

PHYSICALISM, EMERGENCE AND DOWNWARD CAUSATION

Richard J. Campbell and Mark H. Bickhard

1. Physicalism

Physicalism, in one form or another, has been one of the dominant positions in metaphysics in the latter part of the 20th century. But what, precisely, does that position entail? That has been much debated. Rudolph Carnap's early attempt to show how every sentence of psychology could be translated into sentences formulated in physical language is now generally agreed to have been unsuccessful.

The next version to excite attention took off from the idea that, while mental terms cannot be *translated* into physical terms (as Carnap's program had required) nevertheless, it does not follow that these terms *refer* to non-physical states. So, it was proposed, mental phenomena can be identical with states and processes in the brain, even though psychological descriptions have meanings different from physical descriptions. The official line was that it is an empirical question for science to determine whether such identities in fact obtain. Nevertheless, the advocates of this 'contingent identity' thesis left no doubt as to what they believed the answer would be.

In that context, which pitted the 'mental' over against the 'physical', there was no shortage of philosophers prepared to embrace a physicalist position with enthusiasm and without any compromises. For these, the only enlightened response to modern science is the hard-headed one: every regularity described in the language of higher levels, including those of biological function and human intentionality, is causally redundant. All the causal work occurs at the fundamental level of concrete physical particulars. According to this program, the role of a philosopher is to explain how the concepts we ordinarily use to describe both ourselves ('folk psychology') and the world ('naïve realism') can be reconciled with this metaphysical picture. The success of the program in working out, hand-in-hand with scientific research, the details of the required 'reductions' would be the vindication of this physicalist ontology.

Not all philosophers with physicalist sympathies, however, were convinced that such 'reductions' could actually be carried out, even in principle. Two lines of argument against the identity theory emerged. Firstly, it seems at least possible that high-level functional properties could be realized in radically different constituents. Perhaps psychological phenomena can be paired not only with neural properties (e.g., in brains), but equally well with electro-mechanical properties (e.g., in computers). In that case, there would be no justification for identifying the mental with one, rather than the other. Secondly, it was argued that scientific research is unlikely ever to produce any strict psychophysical laws linking *types* of mental phenomena with *types* of physical

phenomena, so the bridging principles required by the contingent identity theory could never be produced.

While acknowledging these difficulties, these philosophers did not want to give up physicalism altogether. So, they sought to develop non-eliminative, non-reductionist understandings of mentality – either some form of functionalism or ‘anomalous monism’ – which would still preserve a minimal physicalism. In the functionalist version, this compatibility was achieved by maintaining that mental phenomena are ‘realized’ in physical properties and processes, while in the anomalous monist version, it was achieved by holding that every event that can be given a mental description can also be given a physical description. In either version, even though there are no scientific laws by which mental phenomena could be ‘reduced’ to physical phenomena, the underlying causality of the world remains entirely physical. According to these alternative versions of a non-reductive physicalism, philosophers need only be committed to this minimal causal claim. In that way, physicalism is compatible with our ordinary way of explaining bodily movements by citing wants and wishes, beliefs and fears, hopes and dreams. It is just that these psychological states can be re-described in physical terms.

The tricky part was to explain how it is that mental properties and events are not reducible to physical properties and events, but are nevertheless dependent upon them. Here the idea of *supervenience* suddenly seemed attractive. Mental concepts may not be able to be defined in terms of physical concepts, but what might look like an emergent, causally efficacious property at the mental level actually ‘supervenies’ on a person’s physical properties and powers. The central idea of supervenience seems disarmingly simple: “no mental change without accompanying physical changes”. But how to render this idea precise and plausible became increasingly complicated and technical, as physicalists debated alternative definitions of ‘weak’, ‘strong’, and ‘global’ supervenience.

In our view, the most incisive and powerful analyses of these issues have been provided by Jaegwon Kim, in an influential series of papers.¹ We therefore propose to discuss the plausibility of physicalism by way of commenting upon his penetrating

¹ ‘The Myth of Nonreductive Materialism’, *Proceedings and Addresses of the American Philosophical Association*, 63 (1989) pp. 31-47; ‘Supervenience as a Philosophical Concept’ *Metaphilosophy* 21, pp. 1-27 (1990); ‘Epiphenomenal and Supervenient Causation’ in D.M. Rosenthal (ed.) *The Nature of Mind* (Oxford UP, Oxford, 1991) pp. 257-265; “Downward Causation” in Emergentism and Non-reductive Physicalism’ in A. Beckermann, H. Flohr, J. Kim (eds.) *Emergence or Reduction? Essays on the Prospects of Nonreductive Physicalism* (Walter de Gruyter, Berlin, 1992); ‘Multiple Realization and the Metaphysics of Reduction’, *Philosophy and Phenomenological Research* 52, 1992, pp. 1-26; ‘The Non-Reductivists’s Troubles with Mental Causation’ in J. Heil & A. Mele (eds) *Mental Causation* (Oxford UP, Oxford, 1993). All these papers are reprinted in his *Supervenience and Mind* (Cambridge UP, Cambridge, 1993). See also his ‘What is the Problem of Mental Causation?’ in M.L.D. Chiara, K. Doets, D. Muncini, & J. van Benthem (eds): *Structures and Norms in Science* (Kluwer Academic, Dordrecht, 1997) pp. 319-329, and ‘Does the Problem of Mental Causation Generalize?’ *Proc Arist. Soc.*, 1997, pp. 281-297.

analyses. Kim's central argument has been that the fashionable position – that one could be a physicalist without reducing the mental to the physical – is untenable. In his recent book, *Mind in a Physical World*, Kim has refined these analyses further. We will argue that he has thereby highlighted assumptions intrinsic to physicalism that are inconsistent with contemporary developments in physical theory. In effect, Kim has identified a *reductio ad absurdum* of physicalist metaphysics in general. Clearly grasping why physicalism is untenable, however, opens the logical space for a fecund notion of genuine emergence.

2. The Failure of Non-Reductive Physicalism

Many non-reductive physicalists avail themselves of both supervenience and realization. Often, they articulate the idea of supervenience in terms of 'global supervenience'. LePore and Loewer, for example, define that idea as:²

If two nomologically possible worlds are exactly alike with respect to fundamental physical facts ... then they are exactly alike with respect to all other facts.

But arguably, this is too strong. It makes physicalism hostage to the ontological mysteries raised by talk of 'possible worlds'. Also, it obscures the fact that physicalism is not a claim about *every* possible world, but only a claim about *our* world, to the effect that its physical nature exhausts all its nature. For these reasons, and wishing to maintain global supervenience as a *contingent* thesis, Frank Jackson, for example, formulates the physicalist thesis as: "Any world which is a *minimal* physical duplicate of our world is a duplicate *simpliciter* of our world."³

Kim, however, has put his finger on what is so unsatisfactory about defining supervenience in global terms, without any accompanying *strong* supervenience (or covariance) claim. Firstly, such a claim permits worlds that differ only in the most minute physical detail, e.g., an extra hydrogen atom in some remote location in the universe, to differ drastically in their globally supervenient properties; one may contain creatures with full sentience, while the other has no mentality at all. This objection can be got around – e.g., by weakening indiscernibility to similarity, or by defining it over spatio-temporal regions rather than worlds – but the proposed solutions are inevitably *ad hoc*. Secondly, and more seriously, global supervenience offers no explanation of *how* some mental phenomenon supervenes upon physical facts. Its advocates ask us to accept a sweeping claim about the nature of our world, and at least some other possible worlds,

² E. LePore and B. Loewer: 'More on Making Mind Matter', *Philosophical Topics*, **17**, 1989, pp. 175-191.

³ F. Jackson: *From Metaphysics to Ethics: A Defence of Conceptual Analysis* (Clarendon Press, Oxford, 1998) p.12. Jackson is a reductive, or 'eliminative' physicalist, but that difference is not relevant at this point.

purely on faith. This criticism applies to both of the proposed characterizations of global supervenience above. We agree with Kim's trenchant objection:⁴

In the absence of specific psychophysical correlations, and some knowledge of them, such a supervenience claim should strike us as a mere article of faith seriously lacking in motivation both evidentially and explanatorily; it would assert as a fact something that is apparently unexplainable and whose evidential status, moreover, is unclear and problematic.

That presumably is why in his latest book, Kim opted for the following way of expressing the relevant sense of supervenience:

mental properties supervene on physical properties in the sense that if something instantiates any mental property *M* at time *t*, there is a physical base property *P* such that the thing has *P* at *t*, and anything with *P* at some time necessarily has *M* at that time.⁵

He then argues that the initially attractive idea that mental properties are realized in and by physical properties *entails* the supervenience of the mental on the physical. Kim spells out the idea of realization, focusing on functional realization, as follows. Functional properties are defined as properties specified in terms of their roles as causal intermediaries between sensory inputs and behavioural outputs. (In section 5 below we will support a more precise, and stronger, definition of realization.) Then, if we understand mental properties and powers to be functional properties, a physicalist can maintain that the only potential occupants, or 'realizers', of these causal roles are physical properties.

The burden of Kim's argument is that explaining mentality in these terms means that supervenience lapses into epiphenomenalism. According to this way of explaining functional properties, the dormitivity of a Seconal pill, for example, is 'realized' by its chemical properties, so that it is the chemical properties, and not its dormitivity, that do all the causal work. There seems, however, to be nothing that restricts this understanding of realization to properties like dormitivity; it pertains to functional properties in general, at any level. In particular, the same point applies in the case of mental properties; there can be no serious causal work for any mental event or property to do – that work is already effected by its physical 'realizer'. So, any 'supervenient' mental

⁴ 'Supervenience as a Philosophical Concept', *Metaphilosophy*, p. 27.

⁵ J. Kim: *Mind in a Physical World*, (MIT Press, Cambridge, MA, 1998) p. 9. Unless otherwise stated, subsequent references to Kim's writing are to this book. In similar vein, to meet the objection that global supervenience fails to account for local dependency relations, Oron Shagrir has recently proposed that a *combination* of global and weak supervenience is a serious option for non-reductive materialists. See his 'More on Global Supervenience', *Philosophy and Phenomenological Research*, LIX, 1999, pp. 691-701.

property or power, on these assumptions, is superfluous, causally impotent. And the basic objection to epiphenomenalism was well expressed by Samuel Alexander in 1927:⁶

It supposes something to exist in nature which has nothing to do, no purpose to serve, a species of noblesse which depends on the work of its inferiors, but is kept for show and might as well, and undoubtedly would in time, be abolished.

Kim's argument turns on the causal closure of the physical domain, that is, the belief that tracing the causal ancestors and consequents of any physical event will never take us out of the physical domain. The argument is that if physicalists accept causal closure in this sense, they have no option but to renounce non-reductive physicalism. Only in that way can they account for the possibility of psychophysical causation.

The basic argument, repeated with only minor variations in many of Kim's papers, goes like this. Suppose *M* is a mental property, with causal powers, and that some instance of it is causally efficacious in bringing about an instance of another mental property, *M**. But, *ex hypothesi*, *M** is physically realized in its physical base *P**. Without the presence of *P**, *M** would not be present. So, *M* must have brought about *P**. The only coherent story is that the instance of *M* caused *M** to be instantiated *by causing its realization base, P*, to be instantiated*. So, a non-reductive physicalist is committed to 'downward causation'.⁷ But *M* has its own physical realization base, *P*. The presence of *P* is sufficient for the presence of *M*. It follows by causal transitivity that if *M* is causally sufficient for the presence of (an instance of) *P**, and thereby *M**, *P* is causally sufficient for both *P** and *M**. Accordingly, the hypothesized causal efficacy of *M* is superfluous; its physical realization base is what does all the causal work.

Consequently, Kim asks, what reason is there for not bypassing mental causes altogether and treating them as mere epiphenomena? The negative answer to this rhetorical question is firmed up if one accepts, as a physicalist is bound to, the causal closure of the physical world; mental events or properties can only bring about physical changes if they are realized in some physical base-properties. The purported mental causes are superfluous. So, either physicalists must give up their anti-reductionism, or else they must reject the possibility of psychophysical causal relations.

This argument is quite general. It is equally effective if we substitute "supervenient upon" for "realized by", and also, so Kim has argued⁸, if instead we substitute "emerges from". Higher states inherit their causal powers from the physical states that realize them, or upon which they are supervenient, or from which they are supposed to emerge. It is their 'realization base', or 'supervenience base', or 'emergence base' that does the work. Kim's diagnosis is that the trouble arises from the way non-

⁶ *Space, Time and Deity*, vol. 2 (Macmillan, London, 1927) p. 8.

⁷ The term 'downward causation' is from D.T. Campbell: 'Evolutionary Epistemology', in P.A. Schilpp, ed., *The Philosophy of Karl Popper* (Open Court, LaSalle, IL, 1974). Pp. 413-463.

⁸ In "'Downward Causation' in Emergentism and Non-reductive Physicalism' in *Emergence or Reduction?*

reductive physicalists try to combine two appealing ideas: that mentality emerges out of, and in that sense depends upon, the physical; and that, in spite of this ontological dependence, it begins to lead a causal life of its own, with a capacity to influence that which sustains its very existence. In short, the difficulty arises from combining ‘upward determination’ with ‘downward causation’.

Essentially the same argument also holds against anomalous monism. If every event that can be given a mental description can also be given a physical description, and the physical domain is closed, there is no significant causal work for mental events to do that is not already effected by those same events under their physical descriptions. Accordingly, citing events under their ‘mental’ description in causal explanations is misleading.

Kim then points out that there are two ways in which psycho-physical causation can be denied: one is to deny that there are mental events; the other is to keep faith with mental events but concede that they never enter into causal transactions with physical processes. So either a physicalist has to espouse so-called ‘eliminative materialism’, or else to move further in the direction of dualism, a dualism that posits a realm of the mental in total causal isolation from the physical domain. Kim’s argument leading to this conclusion therefore poses a physicalist with an invidious choice: either there is no emergence of causally efficacious properties above the base level, or the physical domain is not closed. For any serious physicalist, the latter is not an option. So, if one wants to be a physicalist, one cannot be a non-reductive physicalist; one has to embrace some kind of ‘eliminative’ or ‘reductionist’ program.

3. The Failure of Reductive Physicalism

This conclusion confronts any physicalist with a choice between reduction and emergence. For a physicalist, both options are unpalatable.

Suppose the physicalist accepts reduction. It follows that no science other than physics has a distinctive region of entities with their own properties and causal powers, which it is the proper province of that science to investigate. Or, at least, if it is conceded that chemistry, molecular biology, physiology, psychology, etc. have their own work to do, that is merely a convenient division of labour – necessary in our current state of ignorance about how to ‘reduce’ them to physics. But it is not an intrinsic division in the ontological map. In principle, physics is the only science. This position is likely to generate understandable tensions amongst scientists who would like to believe that they are practising different disciplines! (Some physicalists, including Kim, try to evade this conclusion by complaining that to confine the ‘physical’ to the particles and properties of physics is an excessively, and unjustifiably, narrow conception of the physical domain. They want to include the chemical, the biological – in fact everything dealt with by the natural sciences up to, but not including, the psychological – as ‘physical’. We will return to this manoeuvre in section 5.)

More seriously, this position is also more than a little embarrassing. For while reduction treats the psychological (and perhaps the biological) as epiphenomena, physicalists know that Carnap's program has failed, and they cannot point to any plausible bridging laws that are effective in explaining such phenomena. In the absence of relevant explanations, their doctrine is mere dogma.

Further, for a reductionist to deny some sort of emergence seems foolhardy – even for a philosopher! Emergence appears to be ubiquitous throughout the world. Just about everything that exists now, which is at least a potential topic of scientific interest, has emerged since the Big Bang. So, any purported scientific model of any phenomenon must be able, at least in principle, to account for the ontological and historical emergence of that phenomenon since the Big Bang. Any model that does not provide such an account has to be reckoned as thereby incomplete. In the light of this, we can turn around physicalism's exclusion of the very possibility of novel emergence and use such exclusions as a powerful negative criterion for assessing scientific theories. That is, *any purported model of X that makes the emergence of X impossible is thereby self-refuting*.⁹

On the other hand, accepting emergence, in the sense of higher-level, causally efficacious powers that are not explicable in terms of the lower-level powers of physical constituents, is to give up physicalism. Unless there is some way of explaining how higher-level causal powers are derived from lower-level powers, a physicalist is at a loss to explain how mental events can cause physical changes. It would seem that mental events can be causally efficacious in themselves. For a physicalist, that amounts to embracing dreaded dualism.

These considerations are what sharpen the alternatives facing physicalism, turning the choice between them into a dilemma: each option seems intolerable. How is a physicalist to decide between them? Now, whenever thought finds itself boxed in like this, facing an unbearable dilemma, it is time to re-examine the terms in which it has been posed.

Reflecting on the complex debates that have swirled around this issue, one feature stands out. On all sides, the presumption has been that the fundamental level of the natural world consists of micro-physical entities of some sort, with their primary properties. These physical entities are clearly some sort of particular. But what sort? Some physicalists believe that it does not matter, that the notion of a physical particular might be defined as an object, a concrete event, or whatever.¹⁰ However, the key commitment of physicalism is to some kind of *basic particulars*, which are the fundamental constituents out of which everything in the world is composed. Even those who argue for a wide sense of the 'physical' erect their definitions of that term upon basic

⁹ This is argued at greater length by Mark Bickhard, with D.T. Campbell, in 'Emergence' in *Downward Causation*, ed. P.B. Andersen, C. Emmerche, N.O. Finnemann & P.V. Christiansen (Aarhus University Press, 2000) p. 325.

¹⁰ So F. Jackson: *From Metaphysics to Ethics*, p. 6, fn. 5.

physical constituents. That is, however generous the definition, physicalists take as fundamental *elementary particles*, in the loose sense of the word “particle” commonly encountered in descriptions of quantum mechanical phenomena.

But recent developments in physics tell against that presupposition. What our best contemporary physics reveals is that *there are no elementary ‘particles’*, fundamental events, or some such particulars. There are only *processes* of various scales and complexity. That this is so has not been easy to see, since the behaviour of phenomena at the sub-atomic level has seemed to defy description in coherent and intelligible terms. For more than a century, physicists have struggled to reconcile the facts that these phenomena exhibit, in different experimental circumstances, both particle-like and wave-like behaviour. With so much being discovered that physicists themselves confess to finding weird, the habitual ways of talking about particles persist even though its use is confused and confusing. Even reputable physicists often give explanatory descriptions in terms of ‘particles’ in a way which, if taken seriously, would be incompatible with the physics they are trying to explain. Further, physics throughout the 20th Century has been dogged by a series of deep theoretical inconsistencies that are not yet fully resolved. But enough has now become clear for it to be evident that an ontology of elementary ‘particles’ – basic particulars – can no longer be sustained.

To make this clear, it will be helpful to review briefly the major shifts in physical theory in the 20th Century. The new era in physics was ushered in when Einstein proposed his special theory of relativity to overcome the contradiction between Newton’s laws of motion and Maxwell’s equations of electromagnetism. But that solution immediately generated another contradiction. According to special relativity, no object, and no force, can travel faster than the speed of light; yet in Newton’s theory, universal gravity is a force that is transmitted instantaneously. Again, it was Einstein who proposed the solution to this second conflict, with his general theory of relativity. Special relativity had held that the motion of one entity relative to another influences space and time. Now, according to general relativity, space and time can warp and curve in response to the presence of matter or energy. Such distortions spread out from one place to another, not instantaneously but at the speed of light, becoming more attenuated with distance.¹¹

The general theory of relativity provided for the first time an explanation of what gravity actually is, but until general relativity is integrated with quantum mechanics (as superstring theory now aims to do) ‘particle’-based physics has to treat gravity as an ubiquitous mystery. Indeed, far from overcoming the embarrassment that the apparent action at a distance of gravitational force causes for Newtonian corpuscular physics, recent experimental work has actually produced even worse anomalies for an elementary ‘particle’ metaphysics. For it has discovered a number of non-local phenomena that do not involve the transmission of force. For just one example, the Pauli exclusion principle

¹¹ Brian Greene: *The Elegant Universe* (Norton, New York & London, 1999) p. 5-6.

forbids two leptons (of which electrons are one sort) from being in the same quantum state. Two electrons in the same carbon atom cannot be in the same quantum state. This fact underlies all of chemistry, yet it is an inherently non-local, *relational* constraint. Why should that be? Why should this ‘state’ constraint hold for electrons within a single carbon atom, but not for electrons in two different carbon atoms? And why should these problems go away when we go beyond the quantum level? There are good mathematical explanations for the Pauli exclusion principle, but such non-local phenomena must seem utterly mysterious from the perspective of particle metaphysics – just as Newton admitted that the non-locality of gravity was a deep conceptual anomaly in his physics.

Many aspects of quantum mechanics are well confirmed, but quantum mechanics and relativity theory – both special and general – cannot both be correct. Not only is quantum mechanics incompatible with *general* relativity unless it is transformed in a way that can account for gravity, *special* relativity plus conservation of energy forces a *field* physics, and, therefore, a *field metaphysics*. This is because any transmission of energy from one ‘particle’ to another has to be across some space and will, according to special relativity, take some time. For example, if an electron oscillates in one place and thereby elicits a force on an electron in another place, that force can be felt only after a delay that honours the limitations of the speed of light. But given the law of conservation of energy, it makes no sense to suppose that that energy simply disappears as it leaves the first electron, only to re-appear sometime later, when it arrives at the second. So, there must be a field encompassing both places, *through* which the energy travels. Furthermore, quantum *field* theory involves its own well supported empirical consequences, such as the Casimir effect.¹² It is not just a mathematical but physically irrelevant fix.

In fact, the incompatibility between quantum mechanics and special relativity is overcome by quantum field theory, and superstring theory now aims to overcome the incompatibility between quantum mechanics and general relativity. But despite these achievements, a satisfactorily unified physics is still wanting. A field version of superstring theory has still to be developed, yet quantum fields are still needed to preserve special relativity together with conservation of energy, and to provide for the possibility of explaining non-local phenomena.

Although there is much for physics yet to resolve, the general direction is becoming clear enough for some conclusions of metaphysical significance to be drawn. According to our best science, there are no elementary ‘particles’, or basic particulars, at all; *everything* is composed of quantum fields.¹³ Quantum field theory shifts the basic

¹² Aitchison, I. J. R.: ‘Nothing’s Plenty: The vacuum in modern quantum field theory’, *Contemporary Physics*, 26(4), 1985, pp. 333-391; Sciama, D. W.: ‘The Physical Significance of the Vacuum State of a Quantum Field’, in S. Saunders, H. R. Brown (eds.): *The Philosophy of Vacuum* (Oxford University Press, 1991) pp. 137-158.

¹³ H.R. Brown & R. Harré: *Philosophical Foundations of Quantum Field Theory* (Oxford University Press, Oxford, 1988); Paul Davies: ‘Particles do not Exist’ in *Quantum Theory of Gravity*, ed. S. Christensen (Adam Hilger, Bristol, 1984) pp. 66-77; S. Weinberg: ‘The Search for Unity – Notes for a History of Quantum Field Theory’, *Daedalus*, 106(4) 1977, pp. 17-35;

ontology of the universe from micro-particles to quantum fields. What have seemed to be ‘particles’ are now conceptualized as particle-like processes and interactions resulting from the quantization of field processes and interactions. Those are no more particles than are the integer number of oscillatory waves in a guitar string. Each of the apparent particulars assumed by a physicalist ontology is a quantized field process.

It is just as well that the physicalist’s metaphysical model, which would reduce everything to interactions amongst elementary ‘particles’, is not supported by recent physics, since it is arguably incoherent. Either the basic particulars (particles) have extension, or they do not. If they have no extension, then a field view is forced in order to account for particle interactions, since the probability of such particles ever actually hitting each other is zero. If, on the other hand, these elementary particles do have finite extension, they pose intractable problems for physical theory. They could not be compressible; the velocity of transmission of force through their diameter would therefore be instantaneous; there would be extreme difficulty in explaining different kinds of interactions (gravity, electricity, etc.); and so on. If a move is made to a combination of particles and fields (the typical contemporary semi-sophisticated view), then all the significant issues are already granted anyway in the granting of fields. There are no ‘particles’, but, even if there were, so long as fields are granted at all, the micro-reduction motivation fails – and a strict particle view is not only factually false, but conceptually incoherent as well.¹⁴

Once we have made the conceptual shift required to free ourselves of the age-old prejudice of a substance metaphysics – of which particle metaphysics is the most recent manifestation – it is far from clear that there *is* any basic level. That is, there might well be no fundamental plane of organization, ‘lower’ than which it is not possible to go. Still, this much at least is clear: if the organization at some level cannot be shown to be *necessary*, then one could never have *logical* grounds which ensure that that level of organization was not itself emergent from a yet more basic one. And how would one demonstrate that necessity? One can always sensibly ask the question, why is it so? Therefore, even if there were a level of organization that is the lowest, that would be a contingent fact. Without any assurance of such a basic level upon which to ground his position, the physicalist’s claims become rather hollow.

A reductive physicalist might reply to all this that he has always been open-minded about which particular theory, or set of theories, physical science will eventually accept as the best confirmed. His position is simply that, whatever that account should turn out to be, every phenomenon will, in principle, be determined by the properties and

The Quantum Theory of Fields, Vol. I, Foundations (Cambridge) 1995; *The Quantum Theory of Fields, Vol. 2, Modern Applications*, (Cambridge) 1996; S. Saunders and H.R. Brown, eds: *The Philosophy of Vacuum*, 1991; Cao, T. Y.: ‘Introduction: Conceptual Issues in Quantum Field Theory’ in T. Y. Cao (ed.): *Conceptual Foundations of Quantum Field Theory*(Cambridge University Press, Cambridge, 1999) pp. 1-27.

¹⁴ Bickhard & Campbell: ‘Emergence’ in *Downward Causation*, p. 332, fn. 11.

powers of the basic particulars posited by that account. So, if some refinement of quantum field theory should prove to be the best that physics can provide, why can he not accept quantum fields as the ultimate reality, and regard everything else as supervenient upon and reducible to their behaviour? By adjusting his physics, he retains his physicalist metaphysics.

The adjustment in one's physics, however, is not so simple; the particle view is deeply related to the micro-reductivist position. Particles do not *have* a configuration. They are points. But they do *participate* in configurations relative to each other. Particles are the purported locus of causal power, and the configurations in which they participate are 'merely' the stage setting for the working out of the particle causal interactions. That is, causal power is resident in entities that are not configurational, and configuration or organization is merely a stage setting, with no causal power of its own. In this view, organization is factored out as a legitimate locus for causal power – it's just stage setting – and such delegitimation succeeds because there is a non-configurational candidate available to be such a causal locus: particles. The particle view supports the reductionist view by motivating the elimination of configurations as legitimate loci of causal power. So, abandoning a particle physics in favour of a quantum field physics, a process metaphysics, is not an innocent choice with respect to the issues at hand.

The critical point is that quantum field processes have no existence that is independent of their configurations: quantum fields *are* processes, and can *only* exist in various patterns. Those patterns come in many sizes, of many different physical and temporal scales, some as large as a human person, or a social institution – but they are all equally patterns of processes. There is no 'bottoming out' level in quantum field theory – it is *patterns* of process all the way down, *and* all the way up.

That is the rub. To be a reductive physicalist (or an 'eliminative materialist') at all, is to believe that 'higher-level' entities are nothing other than complex configurations of lower-level entities, in such a way that the higher-level properties and powers are explicable in principle in terms of the properties and powers of the lower-level entities (or at least, determined by them). Consequently, some entity is reducible just in case it is a configuration of lower-level entities. But now the supposed base-level entities are nothing but configurations of process as well! If there is no 'bottoming out', there are no bases to which all other phenomena can, even in principle, be reduced. Our reductive physicalist has lost the ground on which he wants to stand. *If being configurational makes a property or power epiphenomenal, then everything is an epiphenomenon.* That is the *reductio ad absurdum* of this position.

A reductive, or eliminative physicalist might have one last try. He might accept that the supposed base-level entities are nothing but configurations, but try to rescue physicalism along the following lines. Suppose we restrict our view just to those processes discernible at the finest level of resolution, however small that level might be. Why cannot a physicalist claim, in the light of contemporary physics, that these fine-grained processes are what fulfils the role of basic particulars? The physicalist issue

simply is: Are all the truths determined at that level sufficient, in principle, to determine all the truths in the world? If so, that is all the physicalist needs.

This riposte is deeply ambiguous; the answer to the question depends upon whether those truths determined at the level of finest resolution include all the *relational* truths. The physicalist is committed to this being so. But not all the relevant relational truths are discernible at the ‘finest level of resolution’; some, probably most, of the patterns that are causally significant are of a larger scale than that. If we are restricted to just those that are of a small enough scale to be *discernible* at such a fine grain, then the answer to the question must be negative. Reference to ‘all the truths determined at the finest level of resolution’ glosses over this crucial issue of the *scale* of causally efficacious patterns. If these truths are not restricted to patterns of a small enough scale to be discernible at such a fine grain, we may include relational truths at the scale of a whole human person – or larger. Then the answer to our question is clearly yes, but the position cannot claim any longer to be a form of *eliminative physicalism*.

Now, if everything is configurational, there is no reason not to accept that complex configurations can generate properties and powers that are genuinely emergent. Reductive physicalism has failed, just as non-reductive physicalism failed. That means that a physicalist ontology is untenable in any form. The choice posed by the dilemma has been determined. Any metaphysics that respects contemporary physics has to accommodate genuine emergence, and not treat it as merely ‘apparent’.

Does this amount to a vindication of dualism, as the physicalist alleges? Only if one doggedly persists in thinking in Cartesian terms! The readiness with which any challenge to physicalism is turned into an accusation of dualism shows how much the physicalist program is formulated within a Cartesian framework. For all its guise of contemporaneity and scientific seriousness, physicalism is simply Cartesianism with one of his two kinds of substance lopped off. That is just another manifestation of how unfortunate has been the framing of issues about basic ontology in terms of a Cartesian mental/physical dichotomy. Further, it shows how deeply a substance metaphysics is entrenched, so that it has become difficult even to conceive how science could proceed if explanations ultimately do not refer to kinds of basic particulars (particles) and their properties.

4. Emergent ‘Levels’ and Causal Drain

It is clearly time to call into question this loose talk of the ‘mental’ and the ‘physical’. The ‘mind-body problem’, and philosophers’ attempts to deal with it, perpetuate an unfortunate dichotomy, as if Nature is bipartite, divided into two discrete realms: the ‘physical’ and the ‘mental’.¹⁵ All those who address issues posed in those terms are operating in a Cartesian framework, even if they think of themselves as

¹⁵ As William Lycan has rightly complained. See his *Consciousness* (MIT Press, Cambridge MA, 1987) p. 38f.

‘physicalists’, or ‘materialists’. In this scheme, the biological, including the physiological, is labelled as physical, as if that label was unproblematic. Yet, one does not need to know much biology to recognize that this scheme has gone seriously wrong. Just a little reflection on the differences between a living body and a dead corpse calls into question this lumping together of everything non-mental as ‘physical’. We all know that right from the time of death (whatever are the criteria for determining that time, and whether or not death is instantaneous or occurs over a shortish period) the corpse begins to disintegrate. It literally disappears, unless the natural process of disintegration is artificially suspended – by freezing, embalming, or placing in formalin.

Obviously, there is something to being a living organism that is more than the assemblage of atoms and molecules. (In which sense there is ‘something more’ is yet to be determined – recognizing this is not the same as committing ourselves to Bergson’s *élan vital*.) The development of the physical and biological sciences since the time of Descartes has replaced his dichotomy with a multi-layered model of the world as stratified into different levels, in a micro-to-macro hierarchy. Consequently, the bipartite Cartesian model has long been outdated, and in the new model of Nature, entities, characterized by their distinctive properties and processes, emerge (in some sense) out of the entities, properties, and processes of the levels below it.

Kim correctly points out that the way this layered picture of the world has been standardly understood in debates on a variety of issues in metaphysics and philosophy of science is not as innocent as it looks. He suggests that most thinkers will acknowledge a familiar picture. At the bottom is a level consisting of elementary particles, or whatever our best physics is going to tell us are the basic bits of matter out of which all material things are composed. As we go up the levels, we successively encounter atoms, molecules, cells, multi-cellular organisms, human beings, social groups and institutions, etc. But what is the ordering relation that generates these levels, and how is that relation to be understood? Kim suggests that this seemingly uncontroversial picture assumes (tacitly, if not explicitly stated) that this hierarchy is generated by the mereological (part-whole) relation. That is,

entities belonging to a given level, except those at the very bottom, have an exhaustive decomposition, without remainder, into entities belonging at the lower levels. Entities at the very bottom have no physically significant proper parts.
(p.15)

Now, in the usual physicalist’s scheme, what looks like ‘emergence’ is glossed as a mereological kind of supervenience, which, as we have seen, is taken to lapse inevitably into epiphenomenalism. Physical ‘realizers’ pre-empt the causal powers of the second-order, or functional properties that they realize. Accordingly, if the physicalists’ story is right, this hierarchical model of Nature, layered mereologically in terms of constituents at each level, seems to entail that causal powers at any level are ‘realized’ by the micro-properties of the constituent entities. Consequently, it looks like all causation seeps down to the level of the basic particles, as the only locus of causal powers. Worse than

that: if, as has sometimes been suggested, there is no rock-bottom level, causal powers would drain away into a bottomless pit, and there would not be any causation anywhere (p. 81)!

In *Mind in a Physical World* Kim, rightly in our view, takes this prospect as a serious challenge to the plausibility of physicalism. He has a twofold strategy to meet it: he introduces a crucial distinction between higher-*order* and higher-*level* properties to plug the threatened drain of causation down the hierarchy of levels; and he tries to reinstate the simple physical/mental dichotomy.

Higher-*level* properties are ascribed to entities of a higher ontological level than the entities to which lower-level properties are ascribed. Given the mereological way of structuring the hierarchy of levels, entities at a higher level are composed of micro-entities at a lower level. Usually, the functional role/realizer relation has likewise been taken to generate a hierarchy of levels, both of explanation and ontologically. But Kim now argues that the realization relation does not track the micro-macro relation (p. 82). His new proposal is to call second-*order* properties those functional properties that are realized in base (or first-order) properties of the *same* entities and systems. ‘Orders’, therefore, do not generate any new ontological levels, and the causal powers of higher-order properties devolve only to their designated first-order property or properties on the same level. That is, second-order properties merely specify or pick out particular base properties. For example, if the base set comprises colours, then the property of *having a primary colour* would be a second-order property, attributable to the same entity as that to which colours could be attributed (p. 20).

Kim now argues that when we speak of microphysical or mereological supervenience, we are speaking often quite loosely. Suppose we say that some property *P* of something *s* ‘supervenies’ on the microphysical properties or facts about *s*. What we mean to say, he suggests, is that whether or not *s* has *P* is fixed once the micro-constituents of *s* and the properties *and relations* [our emphasis] characterizing these constituents are fixed. That means that the base property on which *P* supervenes is a micro-based property, “the property of having such-and-such proper parts that have such-and-such properties and are configured by such-and-such relations” (pp. 85-86). Although this is a micro-*based* property, it is a *macro*-property of *s*, which does not belong to any of its constituents. Such a supervenient property and its supervenience base, therefore, are both macro-properties of the same (level of) entity.

For example, it is one and the same Seconal pill that a person suffering from sleeplessness might ingest that has both dormitivity and the chemical property which realizes dormitivity; and the person who is in pain is the same entity who has its c-fibres firing. Hence, Kim argues, functional properties and their realizers are at one and the same ‘level’; they do not exist at different levels in a hierarchy of the sciences. So, when we talk of second-order properties and their realizers, Kim concludes, there is no movement downward, or upward, in the hierarchy of entities and their properties ordered by the micro-macro relation. Therefore, the realization relation does not ‘drain’ causation

down to entities of the lowest level, or perhaps away altogether. Rather, causation remains at the same level.

5. Kim's Endorsement of Emergence

Let us now explore what Kim has committed himself to by this defence. The first point to note is that he does not accept that the levels in the standard hierarchical model are generated by mereological supervenience. While he offers a clear definition of a second-order property, he is never quite so clear about how he thinks we should now understand levels *per se*, but one sense of level is that of levels of micro-based properties. A micro-based property of some object is a macro-property that is constituted by its micro-constituents – that is, by its non-overlapping proper parts, each with its own appropriate property, and with some appropriate *relation* holding among them all (p. 84).

In so doing, he endorses the emergence of causally efficacious properties:

macroproperties can, and in general do, have their own causal powers, powers that go beyond the causal powers of their micro-constituents. (p. 85)

He does not argue for this, nor offer any general explanation of how these emergent causal powers arise, but establishes the point with examples. Having a mass of ten kilograms is a causally efficacious property of a table but of none of its constituents. Water has causal properties different from those of hydrogen and oxygen. A neural assembly consisting of many thousands of neurons has properties whose causal powers go beyond those of its constituent neurons, or sub-assemblies. And humans have causal powers that none of our individual organs have.

For these reasons, Kim now argues that these micro-based properties do not *supervene* on their constituent properties taken individually or as a group. Rather, he says,

they supervene on specific mereological *configurations* involving these micro-properties – for a rather obvious and uninteresting reason: they *are* identical with these micro-configurations. (pp. 117-118)

What he now allows is that appropriate configurations of micro-based properties can generate new causal powers:

the fact that we can micro-structurally explain why a micro-based property has a certain set of causal powers does not mean that these causal powers are identical with the causal powers of its micro-constituents. Micro-reductively explainable causal powers may be new causal powers, net additions to the causal structure of the world.... It follows then that we must grant novel causal powers to micro-based properties at higher levels – novel in the sense that these causal powers are not had by any lower-level properties that constitute them. And, as we saw, the supervenience argument does not apply to them, and their causal roles are not threatened by the supervenience argument. (pp. 117-118)

We believe that Kim has reached an important insight in recognizing that *configurations of constituents are what generate the emergence of higher-level causal powers*, as we will shortly elaborate, although he has not fully thought through its implications. His concern is solely to prevent all higher-level causation from being drained of causal potency. Kim himself now can speak of ways “in which a macro-property may be emergent” (p. 117). But he invokes this insight only to argue that the case of micro-based properties is not at all parallel to the case of supervenient psychological properties:

So microphysical, or mereological, supervenience does not track the micro-macro hierarchy any more than the realization relation does; the series of supervenient properties, one mereologically supervenient on the next, when we go deeper and deeper into the micro, remains at the same level in the micro-macro hierarchy, just as the properties ordered by the realization relation stay at the same level. This means that the supervenience argument, which exploits the supervenience relation, does not have the effect of emptying macro-levels of causal powers and rendering familiar macro-objects and their properties causally impotent. (p. 86)

Thereby he has given away most of the issues at stake in his earlier arguments against the possibility of such emergence, without ever acknowledging that he has done so. We surmise that what has prevented Kim from recognizing the full significance of his endorsement of emergence is that he remains committed to physicalism:

None of this is in conflict with the basic commitments of physicalism. Physicalism need not be, and should not be, identified with micro-physicalism. (p. 117)

In particular, his thinking about these issues is dominated by his long-standing preoccupation with ‘the mind-body problem’ and the concept of ‘mental causation’. In the case of supervenient psychological properties,

the physical base properties, presumably certain neurobiological properties, are at the same level as the psychological properties: they are both had by human beings and other sentient creatures. This is part of what generates the problem about mental causation: the causal role of a mental property had by me is threatened with pre-emption by another property, a neural property, also had by me. My causal powers seem fully explicable not only *in terms of* but also *as* the causal powers of my neural/biological/physical properties. (p. 117)

The last-mentioned point expresses Kim’s view that we should think of functionalization as reduction, and insofar as mental properties are functionalizable, they are subject to the supervenience argument. As he sees it, the only way to save mental properties from being rendered epiphenomenal and superfluous is to identify them with their ‘physical realizers’. That way, functionalizable mental properties will turn out to have causal powers – but only by virtue of their identity with these ‘physical’ properties – and mental properties, thus understood, bring no *new* causal powers into the world (p.

118). (The phenomenal properties of conscious experiences, he notes, may not be functionalizable, and therefore, there might be no way to account for their causal efficacy within a physicalist scheme.)

Underlying this train of thought is his continuing insistence on a crucial premise of his earlier arguments: the causal closure of the physical domain. This leads him to argue that it is important that micro-based properties be counted as physical “for otherwise the physical domain won’t be causally closed” (p. 114). The only way he can hold on to this, consistent with his endorsement of emergence, is to defend a wide sense of the word “physical”, to include the biological and neuro-physiological, indeed every emergent property short of conscious experience. Accordingly, he takes all of the following as ‘physical’: first, any entity aggregated out of physical entities; second, any property that is formed as micro-based properties in terms of entities and properties in the physical domain; third, any property defined as a second-order property over physical properties; and perhaps conjunctive properties, if we want them (pp. 114-115).

It is illuminating to compare the position Kim has arrived at with that propounded by the classical British emergentists: Samuel Alexander, C. Lloyd Morgan, and C.D. Broad. They too developed a hierarchical view of natural processes, with novel ‘qualities’ emergent at each higher level. An emergent quality comes into being specifically as a result of the complexity of the organization of constituents. For Alexander, “the new quality and the constellation to which it belongs are at once new and expressible without residue in terms of the processes proper to the level from which they emerge.”¹⁶ Thus, “life is at once a physico-chemical complex and is not merely physical and chemical, for these terms do not characterise the new complex which in the course and order of time has been generated out of them.”¹⁷ Similarly, Morgan calls ‘atomicity’ the quality that emerges from “a swarm of atoms *with relational structure*”. On one line of advance, when atoms and molecules “are *grouped in new relations*” their expression is crystalline form. On another, we find organisms with a different kind of natural relatedness, which gives the qualities of vitality and mentality.¹⁸

In earlier papers, Kim had examined the views of these British emergentists with some care. The kind of emergence advocated by these philosophers has always seemed to be the notion that lurked behind his earlier arguments that non-reductive physicalism is untenable. For example, in his 1993 paper he identifies that position as “the first systematic articulation” of the non-reductivist approach and argues that the latter “is best viewed as a form of emergentism”. Since he takes his supervenience argument to be equally effective against this emergentism, how are we to understand his more recent endorsement of emergence? His proposal that configurations of constituents generate the emergence of novel causal powers now seems very close to the British emergentism he

¹⁶ *Space, Time and Deity*, p. 45.

¹⁷ *Space, Time and Deity*, p. 46.

¹⁸ *Emergent Evolution* (Williams & Norgate, London, 1923) p. 35, our emphases.

once attacked, despite his denial that the new causal powers that he now agrees are emergent are so “in the sense of classic emergentism”.¹⁹

Perhaps Kim’s new position still differs from that of the British emergentists in the following way. An essential part of their position was that emergent qualities are not derivable from the qualities of their lower-level constituents; the *organization* of the latter is what is crucial. Broad was the first to give a detailed explication of the concept of an emergent property. In his analysis, emergent theories differ from ‘mechanist’ theories in that the former held – and the latter denied –

that the characteristic behaviour of the whole could not, even in theory, be deduced from the most complete knowledge of the behaviour of its components, taken separately or in other combinations, and of their proportions and arrangements in this whole.²⁰

That a certain organization of lower-level constituents gives rise to some emergent quality is simply a brute fact of nature. Its existence cannot be denied, but nor can it be explained. It has simply to be accepted, in Alexander’s memorable phrase, in ‘an attitude of natural piety’.

Now, while Kim does not address the issue of derivability, we can assume that his micro-based properties *are* supposed to be derivable from the micro entities, plus their properties *and their relations*. After all, they are decomposable into them, and determined by them. That, we conjecture, is why he still sets aside qualia as perhaps requiring ‘emergence’, presumably of the British emergentist sort. All else could be explained via various levels and orders, but qualia might resist functional explanation, thus resist inclusion in the kind of physicalism that he outlines. (This, of course, assumes that functionalism is the only way to approach such issues.) If this is right, we can read Kim as still differentiating himself from the British emergentists, and as taking their positions to be definitive of the classical model of emergence. Kim wants to be a physicalist in his very broad sense, but does not recognize the process/relational derivability version of emergence as being real emergence of a kind worth exploring. He deploys it to block causal drain, but does not pursue it further.

To prevent misunderstanding, we should point out that we have no interest in trying to revive British emergentism; it fails to be naturalistic, not to mention that it is ad hoc and unreal. Furthermore, for what reason would anyone want an emergentism of that sort? What does it save or preserve? Nothing of importance. Yet, other models of genuine emergence are possible. Science has moved on a great deal from where it was in the 1920s. We are now able to model how organizations of processes give rise to

¹⁹ ‘Does the Problem of Mental Causation Generalize?’, p. 292, fn. 13. One obvious difference is that Kim continues to hold out against conscious mentality being emergent.

²⁰ *Mind and its Place in Nature* (Routledge & Kegan Paul, London, 1925) p. 59. For a helpful discussion of this, see A. Beckermann: ‘Introduction’ to *Emergence or Reduction?*, pp. 15-18.

emergent properties and powers in ways that are more plausible and detailed than the vague assertions of the British emergentists.

For this reason, we believe it important to observe that, within the conceptual framework developed in *Mind in a Physical World*, Kim opens up the possibility that mental properties too might be emergent macro-properties, precisely because they *are* ‘configurations’ – organizations of neuro-physiological processes. And he does accept that neurobiological processes are emergent macro-properties. There is nothing in the model of emergence that he now accepts which would debar functionalizable mental properties from being genuinely causally powerful, and with novel causal powers. That is, apart from his wish to preserve the mental/physical dichotomy, mental properties could be genuinely emergent, with novel causal powers, just in case they can be modelled in terms of lower level ‘parts’ *plus* their ‘configurations’. With a caveat against the static characterizations involved in this way of expressing the point, this is, we would urge, precisely what constitutes emergence, including mental emergence.²¹ Mind, or mental properties, emerge in specific *organizations* of lower level processes.

Kim has prevented himself from recognizing the far-reaching significance of this possibility because of his insistence that neurobiological properties are at the same level as psychological properties. As he loosely (and misleadingly) puts it, they both are ‘had’ by human beings and other sentient creatures. Now, one aspect of this insistence is Kim’s explication of function in terms of functional properties as intermediary causal roles. This focus on singular functional properties is what permits him to run the supervenience argument against functions relative to their (physiological) causal realizers, since supervenience too is restricted to singular properties (see below). Nevertheless, the focus on singular functions prevents Kim from recognizing that his own framework would seem to permit the possibility that (complex) configurations of functional relations could yield novel, emergent, properties and causal powers. That is, Kim can take function to be of higher order, not higher level, relative to physiology only because he has only defined *singular* functional properties, and has not paid any attention to functional *organizations*. The emergents from functional organizations, plausibly, might include psychological properties that are macro-based, emergent, relative to neural processes. We also note that Kim has not addressed *normative* function at all in his discussion of these issues.²²

Let us now step back to review Kim’s chain of argumentation. The first point to note is that he begins his analysis with an understanding of the ‘realization’ relation as

²¹ See, for example, Mark Bickhard: ‘A Process Model of the Emergence of Representation’, in G. L. Farre, T. Oksala (Eds.) *Emergence, Complexity, Hierarchy, Organization*, Selected and Edited Papers from the ECHO III Conference. Acta Polytechnica Scandinavica, Mathematics, Computing and Management in Engineering Series No. 91, Espoo, Finland, August 3 - 7, 1998, 263-270; ‘Emergence’ in *Downward Causation*, p. 327; ‘Levels of Representationality’, *Journal of Experimental and Theoretical Artificial Intelligence*, 10, 1998, pp. 179-215.

²² See, for example, M. Bickhard ‘Emergence’ in *Downward Causation*, p. 327; ‘Autonomy, Function, and Representation’, *Communication and Cognition*, in press.

being the *converse* of the ‘emergence’ relation, i.e., that *P* realizes *M*’ is true if and only if ‘*M* is emergent from *P*’ is true. That being so, we should expect his supervenience argument to count equally against any notion of emergence entirely determined by the behaviour of particular constituents. And he does so argue. Suppose some entity or property were emergent from some set of base conditions, and were to have emergent causal powers. Kim’s asks: why is not that emergent entity or property causally redundant? In other words, why cannot those base conditions simply supplant that emergent entity or property as the cause of any effects that the latter is supposed to bring about? Why do the base conditions not do all the work in bringing about the supposed effect, and suffice as an explanation of why the effect occurred?²³ He then re-presents his general argument, with an ‘emergence base’ in place of the ‘realization base’, for the conclusion that if emergent properties exist, they are causally inert and hence largely useless for theoretical scientific purposes.

The puzzle is why does Kim not take this argument to demolish his own endorsement of emergence. In this book, the only characterization of micro-based properties that would block the supervenience argument is the critical involvement of relations – ‘configurations’ – in the emergence of micro-based properties. What enables him to posit the possibility that relations are causally potent is his maintaining that micro-based properties are based on *specific mereological configurations* of constituents, but not on those constituents themselves, taken individually or as a group. In this new perspective, emergence and realization are no longer converses.

That is, on one hand, Kim has defined the supervenience relation as holding between properties – singular properties (e.g., 9-10, 17-18, 39, 41-46). One property is supervenient on another property; a property cannot, according to this definition, be supervenient on a relation or configuration or organization. The micro-macro-hierarchy, on the other hand, *is* defined in terms of relations, configurations. So, mereological supervenience turns out to be supervenience on a singular macro-property, a macro-property that happens to be micro-based. But the supervenience relation is not transitive through relational levels; something can be supervenient on a micro-based macro-property *without* thereby being supervenient on the micro-entities and properties and *relations* upon which the macro-property is based. Thus, supervenience does not track the micro-macro-hierarchy, and causal drain is blocked.

Kim has (apparently) evaded causal drain, but there are serious problems. First, why should we accept his definition of supervenience, complete with his exclusion (by omission) of relations? It blocks causal drain, but only via definitional stipulation, and stipulation that violates the basic sense of supervenience: no higher level differences without accompanying differences at the lower level, *including relational differences*. It is an arbitrary definitional stipulation that is motivated as a solution to the problem of causal drain, but it is an ad-hoc and merely apparent solution.

²³ This adaptation of his general argument is in his ‘Making Sense of Downward Causation’ in *Downward Causation*, p. 318.

The arbitrariness and ad-hoc character of these definitions, and their ultimate failure, is even more strongly highlighted by the realization that the positions and arguments that Kim bases on them are, nevertheless, *still* vulnerable to causal epiphenomenality and causal drain style arguments that Kim has developed in previous publications. In particular, elsewhere²⁴ he argued explicitly that higher-level configurations of micro-particles could generate novel causal *regularities* in virtue of those relations, but that such regularities would themselves be *no more than the working out of the causal interactions of the basic particles involved in those configurations*. Such configurations, therefore, would not generate any novel, emergent, causal *powers*.

That is, the regularities would be manifest, but epiphenomenal. In these earlier arguments, mereological supervenience *was* taken to include relations in the supervenience base,²⁵ and the standard supervenience argument was simply run on that definition, with the standard conclusion of the epiphenomenality of macro-causation. The only characteristic of Kim's book that blocks that argument and its conclusion – and thereby permits his avoidance of causal drain, *and* his endorsement of a configurational kind of emergence – is the omission of relations from the definition of a supervenience base. But, again, why should that be accepted, and what blocks running his earlier argument in spite of his definitions? Higher level causal novelty is just the working out of lower level causal processes in whatever configurations they happen to occur, and all genuine causality is at such lower levels: all macro-causality is epiphenomenal.

A technical perspective on these points derives from the question of whether or not a micro-based macro-property is supervenient – it is clear that it can be a supervenience *base*, but is it itself supervenient? If it is, then: 1) here is an instance of the supervenience relation that *does* track the micro-macro-hierarchy; and 2) what is its supervenience base?

If it is *not* supervenient, then we have a clear violation of standard notions of supervenience – and a violation that is unmotivated in Kim's discussion (though it seems to be the position taken in the book). Worse, it is not only unmotivated, but it seems clearly unacceptable relative to earlier definitions and arguments for causal epiphenomenality and causal drain. Causal epiphenomenality is not just a semantic problem, to be defined away by 'better' definitions of supervenience. If, on the other hand, it *is* supervenient, then the only candidate for its supervenience base is the *configuration* of micro-properties and elements,²⁶ but, if that is the case, then either causal

²⁴ e.g., Kim, J.: 'Epiphenomenal and Supervenient Causation' in D. M. Rosenthal, ed.: *The Nature of Mind*. (257-265). (Oxford University Press, 1991) pp. 257-265, reprinted in Kim, J.: *Supervenience and Mind* (Cambridge University Press, 1993) pp. 92-108.

²⁵ Kim, J.: *Supervenience and Mind*. (Cambridge University Press, 1993) p. 101.

²⁶ Assuming a property cannot be supervenient on itself. But, even if supervenience is held to be a reflexive relation (how could that make sense?), that only introduces a technical complexity into the issues, and does not alter the basic points developed above.

drain is reinstated, or else there is some unanalyzed special nature to relations that blocks the standard argument.

We, of course, hold the latter position, but, as discussed above, that position is unmotivated and indefensible within Kim's framework – within a particle metaphysics – within the framework of physicalism. The special nature of relations holds only within a process metaphysics, such as is forced by quantum field theory. So, on this analysis, Kim has partially understood the importance of relations, but was unable to develop that insight in a thorough-going way because he remained caught in an incompatible particle metaphysics.

To recapitulate, in this book, Kim does not address and does not justify this switch in definitions. But the argument about the working out of the causal interactions of the lowest level particles in higher levels of organization would seem to hold just as strongly for the framework in this book as ever. Just because Kim is now using supervenience in a strict singular property sense, rather than in a stronger sense, does not mean that the old arguments are no longer valid, or no longer have purchase. We need merely to introduce “not simply property” supervenience and then rehearse his own previous arguments for the conclusion that micro-based properties do not generate any new or novel causal powers at all. There is nothing in his argument (reported in section 2 above) that justifies restricting supervenience to singular property supervenience – to mereological supervenience as defined in this book. On the contrary, all that the argument requires is that higher-level properties always be realized in physical bases, which are composed of ‘constituents’ and their properties. For all Kim says in this book, micro-based properties only yield novel causal *regularities*, not novel causal powers, in the working out of the causal interactions among the basic particles. That is, the differing conclusions he has now drawn are the result of a semantic sleight of hand.

In a significant footnote in a recently published paper, Kim betrays some uneasiness on this issue. In that paper, he repeats his general argument, rehearsed in section 2 above, adapted in order to challenge the very possibility of causal emergence.²⁷ He allows that complex systems can bring new causal powers into the world, powers that cannot be identified with causal powers of more basic, simpler systems. But he believes it debatable whether these new causal powers coincide with ‘emergent’ properties “as standardly conceived”. But as a precaution, he now would like “to restrict the argument to consciousness and other standard examples of emergent properties”.²⁸

²⁷ The paper is his ‘Making Sense of Downward Causation’ in *Downward Causation*, p. 318.

²⁸ Even within this restriction, however, Kim faces unpalatable consequences. If mental properties (perhaps with the exception of qualia) are functionalizable, given Kim's understanding of the ‘realization’ relation, they cannot be emergent relative to the neuro-biological level. But is that relation – realization – the converse of emergence? Ansgar Beckermann has argued that Kim has gone wrong here (‘Introduction’ to *Emergence or Reduction?*, p. 18). Beckermann offers a stronger definition of the ‘realization’ relation:

We have identified a number of reasons that might explain why Kim seems to have missed these salient points:

- He has concentrated his focus so much on related but secondary issues, such as reduction, realization, functionalization, etc., that this basic shift regarding the place of relations in the definition of supervenience and its importance is overlooked;
- His earlier opposition to the very strong British emergentist version of emergence obscures the fact that he has himself endorsed a very powerful model of emergence here – one different from the British emergentist model – and arguably has endorsed all the emergence that anyone needs to account for mental causation;
- He is distracted from that by his worry that qualia might resist functionalization, and thus reduction. That is why, in the final section of the book, he poses a dilemma to those who hold out on functionalization for at least a selected group of mental properties: either embrace a serious form of dualism, or stay with physicalism (at least in the sense of causal closure). The latter option, however, would require a physicalist either to retain supervenient and yet irreducible mental properties – say, qualia – but accept their causal impotence, or else to deny the reality of these irreducible properties. In this way, he devolves the British emergentist position and all of its issues, including the supervenience argument, upon the status of qualia. The basic thrust of our analysis of his position remains untouched by this manoeuvre;
- His thinking is dominated by ‘the mind-body problem’ and so he remains trapped in an essentially Cartesian conceptual framework. This results in his continuing commitment to ‘the causal closure of the physical domain’. Accordingly, he can only allow psychological properties to be causally efficacious if they are functionalizable and realized in physical micro-structure (in his broad sense of “physical”). Mental properties (apart perhaps from qualia) supervene upon neuro-biological properties, but are at the same level as them, and so not emergent from them. The neuro-biological properties are emergent in his new micro-based property sense, but because he has redefined the physical to include the full micro-macro-hierarchy, all cases of the kind of emergence he does admit remain within the domain of the physical, and so do not disturb his basic physicalist picture. In particular, the dichotomization and diremption of the physical from the (Cartesian) mental remains;

If a system S is in a (mental) state F at time t , F can be said to be realized at t by the (physical) state G if and only if S is in G at t and G has in S all the (monadic, relational, etc.) features that are characteristic for states of kind F .

Not only is this a more precise definition, it has the virtue that it makes realization the exact *complement* of emergence. (We must, however, enter a caveat against speaking of ‘states’, when both mental activity and the neuro-physiological level which ‘realizes’ it are *processes*.) A process or property F of a system S will then be emergent if and only if it is *not* realized by a lower-level state of S . Kim’s revised position, if it is to protect emergent powers from his own reductive argument, seems to require this stronger notion of realization. But then, if mental powers are emergent from *configurations* of neuro-physiological processes, they are not *reducible* to those particular processes.

- Also, the whole notion of micro-based properties seems introduced solely in order to block the generalization argument and the related drain of causal power down to the lowest levels. His preoccupation with that challenge seems to obscure his recognition of the full significance of what he has done in constructing this block – in particular, to endorse configurations as legitimate loci of (potentially emergent) causal power.

Since we would never wish to endorse the kind of ‘emergence’ proposed by the British emergentists, the core issues for us remain also untouched by Kim’s shifts here, but for very different reasons from his own. For us, the diagnosis of the original problem is the delegitimation of organization as a potential locus for (emergent) causal power. Nothing in this book resurrects organization as a legitimate potential locus of causal power against Kim’s own original arguments. What he has now done is to insert relations – organization – as a locus of emergent causal power, but he never addresses his own previous arguments that would debar that move. He introduces relations by definition and by example, not by argument. Once we recognize the artificiality of the restriction of supervenience to strict property supervenience, and the related introduction of relations in his definition of micro-physical properties, those definitions and examples do nothing to block the drain of causal power down to the lowest levels.

The crucial point remains that, assuming causal closure of a particle world, higher level organization is causally superfluous relative to the working out of the causal powers of the most basic constituent particles – unless organization, configurations, or relations can themselves be legitimate loci of genuine causal power. If they can, then new organization can yield emergent causal power, but, in that case, not all causal power is resident in particles, or in whatever micro-particulars a physicalist prefers. Kim has given no reason why organization *can* be a locus of causal power, given the particle metaphysics that he assumes. He has given no reason why any causal regularities at a higher level of organization are not merely ‘causally epiphenomenal’.

Further, when we consider Kim’s endorsement of emergence from configurations together with his generous definition of “physical”, it turns out, as Kim acknowledges, that mental processes should count as both emergent and yet ‘physical’. (By extension, the same can be said for the behaviour of social groups and institutions.) Mental properties, modelled in this way, would be novel in the sense that their causal powers are not determined by any lower-level properties that constitute them. The supervenience argument would not apply to them, so it does not threaten their causal roles. In fact, there would be nothing in the world that is not physical *in the sense defined*, yet the argument has conceded everything that an anti-physicalist would want – genuine emergence, irreducibility, causally efficacious mental properties, downward causation – everything short of Cartesian mental substance. And why should anyone still want that? This result empties the term “physicalism” of any serious metaphysical significance.²⁹

²⁹ In addition, C.M. Hansen (in ‘Between a Rock and a Hard Place: Mental Causation and the Mind-Body Problem’, *Inquiry*, 43(4), 451-491) points out that Kim has salvaged the causal

6. Non-linearity and Downward Causation

Clearly, if we are to find a way forward, we need an alternative and scientifically justifiable concept of emergence. Simply to put ‘configurations’ into the specification of ‘base’ properties, and leave the rest of the physicalists’ ontological ingredients untouched, is too arbitrary and *ad hoc* to provide a credible concept of emergence.

We note also that where there are ‘emergent’ properties and powers, it seems that ‘downward causation’ should also be possible. The latter idea is that complex systems, in virtue of their macro-level properties, are able to bring about changes at lower levels. Certainly, there seems to be no shortage of examples of downward causation. Certain psychological states (e.g., prolonged anxiety, embarrassment) can cause physiological effects (heightened blood pressure, eczema, blushing) in a human body. McClelland’s experimental studies of human motivation showed that affiliative motives (the capacity to love and be loved) promote better health. For example, the salivary immunoglobulin A levels of subjects were significantly increased when they view a film of Mother Teresa designed to arouse affiliative motives.³⁰ Again, the functional molecules (DNA, proteins, fatty acids, etc.) within a cell are fabricated within internal processes of the cell itself; they are generated through the web of interactions of the whole system.³¹ That downward causation occurs is a fact; how to understand the phenomenon is the contentious issue.

Our discussion of emergence thus far could easily give the impression that it is a unitary phenomenon, although realized in many different ways at many different levels. That would be misleading. Clearly, not all properties that derive from higher levels of organization of processes warrant being called emergent. Nor is *novelty* sufficient: every new organization is an instance of the higher level property of having that particular organization. In our view, causality is what is crucial, and provides the criterion of non-trivial emergence, namely, the emergence of novel causal properties.³² There is one major divide in kinds of process organizations by which we can begin to identify those process organizations that yield emergent causality: the distinction between linear and non-linear processes and interactions.

Linear processes yield consequences that are simply the additive sum of the influences of the lower level consequences. Standard examples of holistic properties generated by these sorts of processes are simple aggregates. For example, a brick has a

closure of the physical world with his broadened definition of the physical, but has done so at the cost of the causal closure of the *micro*-physical world.

³⁰ D. McClelland: *Human Motivation* (Cambridge Uni Press, Cambridge, 1987), pp. 366-368.

³¹ A. Moreno & J. Umerez: ‘Downward Causation at the Core of Living Organisms’ in *Downward Causation*, pp. 99-117.

³² The following discussion summarizes points made by Bickhard in ‘Emergence’ in *Downward Causation*.

certain mass, which is a property of the whole brick, and none of its parts has that mass. Aggregates of this sort do not provide interesting cases of emergence; even a reductive physicalist can comfortably accommodate them, since the brick's mass results simply from summing the mass of each part.

Nevertheless, even simple aggregates have been taken to raise the question of whether some of their macro-properties could exert downward causation upon their constituent parts. One much-discussed example, because it seems to involve reflexivity, is a wheel. The trajectory of some molecule in an iron wheel cannot be explained other than by taking account of its location in that wheel.³³ The motion of the wheel as a whole, as it rolls along, is clearly what determines the curving trajectory of that single iron molecule. Yet, the wheel is made up of molecules like this one. It seems as if the properties of the whole depend upon the properties of its parts, yet the movement of each part depends upon the movement of the whole. Kim has cogently argued that there is nothing in this that is puzzling or metaphysically challenging, provided we understand this reflexive 'downward causation' diachronically. That is, any appearance of incoherence is avoided if there is a time lag – even if only tiny – between the whole's acquiring some property and its causing some change in its parts.³⁴ So, while linear processes do yield examples of downward causation, the kind of emergence they support is not sufficient to trouble physicalism.

On the other hand, William Wimsatt has proposed four different conditions, each of which provide a different sense in which a system property may be said to be 'aggregative', that is, to result from aggregating the properties of its parts.³⁵ Observing that it was common during the dominance of logical positivism to treat reducibility as the universal solvent for conceptual problems in the sciences, he proposes that, instead of reducibility, we should take non-aggregativity as the criterion of emergence. Then, not only do we get a non-trivial and intuitive concept of emergence, but we also get a classification of different kinds of emergence as a function of which conditions fail to be met.

Non-linearity is crucial to causal emergence. By definition, every instance of non-linearity is an instance of something whose causal properties cannot be derived aggregatively from lower level consequences. In that sense, every instance of non-linearity is an instance of emergence.

When we turn to consider cases of non-linearity, which provide the interesting cases of emergence, a further set of important distinctions needs to be made. These distinctions demarcate central cases of emergents; indeed some might take them to be

³³ This was a favourite example of the neuropsychologist Roger Sperry. See 'A Modified Concept of Consciousness', *Psychological Review*, **76**, 1969, pp. 532-536.

³⁴ In 'Making Sense of Downward Causation' in *Downward Causation*, pp. 305-321.

³⁵ William C. Wimsatt: 'Forms of Aggregativity' in *Human Nature and Human Knowledge*, ed. Alan Donagan, Anthony Perovich & Michael Wedin (Reidel, Dordrecht, 1986) pp. 259-291.

criteria for emergence. (There is a certain semantic arbitrariness here, which is not crucial to the argument.) Consider the following kinds of downward causation:³⁶

- a) Consequences outside of a system that are non-linear with respect to the lower levels of the system, but which nevertheless influence lower level external processes, would constitute a downward causation – this is among the weakest kinds. Example: disturbances in the flow of air around a kite.
- b) Constraints internal to a system that are non-linear consequences of the *organization* of the system would be a more powerful case. System stability, whether of energy well or far-from-equilibrium form, would be examples.
- c) Non-linear constraints *internal* to the constituents of a system – that is, one level down from those mentioned in b) – would be a still more powerful case. These present some of the most interesting cases. For example, the processes internal to cells are strongly constrained by the overall processes of the organism.³⁷ Such ‘meta-internal’ downward causations can extend to the existence of complex molecules that would not exist otherwise. The influence of surroundings on the internal processes of a computer chip would be another example.³⁸
- d) Constraints on the generative processes – sources of constructive variation – as well as the activities *per se*, of lower levels. For example, changes in the organization of an ecosystem can alter the selection pressures on the constituent organisms. Similarly, alterations in the biosphere of the earth can change the selections, and, at least indirectly, the variations, with respect to the species and ecosystems at constituent levels. Such downward causation via selection is among the strongest kinds of emergent causation.

Our contention is that the difficulties that physicalist philosophers have found in admitting the very possibility of emergence results from their inadequate, and according to the best science, false metaphysical framework. Within a more plausible process-based metaphysics, the mere possibility of emergence need no longer be regarded as problematic.

6. A Process Model of Emergence

In the light of these considerations, we can sketch the outlines of a different model of emergence. Let us introduce this via a couple of negative points. The way that Kim characterizes the standard multi-layered hierarchy as ordered by the mereological relation

³⁶ These cases are also discussed in Bickhard: ‘Emergence’ in *Downward Causation*.

³⁷ Moreno & Umerez: ‘Downward Causation at the Core of Living Organisms’ in *Downward Causation*.

³⁸ R. van Gulick: ‘Metaphysical Arguments for Internalism and Why they Don’t Work’ in S. Silvers (ed.): *Rerepresentation* (Kluwer, Dordrecht, 1989) pp. 151-159.

is prejudicial. There are two assumptions built into this seemingly innocuous account, both of which deserve to be challenged.

The first is that by generating the hierarchical structure by the part-whole relation, the levels are characterized primarily in terms of entities, which can be exhaustively decomposed into other entities that are their proper parts. The picture is of *things*, composed of smaller *things*. At the bottom, (if there is one) there are the tiniest things: elementary particles. The account is loaded in favour of a particle-like metaphysics.

Secondly, and not surprisingly, this standard picture assumes that the *relations* amongst things play no constitutive role (although, as discussed, Kim has now departed from the standard picture on this point). It supposes that molecules, for example, can be decomposed, without remainder, into atoms. Of course, no-one wants to deny that some force holds the atoms in a molecule together. Nevertheless, the implicit claim is that the proper parts of, say, a molecule of water, are two atoms of hydrogen and one atom of oxygen, *and nothing else*. (A more sophisticated and informed version of this claim would acknowledge the forces within the molecule, but give them a particle interpretation. Thus, the proper parts of a molecule of water would be two atoms of hydrogen and one atom of oxygen, plus the elementary particles whose exchanging holds the molecules together.) So, this apparently harmless explanation of the multi-layered hierarchy of the sciences already implies a clear denial that the spatio-temporal *organization* of those atoms has any role to play – other than being boundary conditions – in the emergence of the characteristic properties of water. Accordingly, higher-level emergent properties are rendered causally superfluous in this ontology; they have been designed out from the start.

Recognizing this leads us to re-examine the powerful argument deployed by Kim to force the choice between emergence and reduction. It seemed to show that any emergent properties or powers that are supposed to exert causation either at their own level, or ‘downward’, is superfluous, enforcing a reductionist conclusion. (The one concession he now makes is that perhaps causes might be conceptualized or described in terms of concepts that are at a higher level in relation to the concepts in which their emergence bases are conceptualized or described.) But the reductionist physicalism it enforces is false. Something has gone wrong.

When an argument, powerful in its own terms, leads to a false conclusion, suspicion has to fall on the terms in which that argument is couched. So it must here. The problematic step is the supposition that where there are genuinely emergent properties and powers, it is reasonable to infer that any causal consequence of a higher-level organization is eliminable in favour of the working out of the causal consequences of the constituents in that organization’s emergence base. This concept of a lower-level ‘emergence base’ smuggles in particle-physicalist commitments. Reductionist physicalism is false precisely because there are no elementary physical particles that can serve, in the required sense, as ‘emergence bases’. Since everything is an organization of quantum field processes, more complex organizations of processes can yield emergent

properties and powers, but it is the higher-level organization itself that is doing the work, not whatever might be its ‘constituents’. (That is why we earlier emphasized Kim’s invocation of relations and configurations.) Any field view destroys the seduction into a micro-particle reduction because configurational and organizational properties make differences in causal power, not just in the working out of lower-order causal power.

Processes exist *only* in some organization or other. Some organizations of process are fleeting, such as Newton’s legendary falling apple. Others are stable – or at least relatively stable – and may persist for eons. Stability over time and against perturbation manifests the *cohesion* of an overall organization of process.³⁹ The articulation of the required model, which we can only outline here, begins with the observation that there are fundamentally two forms of process stability: 1) energy well stability; and 2) far-from-equilibrium stability.

The former is exhibited when some process remains in or near thermodynamic equilibrium for a significant period. Such ‘energy wells’ require a certain level of energy to change their organization, and they can only be disrupted by a higher level of energy than they typically encounter in their ambient environment. Hence, they are typically very stable and robust. Atoms are straightforward examples; they are a furious process of electron waves around an even more furious dance of quarks and gluons, a process that takes a great deal of energy to destabilize.

Combinations of such stable ‘energy well’ processes exist at the macroscopic level, and some of the properties that such combinations manifest are the resultant of aggregating the properties of more microscopic processes that are their constituents. (Energy well stable process organizations can themselves also manifest emergent properties: e.g., van der Waals forces among molecules, or differing molecular properties that manifest quantum interactions within the molecules.) Note that aggregation is itself a form of spatio-temporal organization, and that more is involved in being a causally effective aggregate than simply the logical sum (a bare conjunction) of its constituents. The components have to *stick together*.⁴⁰

More remarkable is the second kind of stable organization. *Far-from-equilibrium stability* occurs in an organized process when it is not in thermodynamic equilibrium, and

³⁹ W.D. Christensen, J.D. Collier and C.A. Hooker ‘Autonomy, Adaptiveness, Anticipation: Towards autonomy-theoretic foundations for life and intelligence in complex adaptive self-organising systems’ MS, 1999; W.D. Christensen and C.A. Hooker: ‘From Cell to Scientist: Towards an organisational theory of life and mind’ in J. Bigelow, ed.: *Our Cultural Heritage* (Australian Academy of the Humanities, Canberra, 1998; ‘Autonomy and the Emergence of Intelligence: Organised interactive construction’ *Communication and Cognition*, in press; J.D. Collier: ‘Supervenience and Reduction in Biological Hierarchies’ in M. Matthen, B. Linsky, eds.: *Philosophy and Biology*: Supp. Vol. 14 of the *Canadian Journal of Philosophy*, 1988; J.D. Collier: ‘Autonomy in Anticipatory Systems: Significance for Functionality, Intentionality, and Meaning’ in M.D. Dubois, ed.: *Proceedings of CASYS’98, The Second International Conference on Computing Anticipatory Systems* (Springer-Verlag, New York, 1999).

⁴⁰ On this, see W.C. Wimsatt: ‘Forms of Aggregativity’.

yet it persists for some significant period without moving to equilibrium. The most primitive kind of stable system in which far-from-equilibrium processes persist without collapsing is one that is kept going artificially, entirely by external means. Consider, for instance, a chemical bath in which important processes are kept going by external pumps that maintain a flow of the required chemicals, from external reservoirs into the bath, while other pumps remove waste products. Until such a system is switched off, or runs out of chemicals, it is sustained, but the stability of the chemical bath is completely dependent upon its environmental conditions: the pumps and the supply contained in the reservoirs. The analysis that follows takes the minimal stability of this chemical bath as a *point de départ*, as we move step-by-step towards progressively more autonomous far-from-equilibrium systems. (Of course, combinations of far-from-equilibrium systems can also manifest aggregative properties, e.g., mass, but they are not what is remarkable about such systems.)

What enables any far-from-equilibrium system to survive is the interaction of its intrinsic processes with its ambient environment, such as pumps and reservoirs, that counter the physical tendency to move toward equilibrium. A candle flame, for another example, can manifest persistence and stability for hours on end, but it can do so only as long as fuel and oxygen continue to be fed into the process. Cut off either, and the flame is put out; it ceases to exist. In short, the persistence of far-from-equilibrium processes is possible if and only if such crucial interchanges with the environment (are able to) persist: they are *necessarily open* processes.

A candle flame illustrates the next feature of this model also. For a candle flame is a familiar example of a process that makes several active contributions to its own persistence. It maintains its temperature above the combustion threshold; it vaporizes wax into a continuing supply of fuel; and in standard atmospheric and gravitational conditions, it induces convection currents, thus pulling in the oxygen it needs and removing the carbon dioxide produced by its own combustion. Processes like this tend to maintain themselves; they exhibit *self-maintenance*. The ability to be self-maintaining is an *emergent causal power* of the organization of the candle flame; it cannot be explained simply as the physical resultant of the causal properties of its distinct constituents. Of course, in one sense its persistence is dependent upon its constituents: when the candle flame has burnt all its fuel, or it is deprived of oxygen, it ceases to be. But so long as the boundary conditions are fulfilled – its external requirements for fuel and oxygen continue to be satisfied – it continues to contribute to its own persistence. It succeeds in maintaining its own process of burning.

A further level of complexity is exhibited by systems that can maintain stability not only in certain ranges of conditions, but also within certain ranges of *changes* of conditions. That is, they can switch to deploying *different* processes depending on conditions they detect in the environment. A relatively simple example is a bacterium that can swim up a sugar gradient, but tumble if it finds itself to be swimming down a sugar gradient. These two kinds of activity – swimming and tumbling – are different ways for the bacterium to act appropriately to its environmental conditions – appropriate

in the sense that each contributes to its self-maintenance in the differing circumstances. The bacterium's ability to detect sugar-gradients, and to respond by switching between its two modes of action, means that it exhibits a kind of maintenance of its own ability to be self-maintenant; it is able to switch between its self-maintenant processes as the environment changes. That is, it exhibits *recursive self-maintenance*. Note that even in these relatively primitive examples, the description of these systems of organized process has to use *self-reflexive* locutions and to speak of its *abilities* and *actions*.

These considerations entail a radically different ontology. Biological systems – including humans – are *not* substantial entities ('things' in the thick sense) whose constituents are cells (smaller things), which in turn (after a few more reductions) are constituted out of elementary particles. They are open, organized action systems, in *essential* interactions with their environments, such that we cannot say what they *are* without taking those interactive processes into account.

This model of emergence enables us to see clearly what has gone wrong with the concept of an 'emergence base'. Consider first the case of stable 'energy wells'. These are processes that are relatively self-contained, and which do contain smaller but likewise stable 'energy well' processes within them. It makes sense to speak here of 'constituents'. If the concept of an 'emergence base' of constituents were to have valid application, it would be to these cases. However, to take one of the simplest examples, there is more to an atom of hydrogen than an electron and a proton; there is the *pattern* of the relationship between them, and that pattern of the process, its organization, is what is crucial to the emergent properties of hydrogen. Consequently, explaining the emergence even of stable energy wells requires attributing causal power to the organization of their constituents; the constituents alone, with their properties and powers, are not sufficient. So, while we can make sense of the notion that a higher-level 'energy well' process, like a hydrogen atom, has lower level constituents, the concept of the atom's 'emergence base' is far from clear once we take the organization itself, a relational fact, into account – the atom *is* that organization of processes between the nucleus and the electron.

When we turn to the other kind of stable system, far-from-equilibrium systems, the unsoundness of the inference is even more obvious. For the latter kind of systems can maintain their stability *only* by dint of their interactions with their surroundings. Their very existence, and their persistence, are dependent upon their relations with *external* factors in that environment, from which they keep drawing sustenance.

Consider once more a candle flame. What is its 'lower-level emergence base'? One answer (somewhat simplified) we might try is: the molecules of vaporized candle wax and of oxygen. But those molecules that are present at any one time within the visible area of the flame are soon consumed, and the products of that combustion – mainly heat, light, carbon dioxide, and water – are largely either radiated or carried away by convection. All that *persists* within the flame is the wick, but it is not, in any relevant sense, a basal constituent *of* the flame, and it too is progressively consumed. It is true that some of the energy released by the burning at any one time is used to cause the

combustion of succeeding molecules of molten wax and oxygen. But if this argument were sound, the *molecules* of wax and oxygen within the area of the flame at any one time would be nomologically sufficient for (i.e., be causes of) the combustion of other molecules within the flame, at a later time.

That is nonsense. It is the *flame*, the *process* of burning, that causes the molecules even to be present within its own area for a brief period, before they are in turn burnt. The flame has *no* stable set of constituents that could sensibly be called *its* emergence base (not even the wick). There is just an inflow of waxes and oxygen molecules, which are consumed, and an outflow of carbon dioxide and water molecules (ignoring other trace by-products, which do not affect the argument), together with a release of energy.

At bottom, the problem with the argument we are considering is its being framed in terms of entities and their properties, at both higher and lower levels. Its terminology of emergence bases and higher-level properties necessarily envisages only internal structures, not open processes. This is necessarily so, because an emergence base could not include any relations external to the system. To see this, consider the property of being the longest pencil in the box. The pencil's having this property has nothing to do with the molecules and internal relations that make up that pencil.⁴¹ It is a relational property of that pencil, which it would lose if an even longer pencil were added to the box. Now, adding another pencil to the box would not alter in any way the molecules and internal relations of the former pencil. The property of being the longest pencil seems trivial, but the logic of the case would apply equally to any extrinsic, relational property. So, while any distinction between relational and non-relational properties seems context-dependent, and not purely formal, on any account extrinsic, relational properties such as these could not be part of any 'emergence base' of the type presumed by the argument.⁴²

On the other hand, we cannot say what a candle flame *is* without mentioning its *relations* with *external* elements in its ambient situation. The very being of the flame, then, is a function of these external relations. These *physically external* relations are *logically internal* to any flame; they are constitutive of its being. If, say, the temperature of the atmosphere around the candle were gradually raised (independently) towards that of the flame itself, the convection currents required to suck in new oxygen and remove carbon dioxide would progressively become less effective. Either the flame would go out, smothered by the carbon dioxide it had been producing, or the entire candle would first melt and then vaporise. "– Oh", might come the reply, "that is just a matter of the boundary conditions; no-one would seriously suggest that the temperature of the ambient atmosphere is a constituent!" But a significantly lower temperature of the ambient

⁴¹ Bickhard, 'Emergence' p. 333. The example comes from P. Teller: 'A Contemporary Look at Emergence' in *Emergence or Reduction?* pp. 139-153.

⁴²

Positing a framework of a global supervenience can address such issues of external relations, in effect by making them all internal to the global universe, but at the cost of trivializing the notion of supervenience and encountering the problems discussed above in section 2.

atmosphere *is* a necessary part of the nomologically sufficient conditions for the existence of the candle flame; fail to take those conditions into account, and the argument we are assessing collapses. A candle flame is *necessarily* open, lest it be snuffed out. Physically external relations are essential to the flame's ability to maintain itself, an ability that is a genuinely *emergent* causal power.

The same applies in the case of any stable far-from-equilibrium process systems, such as biological systems – from plants to humans. And this is why it has been so easy for philosophers to become confused. For at any one time, a biological system does seem to have a stable set of constituents: the cells and complexes of cells that constitute its body. Nevertheless, the relevant principle is the same as in the case of the flame. In general, a biological system turns over the cells that constitute its body a number of times during its lifetime. (Humans generally replace all the molecules in their body parts, except for those in the lens in their eyes, many times during their lifetime.) Philosophers who have taken note of this has tended to assimilate it to the case, much discussed as a puzzle about identity, of the boat whose planks are replaced one by one while remaining afloat. To think about the turnover of the cells in our bodies in the terms of that famous example, however, is to miss the most significant aspect: the significance, the *necessity*, of the external interactions of the body. In this respect, biological systems are like candle flames; as complex organizations of processes, they persist *only so long as* they are able to maintain appropriate interactions with their environment, by which to sustain their existence. They eat and drink (and in many cases, breathe) – and perspire and excrete. If they stop those activities, they die.

So, we have to dismiss the argument that if there were emergent entities and properties, they could have no causal powers of themselves. This argument takes lower-level emergence bases to be causal factors in a way that belongs with the particle metaphysics rejected in section 3. It ignores the crucial role of organization in the emergence of properties and powers from systems of lower-level processes. It also ignores the essential role of external relations in the self-maintenance of stable far-from-equilibrium process systems, from flames to human beings. Yet stable far-from-equilibrium process systems are the ones that have given rise to the most interesting and intriguing of emergent causal powers: life, consciousness, and self-consciousness.

Within the process-based metaphysics sketched here, the possibility of emergence is no longer problematic. The hard work before us is in creating theoretically rigorous and empirically well-confirmed models of actual emergents.

Richard.Campbell@anu.edu.au	mark.bickhard@lehigh.edu
The Australian National University Canberra, ACT 0200 Australia	Lehigh University Bethlehem, PA 18015 USA