CHAPTER 6

Can the panpsychist get around the combination problem?

Philip Goff

Consciousness, understood as the property of *being a thing such that there is something that it is like to be that thing*, is not an invention of philosophers. We ordinarily suppose that there is something it is like to be a normal functioning human being or animal (at least an animal above a certain level of complexity). But everyday thought restricts attributions of consciousness to organisms. We do not ordinarily believe that there is something that it is like to be the little bits that make up our brains.

Many philosophers find this commonsense position problematic. It holds that organisms are made up of things which entirely lack experience, and yet somehow, at some level of complexity, experience magically emerges. Why do the interactions of several billion non-conscious things result in the emergence of conscious experience? Why don't we just get a complicated, non-conscious system? Viewed from certain angles, the emergence of consciousness from non-consciousness can seem like nothing short of a miracle.

Of course, there are various ways in which philosophers try to *dissolve*, rather than solve, this philosophical difficulty. Why should the emergence of consciousness from non-consciousness be any more problematic than the emergence of life from non-life, or the emergence of liquid from molecules that are not themselves wet? But many remain unconvinced by such analogies. It seems prima facie that I can conceive of my zombie twin: an atom for atom duplicate of me that lacks conscious experience. In contrast, it is far from clear what it would be to conceive of an atom for atom duplicate of me which is not alive, or an atom for atom duplicate of Lake Geneva which is not wet. For these reasons the emergence of consciousness seems philosophically problematic in a way in which the emergence of life or water is not.

One explanation of the emergence of consciousness, powerfully advocated in recent times by Galen Strawson (2006), is panpsychism. Panpsychism is the view that the ultimate constituents of the physical world are conscious; that there is something that it is like to be the ultimate constituents of the physical world. If the littlest bits that make me up are themselves conscious, then arguably we no longer have the mystery of how little non-conscious things come together to constitute something with conscious

experience. It seems like we don't need to explain where consciousness came from if it was there all along.

1. The combination problem

There is a significant difficulty facing the attempt to explain the consciousness of organisms in terms of the consciousness of their ultimate constituents, a problem which is often referred to as 'the combination problem.' The problem is that subjects of experience, i.e. things which have consciousness (things such that there is something that it is like to be them), just don't seem to be the kind of things that can 'sum together' to make other subjects of experience. The problem was vividly articulated by William James:

> Take a hundred of them [feelings], shuffle them and pack them as close together as you can (whatever that may 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 feelings 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 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. (1890/1950: 160)

Small objects with certain shapes, e.g. lego bricks, can constitute a larger object with a different shape, e.g. a lego tower. But it is difficult to see how, say, seven subjects of experience each of which have a visual experience as of seeing one of the colors of the spectrum (and are such that between them they instantiate visual experiences of all seven colors of the spectrum), could constitute a distinct subject of experience having a visual experience as of seeing white.

The most tempting response to the combination problem is to claim that we are simply ignorant of the way in which experiences sum, and that this is no good reason to think that they don't. However, I think there is good reason to think that at least some of the motivation for the combination problem is rooted, not in ignorance, but in a priori knowledge concerning the nature of subjects of experience. Specifically, I take the following to be a principle we can reasonably take ourselves to know a priori:

> **No Summing of Subjects** (NSS): The existence of a group of subjects of experience, $S_1 \ldots S_N$, instantiating certain phenomenal characters, never necessitates the existence of a subject of experience T, such that what it is like to be T is different from what it is like to be any of $S_1 \ldots S_N$.

We can understand this principle by contrasting it with the case of spatial objects. Take the case of seven lego cubes placed on top of each other to make a rectangular

tower. The mere existence of those bricks, each having a specific shape and location, necessitates the existence of the tower having the shape and location it has. We could not coherently conceive of the seven bricks being piled on top of one another in the way that they are in the absence of the tower. In contrast, it is eminently possible to conceive of our seven subjects of experience experiencing the colors of the spectrum, existing in the absence of a subject of experience having an experience of white. The existence of a group of spatial objects, $O_1 \ldots O_{N,}$ with certain shapes and locations, can necessitate the existence of a spatial object with a shape and location different to the shape and location of each of $O_1 \ldots O_N$. It does not seem that subjects of experience, merely in virtue of their existence, can stand in this kind of necessary relation.

How could this principle be objected to? NSS seems to clearly hold for all subjects of experience of which we can conceive. To take another example, ten subjects all feeling slightly pained do not necessitate the existence of a very pained subject. But perhaps it might be claimed that we have no reason to think NSS holds for *all* subjects of experience, including those of which we have no conception. Without doubt, there are many kinds of subjects of experience which we cannot conceive of. As has been pointed out before, we are not able to conceive of what it is like to be a bat.

But any qualitative difference between two subjects of experience, qua subjects of experience (i.e. considered simply as things with consciousness), is merely a matter of a difference in the phenomenal characters that characterize their experience, a difference in what it is like to be those subjects. NSS holds for any group of subjects of experience we can conceive of, regardless of what it is like to be them. The principle seems to hold independently of what it is like to the subjects it concerns. NSS seems to be a conceptual truth concerning the determinable property of *being a subject of experience*, rather than any specific determinates of it.

I do not know how to demonstratively prove that there is not a possible set of subjects of experience which constitute a counterexample to NSS: i.e. a group of subjects of experience which, by their mere existence, necessitate the existence of some distinct subject of experience. But reflection shows NSS be true with regards to all the many varied subjects of experience we are able to conceive of, in a way that doesn't seem dependent on the specific phenomenal characters they instantiate. I take it, therefore, that NSS is a principle we can reasonably take ourselves to know.

2. Making sense of experiences summing

What implications does NSS have for the summing of experiences? It follows from NSS that a certain set of subjects of experience cannot sum *merely in virtue of their existing (and instantiating the specific phenomenal characters they instantiate)*. But it does not imply that a certain set of subjects of experience cannot exist and be involved in some *state of affairs* which necessitates the existence of some distinct subject of experience. There is nothing in the principle which rules out the possibility of there being some

state of affairs of a certain set of subjects of experience *being related in some specific way,* which necessitates the existence of some distinct subject of experience.

To put it another way, NSS implies that there is no state of affairs of the form <subject of experience S_1 exists with phenomenal character x, and subject of experience S_2 exists with phenomenal character y> which necessitates <subject of experience S_3 exists with phenomenal character z>. But it does not imply that there is not some state of affairs of the form <subject of experience S_1 with phenomenal character x bears relationship *R* to subject of experience S_2 with phenomenal character y> which necessitates <subject of experience S_3 exists with phenomenal character z>. Such a sense of experiences summing is not ruled out by NSS.

Neither introspection nor perception affords us experience of any such relation, call it 'phenomenal bonding,' which bonds subjects of experience together to constitute other subjects of experience. Indeed, in line with what James says above, I don't think we have experience of *any* natural relation between subjects of experience qua subjects of experience. In so far as we can think of subjects of experience as spatially located (perhaps in people's heads), so we can conceive of spatial relations between them. But spatial relations are not phenomenal bonding relations. Just as the mere existence of a certain group of subjects of experience does not necessitate the existence of some distinct subject of experience, so the existence of a certain group of subjects of experience standing in certain spatial relations to each other cannot necessitate the existence of some distinct subject of experience.

But it is hardly surprising that we can have neither introspective nor perceptive experience of relations between subjects of experiences qua subjects of experience. We are unable to perceive relations between subjects of experience (qua subjects of experience) through the senses, simply because we are unable to perceive subjects of experience (qua subjects of experience) through the senses. If you examine my brain, you will not be able to see it instantiating phenomenal properties. I have epistemic access to only one subject of experience qua subject of experience, i.e. the subject of my own experience accessed via introspection. It follows from the fact that we can introspect only one subject of experience, that we cannot introspect how subjects of experience qua subjects of experience are related, for to introspect how subjects of experience qua subjects of experience are related we would have to be able to introspect more than one subject of experience. Given that we can experience subjects of experience qua subjects of experience only via introspection, and we have introspective access only to one subject of experience, it follows that we cannot experience subjects of experience qua subjects of experience as related.[1]

Locke, Berkeley and Hume held that experience provides all our ideas, which in turn provide meanings for our words. Because of this, they would take the fact that we cannot experience phenomenal bonding, either through the senses or through intro-

1. This explanation of why we are unable to experience relations between subjects of experience, is reminiscent of McGinn's (1989) explanation of why we are constitutively incapable of understanding how consciousness emerges from the physical.

spection, to imply that the term 'phenomenal bonding' is literally meaningless. But this strict meaning empiricism was based on a very crude philosophy of language. Nowadays philosophers do not take, say, our lack of experience of a four-dimensional object to imply that four-dimensionalism is an unintelligible view.

In the same way, it seems that we can intelligibly suppose that subjects of experience, qua subjects of experience, may bear relations to each other, even though we have no experience of these relations. Assuming subjects of experience do bear relations to each other, I can find no principled reason against supposing that there is some way of being related in which a group of subjects of experience can stand to each other in virtue of which they constitute a state of affairs which necessitates the existence of another subject of experience. In this way, contrary to views I have expressed in earlier work (Goff 2006), I believe that the panpsychist can make good sense of subjects of experience summing, and hence can get round the combination problem.[2]

3. The problem with this solution

Although I think the above solution is a coherent way for the panpsychist to avoid to combination problem, I think it leaves the panpsychist with a difficulty. She ends up ontologically committing not only to the conscious experience of particles, but also to the phenomenal bonding relation which unites the mini-subjects of experience into 'larger' subjects of experience. Whilst we may have a clear idea of what it would be for particles to be subjects of experience, there is a clear sense in which our understanding of the phenomenal bonding relation, and subsequently of the state of affairs of *a group of subjects being related in the phenomenal bonding way*, is incomplete.

We can define phenomenal bonding as 'that relation such that when subjects of experience bear it to each other the existence of a different subject of experience is necessitated,' and form an understanding of the phenomenal bonding relation in these terms. However, perhaps because we lack any experience of such a relation, we are unable to understand the state of affairs of *a group of subjects being related in the phenomenal bonding way* independently of what that state of affairs (if it exists) necessitates. Contrast with the case of spatial relations. We understand what it is for seven lego bricks to be on top of each other even if we are not thinking of them in terms of the tower they form. We cannot understand the state of affairs of *a group of subjects being related in the phenomenal bonding way* without understanding it in terms of the subject of experience which (if it exists) it necessitates.

---

2. The picture of subjects summing I have outlined here might be more similar to the spatial case than I have seemed to suggest. If relationalism about space is true, then spatial objects having the locations they do is a matter of their relational properties. Thus, in both the case of subjects summing and the case of spatial objects summing, relational properties are an essential ingredient of the summing.

But in the same way we might define a slightly different relation, call it 'physical-to-phenomenal bonding' as 'that relation such that when non-conscious physical particles stand in it to each other the existence of a subject of experience is necessitated.' We understand this relation as much and as little as we understand the phenomenal bonding relation. We understand 'physical-to-phenomenal bonding' in the sense that we can define it in terms of what the state of affairs of *a group of non-conscious physical particles being related in the physical-to-phenomenal bonding relation* necessitates. But we do not fully understand it in the sense that we cannot think about that state of affairs other than in terms of what (if it exists) it necessitates.

Just as our lack of full understanding of phenomenal bonding is no reason to deny the possibility of such a relation, so it seems to me our lack of full understanding of physical-to-phenomenal bonding is no reason to deny the possibility of this relation. We do not fully understand (in the sense I have specified above) how non-conscious particles could bond in some special way to form subjects of experience, but nor do we fully understand how subjects of experience could bond together to form different subjects of experience. It seems to me then that the panpsychist has the difficulty of answering the following question: why should we suppose that our conscious experience is the result of phenomenal bonding relating conscious particles rather than of physical-to-phenomenal bonding relating non-conscious particles?

Perhaps the panpsychist could claim that it is a lot more natural to suppose that conscious things emerge from other conscious things, rather than from non-conscious things. I think there may be some force to this point. But there are clear advantages to the opposing view too. On the panpsychist view, we are ontologically committed to *both* a relationship we don't fully understand *and* the conscious experience of particles. On the non-panpsychist alternative under consideration, we only have to believe in a relationship we don't fully understand. We save ourselves from a very demanding, and arguably counterintuitive, ontological commitment.

But doesn't panpsychism dissolve the mystery of the emergence of consciousness? If consciousness is there all along, then surely we don't have to worry about where it came from. The problem is that, by including a relation we don't fully understand, i.e. the phenomenal bonding relation, in her hypothesis, the panpsychist has admitted that the emergence of consciousness (or more precisely the emergence of *human and animal consciousness*, the consciousness of ultimate particles was of course there all along) is something we don't fully understand. The non-panpsychist theorist who postulates the physical-to-phenomenal bonding relation to explain consciousness must confess to a certain degree of ignorance as to how exactly non-conscious particles sum together to make subjects of experience. But similarly the panpsychist who commits to the phenomenal bonding relation must confess to a certain degree of ignorance as to how exactly little subjects of experience sum together to make human and animal consciousness, which is after all the kind of consciousness we have a pre-theoretical need to explain. It is not obvious that the former kind of ignorance is any greater than the latter.

What about zombies? If I can conceive of a physical duplicate of mine which lacks conscious experience, doesn't this entail that any merely physical duplicate of me is going to lack conscious experience? The theorist who postulates the physical-to-phenomenal bonding relation to explain my conscious experience can agree with this.[3] A purely physical duplicate of me would lack conscious experience, but a physical duplicate such that some of its fundamental constituents are related in the physical-to-phenomenal bonding relation cannot, by the very definition of the physical-to-phenomenal bonding relation, lack conscious experience.

4. Conclusion

I don't think that the combination problem signals the end of panpsychism. There is at least one coherent way in which panpsychist can get around this problem. But getting round the combination problem does, I believe, involve the panpsychist in some degree of mysterianism. The panpsychist, because she must confess to not fully understanding the phenomenal bonding relation, ends up with a view whereby the emergence of human and animal consciousness is something of a mystery. This results in a problem with the motivation for the view. Once the panpsychist introduces a degree of mysterianism into her view, she then has the challenge of showing why her view should be preferred to non-panpsychist mysterian alternatives.

There is no reason to think that the panpsychist cannot show a mysterian version of her view to be theoretically superior to non-mysterian alternatives. As we continue to theorize about the correlations between physical states (in the sense of states which physical science reveals to us) and conscious states, it may well be that the best theory to explain these correlations will predict that consciousness is more widely distributed in the world than ordinary thought supposes. Nevertheless, the panpsychist is obliged to make this case. Introducing an element of mysterianism into the view, which I believe to be inevitable if the panpsychist is to get around the combination problem, gets rid of any *obvious* advantage panpsychist accounts of consciousness might have been thought to have over non-panpsychist rivals. The panpsychist can get round the combination problem, but in doing so she is left with a lot of work to do in motivating her view.

3. I am understanding 'physical duplicate' here such that x is a physical duplicate of y iff physical science could not discern a difference between x and y.

CHAPTER 7

Universal correlates of consciousness

Stephen Deiss*

Science is finding it difficult to explain how and why a physical system like a brain can have conscious experience. We know a lot about the neuroanatomy and neurophysiology of color vision for example. However, we cannot explain why the sensation of color happens above and beyond the raw senseless physical process neuroscience can measure. In the following I offer a analysis of key concepts that continue to mislead us in efforts to explain consciousness. These concepts include sensation, perception, consciousness, ego, self, causality, mechanism and laws of nature. This analysis not only explains how the problem arises, but also presents a new rationale for why consciousness is a universal panpsychic process.

I propose a very intuitive explanation of human consciousness which I define as a process of *interpreting sensations*. Interpretation is a matter of finding *meaning*. The meaning resides in the *expectations and predictions* we attach to qualitative sensory contrasts using *associative memory*. These memory-based inferences are further sensations we have that typically involve visual or auditory imagery, in conjunction with our own thoughts and inferences – the 'voices in our heads.' From initial sensations we derive many more by association to complete our experience with our expectations of what lies beyond them. All these sensations fit together to comprise an integrated sensory interpretation which is our perception of reality.

Our reality, even that which we think we directly perceive, is a kind of informal but vivid *theory* derived from qualitative sensory contrasts. If we step away from the assumption of blind mechanisms in nature, it quickly becomes obvious that there is nothing radically different about brains in this regard, other than their advantage of a large associative memory. All systems that we study in science involve processes that are analogous to what goes on in our brains. Instead of senseless passive systems ruled by causal mechanisms and laws, we can view systems as active sensing agents that change and behave by constraint satisfaction. If all systems have sensations which they interpret as constrained by their state, which is a kind of *memory*, then *all* systems become sensation interpreters, and they have at least a modicum of consciousness.

---

* Thanks to Jerry Josties, Paul Lincolnhol, K. L. Mercer, and Natika Newton for helpful discussions and encouragement.