7 Explaining the Ontological Emergence of Consciousness

Philip Woodward

1 Introduction: Can Emergentism Be Explanatory?

A family of so-called "anti-physicalist" arguments have been widely discussed over the last four decades.[1] Each of these arguments purports to demonstrate that phenomenal properties—the felt qualities of conscious experience—are not identical to, constituted by, or realized in non-phenomenal properties, but rather that some of them are *ontologically fundamental*.[2] Let us suppose for present purposes that at least one of these arguments is successful. What *positive* account of phenomenal properties can be given? Where, that is, do phenomenal properties come from, and how are they related to the rest of concrete actuality? Here, there is a divide between two schools of thought: *panpsychists* maintain that phenomenal properties are instantiated by the most basic building-blocks of reality; *emergentists* maintain that they emerge from reality once those building-blocks are suitably arranged. In this paper, I develop the latter option.

Some have suggested that emergentism is no positive account at all. According to this line of criticism, emergentists *appear* to be saying something positive—viz., that consciousness "emerges" from physical reality—but the only substantive way to unpack their claim is in terms of the rejection of other explanatory proposals (identity-theory, non-reductive physicalism, panpsychism, occasionalism, etc.). So, at best, emergentists have no explanation of phenomenal properties to offer.[3] At worst, emergentists rejects the very possibility of such an explanation. According to Thomas Nagel, for example, emergentism implies that phenomenal properties "are not explainable in terms of any more fundamental properties, known or unknown, of the constituents of the system."[4] If emergentists were right, then reality would contain fundamentally unintelligible aspects. This is a consequence Nagel, a friend of the Principle of Sufficient Reason, finds alarming.

But it is no part of emergentism to reject the possibility of, or simply to remain silent about, the explanation of consciousness. As Elanor Taylor (2015) has convincingly argued, that some phenomenon is emergent does not entail that it is inexplicable, but only that it can't be explained in the familiar, scientific way. So it is open to emergentists to provide explanations of phenomenal properties. The aim of the present essay is to contribute to that task.

What would a positive, emergentist explanation of phenomenal properties look like? To begin, I propose that we think of the connection between the emergence-base and that which emerges from the base in *causal terms*.[5] Causation, I will assume, is a fundamental aspect of the world. It occurs when causal powers are manifested. A causal power is a dispositional property whose nature consists entirely in (a) its proprietary manifestation, i.e. the effect at which it aims, in connection with (b) the conditions under which it manifests. *Fundamental* causal powers are those causal powers that are not identical to, constituted by, or realized in any other causal powers. These causal powers are the ultimate explainers of the causal dynamics of the world. In rare cases, fundamental causal powers manifest in an isolated fashion, e.g. in cases of radioactive decay. Typically, many instances of fundamental causal powers manifest jointly. (This is what is going on in collision-mechanics of any degree of complexity.) Joint manifestation can result in cancelation, where the resulting effect is not that at which any of the powers is aimed. Any static physical system is the result of such cancellation. But joint manifestation can also result in amplification, where the resulting effect is greater than that at which any of the powers is aimed. Any macro-level motion is the result of such amplification.

Emergence is the result of a special sort of joint power-manifestation. Consider, by way of analogy, the difference between my ability to contribute causally to the lifting of a car, on the one hand, and to the formation of a club, on the other. The first is an example of the ampliative, joint manifestation of (non-fundamental) causal powers. That is, if I coordinate my efforts with others, we together have the power to lift a car because each of us has the power to lift a part of the weight of the car. My contribution to the formation of a club is not quite like this, though. If I coordinate my efforts with others and we together form a club, this is not because I have on my own the power to form a part of a club. Instead, whatever power I have to form a club is *essentially* collective. It is not just a jointly manifested power, it is a jointly manifested *collective* power. Returning to the matter at hand: the emergence of consciousness is a matter of the joint manifestation of a collective consciousness-generating power had by the ultimate physical constituents of reality (henceforth, the UPCs).[6]

That is, at any rate, a first pass at a positive, causal theory of ontological emergence. But it is only a first pass, and there are a number of further explanatory challenges that face the theory. In the remainder of this paper I will discuss four such challenges:

1 *The Collaboration Problem*: How do UPCs jointly manifest their collective consciousness-generating power?
2 *The Threshold Problem*: Under what circumstances do UPCs jointly manifest their collective consciousness-generating power?
3 *The Subject Problem*: Which object is the bearer of emergent phenomenal states?
4 *The Specificity Problem*: What determines which specific phenomenal state is generated?

2 The Collaboration Problem and the Threshold Problem

The Collaboration Problem is the problem of accounting for how a bundle of UPCs can coordinate their causal efforts, so as to manifest their collective consciousness-generating power. The Threshold Problem is the problem of accounting for why some bundles of UPCs manifest a collective consciousness-generating power, but not all do so. These two problems are closely related, and it is impossible to provide an adequate solution to one without providing an adequate solution to the other. Together, they amount to what I believe is the most difficult explanatory challenge for ontological emergentism.

We now need to confront some disanalogies with the club-formation case mentioned above. When I join with others in forming a club, what makes possible the joint manifestation of our collective power is a certain shared intention to do so, and this shared intention is possible because of communication among us. I do not want to say that an aggregate of UPCs can share an intention, nor even that they can communicate with one another, in anything like the sense of 'communicate' relevant to the club-formation case. So, some other account is needed of the nature of coordination among the UPCs. And I think that this account has to underwrite a pretty strong form of unity among the UPCs. Precisely because the causal power in question is a *collective* power, the coordinated UPCs have to act *as one*; they are not each generating a part of phenomenal states but rather are jointly, synchronously generating entire phenomenal states. Fortunately, the idea of spatially separated entities acting as one is not an idea totally foreign or repugnant to contemporary physics (as it would have been 150 years ago). Quantum-entangled entities can act as one, even when separated by great distances. I tentatively speculate that quantum-entanglement in the brain is what makes coordination possible.[7] If this speculation proves unworkable, some heretofore unknown unification-relation will need to be posited in its place. (I note in passing that Paul Humphreys (1997) points to quantum entanglement as itself an example of an emergent property—the best example we have of one, in fact. Someone might worry that an explanation of one emergent phenomenon in terms of another is defective in some way. But I don't think so. It would surely be nice to know why entanglement occurs, and knowing this would help shed light on the Collaboration Problem. But I have no reason to expect that the explanation of entanglement will look anything like the explanation of consciousness. These are very different phenomena. It's entirely appropriate to appeal to one to explain the other.)

This cannot be the whole story, assuming that some entangled systems are not conscious. In order to solve the Threshold Problem, there must be a further condition on the manifestation of consciousness-generating powers. The Threshold Problem is the problem of describing the types and degrees of complexity that a system has to exhibit in order for its constituent UPCs to jointly manifest their consciousness-generating power. This is a contingent matter: the world could have been such that the threshold for the emergence of consciousness was much lower or much higher than it actually is. Thus, it is reasonable to approach the

Threshold Problem as an empirical problem rather than a philosophical problem. And that is more or less how I will treat it. But there are some conceptual constraints on the answer that are worth noting from the outset.

There are two extremes into which a proposed solution to the Threshold Problem could fall that would render the solution, if not beyond the pale, at least implausible. The first extreme sets the bar too low, such that any system that bears gross functional similarities to our brains—any system, say, that contains feedback loops—gives rise to consciousness. Such a theory carries the vice of *promiscuity*—of attributing consciousness to all sorts of things that common sense deems non-conscious. (As a colleague of mine likes to say, any theory that implies that a toilet is conscious is a bad theory.) The second extreme sets the bar too high, such that only systems that bear fine-grained structural or functional similarities to our brains—for example, systems that have distinguishable cortices and thalamuses that are functionally connected in complicated ways—give rise to consciousness. Such a theory carries the vice of *arbitrariness*—of attributing consciousness only to things with weirdly specific features. Arbitrariness, in the present context, is vicious for two reasons. First, the emergence of consciousness is the manifestation of a *fundamental* causal power, I am maintaining. But we would expect a fundamental causal power to have tidy and elegant manifestation-conditions rather than messy and ugly manifestation-conditions. Arbitrariness, in other words, is an indication that we haven't yet reached explanatory rock-bottom. Second, as Nagel (2010) stresses, a full explanation of consciousness includes an explanation of its evolutionary etiology. And to do this, it's not enough that we specify conditions such that, *once the universe satisfies them*, consciousness appears. Rather, we need to show how it was likely (or at any rate not massively unlikely) for the universe to produce those conditions. Arbitrary conditions are ones that we have no good reason to expect evolution to ever have produced.

Attractive solutions to the Threshold Problem will avoid the extremes of low-bar promiscuity and high-bar arbitrariness. And these are formidable conceptual constraints. It isn't obvious that *any* proposals will avoid both vices, let alone that empirically plausible proposals will. As we'll see, I'm pretty sure that my proposal avoids promiscuity, but I'm less sure it avoids arbitrariness.

After several decades of active, scientific research on the biological correlates of consciousness, broad consensus exists only with respect to a few general claims. First, all mammals exhibit consciousness of some form or another; probably some other animals do, too, beginning with birds. (Some researchers have concluded that practically all animals are conscious.) Second, conscious organisms have complex central nervous systems. Particular types of conscious contents—e.g. noticing a printed word—are associated with localized neural activity. The richness and sophistication of an organism's capacities for cognition and consciousness are loosely correlated with the size of its brain (both absolute size and also relative size, i.e., the proportion of the mass of the organism taken up by its brain). Many researchers take it for granted that the features of nervous systems most directly relevant to consciousness are their ability to transmit electrical signals through neural networks, though, notably, some have drawn attention to features

at a higher level (such as brain-wave patterns resulting from synchronous electrical activity in populations of neurons) or at a lower-level (e.g. quantum coherence in the microtubules of neurons). Third, throughout evolutionary history, central nervous systems develop in complexity roughly in keeping with the complexity and versatility of the bodies that house them. The more ways for the body to interact with its environment, the more extensive a nervous system is needed to direct and coordinate its movements.

At this point, if we want a more detailed picture of the biological correlates of consciousness, we leave the domain of consensus and enter ongoing debates within the science of consciousness. Conveniently, we find that there are roughly two camps within the debate: those who argue for what Peter Godfrey-Smith (2016) calls "transformation" theories of consciousness versus those who argue for "latecomer" theories of consciousness. On transformation theories, consciousness can be found quite low on the tree of life (at least as far down as bees and shrimp), albeit primitively, in the form of basic perceptual and somatic sensations. Mammalian consciousness, while much more sophisticated, is nevertheless an evolutionary outgrowth of these early forms of consciousness. (Evidence for these theories comes in the form of recognizable pain-behavior, such as nursing damaged body parts, and cognitively sophisticated behavior, such as a hermit crab's willingness to endure electrical shocks in order to remain in a high-quality shell.) On latecomer theories, in contrast, consciousness only shows up once more sophisticated cognitive apparatuses are in place. Advocates of latecomer theories point to recent studies that show just how much of human behavior is controlled unconsciously, from the non-reflective way skilled athletes prepare their next play, to elaborate behaviors, such as sleep-walking and sleep-driving, performed by thoroughly non-conscious subjects.

It seems to me that these recent studies of unconscious behavior do undercut the evidence in favor of the transformation theories somewhat: if complex, apparently goal-directed behaviors in humans can be unconsciously caused, we don't have a very strong reason to attribute conscious causes to complex, apparently goal-directed behaviors performed by tiny non-humans. (This is not to say that these studies *falsify* transformation theories, which they certainly do not.) Moreover, I suspect that transformation theories are going to spell promiscuity trouble for ontological emergentism: whatever features bees and shrimp have that advocates of transformational theories treat as the key to their being conscious, robots and eco-systems may very well share them. So, for these reasons, I opt for a latecomer-inspired solution to the Threshold Problem.[8]

I base my (admittedly speculative) latecomer-style solution to the Threshold Problem on Ezequiel Morsella's (2005) "supramodular interaction theory" of consciousness. Many empirical theories of consciousness, especially of the latecomer variety, start from the idea that consciousness plays the role of integrating otherwise functionally separated cognitive processing in the brain. Morsella develops a version of this idea.[9] He takes his cue from an observation about two different types of cognitive "conflict" that regularly occur. The first involves the pre-conscious reconciliation of stimuli that do not accord with one another in their

usual way. Examples include binocular rivalry (distinct images are presented to each of our eyes, but the stimulus from only one retina at a time reaches consciousness) and ventriloquism (speech is both heard and seen, but the auditory stimulus comes from a slightly different angle from that of the visual stimulus, and the "error" is corrected pre-consciously). The second type of cognitive conflict involves conflicting behavioral urges. For example, when carrying a hot plate of food from the kitchen to the dining table, one feels impelled to let go of the plate, but also to get the food to its intended destination.

What explains the fact that conflicts of the former type are typically resolved via non-conscious processes, but conflicts of the latter type are typically resolved in consciousness? Morsella's answer is that the former involve conflicts among the outputs of low-level cognitive modules—visual vs. auditory processing, for example—whereas the latter involve conflicts among the outputs of aggregates of such modules. Paradigms of such "supramodules" are various incentive systems, e.g. those that motivate behaviors we ordinarily associate with pain, fear, hunger, and thirst. (Morsella also seems to treat perceptual recognition and perceptual binding—the attribution, in perceptual experience, of multiple features, within and across sense modalities, to the same object—as the work of supramodules.) It is precisely the outputs of such systems that come into conflict in cases such as carrying a hot plate of food. If we suppose that an evolutionarily old function of consciousness was to resolve such conflicts, then we should expect that the contents of consciousness include the outputs of supramodules, rather than the outputs of the smaller modules that make them up. Now, there are plenty of conscious episodes that do *not* include conflicting impulses. But just as most traffic lights run through their cycles whether or not cars are present at the intersection, so consciousness is always "on," broadcasting the outputs of supermodules, in case a conflict arises.

Suppose Morsella is right about the cognitive function of consciousness. We can still ask *why* consciousness has that function. Why aren't lower-level cognitive conflicts brought to consciousness? Why aren't supramodular cognitive conflicts resolved unconsciously? One answer is that evolution just happens to have selected consciousness to fulfill this function, though it needn't have.[10] This answer is unsatisfying because it fails to account for our intuitive sense that consciousness is a non-negotiable aspect of being human, of having the minds we have. And I think a better answer is available. Consciousness affords a cognitive asset that cannot be acquired without it: *narrow content*. While non-conscious systems are able to make use of internal signaling to track features of the external world and to monitor their own internal functioning, such signaling never means anything intrinsically; its representational content comes in the form of external, tracking relations. But the representational content of consciousness is not exhausted by such signaling. In consciousness, we are able to grasp semantic contents, at least of a subset of the constituents of our thoughts, at least to a certain degree of specificity.[11] And it makes all kinds of sense why narrow content should show up right when an organism needs to settle conflicts among incentive-systems: it's impossible to adjudicate between two action-plans without

understanding what those actions involve and without having a *grasp* of one's environment, one's options, the likely consequences of those options, etc. All of this needs to be directly available to the adjudicator (again, at least to a certain degree of specificity).

Some readers are likely to be surprised by the suggestion that the baseline function of consciousness is to adjudicate among consciously grasped action-plans. Consciousness sure seems to come in a much more primitive form than this, viz., the raw feel of perceptual sensation, emotion, bodily pain and pleasure, and so forth. Is something as cognitively fancy as *weighing one's options* really the scaffold upon which consciousness is built, rather than the edifice itself? In response to this concern, I think that we can understand *the point* of primitive sensations only in the context of richer conscious goings-on. Consider the fact that reflex actions aren't mediated by pain-sensations: if you run your foot against a sharp object, your leg recoils before you feel anything at all. It's only once an organism is in a position to behave in a variety of ways, in response to potentially conflicting incentives, that it makes sense to *feel the pull* (or push) of those incentives. Likewise, only once an organism is in a position to use perceptual information to guide action does it make sense for such information to be consciously presented.

At long last, I present my proposed solution to the Threshold Problem. A system gives rise to consciousness only when, first, it is the functional equivalent of a central nervous system—which is to say, when its parts are causally hooked up in a way such that (a) environement-driven (i.e. sensory) changes at its outer edges modulate system-driven (i.e. behavioral) changes elsewhere in its outer edges, with something like signal-processing going on in between, and (b) multiple such input-output channels intersect and influence one another (feedback loops may be necessary here as well); and, second, this functional structure settles into more-or-less discrete output-modulating sub-systems that can—or perhaps, have begun to—*clash* with one another.

How many input-output channels do there have to be? How many clashing output-modules do there have to be? I don't know. The details matter, of course, if I am going to avoid arbitrariness. (*Two* or *three* would be tidy answers. *Fourteen* would not be a tidy answer.) My proposal may run afoul of vagueness-worries, too: what exactly is it for "something like signal-processing" to mediate between inputs and outputs? I'm not sure what to say. It is known that complex systems can undergo dramatic, sudden transitions when they reach certain levels of complexity.[12] It's possible that the sharp boundaries marked by these "non-linear" transitions could resolve worries about arbitrariness and vagueness that are likely to come up once the details are fleshed out.[13]

3 The Subject Problem

Let us suppose that what I have said so far is adequate: I have accounted for *how* UPCs coordinate their causal efforts (via entanglement or something similar) and *when* they do so (when they form a system that contains potentially conflicting behavior-driving supramodules). Under these conditions, UPCs jointly produce

phenomenal states. But where do these phenomenal states "land," so to speak? Which object instantiates them? This is the Subject Problem. Dean Zimmerman and William Hasker have both argued that even if aggregates of UPCs are responsible for the generation of consciousness, *subjects* (i.e., whichever objects instantiate emergent phenomenal states) are not themselves aggregates of UPCs. Hence the ontological emergentist about phenomenal states must be an emergent *dualist* (according to which subjects are non-physical simples), rather than an emergent *materialist* (according to which subjects are physical entities of some sort).

Zimmerman (2010) argues that all material candidates for being the bearers of emergent phenomenal properties are problematically vague. Given emergentism, the instantiation of phenomenal properties—or 'qualia,' as Zimmerman prefers to call them—is not necessitated by the laws of physics, but must be governed by "fundamental laws of qualia generation." These laws specify (a) the conditions under which a quale is generated, (b) which particular quale is generated, and (c) which object instantiates it. Hence, the bearers of qualia must be mentioned in those laws that amount to the explanatory grounds of their emergence. But which objects are these? Here a dilemma opens up for the emergent materialist. On the one hand, qualia-bearers could be what Zimmerman calls "Garden-Variety Objects," i.e., organisms or parts of organisms such as brains or central nervous systems. But all of these objects have vague spatiotemporal boundaries; there is no fact of the matter about where/when they begin and end. And this is a problem, says Zimmerman, because *fundamental laws don't mention vague objects*. On the other hand, qualia-bearers might be none of these familiar, vague objects, but rather unfamiliar sharp objects. But how are we to decide between the many, sharp objects that partly overlap the vague ones? It is hard to imagine a metaphysically respectable criterion.

A second argument comes from William Hasker (2016). Hasker is interested in the apparent mismatch between the unity of phenomenal states and the multiplicity of UPCs that make up material composites. How can a unified phenomenal state be instantiated by a multiplicity of constituents? One option is that the state, Q, is instantiated by the material composite, O, in virtue of *parts* of Q being instantiated by *constituents of* O. But Hasker takes it as a given that this is impossible. Just as your instantiating a phenomenal state and my instantiating a phenomenal state could never combine to form a third, joint phenomenal state—your state remains forever privately yours and mine forever privately mine—so Q could never be the result of some sort of combination of the phenomenal properties instantiated by the various UPCs one by one. Thus Hasker seems to take the unitary nature of phenomenal states to imply that such states are not mereologically composed.[14] A second option is that Q is instantiated by O in virtue of Q's being instantiated wholly by every constituent of O. But this would mean that when Hasker is enjoying Beethoven's Ninth Symphony, so is each quark in his brain (or whatever object O is supposed to be)—and Hasker finds that this idea "strain[s] one's credulity to the breaking point, and beyond." A final option is that Q is instantiated by O as a whole but not in virtue of the properties of its constituents; it is "spread out," as it were, all over O, and only over O. But if this were so, Hasker reasons, then not only would the properties that make up Q fail to be found in every proper part of O, they would fail to be found in *any* proper part of O. And among the proper parts of O is the fusion of all of its constituent UPCs *save for one quark*. The implication would be that in O as a whole, but not in O minus one quark, the properties that make up Q are to be found. Hasker does not provide an argument for why he rejects this result, but perhaps his reasoning is similar to Zimmerman's: making such a cut-off will be metaphysically arbitrary; nature simply doesn't supply a criterion for doing so. The consequence is that there is no way for a composite such as a brain to instantiate Q.

Both of these arguments can be resisted, in light of the account of emergence developed so far. The material object O that instantiates emergent phenomenal states has a sharp boundary, and there is a non-arbitrary criterion for delineating this boundary. Central nervous systems may very well be vague objects: there are UPCs such that it is vague whether they are part of my CNS. It will not, however, be vague which of the UPCs are contributing at any one time to the joint production of my consciousness. This is true even if more of the UPCs contribute to the generation of consciousness than are required to do so. (Compare: the club could have been formed if one of its founding members had not shown up for its founding. Nevertheless, all of those who contributed to its formation are part of it, at least at the time of its formation.) The subject of a conscious state is to be identified, at a time, with that material composite coincident with the set of UPCs causally responsible for the generation of that subjects' conscious state at that time. Emergent Materialism thus survives the criticisms of Zimmerman and Hasker. We can embrace the second horn of Zimmerman's dilemma: O, the bearer of emergent phenomenal state Q, is sharply-bounded. And we can embrace the third horn of Hasker's trilemma: that O instantiates emergent phenomenal properties does not imply that any proper parts of O instantiate emergent phenomenal properties.

Still, there are reasons in the neighborhood of those adduced by Zimmerman and Hasker to worry that this account is unsatisfactory. The concern I have in mind is whether O, sharply bounded though it may be, is the right sort of thing to be the bearer of fundamental properties. O is what we might call a "loose composite": it is an aggregate of UPCs that more or less cohere with one another. Call a "basic bearer" of a property something whose instantiating of a property is not in virtue of any other property-instantiation. (If O instantiates Q, but it does so not because any of O's parts instantiate Q or part of Q, then O is a basic bearer of Q.) There is reason to wonder whether loose composites can be basic bearers of fundamental properties.

Suppose there are three UPCs, A, B, and C, scattered throughout the universe, but exhibiting the following peculiar commonality: wherever one is found, there also is found an instance of a fundamental property—a color, let's say—or better, an instance of the dynamic unfolding of a *sequence* of colors. Let's call the fusion of A, B, and C 'Comp'. Suppose we were asked to specify the basic bearer or bearers of the color sequence. Should we say that there are three basic bearers (A, B, and C), or that there is one basic bearer (Comp)? Without any more

information about the case, there is no reason to say that there is one basic bearer rather than three—and if there is no reason to say so, then there is nothing to *make* it so. If *Comp* itself instantiates the color-sequence, it does so derivatively. Now, couldn't we say in response: *there's just a fact of the matter* as to whether there is one basic bearer or three? I have no conclusive argument against saying this. But neither is it satisfying. A widely held assumption in metaphysics is that we should privilege bottom-up explanation. That is, we should start by assuming that properties of composites are determined by properties of their parts, and give up this assumption only when bottom-up explanation fails. And when bottom-up explanation fails—as it would, were Comp a basic bearer of the color-sequence—such breakdown needs to be explained.

If this is right, it spells trouble for the version of emergent materialism I have so far sketched. The UPCs that constitute my CNS are also "scattered," if only just barely: the fact that they are in close spatial and causal proximity doesn't change the status of my CNS as an aggregate or as composite in only a *loose* way. So we should reach the parallel conclusion regarding the bearer of emergent phenomenal properties: when UPCs conspire to jointly generate phenomenal properties, there is no good reason to say that the generated properties are instanced only once, with the relevant *composite* as their bearer, rather than that they are instanced many times over, as many times as there are UPCs that form the composite.[15]

So even if there is nothing objectionable about the very idea that a composite could instantiate a phenomenal property without any of its parts instantiating that property (or any part of that property), there is a closely related idea that *is* objectionable, viz., that a mere aggregate could instantiate a fundamental property, without any of its parts instantiating that property (or any part of that property).[16] So the emergent materialist needs a way to affirm that subjects are composite, material entities while denying that they are mere aggregates. For emergent materialism to be successful, that is, subjects need to be *strict* composites, i.e. true unities, bona fide individuals, despite having material parts.

What might a strict composite be? A strict composite must include some element or elements that explain its deep, objective metaphysical unity.[17] One account of strict composites is the 'Emergent Individuals' view of Timothy O'Connor and Jonathan Jacobs ([2003] & [2010]). O'Connor and Jacobs follow David Armstrong in holding that fundamental particulars have a complex structure, consisting of (a) one or more immanent universals, and (b) a "thin particular," i.e. an entity which particularizes universals when they inhere in that entity. So-structured fundamental particulars make up the world of our acquaintance, replete as it is with loose composites of all sorts. But under certain circumstances, composites themselves can come to have their own proprietary thin particular. Such composites are the bearers of emergent phenomenal properties, and are themselves the products of emergence: they are materially-composed *emergent individuals*.

I am not sure that the Emergent Individuals view of O'Connor and Jacobs is the only way to account for strict composites. Perhaps we need not invoke a special, new *thing* (such as a thin particular) to do the job; perhaps certain new properties and relations could do the job instead. Perhaps, for example, there are

special, contingent building-relations that hold between UPCs when those UPCs form strict composites. But I confess I am skeptical that such accounts could deliver the goods: shouldn't we just say in such cases that the UPCs form a comparatively tight-knit aggregate, rather than that, by anybody's reckoning including God's, a *new thing* has come to be?

There is no space here to follow these lines of inquiry.[18] The broader point is this: emergent materialists can supply a viable metaphysics of the bearers of emergent phenomenal properties, but they must inflate their ontology a bit: they need to say that psychological subjects are not mere aggregates but bona fide individuals in their own right, to be included in any minimal inventory of what exists. The resulting picture, it must be granted, is very similar in many respects to that of the emergent dualists. Representatives from both camps can agree on the following:

> Psychological subjects depend on and emerge out of a physical aggregate, but are not identical to that aggregate. Psychological subjects are *fundamental* entities: they must be included even in the most austere inventory of what exists.

And if the Emergent Individuals' view is the right account of strict composites, we can add:

> The generation of psychological subjects involves the generation of an entity that is not materially composed [for the dualist, an immaterial subject; for the materialist, a thin particular].

To be sure, emergent materialism is more parsimonious than emergent dualism. But the margin of victory is much smaller than is usually supposed. And thus, my defense of the *viability* of emergent materialism (contra the challenges of Zimmerman and Hasker) can also serve as an argument against the *wild implausibility* of dualism.

4 The Specificity Problem

The Specificity Problem is the problem of accounting for the *dynamics* of consciousness.

Consider the obvious fact that one's consciousness is populated by different phenomenal properties at different times. For example, when I close my eyes, my visual experience changes completely. We might call this the problem of *qualitative* specificity: why are these specific phenomenal properties generated, but not others?

A closely related problem is the problem of *structural* specificity. Consider the obvious fact that one's consciousness contains structure of various sorts. For example, some of my phenomenal properties hang together as parts of a single visual field, while others hang together as part of an auditory field. This is the *multi-modal* structure of consciousness. My phenomenal states present me with intentional content, for example, the simultaneous experience of a red circle and a

blue square. I don't just experience redness, blueness, circularity, and squarehood; rather, the colors are *attributed* in a particular way to the geometrical shapes. This is the *semantic structure* of consciousness. Furthermore, some aspects of my experience are focal and others are peripheral; this is the *attentional* structure of consciousness. The problem of structural specificity is the problem of understanding why the component qualities in a phenomenal state are multi-modally, semantically, or attentionally structured in one way rather than another.

There are roughly two strategies for solving the Specificity Problem (in its various facets): *Bottom-up* strategies and *top-down* strategies. According to bottom-up strategies, the specificity of emergent phenomenal states can be explained in the same way as the emergence of consciousness generally, viz., in terms of the causal powers of the UPCs. According to top-down strategies, the specificity of emergent phenomenal states cannot be explained in terms of the causal powers of the UPCs; an additional "top-down" causal element is required. Views that employ the top-down strategy I'll call "Top-Down Property Emergence," or TDPE; views that employ the bottom-up strategy I'll call "Bottom-Up Property Emergence," or BUPE.

Here is the simplest version of BUPE. Each type of UPC is responsible for one type of phenomenal property: one type of UPC contributes to phenomenal blue; another type contributes to painfulness; and so forth. When a collection of UPCs generates a phenomenal state, some of the UPCs in the collection contribute to particular qualities to the state. Hence, UPCs and their powers are all that is needed to explain the emergence of specific phenomenal states. Call this "atomic" BUPE.

There are two problems with this view. First of all, it does not address the problem of *structural* specificity at all. For example, there is a difference between a visual experience as of a blue square and a red circle, vs. a blue circle and a red square. The difference is not in which qualities are present but in which qualities are attributed to which apparent objects. An explanation of which qualities the UPCs produce is not by itself an explanation of why they are structured the way they are.

A second problem with atomic BUPE is that it comports awkwardly with what we know about the neural correlates of specific phenomenal states; viz., that such correlates are *high-level functional properties of the brain*. Neuroscience has discovered that phenomenal properties of certain types are correlated with (more or less) localizable activation-patterns in the brain. We know, for example, that particular visual experiences are occasioned by retinal stimulation followed by activation of particular regions in visual cortex; particular auditory experiences are occasioned by stimulation of the inner-ear structures followed by activation of particular regions in the auditory cortex, and so forth. We know, in other words, that which type of experience a subject is having is a matter of which processing-streams in the brain are active. But neuroscience has revealed, further, that these sorts of processing-streams are *multiply realized*. First of all, there is considerable variation in how types of processing-streams are implemented, both across brains and in the same brain at different times (in stroke patients, for example). Second, functionally identical neuronal structures can be made of different types of organic molecule.[19]

To see why atomic BUPE has trouble accounting for these facts about the neural correlates of consciousness, let us consider the "perspective," so to speak,

of one of the UPCs—let's call it 'Ult'—that is partly responsible (per BUPE) for some phenomenal state Q of organism O. Ult contributes some quality—phenomenal blue, say—to Q when certain functional states obtain in O's visual cortex. When O's eyes are closed (or were O to go blind), such that activity in visual-processing pathways ceases, Ult likewise ceases to contribute phenomenal blue to Q. This means that Ult needs to be responsive in some way to whether certain neural pathways are active in visual cortex. That is to say: the manifestation-conditions for Ult's phenomenal-blue generating power include the obtaining of high-level functional properties in the brain. Of course, high-level functional properties are *realized* by fundamental physical states. But because they are *multiply realizable*, in order to cash out the manifestation-conditions for Ult's phenomenal-blue generating power in physical terms, one would have to disjoin all the possible realizers. We ought to avoid attributing essentially disjunctive manifestation-conditions to a fundamental causal power if at all possible. But then it looks as though this view cannot supply bottom-up explanations after all since the manifestation-conditions for Ult's power will include the functional structure of the brain in which Ult is embedded.

In light of these difficulties, we might revise BUPE as follows. Individual UPCs do not contribute specific properties to phenomenal states. Rather, just as the generation of consciousness is an essentially collective effort, so the generation of the particular properties that make up states of consciousness is also an essentially collective effort. The idea here would be that specific, structured phenomenal states are brought into being in their entirety, as the result of the exercise of the consciousness-generating power *of the collection* of UPCs, in *its* entirety. And *which* specific, structured phenomenal state gets generated at a time is a matter of the causal relations among the UPCs. That is, the unique ways that the brain is activated at a time are relevant to which phenomenal state the brain generates at that time. Call this "holistic BUPE."[20]

Holistic BUPE certainly handles the problem of structural specificity better than atomic BUPE does: since phenomenal properties are not generated piecemeal but rather are generated as parts of complex phenomenal states, the structural elements of such states require no additional explanation. But problems related to multiply realizability remain. On the present proposal, collections of UPCs have the power to generate specific phenomenal states. We should probably understand this to mean that collections of UPCs have a single, "multi-track" power. (A multi-track power is a power that cannot be fully characterized by a single conditional of the form, "under conditions C, manifestation M occurs." The relevant power cannot be so characterized because it admits lots of types of manifestation—as many types as there are possible phenomenal states for it to generate. Alternatively, we could understand collections of UPCs to have as many *powers* as there are possible phenomenal states for it to generate. But we should avoid multiplying powers in such a fashion.) The problem related to multiple realizability comes in when we try to specify the conditions under which the power manifests in one of its many ways. The conditions under which brains generate phenomenal states of a particular type are not happily described in the language of

physics, but rather in the language of high-level, multiply realizable neural functions. Again, we ought to avoid attributing essentially disjunctive manifestation-conditions to a fundamental causal power if at all possible. The general lesson is this: neural-functional states are relevant, as such, to the specificity of phenomenal states. Because neural-functional states are multiply realizable, no elegant, bottom-up explanation of the specificity of phenomenal states is forthcoming.

So, we should go in for a version of TDPE. What would a "top-down" theory of the emergence of phenomenal properties look like? To start with, the bearers of phenomenal-property-generating powers would need to be at least as "high-level" as neural pathways, since it is the goings-on in such entities that, as a matter of empirical fact, account for which properties are generated. A natural suggestion, then, would be to treat neural pathways as the bearers of the relevant powers. Nor is this an utterly counterintuitive suggestion: it is easy to think of the pathways in the visual system, for example, as taking sensory stimulation *as input* and as generating perceptual experiences (inter alia) *as output*—somewhat akin to the way a radio receives radio-waves as input and generates audible sound as output. But it is hard to work out the details of this suggestion. *Neural pathways* are odd entities—vaguely bounded, frequently morphing, constituted in part by the functions they implement. How are they individuated? Under what circumstances do they come to instantiate phenomenality generating powers? The story is bound to be complicated; it may require treating neural pathways as emergent individuals in their own right (in addition to the emergent individual that serves as the bearer of the properties generated by the neural pathways). Furthermore, going this route would amount to a return to the "piecemeal" approach to the generation of phenomenal states. As we saw in connection with the first version of BUPE discussed above, a piecemeal approach to the generation of phenomenal states doesn't seem to have the resources to account for structural specificity.

So, I think we should bypass such "mid-level" approaches and go truly top-down: it is the emergent subject that is the bearer of a phenomenality generating power. An emergent subject generates and instantiates phenomenal states, in response to states of the collection of UPCs that make it up. The states to which it is responsive can, of course, include "high-level" states such as neural-functional states. Thus, problems related to multiple realizability do not come up for TDPE. Because emergent subjects generate whole phenomenal states, no problems related to structure come up, either. This picture of the relationship between phenomenal states and brain states suggests a unique way of understanding the phenomenality generating power of emergent subjects: it is very much like an *interpretive* power. Emergent subjects generate phenomenal states that amount to interpretations of the goings-on in the brain.[21]

5 Conclusion

I previously said that collections of UPCs share a consciousness-generating power. When this power is manifested, the immediate result is that these UPCs both (a) compose an emergent individual, and (b) generate a phenomenal state instantiated by that emergent individual. We can now see that picture is not quite right. Here is the more refined picture of the ontological emergence of consciousness that has been developed in this essay: collections of UPCs share a *subject-forming* power. When UPCs that are parts of a system complex enough to exhibit supramodular conflict become entangled, they jointly manifest their subject-forming power. The emergent subjects, thereby formed, exhibit a novel causal power: the power to generate phenomenal states, which they themselves instantiate: states that "interpret" what is going on in the brain.

Does this mean that subjects are passive observers of brains, as it were, rather than causal contributors to the dynamics of the systems that give rise to them? It does not. Recall my discussion of the Threshold Problem above: UPCs give rise to conscious subjects precisely at the stage of development of the system at which consciousness can play a crucial functional role, viz., deciding between competing behavioral options. Emergent subjects must, therefore, have the capacity to influence behavioral outputs of the systems that give rise to them. The better the conscious "interpretation" of the brain that the subject produces—the more effectively it makes use of the brain as a transducer of information about the environment and about behavioral options viz-a-viz that environment—the more effective it will become at executing actions that promote the survival and well-being of the system. Thus, subjects with more sophisticated brain-interpretative powers will prove more adaptive. And as subjects get better at using the information the brain provides, more complex, environmentally attuned brains are likely to develop. Once emergent subjects show up on the evolutionary scene, minds and brains evolve together.

Notes

1 Examples include The Conceivability Argument (Chalmers [1996]), The Knowledge Argument (Jackson [1982]), The Explanatory Gap Argument (Levine [1983]), The Modal Argument (Kripke [1980]), The Property Dualism Argument (White [2010]), The Argument from Revelation (Stoljar [2009]), The Structure-and-Dynamics Argument (Chalmers [2003]), and The Unity-of-Consciousness Argument (Hasker [1999] and LaRock [2007]).

2 Some phenomenal properties are identical to, constituted by, and/or realized in *other phenomenal properties*. Thus, not all phenomenal properties are fundamental.

3 Cf. Bennett (ms), Kagan (2012). Bennett and Kagan target dualisms in general, rather than emergentism in particular.

4 Nagel (1979: 186).

5 Here I follow O'Connor and Wong (2005).

6 The view I sketch in section 4 revises this picture slightly.

7 The idea that quantum phenomena in the brain are relevant to consciousness is somewhat new, but is becoming more widely explored. See e.g. Schwartz, et al. (2005) and Fisher (2015). For a skeptical take, see Koch and Hepp (2006).

8 If a transformationist theory turns out to be correct, a solution to the Threshold Problem could still be formulated to accommodate it. Maybe the best transformationist theory of consciousness won't turn out to be promiscuous after all. Even if it is—for example, if it entails that all sensorimotor systems give rise to consciousness—there are other moves that can be made. Note that solutions to the Collaboration and Threshold

problems are dissociable. Perhaps some sensorimotor systems fail to exhibit quantum entanglement of the right sort; perhaps nearly all do. Another possibility is that consciousness is far more widespread than we previously thought, but that most instances of it are *extremely* primitive.

9 Morsella (2005: 1002) lists a dozen or so scholars whose work falls under this "integration consensus."

10 The question of the cognitive function of consciousness seems to have hamstrung many cognitive scientists, because they have assumed either that consciousness has no function or that it has a function nothing else could play. See Van Gulick (1989) and Polger (2007).

11 Cf. Horgan (2013), who argues that consciousness alone explains what is semantically defective about the linguistic processing that goes on in Searle's "Chinese Room." I say more about the connection between content and consciousness in my paper (2016).

12 Kelso (1995).

13 One further point: in order for my solutions to the Collaboration Problem and the Threshold Problem to come together properly to form a sufficient condition for the generation of consciousness, entangled UPCs need to be able to "know," as it were, that the relevant threshold has been met by the system in which they are embedded. One solution is to say that *all* the UPCs in the system need to be entangled, but this won't be necessary. All that is required is the entanglement of enough UPCs that are more-or-less directly involved in the signal-sending functions of the system. For example, the relevant UPCs might be those involved with ionization in axon membranes.

14 Matters are complicated here, because phenomenal states clearly *have* parts: they can consist in part in perception and part in conscious thought; they can consist in part in pleasant sensations and part in painful sensations; perhaps they can be partly representational and partly purely qualitative. I gather that Hasker would grant all of this but be unmoved in his core intuition that phenomenal states cannot have actually existing, numerically distinct phenomenal states as proper parts (or at least that such parts could be metaphysically prior to the whole). Panpsychists will deny Hasker's core intuition. The reason it is not dialectically otiose for Hasker to rely on his core intuition in the present context is his disagreement is with fellow emergentists. The most common reason (though by no means the only one) to find emergentism preferable to panpsychism is because panpsychism requires just this sort of mereological combination of phenomenal states.

15 Does the fact (if it is a fact) that they are *entangled* change their status as an aggregate? I don't see that it does. To say that entangled UPCs *act as one* is not to say that they have literally *become one*. If entangled entities do in fact become one, then entangled entities *eo ipso* form a strict composite, and the challenge I am discussing is disarmed.

16 In conversation, Hasker has criticized what he calls the "magical holism" required by emergent materialists: their view implies that UPCs magically fuse to become a unified subject. So, the criticism I am currently unpacking might be aptly attributed to Hasker, though it does not show up in his paper.

17 Note that providing an account of this element is not the same as providing the conditions necessary and sufficient for a strict composite to exist. When, for example, Peter van Inwagen says that simples compose a new thing if they jointly form a *life*, he is not *explaining* the unity of such composites; he is only telling us when and where to find such composites.

18 Nor is there space to explore the most vexing objection to emergent individuals, viz., that there is no sense to be made of a thin particular's individuating a composite as such. Each of the simples that make up a composite has its own thin particular. Does the newly emergent thin particular somehow encompass all of these others? What relation underwrites this "encompassing"?

19 Aizawa and Gilet (2009).

20 There is also room for a 'molecular' version of BUPE, according to which *clusters* of UPCs—proper subsets of the collection of UPCs that together generate Q—are responsible for particular qualities that make up Q. This version runs into the same trouble as the atomic version when it comes to explaining semantic structure.

21 One might be puzzled by the suggestion that generating phenomenal states is something that subjects *do*. I am a subject; I do things such as think and act and perceive, and so forth; *interpreting my brain* is not among these things that I do. Quite so, yet there *is* a sense in which interpreting my brain is something I do: this is the same sense in which metabolizing and processing sensory inputs and regulating my heart rate are things that I do.