

Panpsychism, Aggregation and Combinatorial Infusion

Abstract

Deferential Monadic Panpsychism (DMP) is a view that accepts that physical science is capable of discovering the basic structure of reality. However, it denies that reality is fully and exhaustively described purely in terms of physical science. Consciousness is missing from the physical description and cannot be reduced to it. DMP explores the idea that the physically fundamental features of the world possess some intrinsic mental aspect. It thereby faces a severe problem of understanding how more complex mental states *emerge* from the mental features of the fundamental features. Here I explore the idea that a new form of aggregative emergence, which I call ‘combinatorial infusion’ could shed light on this problem and bolster the prospects for this form of panpsychism.

1 Deferential Monadic Panpsychism (DMP)

The core claims of any panpsychist doctrine is that mentality, primarily and most especially consciousness, is fundamental and ubiquitous. The fundamentality of the mental entails that it is not reducible to or explicable in terms of entirely non-mental features of the world. Ubiquity requires that the mental be in some sense involved with everything in the world. It is important to stress however that the ubiquity of the mental does not require that everything in the world be possessed of mind. If, for example, there is a finite set of types of fundamental entities from which all other entities, and all phenomena, in the world emerge then panpsychism only requires that these fundamental entities have a mental dimension to them. It does not demand that everything in the non-fundamental domain also exemplify mental properties.

Beyond these basic principles, panpsychism can be articulated in several different ways. Idealism satisfies the core principles but denies that anything at all is non-mental. This is a rather uninteresting version of the theory, albeit one with immense historical significance, which can be regarded as a limiting case.

Much more interesting are accounts of the dual aspect or neutral monistic forms. The former regards both mind and matter as co-fundamental characteristics of reality. Perhaps the most famous proponent of a dual-aspect theory of mind and matter is Spinoza who accepted both the mental and material as equally fundamental, mutually irreducible but perfectly correlated, characteristics of Nature, which he also referred to as ‘God’. That is, at the fundamental level reality is *both* mental and physical (in fact, Spinoza believed that nature possessed an infinity of equally fundamental, mutually irreducible, attributes) at the fundamental level. In my view, Spinoza (and by extension all dual aspect theorists) regards Nature as equally a mental and a physical substance—there are no attributes which lie ‘beneath’ the mental and physical aspects. They provide a basic and direct, although individually incomplete, characterization of Nature as it is in itself.

While obviously closely related, indeed the theories easily shade into one another at their boundaries, neutral monism should be differentiated from dual aspect accounts by the denial of the claim that either the mental or the physical are truly fundamental. William James called the neutral foundation of reality ‘pure experience’ (an unfortunately mentalistic term obviously) and describes the theory in terms of a kind of potentiality or virtuality. He says ‘there is only one primal stuff or material in the world, a stuff of which everything is composed’ (James 1904, p. 478); this is ‘pure experience’ which ‘is only virtually or potentially either object or subject as yet. For the time being, it is plain, unqualified actuality, or existence, a simple that’ (James 1904, p. 485). Both the mental and the physical appear as ‘expressions of’ or ‘constructions from’ some truly fundamental non-mental and non-physical actuality (see Stubenberg 2008). For example, Russell describes the doctrine as involving the claim that ‘both mind and matter are composed of a neutral-stuff which, in isolation, is neither mental nor material’ (Russell 1921, p. 25). Neutral monists sometimes express this in terms of the viewpoint dependence of the mental and the physical: the very same thing, which is in itself neither mental nor physical at the fundamental level, can be regarded as the one of the other from the appropriate standpoint. If neutral monism is to be a theory distinct from dual aspect views, the idea that reality is, with respect to the mental and the physical, ontologically neutral—neither the one nor the other—must be its central claim. Thus, technically, neutral monism stands in theoretical opposition to panpsychism, although it can be articulated in such a way that the most fundamental physical entities (that is, entities which cannot be reduced to more basic *physical* entities or properties) are also fundamental mental entities (that is, entities whose mentalistic features cannot be reduced to other more fundamental *mental* entities or properties)¹.

¹My thanks to an anonymous referee for emphasizing the importance of the distinction between dual aspect theories and neutral monism.

In contrast to these traditional theories, I wish here to explore a panpsychist view I call Deferential Monadic Panpsychism, DMP (I think essentially this view is endorsed by Galen Strawson, who calls it ‘micropsychism’ in Strawson 2006). DMP agrees with all panpsychist views (and many other accounts of the place of consciousness in the physical world) that physical science does not and in a certain sense cannot deal with consciousness. I am inclined to think that the general reason for this inability is that science deals with the relational structure of the world and consciousness is an, or *the*, intrinsic feature of the world². Despite this essential lacuna in the purview of science, DMP agrees that physical science has found (or is close to finding) the basic outline of reality. Fundamentally, the world really is the way physics suggests it is although, needless to say, physics (and its interpretation) is always in flux and it remains highly controversial exactly what the purely physical picture of the world is.

DMP accepts that physics reveals the fundamental structure of the world. It seeks to, as it were, complete the picture by finding a place for consciousness within this structure. It does not seek to amend the picture of the world which science reveals but simply to append consciousness to it in the right sort of way. Hence the term ‘deferential’. ‘Monadic’ appears because DMP regards each fundamental entity in the world (whatever it might be) as the appropriate target for some equally fundamental mental aspect. If, for example, the world is best described physically in terms of a set of interacting fundamental particles then each such particle will be assigned some kind of primitive consciousness by DMP.

2 Problems for DMP

DMP faces two obvious challenges. The first is the complaint that its basic presumption, that consciousness is an intrinsic, non-relational feature of the world, is incorrect. This would open the way for understanding consciousness as a relational or structural feature itself, one presumably amenable to standard scientific explanation, even if we lack at the moment any glimmering of how such an explanation would proceed.

The second problem is that the monadic nature of DMP requires some account of the genesis of complex states of consciousness of the sort we are familiar with and which are presumably very different from the kinds of mentality associated with the fundamental physical entities. This is a familiar problem for panpsychism, sometimes called the ‘combination

²This diagnosis goes back a long way. It can be found in Leibniz’s monadic metaphysics (Leibniz 1714/1989a). Other now historical figures who advanced this view of the fundamental limits of science include Bertrand Russell (Russell 1927) and the physicist Arthur Eddington (Eddington 1928). Recently, a number of philosophers have addressed this issue within the context of the problem of consciousness (see Chalmers 1996, especially Ch. 4; Rosenberg 2004; Stoljar 2001; Strawson 2006).

problem’ (Seager 1995). It is a problem of emergence, but not the straightforward one of explaining how the novel can emerge from the pre-existent. The difficulty is that if we assign some sort of extremely primitive or simple consciousness to the elementary features of the physical world we then need to explain how the *complex* conscious states we are introspectively familiar with arise. This seems to be a form of emergence and then the problem is that if panpsychism itself requires a mechanism of emergence then why not take the theoretically more economical route of letting consciousness emerge directly from the physical basis itself rather than from a mental basis.

To examine these questions, we need to start with the notion of intrinsic properties. What, exactly, intrinsic properties are remains very unclear and there is a lively philosophical debate about the issue which has been going on for some time. Roughly speaking, intrinsic properties are the properties things have ‘in themselves’, properties that are had without the ‘metaphysical help’ of any other thing (Lewis 1983). It is important to note that this characterization does not imply that intrinsic properties are uncaused. The question is about the status of a property once an object has it, not about how it might have gotten the property in the first place; I tried to indicate this with the somewhat bizarre phrase ‘metaphysical help’.

The idea is often explicated in terms of duplication—what duplicates share are their intrinsic properties. A newly produced duplicate of me that sprang into existence would not be an uncle, would not own a Volvo, would not be a Canadian citizen, etc. But it would weigh just what I do, would have the same intrinsic angular momentum and would be having the same kinds of states of consciousness that I am having (or was having at the moment of duplication). But it is very hard to spell out what duplication is without the idea of intrinsic property.

Another intuitively appealing explication is that the intrinsic properties of an object are those that do not depend on the existence of anything else (see Langton and Lewis 2001). That is, consider the thought experiment of imagining what properties an object would retain were it to be the only thing that existed and had ever existed. If I were such a creature, I would no longer count as an uncle, but would have the same shape I do now and would be having experiences just like I am now (so long as my brain continued to function as it does in the actual world which the freedom of thought experiment allows).

Of course, some examples of the intrinsic/non-intrinsic distinction seem clear. The property of being an uncle is clearly not intrinsic—you can’t have it without the help of something else, namely a sibling and his or her offspring. Clear examples of intrinsic properties are harder to find. Mass (or rest mass) might be an intrinsic property and is indeed a traditional example. But this is in fact quite complicated. I’m told mass may arise from an interaction with a quantum field (the Higgs field) in which case things would not have mass ‘in

themselves’ (though presumably there would remain the ‘übermass’—or perhaps it would be better to say, ‘untermass’—of the Higgs particle itself). Another threat to the intrinsicity of mass is the venerable idea that it somehow arises from some kind of global mutual effect that generates inertia. Quantum spin is perhaps a better example of an absolutely intrinsic property, but one unfortunately remote from everyday experience.

However, as we shall see, there may be a good reason why it is hard to find *physical* examples of intrinsic properties. Is there another place to look?

One example seems to stand out in clarity: consciousness itself. States of consciousness seem to be excellent candidates for being intrinsic properties. Descartes noticed, to his distress, that no matter what might be happening apart from or ‘outside’ his conscious mind, his experience could be exactly as it is. The ‘what it is like’ of conscious experience seems to be something ‘in itself’ not logically or metaphysically (though of course in general causally) dependent on any other thing. A duplicate of me would lack all its relational, non-intrinsic properties and yet be conscious all the same. If I was the only thing that had ever existed I would still have conscious experiences (so long as I survived).

The notorious difficulties in providing plausible relational theories of phenomenal consciousness further suggest that consciousness is a domain of intrinsic properties³. For example, a functionally defined system of causally interacting states responsive to, say, electromagnetic radiation of various wavelengths, does not appear to be sufficient to constitute (though it might cause) conscious experience of colours, even if it successfully discriminates colours just as well as I do.

The venerable argument for panpsychism referred to above claims that science reveals only the relational properties of the physical. The most significant of these—perhaps all of these—are of course the causal powers of the physical. This observation undercuts any pretense that physical properties such as charge, spin or mass are truly intrinsic—they are but the nominal shadow of the relation we call causation. The argument further assumes that there must be an intrinsic ground to any existent. It may be that this intrinsic ground explains the relational properties, and hence the causal powers, of things, or maybe not (although the assumption of grounding is stronger if the intrinsics have this role). But given that nothing in science presents us with any candidate for the intrinsic nature of the physical, we can either opt for there being an unknowable sub-physical intrinsic (something along these lines is endorsed by Stoljar 2006 and a form of neutral monism can be developed in this way in which this sub-physical ‘layer’ plays the role of the neutral) or we can advance the idea

³Such problems are most in evidence for functionalism which explicitly attempts to give a purely relational or structural account of the mind and consciousness. For a classic statement of some of the key problems see Block (1980). In the realm of conscious thought, I think the same worry can be found at the heart of John Searle’s infamous ‘Chinese Room Thought Experiment’ (Searle 1980).

that the one known intrinsic feature—consciousness—ought to be regarded as the intrinsic ground of everything.

Versions of this argument are very interesting, but to many they are not very convincing. The option of accepting some unknowable but presumably non-mental physical intrinsic ground is, for many, much more attractive than attributing mentality to electrons and quarks. This view then faces the challenge of explaining the emergence of consciousness from the radically non-mental intrinsic feature just hypothesized—a problem that supporters of panpsychism see as very serious, especially insofar as no relational theory of phenomenal consciousness is forthcoming. As noted above however, it seems that the panpsychist also faces a similar problem of explaining how complex conscious states emerge from the primitive mental states ascribed to the fundamental entities of the world.

What is more, the assumption that everything requires an intrinsic grounding is itself highly debatable. The so-called ‘ontological structural realists’ have attempted to develop an entirely relational account of a scientific realist metaphysics; see Ladyman (2009) for a comprehensive, if highly sympathetic, review (for a more purely philosophical version of a relational metaphysics see Dipert 1997).

3 Maximality and Panpsychism

However, it is possible to approach this argument for panpsychism from another direction which requires only the assumption that consciousness is an intrinsic property of things. This novel approach is perhaps slightly more robust insofar as it does not depend on the hypothesis that all relational or structural features are determined by intrinsic features.

To begin, we must look a little more closely at the nature of intrinsicity, and consider a recent amendment or extension of the idea of intrinsicity proposed by Theodore Sider (see Sider 2001, 2003). Sider notes that a great many properties of ordinary things are what he calls ‘maximal’. That is, these properties are such that they do not characterize any of the proper parts (that is, parts which are not identical to the whole) of the object in question. A proper part of an object *O*, in the relevant sense, is not what is typically meant by a part, which is some functionally or systematic subsystem of *O*. To get the idea, imagine knocking off one molecule from *O*; the resultant thing, call it *O-*, is, when *O* is whole, a proper part of *O*. Sider’s example is the property of being a rock, which does *not* hold of a rock’s proper parts, despite the fact that a great many of these parts are such that, were they to exist on their own, they would properly be classified as rocks. Most rocks, after all, have the feature that if they are split in two, two new rocks appear. Yet in a sense these new rocks were already there, so to speak, ‘within’ the initial rock. These parts had ‘all it takes’ to be a rock save

for the disqualifying feature of being a part of a rock. Failure of maximality leads to a most peculiar proliferation of entities: without maximality, every rock is actually something like a vast assemblage of rocks, all appearing in the same place (or at least extensively overlapping places).

Although commonsense takes it that many everyday properties are maximal, the metaphysics of maximality is not altogether straightforward. Houses are usually maximal, but not always—it is evidently possible to build a house out of other houses where it would be agreed that the house-parts remain houses. Nonetheless, it seems pretty clear that a large range of properties, across a large range of their instances, are maximal.

Human beings (or the property of being human) seem to be maximal in Sider's sense. It seems plainly absurd to suppose that there are actually a vast number of people inhabiting (most of) the space which I inhabit.

Now, Sider has deployed the concept of maximality to argue against the idea that consciousness is an intrinsic property. More specifically, Sider deploys this argument against Trenton Merricks's (1998) use of the view that consciousness is an intrinsic property as part of a larger argument he develops against the general psycho-physical supervenience claim⁴, whose rough outline will be obvious given the foregoing. Sider's reasoning is that things which have maximal properties metaphysically depend upon other things: in particular, whether a certain continuous lump is a rock depends on, among other things, whether or not it is a part of another such lump. If not, it may be a rock, if it is, then the thing of which it is a part gets to be the rock, by maximality. If this reasoning is cogent, then maximal properties cannot be intrinsic.

The property of being a conscious subject appears to be maximal. I am a conscious being and I have myriads of proper parts which have everything it takes to be a conscious being except for the fact that they are parts of me. That part of me which is equal to me minus, say, 100,000 neurons scattered here and there in the brain, would be a conscious being if it were to exist apart from me. In fact, every day, parts such as these rise up to take my place, so to speak, as large numbers of neurons in my brain die of natural causes (Google says maybe 10,000 a day for an adult)—perhaps there is a somewhat novel problem of self-identity lurking here.

On the other hand, the claim that 'being conscious' is non-intrinsic seems to fly in the face of the strong intuitive, introspective and other evidence of 'in itself-ness' which indicate that conscious states are intrinsic states.

If we hold to the intrinsicity of consciousness we can evade Sider's argument in various

⁴Roughly speaking, this claim is simply that the mental supervenes on the local physical state of the subject, that is, every instance of a mental state depends on and is determined by an underlying physical state.

ways. One way is to deny that the subject of consciousness has parts so that maximality is irrelevant. Or, perhaps better, if the subject of consciousness is a ‘metaphysical simple’ (as Leibniz believed) then consciousness is a trivially maximal property (it is such that none of its proper parts are also conscious beings). This at least shows that maximality does not logically exclude intrinsicity.

However, the idea that the property of being conscious applies only to metaphysically simple entities appears to endorse a kind of dualism closely akin to that of Descartes and as such is not a very attractive option. It seems to be entirely at odds with the huge amounts of evidence we already possess that consciousness is intimately associated with large scale dynamical and structural features of the brain. It really does ‘look’ as if consciousness is the large scale product and property of the brain or nervous system. And, of course, this view seems to be at odds with DMP, which takes the complex physical structure of the world as revealed by physics as the ultimate guide as to where to locate both the fundamental and non-fundamental mental properties.

DMP can provide another way to save the intrinsicity of consciousness without denying the complexity of the systems which are conscious, nor the idea that consciousness is somehow an emergent feature of such complex entities. To see how, consider some non-maximal properties of non-simple systems. The property of having mass is not maximal. A rock has mass, and all of its proper parts also have mass. Does this mean that every rock actually weighs millions of times more than our metaphysically blinkered scales reveal? No, because mass is what might be called an *aggregative* property such that the weight of any object is the sum of the weights of its *elementary* or fundamental constituents.

The word ‘sum’ here is misleading—mass adds in peculiar ways. The mass of the hydrogen atom is less than sum of the masses of a proton and an electron because it takes energy to pull them apart and that energy ‘deficit’ turns up in the mass of the composite object. It would be better to say that mass *emerges* from the mass of the elementary constituents and the way they interact. This does not prevent the mass of the atom from being one of its intrinsic properties (modulo the caveats expressed about mass above). In the case of a rock, it has a certain mass and so too does each of its myriads of parts. The mass of each of these is the ‘emergent sum’ (for a rock, pretty much just the sum) of all its elementary constituents. But we don’t get to double count constituents in the aggregate of parts which metaphysicians discover in the rock. Each part contributes its mass to the whole rather than keeping it, as it were, for itself. We could also express this idea by saying that the mass of the whole is a function of the mass of all its disjoint parts given any carving into disjoint parts plus the dynamics of those disjoints.

We might thus call properties like mass, ‘aggregative’ properties; they contrast with

maximal properties. All the physically basic properties are aggregative it seems: charge, mass, spin, and the like.

The argument linking these considerations to panpsychism is now obvious. If consciousness is an aggregative property then we can evade Sider's argument; we can retain the intrinsicity of states of consciousness without giving up the idea that consciousness is an emergent feature of large scale physical objects such as the brain. The price is to add consciousness to the list of elementary features that 'aggregate' together as more complex systems are formed. If we have already bought into the idea of panpsychism, so that we already are willing to grant that all the fundamental physical entities also partake of consciousness, this price is very low, a theoretical bargain in an area where few are to be found.

The nature of the aggregation of consciousness may be very complex. It is part of the general problem of consciousness that we have no idea of how it might work. It certainly does not follow that all physical complexes will be conscious—that will depend on the details of how the elementary features of conscious come together to form more complex states of consciousness. A simple analogy is electric charge: although the basic constituents, quarks and electrons more or less, are all possessed of electric charge it does not follow that all physical objects have an electric charge, simply because positive and negative charge can cancel each other out.

According to both Merricks and Sider, one of the unpalatable consequences of regarding consciousness as intrinsic and an emergent feature of large and complex physical systems is the absurd proliferation of independent consciousnesses, one for each (large and complex enough) part of any conscious system. DMP appears to avoid this result. Aggregative properties contribute their properties to that of the whole but are not perniciously independent. It is interesting that such an intuitively bizarre theory as panpsychism provides a way to maintain two highly plausible or even commonplace views about the nature of consciousness.

4 Panpsychism and Emergence

Nonetheless, one might object that a problem still remains. If we follow the analogy with mass, although each part contributes its mass to the whole in a way that prevents 'double counting', it is still true that each proper part is a massive entity. So, although it may be that consciousness is an aggregative property, it apparently remains true that some of the proper parts of a complex conscious entity will themselves be conscious, albeit that they are contributing, so to speak, their consciousness to the overall conscious state of the whole entity.

It is open to the panpsychist simply to reject this as an objection. The idea that complex

consciousness is ‘built up’ out of simpler forms is hardly foreign to panpsychism. Leibniz explicitly endorsed the idea that every complex physical entity is associated with a hierarchical system of monads, as in the following passage from *The Principles of Nature and Grace*:

... each distinct simple substance or monad, which makes up the centre of a composite substance (an animal, for example) and is the principle of its unity, is surrounded by a *mass* composed of an infinity of other monads, which constitute the *body belonging to* this central monad, through whose affections the monad represents the things outside it... (1714/1989b, p. 207).

On the aggregative model outlined here, any complex consciousness will be supported by a host of conscious parts which form a hierarchy of complexity. In opposition to Leibniz’s view, in which all monads are metaphysical simples, the view proposed here accepts that it is literally true that consciousness emerges from these conscious parts in an aggregative way.

Nonetheless, the objection has some force. In particular, the notion that consciousness is an aggregative emergent has seemed to many extremely implausible, if not incoherent. The most famous proponent of this worry was William James who disparagingly named a version of panpsychism current at the time, and which is very similar to DMP, the ‘mind dust’ theory. His main criticism focused on the difficulties of understanding aggregative emergence in the case of consciousness:

Take a hundred [feelings], shuffle them and pack them as close together as you can (whatever that might mean); still each remains the same feeling it always was, shut in its own skin, windowless, ignorant of what the other feelings are and mean. There would be a hundred-and-first feeling there, if, when a group or series of such feeling were set up, a consciousness belonging to the group as such should emerge. And this 101st feeling would be a totally new fact; the 100 original feelings might, by a curious physical law, be a signal for its creation, when they came together; but they would have no substantial identity with it, nor it with them, and one could never deduce the one from the others, or (in any intelligible sense) say that they evolved it (James 1890, p. 160).

There are two problems that can be discerned in James’s complaint. The first is the obvious worry that one cannot make any sense of the idea that a set of distinct conscious states could aggregate into a single, more complex conscious state. Or, at least, there is no way that such aggregation could occur that was intelligible (unlike the case of the aggregation of mass, or any other familiar aggregative property). We might call this the intelligibility problem.

The second and rather more obscure problem I will call the problem of the irrelevance of the totality. What I mean can again be illustrated with the mass analogy. If we suppose

that some massive body is an aggregate of massive parts, then if we know the positions and masses of the parts we know all there is to know about the mass of the totality. There is nothing novel, with respect to mass and its effects, possessed by the total system.

Newton famously showed that, in the case of a spherical massive body, the gravitational effects of that body are the same as if all its mass were concentrated at a point at its centre. This is highly convenient, since we can then ignore the detailed disposition of the parts. But of course this feature of spherical bodies is a mathematical consequence of the distribution of all the massive parts of the body. The mass of the totality and all of its effects are completely and fully determined, and exhausted by the masses of the parts and their joint effects.

In general, with aggregative properties, the property of the whole is nothing but a reflection of the property as distributed through all the parts of the whole. The ‘totality’ is thus in a sense irrelevant.

Both of these problems loom against DMP. DMP simply *assumes* that consciousness aggregates into more complex forms; it provides no model of how this is supposed to work. As James’s quote illustrates, consciousness cannot aggregate merely by the co-occurrence of some set of conscious states nominally taken to be ‘parts’ of a putative total and more complex conscious state. And while aggregation is not in general a function of mere co-occurrence (not even in the case of mass), it is not unfair to ask how precisely the aggregation of conscious states is supposed to occur.

DMP also faces the objection that even if some robust form of aggregation is assumed, there will be no distinctive role for the aggregate. Instead, all the features of the aggregate state will be determined and exhausted by the communal action of that state’s parts.

There is a traditional answer to these worries, which is to endorse a stronger form of emergentism than aggregative wholes exemplify. Such emergence is sometimes called ontological emergence, sometimes it is known as radical emergence. Such a view is by no means unheard of in the history of philosophy. A version of it was widely championed in the 19th and early 20th centuries by thinkers such as John Stuart Mill (1843/1963), Samuel Alexander (1920), Conwy Lloyd Morgan (1923) and C. D. Broad (1925; Broad in the end did not endorse emergentism but clearly regarded it as a coherent doctrine with much to be said in its favour). For our purposes, the distinctive claim of ontological emergence is that the properties of a whole are not predictable, even in principle, from an exhaustive knowledge of the properties of the parts and their inter-relations⁵. Instead, in principle predictability

⁵Note that one form of predictability is simulatability. It is an important discovery that some systems’ behaviour cannot be predicted in any more efficient way than simulation. One example is John Conway’s ‘game of life’ (see Gardner 1970). It is provable that this special cellular automaton is Turing complete—it can be configured to emulate any Turing machine. If there was an efficient way to predict the behaviour of cellular automata in general then the Halting Problem would be solvable (see Bedau 1997 for an interesting discussion of this property of cellular automata and its relation to the issue of emergence). Hence, the only

would require knowledge of additional ‘laws of emergence’ or ‘laws of complexity’ which are free additions to the laws which govern the fundamental elements of the system and cannot be reduced to such laws (and the initial state of the fundamental elements).

If we add ontological emergence to DMP we can of course reply to William James’s complaint by endorsing his final remark: ‘the 100 original feelings might, by a curious physical law, be a signal for its [the 101st feeling] creation, when they came together; but they would have no substantial identity with it, nor it with them, and one could never deduce the one from the others, or (in any intelligible sense) say that they evolved it’.

While there are those who deny the intelligibility of ontological emergence (see Nagel 1979; Strawson 2006) I think it is hard to deny the bare logical coherence of the notion. I do not think that thinkers such as Morgan or Broad were contradicting themselves when they propounded the doctrine of radical emergence, even though there were doubtless wrong about its actual prevalence in nature⁶.

It is then perhaps possible to evade James’s objection by appeal to ontological emergence. But the cost is high. We lose the aggregative model and replace it with a totally mysterious claim of radical ontological origination. Worse still, the motivation leading to panpsychism in the first place appears to evaporate. If ontological emergence is possible, then why not simply accept that consciousness emerges in this—radical—way from a purely physical, entirely non-experiential, fundamental substrate. One of the core advantages of panpsychism, viewed as nothing more than a philosophical doctrine, was that it promised to avoid radical emergence by postulating that consciousness itself was a fundamental property of the world. If we must even so appeal to radical emergence then any distinctive advantage of panpsychism seems to have disappeared. And then the intuitively attractive idea that the fundamental physical features of the world are completely devoid of mental attributes, coupled with ontological emergence, appears to suffice to at least provide an outline of a solution to the metaphysical problem of consciousness.

5 Combinatorial Infusion

In the face of this dialectic, the panpsychist should search for some other model for the emergence of consciousness, which can retain the advantages of the aggregative model while

way to ‘predict’ cellular automata in general is via simulation. However, simulation is a perfectly legitimate form of prediction. The important point here is that the simulation requires only knowledge of the properties of the parts (e.g. the initial state of the cells a game of life ‘world’) and the rules governing the interactions of the parts. Successful simulation does not require any reference to the complexity of the patterns generated, or any other feature not in principle reducible to and exhausted by the basic interactions of the fundamental entities of the system.

⁶For an excellent exposition of emergentism and a defense of its coherence see McLaughlin (1992).

avoiding James's objections. I believe there is such a model, but one that brings together the rather odd bedfellows of aggregation and metaphysical simplicity.

The final idea I want to explore here has three characteristics:

1. The mental character of the combined or aggregative mental state stems from the mental characteristics of the constituents,
2. The combined or aggregative mental state is a novel state which in some way 'absorbs' or supersedes the mental states of the constituents, but
3. There is no radical or ontological emergence of the aggregative mental state; rather there is an intelligible relation which holds between the mental components and the resulting aggregative state.

This would be a new kind of combination which deserves its own name: *combinatorial infusion*.

Such an idea is not entirely new. Some have seen in certain of the characteristically bizarre phenomena of quantum mechanics a way to model a new kind of 'combination' which could illuminate the emergence of the mental. Entanglement for example presents a case where a composite object has holistic properties not reducible to the separable properties of the system's constituents. Some have seen this as evidence for ontological emergence (see Silberstein and McGeever 1999). If this is correct, then the quantum approach will not meet my desired condition (3).

For my part, I don't see any ontological emergence in quantum mechanics. The entangled state is a predictable result of the basic laws of quantum mechanical systems and their interactions. In fact, of course, it *was* predicted, by Schrödinger, who introduced the term 'entanglement', right at the birth of quantum mechanics. There is no hint of radical or ontological emergence here. Nonetheless, the possibility of the creation of an entangled state most certainly represents a new kind of combination principle which might cast light on mental composition as well⁷.

To assess whether such models can succeed, we need a better understanding of how combinatorial infusion might work. I suggest that the correct (or at least a promising) way to look at conscious states is as what I will call 'large simples'. By this I mean something which is metaphysically simple: utterly lacking parts but which possesses some kind of 'extensiveness'. I do not necessarily mean here *spatial* extensiveness, although this can serve as a good entry to the concept.

The idea of a spatially extended metaphysically simple entity goes back at least to the 17th century with the idea that space itself is such an entity. The English philosopher

⁷I explored this in Seager (1995); it is more extensively examined in Hameroff and Powell (2009) and in Jonathan Powell's doctoral dissertation at Reading University.

and mystic Henry More defended this claim, characterizing space as ‘one, simple, immobile, eternal, complete, independent. . .’ and so on for over twenty attributes (see Koyre 1957, chs. 5-6, for a discussion of More’s views; see also Holden 2004). Recently, the consistency of the notion of an extended spatial simple is given a philosophical defense in McDaniel (2007). As McDaniel notes, the physicist Brian Greene implicitly appeals to the notion when speculating about the ultimate nature of *strings*, writing that (on one possible view) ‘even though strings have spatial extent, the question of their composition is without any content’ (Greene 1999, p. 141).

The simplicity of space is and remains controversial, but at least it provides a reasonably clear model of something possessing extension but which lacks parts. If we grant, for the sake of the example, that space is an existing entity understood in the way Newton understood what he called absolute space (so not reducible to relations between the entities existing, as we say, *in* space), then it is clear that space does not possess parts in any but a purely notional sense. Space contains regions, but these are not parts of space which are ‘removable’ or whose interactions ground the distinctive properties of space itself.

What the model of space does not provide is any understanding of the emergence of a large simple. For that we would need an entity that depends on its constituents and inherits its properties in an intelligible way from its constituents, but which in some sense absorbs or supersedes them. Perhaps we can regard the parts as infusing their properties into the whole and by so doing effacing themselves. Curiously, there is a physical model which provides a beautiful exemplification of something very like this idea: the classical black hole.

When a classical black hole forms it does so from a system of constituents bound together by gravitational collapse. These constituents could be any kind or form of matter whatsoever. Whatever the constituents might be, their particular features are erased by the formation of the black hole. The so-called ‘no hair’ theorems suggest that a classical black hole can be completely and exhaustively characterized by exactly three physical properties: mass, angular momentum and electric charge. The black hole is a kind of ‘elementary particle’, partless yet extended, and exhaustively characterized by its three essential properties (of course, spin and charge can take on a value of zero)⁸.

It seems that in actual fact black holes will not be classical. For example, the information loss paradox suggests that any black hole will somehow ‘encode’ the details of its formation which fact would presumably be in principle detectable (however there is some recent weak evidence that the ‘no hair’ theorems are respected by nature; see Valtonen *et al.* 2010). The presence of a singularity at the heart of the black hole might also be regarded

⁸I do not mean to suggest that black holes are akin to what modern physics regards as elementary particles. That is a thoroughly quantum conception.

as physically impossible, indicating that some successor theory is required to fully deal with this phenomenon.

Such worries are irrelevant here however. The point I wish to make is simply that we are already in possession of a rigorous model of a kind of combination that has the properties demanded by combinatorial infusion. It is therefore possible to deploy this sort of model when thinking about how complex states of consciousness could emerge from simpler ones. If this is a process of combinatorial infusion, then James's objections can be circumvented and we have an interesting viewpoint on DMP.

With regard to James's objections, note that the creation of large simples by combinatorial infusion is determined entirely by the properties and interactions of the constituents. Here it is crucial to distinguish between two forms of emergence: diachronic and synchronic. combinatorial infusion is a diachronic process in which the constituents and their characteristics are 'absorbed' into the new whole. Their identity is lost. Thus combinatorial infusion evades James's worry by providing a way for a new state to arise from the states of the combined constituents. Furthermore, since the constituents themselves are, so to speak, effaced during the creation of the new whole, there is no problem about the efficacy of the new state. Since it is what remains, it has got to be the causal source (and/or sink) of future effects. There is no radical emergence here, but there is nonetheless the creation of a new entity.

Well, this is just a philosophical model of a possible form of panpsychism. Many problems remain for its articulation and defense. One obvious issue stems from the deferential nature of DMP. DMP requires that the mental realm shadow the physical, so we expect to find physical correlates of mental processes, including any case of combinatorial infusion. We don't really have at the present time any firm grasp on the neural correlates of consciousness, although this topic is the focus of considerable research at the present time. What I think it is important to emphasize is that it is *not* a requirement of DMP that a mental instance of combinatorial infusion be accompanied by a physical instance of combinatorial infusion.

I would say there is little evidence that the brain supports any processes that could count as combinatorial infusion at the physical level. It is conceivable that the brain's functioning in some way is essentially dependent on some distinctive quantum process which would be akin to combinatorial infusion. Defenders of quantum theories of consciousness typically do claim that such processes occur in the brain, but this is highly controversial. I note two possible kinds of arguments against combinatorial infusion in the brain. The first depends on the idea that combinatorial infusion in the physical realm will be essentially linked to quantum coherence/entanglement. Given that assumption, it is natural to argue that the hot and noisy physical environment of the working brain is antithetical to the creation and maintenance of entanglement. One well know argument to this effect is Tegmark (2000). A second, not

unrelated, line of argument depends on the assumption that the brain is fundamentally a network based system whose function is to represent both the sensible world and possible active responses to it. The argument is then that the requirement for the maintenance of stable representation given that the network is constructed from intrinsically noisy components will lead the brain to construct a discrete, classical network resistant to lower level disturbances (for a worked out version of this argument see Eliasmith 2000). Arguably, this is at odds with the brain engaging in combinatorial infusion at the physical level.

Thus it is important to recognize the mental combinatorial infusion could be accompanied by less metaphysically peculiar physical activity. It could be, to take one prominent example, that widespread neural synchrony is the relevant correlate to mental combinatorial infusion.

Finally, on a yet more speculative note, it seems to me that the familiar but deeply puzzling feature of the mind we call the unity of consciousness might be illuminated by the idea of combinatorial infusion. Our overall state of consciousness involves a host of particular aspects, some associated with all the sense modalities which we can experience simultaneously, some associated with states of thought, emotion and bodily sensation. These can be conceptually divided and considered as separate, but it would be a gross distortion to claim that our consciousness was a simple aggregate of these phenomenologically distinguishable elements of our mental state. In some way, our minds integrate all that we experience into a single overall experience. Clearly, this phenomenon has affinities with the idea of combinatorial infusion and this presents some hope that this idea will illuminate both the nature of panpsychism and even the structure of consciousness itself.

William Seager
University of Toronto

References

- Alexander, Samuel (1920). *Space, Time and Deity*. Macmillan & Co.
- Bedau, Marc (1997). ‘Weak emergence’. In James Tomberlin (ed.), *Philosophical Perspectives 11: Mind, Causation, and World*, pp. 375–399. Oxford: Blackwell.
- Block, Ned (1980). ‘Troubles with functionalism’. In N. Block (ed.), *Readings in Philosophy of Psychology, Vol. 1*, pp. 268–305. Cambridge, MA: Harvard University Press.
- Broad, C. D. (1925). *Mind and Its Place in Nature*. London: Routledge and Kegan Paul.
- Chalmers, David (1996). *The Conscious Mind: In Search of a Fundamental Theory*. Oxford: Oxford University Press.
- Dipert, Randall (1997). ‘The mathematical structure of the world: The world as a graph’. *Journal of Philosophy*, 94 (7): pp. 329–58.
- Eddington, Arthur (1928). *The Nature of the Physical World*. New York: Macmillan & Co.
- Eliasmith, Christopher (2000). ‘Is the brain analog or digital?’ *Cognitive Science Quarterly*, 1 (2): pp. 147–70.
- Gardner, Martin (1970). ‘The fantastic combinations of John Conway’s new solitaire game “Life”’. *Scientific American*, 223: pp. 120–123.
- Greene, Brian (1999). *The Elegant Universe: Superstrings, Hidden Dimensions, and the Quest for the Ultimate Theory*. New York: W. W. Norton and Co.
- Hameroff, Stuart and Jonathan Powell (2009). ‘The conscious connection: A psycho-physical bridge between brain and pan-experiential quantum geometry’. In David Skrbina (ed.), *Mind That Abides: Panpsychism in the New Millennium*, pp. 109–27. Amsterdam: Benjamins.
- Holden, Thomas (2004). *The Architecture of Matter: Galileo to Kant*. Oxford: Oxford University Press.
- James, William (1890). *The Principles of Psychology*, vol. 1. New York: Henry Holt and Co.
- James, William (1904). ‘Does “consciousness” exist?’ *Journal of Philosophy, Psychology, and Scientific Methods*, 1: pp. 477–91.

- Koyre, Alexander (1957). *From the Closed World to the Infinite Universe*. Baltimore: Johns Hopkins University Press.
- Ladyman, James (2009). ‘Structural realism’. In Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. Summer 2009 ed. URL <http://plato.stanford.edu/archives/sum2009/entries/structural-realism/>.
- Langton, Rae and David Lewis (2001). ‘Defining “intrinsic”’. *Philosophy and Phenomenological Research*, 58: pp. 333–45.
- Leibniz, Gottfried Wilhelm (1714/1989a). ‘Monadology’. In Roger Ariew and Daniel Garber (eds.), *G. W. Leibniz: Philosophical Essays*, pp. 214–24. Indianapolis: Hackett.
- Leibniz, Gottfried Wilhelm (1714/1989b). ‘Principles of nature and grace’. In Roger Ariew and Daniel Garber (eds.), *G. W. Leibniz: Philosophical Essays*, pp. 206–213. Indianapolis: Hackett.
- Lewis, David (1983). ‘Extrinsic properties’. *Synthese*, 44: pp. 197–200.
- McDaniel, Chris (2007). ‘Extended simples’. *Philosophical Studies*, 133 (1): pp. 131–41.
- McLaughlin, Brian (1992). ‘The rise and fall of British emergentism’. In A. Beckermann, H. Flohr and J. Kim (eds.), *Emergence or Reduction*, pp. 49–93. Berlin: De Gruyter.
- Merricks, Trenton (1998). ‘Against the doctrine of microphysical supervenience’. *Mind*, 107: pp. 59–71.
- Mill, John Stuart (1843/1963). *A System of Logic*, vol. 7-8 of *The Collected Works of John Stuart Mill*. Toronto: University of Toronto Press.
- Morgan, Conwy Lloyd (1923). *Emergent Evolution*. London: Williams and Norgate.
- Nagel, Thomas (1979). ‘Panpsychism’. In *Mortal Questions*, pp. 181–95. Cambridge: Cambridge University Press. (Reprinted in D. Clarke *Panpsychism: Past and Recent Selected Readings*, Albany: SUNY Press, 2004.).
- Rosenberg, Gregg (2004). *A Place For Consciousness: Probing the Deep Structure of the Natural World*. Oxford: Oxford University Press.
- Russell, Bertrand (1921). *The Analysis of Mind*. London: George Allen & Unwin.
- Russell, Bertrand (1927). *The Analysis of Matter*. London: K. Paul, Trench, Trubner.

- Seager, William (1995). ‘Consciousness, information and panpsychism’. *Journal of Consciousness Studies*, 2 (3): pp. 272–88. (Reprinted in J. Shear (ed.) *Explaining Consciousness*, Cambridge, MA: MIT Press, 1997.).
- Searle, John (1980). ‘Minds, brains and programs’. *Behavioral and Brain Sciences*, 3 (3): pp. 417–57.
- Sider, Theodore (2001). ‘Maximality and intrinsic properties.’ *Philosophy and Phenomenological Research*, 63: pp. 357–64.
- Sider, Theodore (2003). ‘Maximality and microphysical supervenience’. *Philosophy and Phenomenological Research*, 66: pp. 139–49.
- Silberstein, Michael and J. McGeever (1999). ‘The search for ontological emergence’. *Philosophical Quarterly*, 49: pp. 182–200.
- Stoljar, Daniel (2001). ‘Two conceptions of the physical’. *Philosophy and Phenomenological Research*, 62: pp. 253–81.
- Stoljar, Daniel (2006). *Ignorance and Imagination*. Oxford: Oxford University Press.
- Strawson, Galen (2006). ‘Realistic monism: Why physicalism entails panpsychism’. *Journal of Consciousness Studies*, 13 (10-11): pp. 3–31. (Reprinted in A. Freeman (ed.) *Consciousness and Its Place in Nature*, Exeter: Imprint Academic, 2006.).
- Stubenberg, Leopold (2008). ‘Neutral monism’. In Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. The Metaphysics Research Lab, fall 2008 ed. URL <http://plato.stanford.edu/archives/fall2008/entries/neutral-monism>.
- Tegmark, Max (2000). ‘Importance of quantum decoherence in brain processes’. *Physical Review E*, 61 (4): pp. 4194–4206.
- Valtonen, M. J. *et al.* (2010). ‘Measuring the spin of the primary black hole in OJ287’. *The Astrophysical Journal*, 709 (2): pp. 725–32.