

Supervenience, Emergence, and Epiphenomenalism: Physicalism's Unholy Trinity

Stuart Silvers
Clemson University

Supervenience was once thought to be the “silver bullet” that would slay the demons of dualism and make the world, in particular the mental world, safe for physicalism. Once such hobgoblins were explained away, the way would be clear for deployment of the Positivist grand program of unified science based securely in physical theory. However, to anti-physicalists, whether dualist or not, it has always seemed that the supposed neural correlates of hopes for the metaphysical hegemony of physicalism would continue curiously to be dashed on obstinate and yet apparently non-physicalistic incidences of phenomenal consciousness and intentionality. These doggedly non-physical phenomena are not some trivial anomalies that are merely peripheral to an otherwise conceptually secure physicalist *Weltanschauung*; they are central its very coherence as well as crucial to the coherence of ourselves as persons and agents. So, philosophically speaking, there's a lot riding on whatever the outcome turns out to be.

Then too, from within the physical science community itself, there are developments that both highlight the discrepancies internal to the traditional physicalist project and appear to seriously undermine the basic presuppositions upon which its coherence depends. Against the former anti-physicalist antagonists, Jaegwon Kim has, in a series of works over the last twenty years, mounted a sustained defense of physicalism. To some, perhaps even Kim himself, his arguments fending off the concept of autonomous mental causation have delineated the dreaded physicalist dilemma. To put it crudely here, either accept the idea of an alternative concept of non-physical causation (and surrender the physicalism) or acknowledge that *apparent* mental causes are epiphenomenal.

In a microcosm of an evolutionary arms race, arguments about the efficacy of supervenience escalated (some perhaps might say “regressed”) to arguments about emergence and downward causation. Since there is obviously much more to this debate than can be discussed here, I’ll focus now on Kim’s analysis and some of the critique it has provoked, and conclude with a programmatic but positive nod in the direction of epiphenomenalism. Kim’s analysis of emergence is, on the one hand, designed to justice to the views of the British emergentists (to use Brian McGlaughlin’s characterization) Lloyd Morgan, Samuel Alexander, and C.D. Broad, and on the other hand, to spell out the relationship between the properties of emergent phenomena and related notion of “downward causation.” (Kim 1999, p. 19) As he puts it,

It is not only that emergent properties are to have their own distinctive causal powers but also that they be able to exercise their causal powers “downward” – that is, with respect to processes at lower-levels, levels from which they emerge...The claim that emergents have causal powers is entirely natural and plausible if you believe that there are such properties. For what purpose would it serve to insist on the existence of emergent properties if they were epiphenomena with no causal or explanatory relevance?” (p. 19)

Kim distinguishes two components that comprise the emergent thesis: emergent properties are (a) ‘novel’ in the sense that they are neither *explainable* or *reductively explainable* nor *predictable* from “the most complete and exhaustive knowledge of their emergence base” and (b) “emergents have causal/explanatory powers in their own right, introducing novel, and hitherto unknown causal structures into the world.” (p.6) Recently, Chalmers (2002) distinguished *strong* and *weak* emergence, the former are truths about high-level phenomena “not *deducible* even in principle from truths in the low-level domain,” the latter are truths that are merely “*unexpected* given the principles governing the low-level domain.” (p. 1) Chalmers’ version seems to correspond more to Kim’s and the traditional emergents’ distinction between

genuine irreducible emergents and *resultant* properties, ones which might be surprising in some sense, but are entailed by principles of the low-level domain.

Be that as it may, Kim argues that the reductive inexplicability emergent properties by “the basal conditions from which they emerge” (pp. 9-10) can be accommodated by his *functionalizability* thesis. This is the (familiar) idea that property identity is specified in terms of fulfilling a causal role, that is, by the causal work done. Thus, the DNA molecule turns out to be the vehicle of hereditary transmission. Functionalizability is necessary and sufficient for reducibility, genuinely emergent properties are not functionalizable, ergo, emergent properties are not reducible to the basal conditions from which they emerge. Of course Kim is skeptical that there are any physical yet nonfunctionalizable properties but if there are then “It seems to me that if anything is going to be emergent, the phenomenal properties of consciousness, or ‘qualia’, are the most promising candidates.” (p. 18) Chalmers (2002) agrees but is no skeptic about there being strong emergents. “I think there is exactly one clear case of a strongly emergent phenomenon, and that is the phenomenon of consciousness.” (p. 2)

In general, instances of emergence and downward causation are ontologically benign, that is, cases where causal properties of a whole W are not reducible to the properties of its micro-constituent and causes changes in lower micro-levels. (Kim, p. 26) In contrast, *reflexive downward causation* occurs if, “some activity or event involving a whole W is a cause of or has causal influence on, the events involving its *own* micro-constituents.” (p.26)

Since Kim’s view is that for there to be a difference in properties is for there to be a causal difference, the question is whether there is the kind of emergents we (should) care about, namely those which “bring into the world new causal powers of their own, and in

particular...have powers to control the direction of the lower-level processes from which they emerge.” (pp. 5-6)

The principle of downward causation in dispute is,

“To cause any property (except those at the very bottom level) to be instantiated, you must cause the basal conditions from which it arises (either as an emergent or as a resultant).”

The core of Kim’s critique rests on two competing principles: causal inheritance and causal-power actuality. The former states roughly that the causal powers of a realizer R are identical to the causal power of the property P that it instantiates. The latter, equally roughly, has it that if object O exercises causal/determinative powers at t because of property P, O must already have P at t; and if O is caused to acquire P at t, O doesn’t already have P at t and thus can’t exercise P’s causal powers.

Let Oscar be whole W with a certain emergent property M (experiencing pain) at time t. M emerging at t (Oscar in pain at t) depends on M having a certain micro-configuration at t (some specific set of C-fibers firing at t) including a given constituent neuron a_j firing at rate n (a_j had P_j at t). Synchronic reflexive downward causation asserts that a_j is caused to fire at rate n at the same time t by Oscar (W) being in pain (having M) at t. Still, because Oscar could not experience pain (M could not emerge in W) unless a_j fires at rate n, this seems to be a case of self-causation. If one accepts, as Kim does, the “causal power actuality principle” then synchronic downward causation is unintelligible. Diachronic downward causation, however, is spared this fate but only to fall prey to another. Diachronic downward causation preserves emergentism from the depredations of the causal-power actuality principle in virtue of W having emergent property M at t but exerting causal influence on a_j at a time $t+1$. Importantly, a_j at $t+1$, is not a constituent of W at t.

Nevertheless, emergents exerting diachronic downward causation are, on Kim's analysis, even less than benign, they are epiphenomenal. His view is there is no plausible reason to think otherwise than that it's the basal conditions P from which M emerges that does all the causal work. Emergent properties M have both basal conditions P (from which they arise) and higher-level effects, e.g., M* in virtue of M's causing M*'s lower-level base P*. Kim's claim is that there is nothing blocking the displacement of M's putative causal powers by P*. If causation is nomological sufficiency, then since P is sufficient for M and M is sufficient for P*, then P is nomologically sufficient for P*. And, ex hypothesi, since M is emergent it can't be a causal effect of P. If, contra emergentism M were, then it would be functionalizable and hence reducible.

My makeshift summary does no justice to Kim's assiduous analysis but I hope the gist is clear enough to comprehend the sequel. I distinguish two types of critique of Kim's thesis: one challenges his argument from within the bounds of the presupposed physicalist metaphysics, the other draws upon theoretical developments (advancements) in quantum physics as grounds for rejecting physicalism as a now obsolete metaphysical foundation for explanation of physical phenomena. Representing the former alternative, Welshon (2002) for example, argues that to escape Kim's conclusion of the epiphenomenality of emergent properties, it is sufficient to deny the conjunction of "the functionalization of higher-level property instances and the causal inheritance principle." (p. 47) He and other functionalists think that the multiple realizability of property realizers undermines the causal inheritance principle. The debate here turns on the relative plausibility of the competing intuitions. Kim's (1999) view is that since diverse emergent level realizers are "causally and nomologically diverse" that such causal/nomic heterogeneity at the emergent level makes it "unfit to figure in laws, and is thereby disqualified as a useful

scientific property.” (pp. 17-18) In contrast, non-reductive functionalists argue “that the nomological diversity of a set of realizers and the nomological homogeneity of the realized property together suggest that the causal inheritance principle is false.” (Welshon 2002, pp.48-49) Even allowing for the possibility of ontologically emergent properties, however, does not justify claims about their causal efficacy nor for that matter if there are such properties. He concludes, “Emergent mental properties, even if not epiphenomenal, if they exist, are at this point still only epistemologically emergent.” (p. 49)

Shoemaker (2002) tries to neutralize Kim’s list toward emergent epiphenomenalism by extending Broad’s notion of latent properties to latent causal powers. Roughly, these latent properties and powers of elements are ones that can be known only after observing the results of combining them. Although we can know all the properties of the elements, “we can’t know all of the laws about the results of combining elements with these properties until we have actually combined such elements in what I will call ‘emergent engendering’ ways.” (p. 54) However, since the project is as Kim notes, to make sense of emergence, reliance on the idea of emergence engenderment seems to come threateningly close to begging the question.

Turning now to the second of the two kinds of critique of Kim’s physicalist reductionism, I need to step gingerly in light of Richard Campbell’s the presence on this very panel who together with the Godfather of Interactivism, Mark Bickhard, are the leading figures in the anti-physicalist movement. (Have you noticed the curious preponderance of Interactivist Campbells?) Happily (and I think wisely) my remarks here will be mercifully short and almost politically correct for, as you shall soon see, I emulate their approach in drawing upon developments in the practice of actual, ongoing science.

In their rigorous analysis, “Physicalism, Emergence, and Downward Causation,” Campbell and Bickhard ascribe the failure of both reductive and non-reductive physicalism to accommodate emergent phenomena and concede the epiphenomenality of the mental to, “the presumption...that the fundamental level of the natural world consists of micro-physical entities of some sort, with their primary properties.” (p. 7) They counter with, “But recent development 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...But enough has now become clear for it to be evident that an ontology of elementary ‘particles’ – basic particulars – can no longer be sustained.” (p. 8) They continue that contrary “the age-old prejudice of 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” (p. 10) Moreover, physicalists cannot simply adopt as basic whatever the best physics delivers to us because “There is no ‘bottoming out’ level in quantum field theory – it is *patterns* of processes all the down *and* all the way up.” Campbell and Bickhard conclude that because everything is a configuration of processes without any basic-level entities to which all other phenomena are reducible, physicalism entails that “*everything is an epiphenomenon.* That is the *reductio ad absurdum* of this position. (p. 11)

There is no doubt that the issues implicated in this ontological priority debate are enormous in scope. Campbell and Bickhard claim that respect for contemporary physics requires physicalism to acknowledge “genuine emergence, and not treat it as merely ‘apparent.’” (p. 12) The physicalist would probably reply that acknowledging patterns of processes requires us to (a) jettison causation as a physical relation in favor of an alternate concept whose parameters are far

from clear, (b) might (*pace* Kant, Quine, Kuhn, Nagel and McGinn) defy our cognitive-conceptual capacity, and (c) have us treat typically physical object and their constituents as merely apparent. You will be pleased to know that I won't speak to any of these three points but I do want to make two other related comments. The first harks back to Putnam's (1978) remarking on Einstein's Relativity succeeding Newton's Universal Gravitation that if the object referred to "don't really exist at all, then it is a *miracle* that a theory which speaks of gravitational action at a distance successfully predicts phenomena." (p. 19) The second is that even the anti-physicalist Chalmers 2002, claims there are no other cases of strongly emergent phenomena than consciousness and argues that we postulate fundamental psychophysical laws to explain conscious causation. There is, it seems to me (and other defenders of the physicalist "faith") a real question of whether consciousness, if it is the only clear instance of strong emergence, a case of the emergentist tail, albeit a spectacular one, wagging the physicalist dog. If physicalism is false and is toppled by a recalcitrant phenomenon of consciousness, genuine questions are and should be raised about its miraculous track record of explanatory success.

Finally, since imitation is the sincerest form of flattery, I take my cue from Campbell and Bickhard, to introduce some remarkable results from scientific experiments on phenomena of consciousness and discuss them in relation to Kim's (and others, e.g., Kinsbourne) emergent epiphenomenalism. After all, if contemporary physical theory is a source of the critique of physicalism, then by parity of reasoning, one should be able to appeal to contemporary experimental psychology both as a source of the critique of the causal efficacy of consciousness and fecund insights into adaptive unconscious mentality. Libet's (1985) much-discussed "readiness potentials" experiments that within the design limitations inherent to testing for consciousness causation nevertheless remain an evidential benchmark for what Harvard

psychologist Daniel Wegner (2002) in his book of the same name calls “*The illusion of conscious will.*” I’ll not rehearse this story here. Then there is the February 27, 1997 issue of *Science* that carried a brief report on results of recent experiments into epiphenomenalism that, quite remarkably, caught the attention of National Public Radio News in the United States. The research team of Bechara, Damasio, Damasio, and Tranel, (1997) investigated the theory that the ventromedial frontal cortices hold dispositional knowledge and explored the claim of causal efficacy for conscious, occurrent beliefs vis-à-vis corresponding nonconscious, standing beliefs. The findings are notable because they seem to sustain rather than allay epiphenomenalist suspicions (skepticism) about conscious causation. These suspicions are reinforced insofar as “nonconscious signals...act as covert biases on the circuits that support processes of cognitive evaluation and reasoning” (p. 1294). The evidence suggests that a nonconscious biasing using “neural systems other than those that support declarative knowledge” precedes overt reasoning. Wegner (2002) discusses Delgado’s (1969) brain stimulation research already in 1969 anticipates the Damasio team’s multiple-track brain processing hypothesis. Delgado’s work “suggests that the brain structure that provides the experience of will is *separate* from the brain source of action...This, in turn, suggests...that conscious will is an add-on...” (p. 47) Based on his review of “research to date on the anatomy of conscious will” Wegner argues that “the experience of will may be manufactured by the interconnected operation of multiple brain systems, and these do not seem to be the same as the system that yield action. (p. 49)

Wegner and Wheatley (1999) and Wegner (2002) develop what they call A Theory of Apparent Mental Causation: “*People experience conscious will when they interpret their own thought as the cause of their action.*” (Wegner 2002, p. 64) The theory is, of course, folk psychology writ large and the goal is to show that the notion of conscious mental causation

operative in the belief-desire model of explanation can itself be explained in terms of non-conscious brain processes together which coupled with a theory of self-perception “suggests that the will is a conscious experience that is derived from interpreting one’s action as willed.” (p. 66)

The schematic of Wegner’s analysis bears a striking similarity to Kim’s picture of causal preemption. Apparent mental causation relies on our ignorance of conscious mental processes; “...the real causal mechanisms underlying behavior are never present in consciousness. Rather, the engines of causation operate without revealing themselves to us and so may be unconscious mechanisms of mind.” (p. 97) Wegner test his theory of apparent mental causation against data from a vast array of experiments (that he charmingly augments with anecdotal accounts) and relates it to an analysis of automatism and virtual agency. I think the point to be stressed here is that whatever the fate of Wegner’s theory, and despite the obstacles to testing for the causal efficacy of consciousness, there is, a growing corpus of experimental data on consciousness (for example, by V.S. Ramachandran and Joseph LeDoux, to mention two prominent neuroscientists) that needs to be taken seriously. In a recent lecture to the Society for Philosophy and Psychology, neuropsychologist Marcel Kinsbourne reported that after reviewing the history of consciousness experiments, he has concluded that the evidence favors epiphenomenalism.

In his book, *Strangers to Ourselves*, University of Virginia psychologist Timothy Wilson (2002) explores the idea and scope of what he calls the adaptive unconscious. Wilson too recognizes multiple track neural processing with the adaptive unconscious consisting of “a collection of modules that perform independent functions outside the conscious view.” (p. 49)

This thesis is now well documented by studies of patients with various kinds of brain damage and other often bizarre neurological deficits such as blindsight and hemi-neglect.

“Consciousness, on the other hand, seems to be a single entity” (except, as he notes, perhaps in

cases of multiple personalities). “Exactly how to define it, and exactly how it is related to brain functioning, are not known.” (pp. 49-50) Still, Wilson opts for a functionalist account of consciousness where “we retain some ability to influence how our minds work. Even if the adaptive unconscious is operating intelligently outside our purview, we can influence the information it uses to make inference and form goals. One of the purposes of this book is to suggest ways this can be done.” (p. 48) It might be that Wilson falls prey to the very deliverances of the adaptive unconscious that yields the conscious experience of will that Wegner claims is illusory. While Wilson’s book is a valuable systematic analysis the pitfalls of traditional (and Freudian) first-person epistemology, he fails to address the crucial question of the experienced causal efficacy of conscious events that instantiate underlying neural activities. This is understandable since as a social psychologist, Wilson’s thesis is, to put it crudely, oriented more toward therapy than theory. His account of introspection as self-narrative is important insofar as it speaks to the question the economy of our conscious lives vis-à-vis our lives under the causal control of the mechanisms of the adaptive unconscious. As you might expect, conscious introspection fares badly when tested experimentally. But it’s not quite so bad as portrayed in *Saturday Night Live*’s celebrated case of the Home Headache-Tester, a device you attach to yourself to determine whether or not you have a headache.

The prospect of consciousness epiphenomenalism is universally dreaded for all the well-known reasons: loss of self, agency, free will, meaning of life, etc. But what stops the slide on the slippery evidential slope of greater and greater control of the adaptive unconscious that results in consciousness being a spandrel of it in the way that dreams are (*pace* Flanagan, 2000) spandrels of the sleep cycle? The stakes are high, or so it seems from our egocentric conscious point of view. If we ascribe a point of view to the adaptive unconscious it’s a very different

story, not about self-serving survival strategies of host organisms but about indifferent gene replication strategies. So epiphenomenalism is a dread that is parochial to the cosmic contingency that happens to be our species. The brief evolutionary history of the species would seem to suggest that to believe in the causal efficacy of consciousness is to fall prey to an illusion. Geocentrism turned out to be an illusion, merely apparent and Heliocentrism is the reality, the sun's rising in the eastern sky is an illusion caused by the relative position of the Earth and the sun. Searle (1992), among others, argues that consciousness is not similarly illusory but real. The epiphenomenalism purportedly entailed by physicalism is not that consciousness is unreal but that the causal powers ascribed to it are apparent. As we've seen, physicalists argue that the causal powers of irreducible emergents is apparent and that basic particles and their powers are the reality; conversely, quantum theorists propose that substance physicalism is apparent and it is processes replete with all kinds of emergent phenomena are the reality. If Chalmers is right that the only genuine emergent is phenomenal consciousness then it seems to me that we're dealing not with a question of the fecundity of competing metaphysical presuppositions but with a species-specific problem of the intricacies of reflexive self-conception.

References

- Bechara, A., Damasio, A., Damasio, H., and Tranel, D. (1997, February 28). Deciding advantageously before knowing the advantageous strategy. *Science*, 275, 1293-1294.
- Bickhard, M. and Campbell, R. (MS). Physicalism, emergence, and downward causation.
- Chalmers, D. (MS). Varieties of emergence.
- Delgado, J. (1969). *Physical control of the mind*. New York, Harper and Row.
- Flanagan, O. (2000). *Dreaming Souls*. New York, Oxford University Press.

- Kim, J. (1999). Making sense of emergence. *Philosophical Studies* 95, 3-36.
- Libet, B. (1985). Unconscious cerebral initiative and the role of conscious will in voluntary action *Behavioral and Brain Sciences*, 8, 529-566.
- Putnam, H. (1978). *Meaning and the moral science*. London, Routledge & Kegan Paul.
- Searle, J. (1992) *Rediscovery of the mind*. Cambridge, MA, MIT Press.
- Shoemaker, S. (2002) Kim on emergence. *Philosophical Studies* 108, 53-63.
- Wegner, D. (2002) *The illusion of conscious will*. Cambridge, MA: MIT Press.
- Wegner, D. and Wheatley, T. (1999). Apparent mental causation: Sources of the experience of will. *American Psychologist* 54, 480-491.
- Welshon, R. (2002) Emergence, supervenience, and realization. *Philosophical Studies* 108, 39-51.
- Wilson, T. (2002). *Strangers unto ourselves*. Cambridge, MA: Harvard University Press.