinates of the determinable property *weight*. So such cases are not cases where the property of the whole is a novel property: simpler properties combine to form other properties from what Ernest Nagel calls the ‘additive point of view’.¹³

Richard Spencer-Smith defines the novelty of a property thus:

a property P is *novel* in x if x has P, and there are no determinates P’ of the same determinable as P, such that any constituents of x have P’.¹⁴

This rules out weight, age, and most spatio-temporal properties of objects as candidates for being novel, which seems correct. However, the definition would fail to classify an object’s colour as novel, and this seems wrong. For consider a uniformly blue object O which can be divided into two parts, both of which are blue. Here O has a colour, and two of its parts have a determinate property—blueness—of the determinable *colour*. So colour is not a novel property, according to this definition. Yet a colour is something objects can have only when they reach a certain size, so surely it ought to be classified as novel. Spencer Smith’s

¹³See E. Nagel, ‘Wholes, sums and organic unities’ in D. Lerner (ed.) *Parts and Wholes* (New York: free Press 1963). For a recent discussion of the determinable/determinate relation, see Stephen Yablo, ‘Mental causation’, *Philosophical Review* 101, 1992.

¹⁴Richard Spencer-Smith, ‘Reductionism and emergent properties’ *Proceedings of the Aristotelian Society*, 95 1995, 117. Spencer-Smith does not think that novelty in this sense is sufficient for emergence.

definition would similarly misclassify the *wetness* of a liquid, which ought to be a novel property.¹⁵

Spencer-Smith's criterion for novelty can be improved by replacing 'any' with 'all': the non-novel properties of a whole object would be determinable properties, other determinates of which are had by all of its parts. A object's novel properties would be those determinable properties whose determinates are not had by all of the object's parts. Since not all an object's parts are coloured, and not all of a liquid's parts are wet, *colour* and *wetness* are novel properties.

But is novelty—in this sense—sufficient for a property to be emergent? Not if emergent properties are meant to be distinguished from *reducible* properties. For even the most extreme reductionist can accept that there are novel properties in this sense. J.J.C. Smart, for example, emphasises that 'in saying that a complex thing is nothing but an arrangement of its parts, I do not deny that it can do things that a mere heap or jumble of its parts could not do'.¹⁶ Smart admits—as everyone should—that objects can have properties and powers which their parts do not have. But this doesn't mean that these powers or properties are not reducible to the powers of properties of the parts. The very most it means is that the properties need not be reducible in the 'additive' sense.

So this definition of novelty fails to distinguish emergent properties from reducible properties. Perhaps the emergentist could say that we need a stronger notion of 'emergent property': the notion of a property of a whole whose powers are unrelated to whatever the powers of its parts are. But how can we make sense of this idea? After all, if emergent properties are supposed to be properties of a

¹⁵Assuming that half of a sample of a liquid can be called a 'constituent' of the sample. The contrast here between novel and non-novel properties is called the contrast between 'Empedoclean' and 'Democritean' forms of explanation by Robert Klee ('Micro-determinism and concepts of emergence' *philosophy of Science* 51 1984, 44-63, 50), who follows T. R. Girill, 'Evaluating micro-explanation' *Erkenntnis*, 10, 1976, 387-405. See also John Searle, *The Rediscovery of the Mind* 111-112.

¹⁶J.J.C. Smart, 'Physicalism and emergence' in *Essays Metaphysical and Moral* (Oxford: Blackwell 1987) 248. Samuel Alexander (*Space, Time and Deity* 47) does not help matters by arguing that accepting emergent properties is no more problematic than accepting colours. Compare Spencer-Smith, 'Reductionism and emergent properties', 113.

thing which is *made out of* its parts, then how can it be that they are metaphysically unrelated to the parts in question? For instance, if affecting the parts can in many cases affect how the whole is—i.e., affect the *properties* of the whole—then the properties of the whole are, in these cases, related causally to the properties of the parts. And if putting the parts together in the same way guarantees that the whole will appear in the same way (*ceteris paribus*) then there is a metaphysical relation between the parts and the whole. But these apparently unproblematic part-whole relations would be denied by this strong conception of emergent properties.

This strong conception of emergence seems to require denying the plausible thesis that the emergent properties of a whole are supervenient upon the properties of its parts. This supervenience thesis will not be true for all properties of wholes—for instance, if there are relational properties, then truths about an object's relational properties will not supervene upon truths about the objects and its parts alone. But for someone who believes in *Dependence*, the idea that the truths about the *intrinsic* properties of wholes supervene on truths about the properties of their parts is surely a plausible thesis. For if they do not thus supervene, then it seems somewhat perverse to describe the properties as 'emergent'. Presumably part of the point of this label is to pick out the sense in which putting a thing's parts together gives you something new—but not because you have 'added' something 'from the outside'. If emergentism is to be distinguished from dualism and vitalism (which do add something 'from the outside') then it must reject this strong notion of emergence. The upshot is that a reasonable emergentist thesis is committed to the supervenience of a whole's properties on the properties of its parts. (Historically, this does seem to have been the case: drawing on important work by Brian McLaughlin, Terence Horgan argues some emergentists were explicitly committed to supervenience.¹⁷)

¹⁷Horgan, 'From supervenience to superdupervenience' 558-559.

This will seem strange to anyone who agrees with David Lewis that ‘a supervenience thesis is, in a broad sense, reductionist’, or with D.M. Armstrong that ‘the supervenient is not a feature of the world distinct from the features it supervenes upon’.¹⁸ For on these views, the idea that X supervenes on Y is intended to express the idea that X is nothing ‘over and above’ Y. The idea that everything supervenes on the physical, for instance, is intended to express the *necessary determination* of everything by the physical: the non-physical is nothing ‘over and above’ the physical because, given the physical facts in our world, any world with those facts (and no others) must contain the same non-physical facts.¹⁹ Yet if emergent properties are anything at all, they are precisely something ‘over and above’ that from which they emerge. So how can an emergentist be committed to supervenience?

The solution to this apparent puzzle is to recognise that the mere idea of supervenience does not support the glosses which Lewis and Armstrong place upon it. For it is quite consistent to hold both that the physical facts determine all the facts, and also to hold that the non-physical facts are distinct from the physical facts. In a word: determination of A by B does not imply that A and B are not distinct. Therefore, belief in a supervenience thesis does not require that one’s ontological commitment to the supervenience base exhausts one’s ontological commitments. One might, of course, have other reasons for requiring this—consider, for instance, the reasons which lead Lewis and Jackson to an identity theory²⁰—but the present point is that these reasons are independent of the supervenience thesis itself.

So an emergentist can hold that mental properties supervene on physical properties, yet they are something ‘over and above’ those physical properties. This

¹⁸Lewis, ‘New work for a theory of universals’ *Australasian Journal of Philosophy* 61, 1983, 358; Armstrong, *A Combinatorial Theory of Possibility* (Cambridge: Cambridge University Press 1989) 7. My own sympathies are with Chris Daly’s discussion of this kind of point in ‘Pluralist metaphysics’ *Philosophical Studies* 87 1997, 185-206.

¹⁹See Frank Jackson, *From Metaphysics to Ethics* (Oxford: Oxford University Press 1998) chapter 1.

²⁰See Frank Jackson, ‘Essentialism, mental properties and causation’ *Proceedings of the Aristotelian Society* 95, 1995.

point can be explicitly spelt out by employing Kim's notion of *strong supervenience*.

A family of properties A strongly supervenes on a family of properties B iff

- (S) Necessarily, if anything has property F in A, then there is a property G in B such that that thing has G and necessarily everything that has G has F.²¹

This notion of supervenience does not say anything about whether the A-properties are 'something over and above' the B-properties: (S) is consistent with the distinctness of the A- and B-properties, and also consistent with the identification of each A-property with a B-property. In addition, it is consistent with the A-properties having independent causal powers. So, the strong supervenience of the mental on the physical is consistent with emergentism.

Some philosophers have claimed that the supervenience claim is all that emergentism really amounts to. Kim, for example, claims that:

According to emergentism, higher-level properties, notably consciousness and other mental properties, emerge when, and only when, an appropriate set of lower-level 'basal conditions' are present and this means that the occurrence of the higher properties is determined by, and dependent on, the instantiation of appropriate lower-level properties and relations. In spite of this, emergent properties were held to be 'genuinely novel' characteristics irreducible to the lower-level processes from which they emerge. Clearly, then, the concept of emergence combines the three components of supervenience, namely, property co-variance, dependence and non-reducibility. In fact, emergentism can be regarded as the first systematic formulation of non-reductive physicalism.²²

But as I am understanding non-reductive physicalism, a non-reductive physicalist would disagree with Kim's description here, since it holds that mental properties are supervenient, but not emergent. So what extra feature distinguishes the idea of an emergent property from that of a supervenient property?

²¹See Jaegwon Kim, *Supervenience and Mind* (Cambridge: Cambridge University Press 1993).

²²Jaegwon Kim, 'Supervenience' in *A Companion to the Philosophy of Mind* ed. S.Guttenplan (Blackwell 1994) 576-7.

In the statements of many emergentists and their critics, we find the idea that an emergent property is one whose instantiation in an object is not *predictable* from knowledge of the instantiation of the properties of the object's parts. In discussing the emergence of organic properties, for instance, Broad says that

No amount of knowledge about how the constituents of a living body behave in isolation or in other and non-living wholes might suffice to enable us to predict the characteristic behaviour of a living organism. This possibility is perfectly compatible with the view that the characteristic behaviour of a living body is completely determined by the nature and arrangement of the chemical compounds which compose it, in the sense that any whole which is composed of such compounds in such an arrangement will show vital behaviour and that nothing else will do so.²³

Here Broad endorses the supervenience of the organic on the inorganic, but combines this with the view that facts about the organic cannot be predicted from knowledge of facts about the inorganic constituents alone. He would also say the same about mental and all other emergent properties.

But the fact that we cannot predict (or explain) the behaviour of the higher-level properties from knowledge of the lower-level properties alone does not tell us whether these properties are emergent. For whether or not we can predict the higher-level phenomena will depend on our having a vocabulary in which to describe these phenomena. And it is a familiar fact that this vocabulary cannot be given solely by the science of the lower-level properties. Consider, for example, the case of inter-theoretic reduction, where we may in certain cases identify a property at a lower level with a property picked out in the higher-level vocabulary. In order to derive truths expressed in the vocabulary of the higher-level science from truths expressed in the lower-level vocabulary, we need 'bridge laws': sentences which tell us how to link the two vocabularies.²⁴ So in these cases (e.g. temperature in gases to

²³*The Mind and its Place in Nature* 67-68. The phrase 'nothing else will do so' indicates that Broad will reject the possibility of variable/multiple realisation of the emergent by the subvenient. But this can be safely regarded as inessential to the emergentist's picture.

²⁴See Ernest Nagel, *The Structure of Science* (Indianapolis: Hackett, second edition 1979) chapter 11.

mean molecular kinetic energy of constituent molecules) we cannot predict the macro-phenomena from knowledge stated in the vocabulary of science of the micro-phenomena *alone*, since we need bridge laws to link the vocabularies.

As I noted above, reduction is not emergence. If the impossibility of predicting (from knowledge of the lower-level alone) arises for reductionism, then it cannot be a distinguishing feature of emergentism. And in particular, it cannot distinguish emergent properties from other properties: for if prediction is impossible in the cases where we *identify* properties at different 'levels', then its impossibility in the case of emergent properties does not help us in individuating those properties. The unpredictability arises because there are limits to what can be said in the lower-level vocabulary alone, not because of anything about the nature of the properties at the higher level.

I am here understanding prediction and explanation as epistemic notions. An alternative is to take the notions more metaphysically, and talk of 'predictability in principle'. For the higher-level truths to be predictable from the lower-level truths in this sense is just for there to be some way of deriving one from the other, whether or not anyone will ever know it. We should not object to this as a way to talk—but what does it really amount to? As far as I can see, it is just another way of expressing the thought that fixing the lower-level fixes the higher-level. And this is simply the supervenience thesis to which we have already committed the emergentists.

It turns out, then, that neither predictability in principle nor unpredictability in practice can distinguish emergent properties from non-emergent properties. For insofar as emergentism is committed to the supervenience thesis—fix the base properties and the laws, and the emergent properties emerge—then emergent properties are as predictable 'in principle' as non-emergent or reducible properties

are. But insofar as bridge laws are required for prediction, then emergent properties are as unpredictable in practice as reducible properties are.²⁵

So the notions of novelty, unpredictability and supervenience as such do not distinguish emergent properties from the properties postulated by reductive physicalism. This does not mean that there is nothing to be said about emergent properties. For what does seem to be true is at least this: emergent properties are supervenient properties which are distinct from the properties on which they supervene. If we focus on what their distinctness consists in, we can begin to complete this story.

Why say emergent properties are distinct properties at all? Presumably because we think that they make a difference to an object which has them: the object is different from the way it would have been if it had just had (*per impossibile*) its non-emergent properties. Coming about through a purely natural process, this difference must be capable of manifesting itself in some way; and a theory of these properties will tell us how these manifestations are detectable. In other words, in picking out differences in natural properties, we are (at least) picking out the differences in an object's causal powers (whether or not we treat properties as identical to their causal powers). So an emergent property, on this conception, is one which has causal powers which are distinct from the causal powers of the lower-level properties on which it supervenes. If you give a list of an object's causal powers, listing only the causal powers of the lower-level properties of the objects, then you will not have given a complete list of the object's powers. What is more, some of these properties are capable of affecting the motion of objects: emergent properties are responsible for 'downward causation' from the higher (e.g. mental) to the lower (e.g. microphysical) levels of nature. This, as Brian McLaughlin rightly observes, is one of the distinguishing features of emergent properties.²⁶ But if I am

²⁵In any case, there is something wrong with construing the notion of an emergent property in terms of epistemic notions like predictability, since emergence is supposed to be a metaphysical category. See Spencer-Smith, 'Reductionism and emergent properties' pp.120-121.

²⁶The term 'downward causation' derives from the biologist D.T. Campbell, "'Downward

right that non-reductive physicalism and emergentism have the same metaphysical commitments, then non-reductive physicalists must be committed to downward causation too. Is this right?

3. Downward causation

The idea of downward causation is a simple one: it is the idea of causal influence from the macroscopic to the microscopic levels of nature. How things are at a higher level of complexity affects what happens at a lower level. Take a mental case: I decide to take a drink from the glass in front of me. I move my arm. My arm moves, and so do the molecules in the cells which make it up. A macro-property of me—my decision—affects certain microproperties of my arm—the positions and velocities in the particles which make it up (and much more besides).

The British emergentists were explicitly committed to downward causation.

C. L. Morgan is representative :

when some new kind of relatedness is supervenient (say at the level of life), the way in which the physical events which are involved run their course is different in virtue of its presence—different from what it would have been if life had been absent.²⁷

It is perhaps not so clear that the non-reductive physicalist, on the other hand, is committed to causation of this kind. But in fact it follows straightforwardly from the non-reductivist's denial of the identity theory *plus* their characteristic denial of epiphenomenalism. For the denial of the identity theory means that they cannot say that the molecules in my arm are only caused by some purely physical property of

causation" in hierarchically organised biological systems' in F.J. Ayala and Dobzhansky (eds.) *Studies in the Philosophy of Biology* (Berkeley: University of California Press 1974). McLaughlin ('The rise and fall of British emergentism') argues that downward causation, while not incoherent, presents emergentism with its biggest problems. The upshot of the next section is that if McLaughlin and I are both right, then these are problems for non-reductive physicalism too.

²⁷Morgan, *Emergent Evolution* 16.

my arm; and the denial of epiphenomenalism just is the thesis that mental properties are causes.²⁸

What are the consequences of this commitment to downward causation? Notice first that to believe in downward causation one does not have to believe that actions, like my moving my arm, are identical with bodily movements, or identical with aggregates of events described in the language of microphysics. Such identifications are implausible; but equally implausible is the idea that if my decision genuinely does move my arm, then this decision has nothing to do with the subsequent motion of the parts of my arm. There seem to be only two possibilities: either my decision does move my arm, in which case it has something to do with the simultaneous movement of the arm's parts; or all motion is determined by the microscopic properties of the parts, and the movement of my arm by my decision is an illusion. But this suggests a dilemma: in the first case, it seems that my mind acts immediately an immediate action upon the molecules. But how? And in the second case, my mind makes nothing move at all.

It can seem that this dilemma—magic or epiphenomenalism—is spurious. For the fact that the molecules in my arm move when I decide to move my it might not, in itself, raise any metaphysical worries. The emergentist neuroscientist R.W. Sperry claims that to say that a mental event causes the motion of molecules is as innocuous as saying that 'the molecules and atoms of a wheel are carried along when it rolls downhill'.²⁹ Presumably, the idea here is that the motion of the wheel causes its atoms and molecules to move too. Or consider the case of a gas at constant temperature whose volume is suddenly halved:

²⁸I am assuming here, with most participants in the debate, that properties are causes (rather than events in Davidson's sense). So by epiphenomenalism I mean the stronger thesis of what McLaughlin calls 'type epiphenomenalism'. See Brian McLaughlin, 'Type epiphenomenalism, type dualism and the causal priority of the physical' *Philosophical Perspectives* 3 1989. See also Jackson, 'Mental causation' *Mind* 105, 1996. For a non-reductive physicalist who embraces epiphenomenalism, see Chalmers, *The Conscious Mind*, 165.

²⁹R.W. Sperry, 'A modified concept of consciousness' *Psychological Review* 76, 1969, 532.

if the gas is ideal, Boyle's law entails that when its pressure settles down again it will be twice what it was. That law does not dictate all the interim behaviour of the sample's molecules—except that it must be such as will eventually double the sample's pressure.³⁰

Again, we could cite this as an unproblematic example of downward causation: a macroscopic event—halving the volume of the gas—causes things to happen among the gas's molecules while also causing the sample to double in pressure. So why can't we say that the causation in question takes the same general form when my decision causes my arm to move, thus causing the molecules in my arm to move?

Peter Smith comments, of the gas case, that it 'plainly doesn't cut against the notion that the microcausal interactions, this time of a gas, causally suffice to produce the macrobehaviour (exemplified as pressure and temperature)'.³¹ This is true, but all it shows is that the downward causation so described is not inconsistent with the supervenience of the macro-facts on the micro-facts. And since we have already assumed that both emergentism and non-reductive physicalism are committed to supervenience, this only goes to show that downward causation and supervenience are not inconsistent, without the addition of other assumptions.

What other assumptions must these be? It is sometimes said that such causation is incompatible with the laws of mechanics, the science of motion.³² But, as McLaughlin points out, this is not so. Suppose for the sake of argument that what downward causation requires is what McLaughlin calls 'configurational forces': forces that can only be exemplified by matter which has a certain complexity, or a certain kind of structure. Configurational forces are therefore unlike the gravitational force, which holds between any two particles. To illustrate McLaughlin's point, let's consider the case of the laws of classical mechanics:

³⁰Tim Crane and D.H. Mellor, 'There is no question of physicalism' *Mind* 99, 1990, 190-191.

³¹Peter Smith, 'Modest reductions and the unity of science' in D. Charles and K. Lennon (eds.) *Reduction, Explanation and Realism* (Oxford: OUP 1992) 26.

³²The charge is implicit in David Papineau, 'Why supervenience?' *Analysis* 50, 1990. See also Fodor, 'The mind-body problem' in Szubka and Warner (eds.) *The Mind-Body Problem*.

Newton's laws of motion.³³ When a body acts on another body to produce acceleration, it must conform to these laws. These laws are, in Broad's words, 'general conditions which all motions, however produced, must conform to'.³⁴ That is, they do not tell us everything about how motions are produced, or why things move. When a particular force is exerted on a given object, say the force exerted by a body's electric charge, then the acceleration of the body will be fixed by the laws governing electric charge—e.g. Coloumb's law—and any other forces acting upon the body, in accordance with the general laws of motion, to produce the resultant acceleration. The laws of motion themselves do not place any limit on what kinds of forces can operate on bodies; so if there are forces which can only come into being when matter achieves a certain level of complexity, all that classical mechanics requires is that the motion produced by these forces should *conform* to Newton's laws. So if we understand downward causation in terms of configurational forces, then the existence of downward causation is not incompatible with the laws of mechanics.

Where the might seem to be a conflict is between the existence of downward causation and a more general metaphysical principle, which has been called 'the completeness of physics'.³⁵ This principle has been stated in many ways, but essential to all its statements is the idea that any physical effect (that is, any effect describable in the language of physics) is completely fixed, deterministically or indeterministically, by purely physical causes. If the completeness of physics is true, then all physical effects are fixed (for example) by the behaviour of atoms and their constituent electrons, protons and neutrons. Yet according to both emergentism and non-reductive physicalism, there are physical effects—e.g. movements of the

³³McLaughlin argues in addition that downward causation and configurational forces are not incompatible with quantum machnics, nor with special and general relativity. See 'The rise and fall of British Emergentism' 53-54, and 74-75. But the essential point can be made in relation to classical mechanics.

³⁴C.D. Broad, *Scientific Thought* 177.

³⁵See David Papineau, *Philosophical Naturalism* 16.

molecules which make up the cells in my hand—which are the effects of mental properties. How is this compatible with the completeness of physics?

This is the problem of mental causation for non-reductive physicalists, which has recently generated much dispute about the causal character of the mind. Arguably, the acceptability of downward causation is the very issue which divides reductive from non-reductive views of mental properties. Now my aim here is not to resolve this dispute, but merely to point out that insofar as there is a problem here, it is the *same* problem which challenges emergentism and non-reductive physicalism. And it is the same problem for exactly the same reason: how should we reconcile *Distinctness* with the completeness of physics and the denial of epiphenomenalism?³⁶ If this is the problem, then emergentism and non-reductivism are in the same position. There is not a further problem which attaches to emergentism, as Horgan and McLaughlin claim.

4. The epistemological significance of emergence

I have argued—with Kim, and against Horgan—that the metaphysical commitments of the most plausible versions of non-reductive physicalism and emergentism are the same. So it is a mistake to attempt to define non-reductive physicalism in terms of its denial of emergence.³⁷ Why does this matter? I shall draw two morals: one about the cogency of non-reductive physicalism, the other about the current conception of the mind-body problem.

³⁶For a concise statement of the problem in this form, see Stephen Yablo, 'Mental causation'.

³⁷Though I am sympathetic to Kim's discussion in 'The non-reductivist's trouble with mental causation' (in *Supervenience and Mind*) I draw a somewhat different moral. It might be said that the difference between emergentism and physicalism only becomes obvious when we consider their modal status: physicalism holds that supervenience is metaphysically necessary, whereas emergentism allows that it might turn out to be contingent. There is something unsatisfactory about the idea that the interesting difference between these two doctrines might rest on a their modal status, rather than on the kinds of entities they recognise. But I shall leave discussion of this question to another occasion.

Let me begin with the relevance of my conclusion to non-reductive physicalism. McLaughlin describes the downfall of the British emergentist tradition as follows:

In their quest to discover 'the connexion or lack of connexion of the various sciences' (Broad 1923 pp.41-42) the Emergentists left the dry land of the *a priori* to brave the sea of empirical fortune. (The only route is by sea, of course.) They set off in a certain direction, and for awhile winds of evidence were in their sails; but the winds gradually diminished, and eventually ceased altogether to blow their way.³⁸

McLaughlin's point is that British emergentism failed for empirical, rather than philosophical, reasons: with the advent of (for example) the quantum mechanical explanation of chemical bonding, there was no longer any need to postulate irreducible chemical forces, and the emergentists lost some of their most plausible examples (and consequently, lost their metaphysical nerve).

McLaughlin's claim about the failure of this particular version of emergentism is plausible. But what conclusions should contemporary non-reductive physicalists draw from this claim? Horgan argues that the failure of British emergentism gives the non-reductive physicalist a motive to look for an account of the mind-body relation, an explanatory relation stronger than supervenience, which he calls 'superdupervenience'. The appeal to the notion of superdupervenience is supposed to explain why supervenience claims are true, and therefore give us a satisfactory account of the relation between mind and body. Horgan's point is that without such an account, the non-reductive physicalist is left in the same position as the emergentist.

Now here we have finally located a difference between emergentists and non-reductive physicalists. For emergentists do not believe on *a priori* grounds that no explanation of the connection between the different levels of nature can be given. They believe it is the best conclusion to draw from their empirical investigations.

³⁸'The rise and fall of British emergentism' 90.

What we find when we look at levels of nature are discontinuities and downwards causation. But although we should accept these facts with natural piety, this should not stop us from investigating the connections between levels, to see whether what we have is an emergent property or a 'resultant'. This was Morgan's view:

Cognitive relatedness just emerges, as something genuinely new, at a critical stage of evolutionary advance. That, however, does not preclude—nay, rather, it imperatively demands from us as evolutionists—a resolute attempt to analyse the situation and to trace, if possible, subsidiary stages of emergence, on the understanding that, in evolutionary progress, there is never any breach of continuity in the sense of a gap or hiatus.³⁹

Perhaps when we investigate the relations between levels, all we will find are 'stages of emergence' but perhaps we will find resultants. Whether we do or not is an empirical matter (as McLaughlin says, the only way to go is by sea). It is for this reason that the emergentists should be considered 'naturalists'.

By contrast, the non-reductive physcialist's view can appear less naturalistic:

Resolutely shunning the supernatural, I think it is undeniable that it must be in virtue of *some* natural property of the brain that organisms are conscious. There just *has* to be some explanation of how brains subserve minds. If we are not to be eliminativists about consciousness, then some theory must exist which accounts for the psychophysical correlations we observe. It is implausible to take these correlations as ultimate and inexplicable facts.⁴⁰

These remarks of Colin McGinn's express some of the assumptions which lie behind the recent debate about the problem of consciousness. In articulating the contemporary problem of consciousness, Joseph Levine says: 'we want an explanation of why when we occupy certain physico-functional states we experience qualitative character of the sort we do.'⁴¹ Commenting on this passage, Spencer-Smith remarks that 'we would like more than the neural correlate of an experience,

³⁹C. Lloyd Morgan, *Emergent Evolution* 9.

⁴⁰Colin McGinn 'Can we solve the mind-body problem?' 6.

⁴¹J. Levine, 'On leaving out what it's like' 128.

an explanation which would ideally take the form “x is in pain iff ...” —we want an explanation of how pains feel.’⁴²

It is here that we encounter the deep difference between emergentism and non-reductive physicalism. It is not a metaphysical difference, but a difference in the reactions of the two theories to limitations in our knowledge. Both emergentism and non-reductionism agree that we do not currently understand how the non-mental properties of the brain are related to its mental properties. But they react to this in different ways: the non-reductionists react by claiming that there must nonetheless be an account of *why* ‘we experience qualitative character of the sort we do’, an account which does not just state the complex correlations between the non-mental and the mental. The emergentists deny that this *must* be so. If it turns out that the relation between consciousness and the brain is inexplicable, this ends up being one of the facts that must be accepted with natural piety. As Morgan says, this does not preclude us from looking for levels of emergence, to ensure that there is no ‘breach of continuity’. But levels of emergence are still levels of *emergence*, and at some point in our investigations we may have to accept that the most our scientific investigations will give us are correlations. The availability of the emergentist position encourages us to look with suspicion on the idea that there *must* nonetheless be a philosophical account of these correlations—the sort of account demanded (for various reasons) by Horgan, McGinn, Nagel, Levine and others.⁴³

So one possible approach to the mind-body problem, inspired by emergentism, is that the insistence that there must be a metaphysical ‘explanation of how pains feel’ is misplaced. This is not because consciousness and thought are mysterious properties, unrelated to properties of the brain. On the contrary: anyone who takes a naturalistic approach to these issues is not going to find it surprising or illuminating to be told by John Searle that consciousness and thought are ‘higher-

⁴²‘Reductionism and emergent properties’ 127.

⁴³On this issue, see my debate with William Child in *Proceedings of the Aristotelian Society* 96, 1996.

level feature of the brain' (though they might find it more acceptable to say with Chomsky that 'people think, not their brains, which do not, though their brains provide the mechanisms of thought'⁴⁴). The mind-body problem from this emergentist perspective is not the mystery of how the brain can produce consciousness, since the existence of mental properties, and their dependence on the brain, are accepted with natural piety. Rather, it is what Chomsky calls a 'unification problem': to explain how the mind/brain works, given that 'we have good and improving theories of some aspects of language and mind, but only rudimentary ideas about the relation of any of this to the brain'.⁴⁵ This way of treating the mind-body problem does not assume that there is one explanatory gap, or one hard problem, any more than it assumes that one metaphysical notion like supervenience should be usefully employed in bridging the gap or solving the problem. What it does assume is that we should not say *a priori* when we should take the facts of nature to require further explanation. Hence the need to be open to the attitude of natural piety.

The significance of emergentism, then, lies in the value of the epistemological attitude it recommends to naturalism. Naturalistic non-reductive approaches to the mind-body problem should look more favourably than they have done in recent years upon the attitude of natural piety. It may be that whether emergentism (in the sense explained in §§2 and 3 above) or reductionism is true is still an open question. But the contention of this paper has been that emergentism is metaphysically indistinguishable from non-reductive physicalism, and perhaps a more authentic position for a genuine naturalist to adopt.⁴⁶

⁴⁴See John R. Searle, *The Rediscovery of the Mind* chapter 1; Noam Chomsky, 'Language and Nature' *Mind* 1995, 8.

⁴⁵Chomsky, 'Language and Nature' 11.

⁴⁶For comments and advice I am especially grateful to Richard Holton and Rae Langton, and also to Chris Daly, André Gallois, Giovanna Hendel, Lloyd Humberstone, Neil Manson, Mike Martin and Graham Oppy. This paper was written while I was a visiting fellow at the Research School of the Social Sciences of the Australian National University.