

Emergentism, a Partial Solution to the Mind-Body Problem

By Chris Wright

The mind-body problem—which Jaegwon Kim characterizes as the problem of “finding a place for the mind in a world that is fundamentally physical”¹—has been puzzled over for centuries, and is unlikely to be resolved any time soon. Every possible solution seems to have inadequacies. Anyone familiar with philosophical literature is aware of all the problems with Cartesian substance dualism, reductive physicalism, eliminative materialism, behaviorism and functionalism, non-reductive physicalism and emergentism. One is tempted to agree with Colin McGinn that the cognitive apparatus of humans is intrinsically inadequate to the problem of explaining the relation between the mind (or more specifically consciousness) and the brain.² How something like consciousness can emerge from something like the brain seems totally inexplicable.

I suspect that *complete* understanding of the mind-body relationship is impossible, just as it is impossible with respect to the domains of quantum mechanics and general relativity. “Formally”—mathematically—we may be able to ‘understand’ the latter theories, but “intuitively” they’re a mystery to us. We cannot visualize the worlds they describe, worlds of eleven dimensions or multiple universes or a finite but unbounded cosmos. Our possible experience is bounded, as Kant argued; we cannot intuitively conceive of indeterministic physical structures or of space in anything but three dimensions. Similarly, understanding *how it is possible* for electrochemical neural processes to produce consciousness (intentionality, qualia, meaning, etc.) may, for whatever reason, be beyond our cognitive horizons. It might, however, be possible to represent the causal relationship between brain and consciousness in some sort of vague conceptual sense by means of concepts like emergence, downward causation, and so on. These things are somewhat mysterious, but if we can use them in a “solution” more plausible than any alternative—perhaps by removing their apparent incoherence—then we might be justified in believing we have made some headway in understanding the mind-brain connection.

¹ Quoted in Barbara Montero, “Post-Physicalism,” *Journal of Consciousness Studies*, Vol. 8, No. 2 (2001): pp. 61-80.

² See Colin McGinn, “Can We Solve the Mind-Body Problem?,” in *The Nature of Consciousness: Philosophical Debates*, eds. Ned Block, Owen Flanagan and Güven Güzeldere (Cambridge, Mass.: MIT Press, 1997), pp. 529–542.

In fact, I think that some version of the emergentist approach is the only possible quasi-solution. Nothing else is remotely plausible, as I'll briefly argue below. Unfortunately, as it stands, emergentism is not yet a polished theory; and this will be evident as I look at a couple of recent proposals on how to conceptualize it. I choose Roger Sperry's version as well as the version put forward by Timothy O'Connor and Hong Yu Wong in their paper "The Metaphysics of Emergence"—I have found their defenses of the theory to be fairly sophisticated.

O'Connor has criticized reductive accounts of consciousness, which propose that it "really is" something else, i.e. something physical, deeming them to be deeply counterintuitive. "The claim that 'token' mental states—i.e., particular, concrete mental occurrences—just are complex electrochemical events (which just are complex microphysical events) implausibly denies that there is anything distinctive about mental activity in the world."³ Consciousness cannot be neurons firing since it cannot be something other than *what it is*, i.e. thoughts, sensations, phenomenal experiences, etc. Consciousness is private, while physical, electrochemical events are (at least in principle) publicly observable.

Indeed, any theory that tries either to "eliminate" consciousness or to reduce it to something else—something public, be it physical states, behavior, functional roles, or whatever—flies in the face of incontrovertible intuitions. Despite the efforts of such writers as Daniel Dennett⁴ and Richard Rorty⁵ to deny that there are qualitative, phenomenal, private or first-personal aspects to experience, it remains *intuitively* evident that mental experience must be characterized in mental language because it is fundamentally private and qualitative. Phenomenal experience cannot be adequately described in anything but phenomenal language, and the same is true of intentional experience. No other sort of language does it justice, or is true to what we perceive.

Consider functionalism. Functionalists define (types of) mental states in terms of their functional roles, thus effectively denying the importance of the state's private features, be they phenomenal or intentional. In his article "What is Functionalism?" Ned Block writes that

³ Timothy O'Connor and Hong Yu Wong, "The Metaphysics of Emergence," *Noûs*, Vol. 39, No. 4 (1990): p. 660.

⁴ See, e.g., his article "Quining Qualia," in Block et al., *The Nature of Consciousness*, pp. 619–639.

⁵ See his paper "Mind-Body Identity, Privacy, and Categories" (1965). He argues there that sensations don't exist, and that the only reason we think they do is that to eliminate talk about them from our language would be "impractical." In other words, when he feels pain he doesn't feel pain, but it's useful to talk as if he does.

according to functionalism, the nature of a mental state is just like the nature of an automaton state: constituted by its relations to other states and to inputs and outputs. All there is to ... being in pain is that it disposes you to say ‘ouch,’ wonder whether you are ill, it distracts you, etc.

By reducing the mental (the private) to something else, functionalists try to eliminate the mind-body problem. David Lewis, for example, writes that “The definitive characteristic of any (sort of) experience as such is its causal role, its syndrome of most typical causes and effects.”⁶ But he is wrong: types of consciousness cannot be defined in a non-circular way, a non-mental way. The definitive characteristic of pain is not that it tends to cause a certain behavior, but simply that it is painful. The definitive characteristic of pleasure is that it feels good. In a deep sense, *consciousness is private*, and one cannot define (or “adequately describe”) the private in terms of the public. This is the essence of the “explanatory gap,” the essence of the mind-body problem—which can also be called (and *should* be called, for clarity’s sake) the private-public problem. Functionalists and behaviorists (and eliminative materialists and reductive physicalists) try to solve the problem by eliminating the private or reducing it in some way to the public, but they fail because there is manifestly a private realm of experience with its own *sui generis*, “first-personal” properties and features. None of the verbal legerdemain of contemporary philosophers has been able to get past that simple fact.

Frankly, I have never been able to understand how someone can be, say, a functionalist. Or a reductive physicalist. Does this person not have sensations? Is he not conscious? Is he a “philosophical zombie”? Can he not see that only *he* has the ability to perceive “from the inside” what is going on in his own mind, what thoughts and sensations he is having? To say that “*all there is to being in pain is that it disposes you to say ‘ouch’ ...*,” or that pain is *nothing but* a series of neuronal firings, is to deny that pain is qualitative and private. But this is to deny the existence of pain as each of us understands it.

In short, O’Connor is right that the mind-body problem cannot be solved—cannot even be *formulated* or recognized as a genuine problem—unless one has first acknowledged that mental

⁶ David Lewis, “An Argument for the Identity Theory,” in his *Philosophical Papers Vol. 1* (Oxford: Oxford University Press, 1983).

life is irreducible to publicly observable properties, be they physical, functional or behavioral. But by acknowledging this, one has already thrown out the majority of proposed solutions to it. (Of course, they couldn't be solutions anyway, because they amount to a denial that there is any problem at all. For if consciousness "just is," e.g. electrochemical activity, then there isn't a mind-body problem after all! That leaves it a mystery, though, as to why philosophers have thought for hundreds of years that there is.) The correct theory has to be non-reductive. That leaves only substance dualism, some form of non-reductive physicalism more sophisticated than functionalism, or emergentism (which can in fact, I think, be considered a version of non-reductive physicalism).

It has long been recognized that [substance dualism is problematic](#). The notion of two fundamentally different kinds of substances, which have nothing in common yet somehow interact, is incomprehensible. Its explanatory advantages, if there are any, are far less compelling than its disadvantages. So it has rightly been discredited.

All that's left is some sort of physicalism that acknowledges the non-physical or non-publicly-observable properties of consciousness. This, as I have said, rules out functionalism, for while functionalists admit that mental properties (such as painfulness) are not identical to physical properties (such as C-fiber stimulation), they think that the former *are* identical to *functional* properties (such as causal roles). But this cannot be true. Mental properties, like mental particulars, are private or immediately accessible only to the consciousness that experiences them, whereas the property of fulfilling a causal role can be exhaustively analyzed and understood through public language. In other words, there is nothing it is "like" to be a cause or an effect as such, while there *is* something it is like to be in pain. Pain, then, cannot just be a cause of so-and-so and an effect of so-and-so; there is something more to it than "its syndrome of most typical causes and effects."

Another candidate is Donald Davidson's anomalous monism. Insofar as Davidson adopts the Spinozistic position that a mental event is simply a physical event described in mental language, and vice versa, without explaining what that means and how it can be possible—how, that is, a private, mental event can be identical to a public, physical event—his monism is unsatisfactory. It is incomplete. To be plausible, a physicalist theory (such as Davidson's) has to do more than just say that a mental event is "also" a physical event. It has to give an account of the relation between physical (public) and mental (private) aspects.

Now, only one theory that I know of has offered such an account. That is, only one theory promises to explain *in what sense* a neural event is also a private mental event. It is the only theory that accommodates our Cartesian intuitions while explaining in what way they are wrong. In fact, if one accepts that consciousness is private (subjective, intentional, qualitative) but also that it somehow arises out of the brain (which is *not* private in the same way, but in principle publicly observable—e.g., during surgery or under the microscope), this theory is all that's left.

The theory, as I've said, is emergentism. There is more than one version of it, but common to every version (at least implicitly) is the idea that an event in consciousness can be considered from two perspectives. The first is the perspective from which neurologists consider it, namely as consisting of a series of neuronal firings and electrochemical processes. From this neurological viewpoint, what is looked at is the *aggregate*, the "series," of physical processes that "correspond to" the conscious state. The processes are atomistically "summed up," so to speak: "first *this* happens, then *this*, then *this*...." A given state of pain or pleasure or whatever is considered as being a particular series of electrochemical events.

The other perspective is *holistic*. The neural events are considered together, as mutually contributing to some "larger" state. This holistic state, the *unity* of all the events in their interactions, is precisely the conscious event.

"Holistic" interactions are ubiquitous in biology. The cell is an obvious example. Each cell is constituted by molecules and molecular processes; in a sense it can be reduced to nothing else but these molecules. And yet it is also its own entity, an entity distinct from its micro-level constituents, possessing its own causal powers. (It can interact with other cells, for example.) Holistically, all the molecular events on which it supervenes interact to produce this new thing, the cell, which has causal powers that are not just the sum of the causal powers of its constituents (considered in themselves, isolated from their cellular environment). And so the cell is *emergent*.

But what exactly does that mean? The relevant definition of "emergence" is a matter of some controversy. There is a consensus, though, that various definitions have to be distinguished from each other. In the weakest possible sense, emergence is very common: wherever there is a physical structure that has properties not possessed by its "micro-level" constituents in themselves, there is emergence. The mass of an object, for instance, is not an emergent property, even in this weak sense, for it is merely the sum of the masses of its constituents. By adding up the latter, one arrives at the former. The shape of a wheel, on the other hand, is emergent, in that it depends on a

specific arrangement (or set of possible arrangements) of the particles that constitute it. In this incredibly weak and uninteresting sense of emergence, everyone is an emergentist. As J.J.C. Smart, a reductionist, puts it, “in saying that a complex thing is nothing but an arrangement of its parts, I do not deny that it can do things that a mere heap or jumble of its parts could not do.” “Smart admits,” writes Tim Crane, “that objects can have properties and powers which their parts do not have. But this doesn’t mean that these powers or properties are not reducible to the powers or properties of the parts. The very most it means is that the properties need not be reducible in the ‘additive’ sense.”⁷

It is worth noting that even in these cases, there is, or there can be, a kind of downward causation. (The notion of downward causation—i.e., of “causal influence from the macroscopic to the microscopic levels of nature [such that] how things are at a higher level of complexity affects what happens at a lower level”⁸—is typically thought to be a defining feature of emergence.) Consider Roger Sperry’s example of a wheel rolling down a hill.⁹ The properties of the wheel broadly supervene on the properties of the molecules within the wheel, and yet the former are able to influence the latter. For, while the downward motion of the wheel doesn’t cause “reconfiguration” of the wheel’s molecules *relative to one another*, it does reconfigure them relative to the rest of the world. At one moment a molecule is at a particular location relative to the grass; at the next moment it is at a different location. Thus, there is a kind of downward causation: certain properties of the macro-structure determine certain (“relational”) properties of the micro-structures.

The example of the biological cell is a little different. The cell is a living thing that actively plays a role in its environment by excreting and “ingesting” molecular structures. Likewise, inside the cell are constantly occurring ‘coordinated’ chemical processes. These processes are not purely random and accidental, as in the case of a non-living thing; they are tightly controlled, carefully “monitored” by means of unfathomably complex communicative mechanisms between all the molecular structures involved. The mechanisms and all the activities they regulate are ultimately

⁷ Tim Crane, “The Significance of Emergence,” p. 8, at <http://sas-space.sas.ac.uk/222/1/Significance%20of%20emergence.pdf>.

⁸ *Ibid.*, p. 15.

⁹ See Timothy O’Connor’s paper “Emergent Properties,” *American Philosophical Quarterly* 31 (1994): pp. 91–104. Cf. also Richard Campbell and Mark Bickhard, “Physicalism, Emergence and Downward Causation,” p. 26, where they give a fuller treatment than I do. (The paper is available at <http://www.lehigh.edu/~mhb0/physicalemergence.pdf>.)

caused by one factor alone: the holistic character of the system, the “web of interactions of the whole system”¹⁰ itself. They are caused, that is, by the emergent structure (the cell) itself. This isn’t efficient causation; it is something that is harder to grasp conceptually. Something like “structural” causation, in the language of the Marxist Louis Althusser. The efficient cause of a given cellular event is, of course, some other cellular event, some micro-level event or series of events. But all these events happen only because their totality constitutes an emergent entity, a holistic system (or structure) that itself makes possible the events, by organizing them in a particular way. The properties of the whole, the cell, supervene on what is happening at the level of its constituents, but what is happening there is brought about through the “holistic” influence of the system, the cell.

All this is rather obvious, but it’s hard to conceptualize. It reminds one of the “chicken or the egg?” dilemma. You can’t have the cell without its constituents and their interactions, but you can’t have their interactions without the organizing principle that is the cell. Each factor, in a way, seems to cause the other. No wonder, then, that Jaegwon Kim, in his article “Making Sense of Emergence,” doubts the coherence of this reflexive variety of downward causation:

How is it possible for the whole to causally affect its constituent parts on which its very existence and nature depend? If causation or determination is transitive, doesn’t this ultimately imply a kind of self-causation, or self-determination—an apparent absurdity? It seems to me that there is reason to worry about the coherence of the whole idea.¹¹

What ensures the logical possibility of this sort of downward causation is that it is not quite the kind of causation Kim has in mind. It is not efficient (in Aristotle’s sense); it is not even on the same order of causation to which the question of whether it is “transitive” can be applied. Instead, it is “structural”: its causal influence is manifested through the micro-level states in (the totality of) which the holistic structure consists. Structural determination is, indeed, manifested in every cellular event. What this means is that the web-like structure of the “web of interactions” is

¹⁰ Campbell and Bickhard, “Physicalism, Emergence and Downward Causation,” p. 25.

¹¹ Jaegwon Kim, “Making Sense of Emergence,” *Philosophical Studies* 95 (1999): p. 28.

ultimately responsible for the overall pattern of the interactions. To put the point paradoxically: the web-like structure is responsible for itself.

But not even *this* kind of emergence is the kind directly relevant to the mind-body problem. I have discussed it mainly to show that philosophers' attacks on the notion of emergence are misguided. Roger Sperry is right that emergence abounds in nature, and that it in no way contradicts the causal closure of physics or is a totally mysterious property. An emergent structure's downward causal influence occurs through the activity of the micro-constituents themselves. "Microdeterminism," writes Sperry, "is not so much refuted or falsified as it is supplemented."¹² That's why physicists and biologists have rarely felt the need to invoke the concept of emergence to explain an event: physical events can be explained "reductively," so to speak, even as the reductive explanations would literally have nothing to explain were the phenomena to be explained not embedded in a holistic environment.

To explain consciousness, however, it is necessary to invoke a stronger and more mysterious kind of emergence. The type I've been describing—the type manifested in the cell—is, I believe, essentially what Mark Bedau has in mind when he writes, "[emergence] involves downward causation only in the weak form created by the activity of the micro-properties that constitute structural macro-properties."¹³ He contrasts his conception with O'Connor's stronger version, which he thinks is inexplicable or "magical." I'll look at O'Connor's conceptualization shortly, but first I want to explain what motivates it. That way we'll be able to understand it better, and also what's wrong with it.

First let's recall the intuitive attractiveness of dualism. "The irreducibility of conscious experience and self-determining action already commits one to a kind of dualism, a duality of physical and conscious properties."¹⁴ But this dualism is not of the typical non-reductive physicalist sort, according to which mental features are in some unexplained way "realized in" and supervening on physical processes. Rather, "the dualism we must accommodate is ontological."¹⁵ We cannot just sidestep our intuitions, which are dualistic in a robustly ontological way. Functionalists and their kind are inclined to do just that (i.e. to ignore intuitions) but the result is

¹² Quoted in O'Connor, "Emergent Properties," p. 24.

¹³ Mark Bedau, "Weak Emergence" (1997), p. 3, at <http://people.reed.edu/~mab/papers/weak.emergence.pdf>.

¹⁴ O'Connor, "Groundwork for an Emergentist Account of the Mental," *PCID* 2.3.1 (October 2003): p. 4.

¹⁵ *Ibid.*

an unconvincing philosophy. What we have to do is find a middle way that avoids the extremes of substance dualism and reductionism of any sort, the first of which falls victim to logical puzzles and the second to both logical puzzles and compelling intuitions.

It is useful to keep in mind exactly why Descartes found dualism plausible in the first place. In his *Meditations* he wrote the following:

On the one hand I have a clear and distinct idea of myself, in so far as I am simply a thinking, non-extended thing; and on the other hand I have a distinct idea of my body, in so far as this is simply an extended, non-thinking thing.There is a great difference between the mind and the body, inasmuch as the body is by its very nature always divisible, while the mind is utterly indivisible. For when I consider the mind, or myself in so far as I am merely a thinking thing, I am unable to distinguish any parts within myself; I understand myself to be something quite single and complete. Although the whole mind seems to be united to the whole body, I recognize that if a foot or arm or any other part of the body is cut off, nothing has thereby been taken away from the mind....

He recognized that consciousness is partless or “non-structural,” unextended, unlike the physical objects of our experience. This means that it cannot be composed of physical things, as the cell is. Instead, it has to be emergent in some way. But its “emergence” is very different from the cell’s structural emergence: while the cell is indeed a new “thing” with its own emergent causal powers, it is not a new *kind* of thing. But consciousness is.

I said above that consciousness can be analyzed from two perspectives, viz. the physical and the non-physical. The former is the scientist’s approach when he considers consciousness as a series of neuronal firings and other electrochemical events. His approach is the “additive” one, the atomistic, non-holistic, aggregative one. As far as he is concerned, action potentials and excreted neurotransmitters and so forth are the building blocks of consciousness. He is not entirely wrong. For, while consciousness is not “made out of” them in the way that a cell is made out of molecules (i.e. structurally), when they interact together in a single holistic state they “produce”—or, from the holistic perspective, they *are*—a given conscious state, such as pain. The *neural impulses* are not the mental state, but the holistic, emergent totality of their interactions is. This totality can and

should be considered a physical phenomenon and, in a sense, a neural one. However, its privacy, its intentionality, and its qualitative features distinguish it from apparently every other physical phenomenon in the universe; and this is the sense in which we are committed to dualism. Not *substance* dualism, though, because it is misleading—indeed, meaningless—to call consciousness a substance. Such terminology serves no purpose; it only propagates misunderstandings. Nevertheless, what we are confronted with is a kind of dualism, involving a phenomenon (consciousness) that can be analyzed from both physical and non-physical viewpoints, or additive and holistic viewpoints. This, it seems to me, is the only plausible way to conceptualize consciousness. Whatever difficulties it gives rise to, the greater difficulties of every conceivable alternative make the emergentist approach the most appealing.

No doubt Colin McGinn is right when he says that, insofar as emergentism doesn't conceptually explain how the brain can give rise to consciousness, it is unsatisfactory. It leaves the mystery of how the subjective can possibly come from the objective mysterious. No amount of philosophical elaboration on the theses of emergentism can overcome that fact. It does no good, for example, to adopt Roger Sperry's line:

Instead of following the usual approaches that tried to inject conscious effects into the already established chain of microcausation, the logical impasse was resolved by leaving the microcausation intact but embedding it within higher brain-processes having subjective properties with their own higher-level type of causation, and by which the embedded micro-events are thereafter controlled.

....[E]xcitation of a cortical cell is enjoined into the higher dynamics of passing patterns of cognitive activity. A train of thought with one mental thought evoking another depends throughout on its neurocellular physiology and biochemistry. Nevertheless, like molecules in passing waves in a liquid, the brain cell activity is subject to higher-level dynamics which determine the overall patterns of the neuronal firing, not relative to other events within this particular brain process, but relative to the rest of the organism and its surroundings.¹⁶

¹⁶ Quoted in O'Connor, "Emergent Properties."

However true these ideas may be, they leave the basic mystery of the mind-body problem unsolved. They don't help to bridge the gap between the private and the public. Nevertheless, it would be wrong to reject emergentism for that reason. For it isn't as though the emergentist says *nothing*. His theory is not totally uninformative. While it cannot resolve the intuitive paradox of the mind's emergence from matter, no other theory can either. But its advantage over other theories is that it, so to speak, acknowledges its limitations, unlike, say, functionalism which in effect—by denying the significance or existence of the “private” dimension of experience—pretends the basic mystery doesn't exist.

I have not yet said very much, however, about what this theory entails. Unfortunately this question is not easy to answer—precisely because the relation between consciousness and the body is difficult to understand. Consciousness is emergent, but does that mean it exerts a downwardly causal influence on the electrochemical processes that are responsible for it? The cell's emergence is characterized by downward causation, but consciousness is not a structural entity. It is partless; it is not composed of molecules but rather of awareness, intentionality and qualia which, as O'Connor notes throughout his work, are non-structural phenomena. So if it exhibits downward causation, it must do so in a different way than the cell does. But what way is that? And if it doesn't, is it then non-epiphenomenal? How could mental states have causal powers if not by influencing the neural states on which mental states are known to supervene?

Let's look at this question more carefully. A mental state, I have said, is a neural state—a “radically emergent” one, but a neural state nonetheless. So, when saying that it “supervenes” on neural states, what I am really saying is that it supervenes on neural events considered *additively*, as individual firings of neurons and particular interactions between neurons. In other words, the macro-level holistic state supervenes on the micro-level neural events. Both “levels” consist of neural events, but one is holistic while the other events are “additive.” Given the causal closure of physics, it is through this series of micro-level events that any causality is manifested. The *holistic* neural event (e.g., a particular sensation of pain) cannot be the direct ‘mechanistic’ cause of a succeeding event (such as the act of removing one's hand from the fire), because all mechanistic, physical causation operates through individual “micro-level” neural events. There has to be this micro-causation.

Admittedly, it is conceivable that the act of removing the hand from the fire was “overdetermined,” in that it was caused *both* by a chain of neural events *and* by the sensation of

pain, but this hypothesis is philosophically unattractive. It is uneconomical and counterintuitive. Kim rightly rejects it in his paper “Making Sense of Emergence.” It’s implausible to think that an event has two independent causes that operate at the same exact time. One cause does all the work necessary.

Similarly, it’s absurd to say that the mental state somehow directly causes the neural events on which it supervenes. This form of downward causation is incoherent. It is literally senseless, for it entails causal circularity.

So, then, are mental states epiphenomenal? Are they mere side effects of underlying neural processes, possessing no causal powers of their own? O’Connor tries to avoid that conclusion in “The Metaphysics of Emergence.” He provides the reader with an elaborate diagram meant to show the (probable) relationship between mental states and neurophysiological events, according to which diagram a given mental state is the result of a complex array of neural events and mental states that somehow “work together” to produce the state in question.

As a fundamentally new kind of feature, [an emergent state] will confer causal capacities on the object that go beyond the summation of capacities directly conferred by the object’s microstructure. Its effects might include directly determining aspects of the microphysical structure of the object as well as generating other emergent states.¹⁷

Thus, an emergent state (says O’Connor) can be responsible for another emergent state as well as for the microphysical events on which such states supervene.

In his paper “Emergent Individuals” he describes his position as follows:

Since the...emergent states themselves will help to produce similar subsequent states—possibly resulting in a complex, stratified range of such states—the microphysics alone will *not* determine these later states. Likewise, emergent states will work in tandem with the underlying *micro*-states to determine later micro-states, manifesting a sort of “downward” causation. Hence, the existence of emergent states is contrary to the assumptions of much contemporary metaphysics

¹⁷ O’Connor and Wong, “The Metaphysics of Emergence,” p. 665.

and philosophy of mind, assumptions which typically include the truth of some fairly strong mental-physical supervenience thesis and the causal closure of the microphysical realm. Neither of these assumptions will hold if there are emergent states as here defined.

This conclusion itself makes O'Connor's position unattractive and implausible. For scientists have discovered no evidence to suggest that physics, or the microphysical realm, is causally incomplete. But there is another reason to reject O'Connor's ideas: they entail causal overdetermination. If micro-states *and* emergent states determine micro-states, then Kim's old objection is relevant again: the "activity of the emergent property" seems "redundant."¹⁸ Consciousness cannot *directly* determine a micro-state, in the way that prior microphysical events determine it. Sperry had already rejected such a theory when he rejected the attempt to "inject conscious effects into the already established chain of microcausation."

Indeed, I find Sperry's conceptualization more plausible than O'Connor's. Sperry acknowledges the causal closure of physics: on his version of emergentism, as I quoted, "microdeterminism is not so much refuted or falsified as it is supplemented." The way it is supplemented, argues Sperry, is similar to the way it is supplemented in the case of the cell: while the cell's properties supervene on the properties of its constituent molecules, the latter properties are diachronically determined by the holistic state of the cell. Its overall state in a given moment determines the molecular processes in the succeeding moment, on which processes supervenes the cell's "overall state" in that moment, which determines (through the micro-level events) the molecular processes in the next moment, etc. Similarly, it is plausible to suppose that "individual nerve impulses and other excitatory components of a cerebral activity pattern are simply carried along or shunted this way and that by the prevailing overall dynamics of the whole active process"¹⁹—dynamics that are integrally related to the emergent neural states that are the mental states supervening on the nerve impulses in question. Precisely what the relation is between these "overall [cerebral] dynamics" and the conscious states is unclear—Sperry himself seemed to identify the two—but it is possible that as the science of neurophysiology progresses, that question will become less mysterious.

¹⁸ Ibid., p. 670.

¹⁹ Sperry, quoted in "Making Sense of Emergence," p. 26.

Experiments have shown, for example, that dogs that cannot feel pain willingly stick their snouts into fire and casually withdraw them.²⁰ In general, a creature that doesn't feel pain acts differently from one that does. So sensations appear not to be epiphenomenal. It is plausible to suppose, then, that consciousness itself is not epiphenomenal. The question is *how* it isn't. How does it interact with neural events, and how does it determine an individual's behavior?

Such questions, in fact, bring us to the controversy over "free will" and its meaning. And here we are in a philosophical quagmire from which there is likely no escape. From the perspective of natural science, acts are determined by biological processes; from the perspective of the self, acts are determined by desires and reasons. Can these perspectives be reconciled? One is deterministic, the other not. According to one, there is no place for "self-control"; according to the other, there is. It is hard to see how such opposed viewpoints can be made compatible with each other. Either consciousness and the self have causal power or they don't. The biological sciences at least implicitly deny that they do (because biology rejects attempts to "inject conscious effects into the already established chain of microcausation"), whereas our self-experience at least implicitly affirms that they do. And both frameworks for interpreting ourselves seem to have irresistible power. So we appear to be at an impasse. It seems both obvious and impossible that consciousness is epiphenomenal.

It is possible, though I think unlikely, that in the future scientists and philosophers will somehow find a way out of these difficulties. For now, though, we are compelled to invoke the wisdom of Aristotle: as philosophy begins in wonder, so it ends in wonder.

²⁰ Colin Radford, "Pain and Pain Behavior," *Philosophy*, Vol. 47, No. 181 (Jul, 1972): 189–205.