

Philip Goff

# *The Phenomenal Bonding Solution to the Combination Problem*

## **I – The high probability of panpsychism**

Panpsychism, the view that fundamental physical entities are conscious, is a highly probable theory of the natural world. Appreciation of this requires little more than getting our epistemic situation right.

Physics tells us much about the dispositions of fundamental natural entities, but leaves us completely in the dark about their categorical nature. In knowing that an electron has a certain amount of mass, we know how it is disposed to resist acceleration and attract other things with mass. In knowing that an electron has negative charge, we know that it is disposed to repel other things with negative charge and attract things with positive charge. Everything natural science has to tell us about electrons concerns their behaviour; we learn nothing about what an electron *is* independently of what it *does*. The only thing we know for certain about the categorical nature of natural entities is that at least some of them, for example me and you, are conscious.

We now have a theoretical choice. We can either suppose that the categorical nature of fundamental particles, such as electrons and quarks, is constituted of some form of consciousness, or we can suppose that they have some entirely unknown categorical nature. On the former supposition, the nature of macroscopic things is continuous with the nature of microscopic things. The latter supposition in contrast adds complexity, discontinuity and mystery. The theoretical imperative to form as simple and unified a view as is consistent with the data leads us quite straightforwardly in the direction of panpsychism.

The main objection one comes across to panpsychism is that it is ‘crazy’ and ‘just obviously wrong’. It is thought to be highly counterintuitive to suppose that there is something that it is like to be an electron, and this is taken to be a very strong reason to doubt the truth of the panpsychism. But the view that time slows down at high speeds, that particles have determinate position only when measured, that the Earth goes round the sun, or that our ancestors were apes were (indeed still are) also highly counterintuitive, to many ‘just obviously wrong’. And yet the counter-commonsensicality of these views gives us little or no reason to think them false. It is hard to see why the fact that most Westerners living today happen to be pre-theoretically inclined to think panpsychism false constitutes a reason to think that it is false.

Probably the willingness of contemporary philosophers to accept special relativity, natural selection and quantum mechanics, despite their strangeness from the point of view of pre-theoretical common sense, is a reflection of their respect for the scientific method. We are prepared to modify our view of the world if we take there to be good scientific reason to do so. But in the absence of hard experimental proof, philosophers are reluctant to attribute consciousness to electrons.

However, whilst there is no observational data that supports panpsychism, there is a hard datum which counts in its favour: the existence of consciousness. The reality of consciousness is more evident to us than any empirical postulation. The existence of consciousness does not entail the truth of panpsychism, but it counts in its favour in the sense that panpsychism is the most unified picture of the world which is consistent both with its existence and with our observational knowledge. Compare the datum that the speed of light is measured to be the same in all frames of reference. This datum does not entail the truth of special relativity, but it counts in its favour in the sense that special relativity is the most elegant picture of the world consistent with it. The evident existence of consciousness supports the truth of panpsychism in much the same way that measurements of light support special relativity.

Whilst in the mind-set of thinking that physics is on its way to giving a complete picture of the fundamental nature of reality, panpsychism seems improbable, as physics does not attribute conscious states to fundamental particles. But once we realise that physics leaves us completely in the dark about the categorical nature of the entities it talks about, and indeed that the only thing we know for certain about the categorical nature of the universe is that some of it is taken up with consciousness, things look very different. All we get from physics is this big black and white abstract structure which we metaphysicians must somehow colour in with real categorical nature. Assuming the falsity of substance dualism, we know how to colour in one bit of it: the brains of organisms are coloured in with consciousness. How to colour in the rest? The most elegant, simple, sensible option is to colour in the rest of the world with the same pen.<sup>1</sup>

## II – Introducing the combination problem

Despite its obvious attractions, panpsychism suffers from a serious problem: the so called ‘combination problem.’ It is natural to suppose that my mind, the subject of my consciousness, is not a microscopic entity. Assuming the falsity of emergentism and substance dualism, my mind is a macroscopic entity which derives its nature from the microscopic entities which compose it, ultimately from the entities that fundamental physics talks about, which the panpsychist takes to be conscious subjects. Somehow little subjects, such as electrons and quarks, come together to produce big conscious subjects, such as human brains. The combination problem is given by the fact that it’s hard to make sense of this kind of combination.<sup>2</sup>

The inspiration for the combination problem is the following much quoted passage from William James:

Take a hundred of them [feelings], shuffle them and pack them as close

---

<sup>1</sup> Of course there are a number of options I am implicitly ruling out here, such as physicalism about consciousness in conjunction with dispositional essentialism about physical entities. I argue against these views elsewhere (Goff Forthcoming 2011a/Forthcoming a/MS). Here I would like to concentrate on exploring the details of panpsychism.

<sup>2</sup> Chalmers (This volume) spells out a whole series of combination problem, grouped around the three themes of subjective character, qualitative character and structural character. I think the central problem pertaining to subjective character, that is, the difficulty I explore here of making sense of subjects summing, is recognised to be the central ‘combination problem’, and the one that threatens to undermine the coherence of the view from the off. The problems pertaining to qualitative and structural character strike me more as challenges that need to be responded to in working out the details of a specific panpsychist view (as Coleman does in this volume), rather than potentially yielding knock down arguments against panpsychism.

together as you can (whatever that may mean); still each remains the same feelings 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 101<sup>st</sup> 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, not it with them, and one could never deduce the one from the others, nor (in any intelligible sense) say that they *evolved* it.<sup>3</sup>

Many philosophers, under the influence of this passage, claim to find some special conceptual difficulty in the idea of *feelings* or *subjects* combining. In fact, closer examination of the text surrounding this passage reveals that James' resistance to the summing of mental entities is grounded in a general resistance to the idea of *anything* combining:

...no possible number of entities (call them as you like, whether forces, material particles, or mental elements) can sum *themselves* together. Each remains, in the sum, what it always was; and the sum itself exists only *for a bystander* who happens to overlook the units and to apprehend the sum as such; or else it exists in the shape of some other *effect* on an entity external to the sum itself. Let it not be objected that H<sub>2</sub> and O combine of themselves into 'water,' and thenceforward exhibit new properties. They do not. The 'water' is just the old atoms in the new position, H-O-H; the 'new properties' are just their combined *effects*, when in this position, upon external media, such as our sense-organs and the various reagents on which water may exert its properties and be known.<sup>4</sup>

Without much argument, James takes it to be evident that entities 'combine' only in the sense that their acting in concert gives rise to some distinctive perception in observers. Combinations only exist in the eye of the beholder. The 'combination problem' according to James goes as follows:

1. There are no combinations.
2. Therefore, there are no mental combinations.

However, the more suggestive elements of the former paragraph have made more of an impression on contemporary philosophers than the argument of the latter paragraph. There does seem to be some deep difficulty making sense of distinct subjects combining to produce a greater subject; a difficulty we don't seem to face making sense of distinct objects in space combining.

In earlier work, I said the following:

Small objects with certain shapes, e.g. lego bricks, can constitute a larger

---

<sup>3</sup> James 1890: 1. 160.

<sup>4</sup> James 1890: 1. 158-9.

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 has a visual experience as of seeing one of the colours of the spectrum, could constitute a distinct subject of experience having a visual experience as of seeing white....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 colours 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, O1....ON, 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 O1....ON. It does not seem that subjects of experience, merely in virtue of their existence, can stand in this kind of necessary relation.<sup>5</sup>

In the above passage I subscribe to the following epistemic principle:

*Conceptual Isolation of Subjects (CIS)* – For any group of subjects, instantiating certain conscious states, it is conceivable that just those subjects with those conscious states exist in the absence of any further subject.

from which I draw the following metaphysical conclusion:

*Metaphysical Isolation of Subjects (MIS)* – For any group of subjects, instantiating certain conscious states, it is possible that just those subjects with those states exist in the absence of any further subject.

In contrast, I rejected the following epistemic principle:

*Conceptual Isolation of Lego (CIL)* – For any group of lego bricks, at certain locations, it is conceivable that just those objects at those locations exist in the absence of any further lego object.

and hence rejected the following metaphysical conclusion:

*Metaphysical Isolation of Lego (MIL)* – For any group of lego bricks, at certain locations, it is possible that just those objects exist at those locations in the absence of any further lego object.

I now think that in this earlier work I didn't get the disanalogy between the phenomenal case and the

---

<sup>5</sup> Goff 2009a: 130-1. I further defend CIS (although not under that name) using panpsychist zombies in Goff 2009b. Chalmers discuss this argument in his article in this volume.

lego case quite right. Whether or not James' nihilism about composite objects is plausible, it does not seem to be *incoherent*. One is not contradicting oneself when one claims that the lego bricks do not really combine when arranged 'tower-wise' by a child, or that they combine only in the weak sense that they produce a distinct visual impression in the child playing with them. Perhaps such a view is counter to common sense, but it seems coherent.<sup>6</sup> Thus, it now seems to me that *both Conceptual Isolation of Subjects* and *Conceptual Isolation of Spatial Objects* are true.

The crucial difference between lego combination and subject combination arises when we try to move from conceivability to possibility. It would be helpful at this point to take a digression into the relationship between conceivability and possibility.

### III – Conceivability and possibility

Since Kripke moving from conceivability to possibility has not been so straightforward. In *Naming and Necessity* Kripke argued that there are *a posteriori necessities*, propositions which are necessarily true, but can only be known to be true through observation or experiment.<sup>7</sup> The proposition water is H<sub>2</sub>O is one example: we cannot know sitting in the armchair that water is H<sub>2</sub>O, and yet there is no possible world in which water exists with some other chemical composition (in what follows I will refer to concepts and propositions with underlined words). An *a posteriori* necessity is conceivably false, in the sense that we cannot know it to be true *a priori*, and yet is not possibly false. Thus, if there are *a posteriori* necessities, then a traditional principle of philosophy is false:

*Conceivability Principle (CP)* – If P is conceivably true, then P is possibly true.

The work of David Chalmers is perhaps best seen as an attempt to defend a somewhat traditional conception of philosophy in the light of Kripke's work. Chalmers accepts Kripke's examples of *a posteriori* identities, and hence rejects CP. However, he defends a more subtle principle linking conceivability to possibility:

*Two-Dimensional Conceivability Principle (2D-CP)* – If P is conceivably true (upon ideal reflection), then there is a possible world W, such that P is true at W considered as actual.<sup>8</sup>

Twin Earth is the genuine possible world which is indiscernible from our world except that the colourless odourless stuff that fills oceans and lakes and falls from the sky is XYZ rather than H<sub>2</sub>O. Chalmers agrees with Kripke that in our normal way of thinking about a possible world, considering it as *counterfactual*, i.e. as a *way things might have been*, the proposition water is XYZ is false at Twin Earth. Water is the *actual* colourless, odourless stuff that fills oceans and lakes and falls from the sky; the actual stuff in oceans and lakes is H<sub>2</sub>O; and hence water is H<sub>2</sub>O in all possible worlds (even ones where oceans and lakes and filled with XYZ).

However, Chalmers thinks that there is another way to think about a possible world: as *actual*, that is to say, as a *way things might actually turn out*. When Twin Earth is considered this way, the proposition water is XYZ comes out true at Twin Earth. Water is the actual stuff in oceans and lakes; if

---

<sup>6</sup> Theodore Sider (2013) has recently defended the non-existence of composite objects.

<sup>7</sup> Kripke 1972.

<sup>8</sup> Chalmers 1996/2002/2009.

the actual stuff in oceans and lakes turns out to be XYZ, then water is XYZ. Hence, although necessarily false, there is a genuine possibility corresponding to water is XYZ in the sense articulated by 2D-CP: when that possibility is considered as actual, water is XYZ come out true.<sup>9</sup> Chalmers holds that every conceivably true proposition corresponds in this way to some genuine possibility.

I have argued in other work that 2D-CP rests on highly contentious semantic assumptions, for which Chalmers has not provided adequate defence.<sup>10</sup> However, I think that a much simpler and less contentious principle linking conceivability to possibility is both defensible and consistent with Kripke's cases: *conceivability entails possibility when you completely understand what you're conceiving of*. In the current context, we can partially explicate this principle in terms of the following:

*Transparency Conceivability Principle (TCP)* – For any proposition P, if (A) P involves only quantifiers and predicates expressing transparent concepts, (B) P is conceivably true (upon ideal reflection), then P is possibly true.

Understanding TCP requires understanding the distinction between *transparent* and *opaque* concepts, which I will try to express in the following.

For something to be water is for it to be constituted of H<sub>2</sub>O molecules, but our ordinary concept of water does not reveal this. Our ordinary concept water is *opaque*, in the sense that it is not a priori (for someone possessing the concept, and in virtue of possessing the concept) what it is for something to be water. Due to the opacity of the concept water, one is able to conceive of scenarios involving water *without fully understanding what is being conceived of*. It is this lack of transparent understanding of what is being conceived of which blocks the move from conceivability to possibility.

In contrast to water, the concept million-sided object is *transparent*: it is a priori (for someone possessing the concept, and in virtue of possessing the concept) what it is for something to have a million sides. Hence when one conceives of a million-sided object one completely understands, or is in principle able to reason one's way to a complete understanding of, the situation being conceived of. In conceptions involving only quantifiers and predicates expressing transparent concepts, e.g. a conception of there being a million-sided object, it is a priori for the conceiver what it is for the state of affairs they are conceiving of to obtain. In such cases, I believe that we can move from the conceivability (upon ideal reflection) of the states of affairs so conceived, to its genuine possibility.

Suppose this were not the case. Suppose the existence of a million-sided object were just 'brutely impossible', in the sense that, (A) a complete understanding of what it is for there to be a million-sided object, coupled with ideal reflection upon that state of affairs, cannot reveal any incoherence in there being a million sided object, and yet (B) it is impossible for there to be a million-sided object. Further imagine an omnipotent and perfectly rational being tries to create a million-sided object. She examines the notion of such a thing from all angles, and finds no bar to its existence. And yet when she tries to create such a thing, she finds herself unable. She is unable not because of any limit of power on her part, but simply because it is impossible – in such a way that nobody could ever make

---

<sup>9</sup> Ultimately Chalmers grounds all this in what it is rational to suppose. On the assumption that the stuff in oceans and lakes is XYZ, it is rational to suppose that water is XYZ.

<sup>10</sup> Goff & Papineau Forthcoming, Goff 2011/MS.

intelligible sense of its being impossible – for there to exist a million-sided object.

When metaphysical possibility is so radically divorced from conceptual coherence, as vividly expressed in the last paragraph, I start to lose my grip on what metaphysical possibility is supposed to be. Moreover, a radical separation between what is conceivable and what is possible has the potential to make our knowledge of possibility problematic. If David Lewis is right that possibilities are concrete worlds spatio-temporally distinct from our own, then it is hard to account for our knowledge of what is or is not possible.<sup>11</sup> But if metaphysical possibility is just a special kind of conceivability, then knowing what is possible is just a matter of knowing what is conceivable in the requisite sense.

Furthermore, panpsychism is very often grounded in an opposition to physicalism, which is in turn very often grounded in some kind of conceivability argument. Hence, the panpsychist has strong motivational reason to accept some kind of principle linking conceivability to possibility. In defending panpsychism in the first section of this paper, I was implicitly assuming the falsity of certain alternatives, such as physicalism about consciousness in conjunction with dispositional essentialism about physical entities. In a more extended defence of panpsychism, such alternatives would need to be ruled out.<sup>12</sup>

I take it, then, that the panpsychist has good reason to accept TCP, and I will assume it in what follows.<sup>13</sup>

#### **IV – Back to the combination problem**

The predicate ‘is a lego brick’ does not express a transparent concept. A lego brick is essentially composed of a certain kind of plastic. That plastic is essentially constituted of a certain arrangement of atoms of certain kinds; which atomic kinds and arrangements constitute the essence of plastic cannot be known a priori. Therefore, although I can conceive of lego bricks being arranged ‘tower-wise’ in the absence of a lego tower, in doing so I do not fully understand the nature of the state of affairs I am conceiving of. The real nature of a lego brick is not apparent to me in my thoughts about lego. On account of this, I cannot use TCP to move from the genuine conceivability of lego bricks arranged tower-wise without a lego tower (*Conceptual Isolation of Lego*) to the genuine possibility of lego bricks arranged tower-wise without a lego tower (*Metaphysical Isolation of Lego*).

In contrast, direct phenomenal concepts are plausibly transparent. One forms a direct phenomenal concept when one attends to a conscious state one is currently in, and thinks about it in terms of how it feels.<sup>14</sup> When I attend to a pain, it is directly revealed to me what it is for something to feel that way. When I attend to my experience of orange, it is directly revealed to me what it is for something to instantiate an experience of that kind. Many physicalists deny the thesis that direct phenomenal concepts are transparent (*Phenomenal Transparency*), holding that phenomenal

---

<sup>11</sup> Lewis 1986.

<sup>12</sup> I try to rule out these alternatives in Goff Forthcoming a/Forthcoming b/MS.

<sup>13</sup> I provide further defence of a less restricted version of TCP – able to deal with propositions involving singular terms – in Goff & Papineau Forthcoming and Goff MS.

<sup>14</sup> The notion of a direct phenomenal concept is from Chalmers 2003.

concepts are entirely opaque concepts which turn out to denote brain states.<sup>15</sup> However, on the assumption that phenomenal concepts are opaque, it is hard to make sense of our a priori knowledge concerning our conscious states, e.g. that phenomenal red is similar to phenomenal orange, or that pain has moral significance.<sup>16</sup> Moreover, as discussed above, most panpsychists are motivated by an opposition to physicalism, commonly grounded in conceivability arguments. But if phenomenal concepts are opaque, then the kind of moves from consciousness-involving conceivability to possibility involved in the anti-physicalist arguments, facilitated by principles such as 2D-CP or TCP, are blocked.<sup>17</sup> I will take it, then, that a panpsychist has good reason to accept *phenomenal transparency*.

Given *phenomenal transparency*, when I conceive of a group of subjects existing in the absence of a further subject, at least in so far as I have direct phenomenal concepts of the conscious states involved in my conception, I completely understand what I am conceiving of. It follows from *phenomenal transparency* that the conception we reach when we reflect on the famous James passage involves (or at least can involve) only quantifiers and predicates expressing transparent concepts.

Thus, we reach the heart of the combination problem: given the lack of opacity in the relevant conception, we are licenced to infer from *Conceptual Isolation of Subjects* to *Metaphysical Isolation of Subjects*. We may press the difficulty with the following argument:

The No Summing of Subjects Argument (the heart of the combination problem)

1. *Conceptual Isolation of Subjects* – For any group of subjects, instantiating certain conscious states, it is conceivable that just those subjects with those conscious states exist in the absence of any further subject.
2. *Transparency Conceivability Principle* – For any proposition P, if (A) P involves only quantifiers and predicates expressing transparent concepts, (B) P is conceivably true upon ideal reflection, then P is metaphysically possibly true.
3. *Phenomenal transparency* – Phenomenal concepts are transparent.
4. *Metaphysical Isolation of Subjects (MIS)* – For any group of subjects, instantiating certain conscious states, it is possible that just those subjects with those states exist in the absence of any further subject (from 1, 2 and 3).
5. For any group of subjects, those subjects with those conscious states cannot account for the existence of a further subject (from 4).
6. Therefore, panpsychism is false (from 5).

---

<sup>15</sup> McLaughlin 2001 and Papineau 2006 explicitly deny phenomenal transparency. Most contemporary physicalists adopt some kind of semantic externalist account of the reference of phenomenal concepts, which seems to entail the falsity of phenomenal transparency: if reference to a given conscious state is determined by facts external to our understanding, then it's hard to see how we could have a priori to the essence of consciousness.

<sup>16</sup> I argue at length against physicalist views of phenomenal concepts, on the basis of these kinds of consideration, in Goff Forthcoming a/MS.

<sup>17</sup> In relation to 2D-CP, moves from conceivability involves a concept having identical primary and secondary intensions (i.e. having the same referent at a world whether it is conceived of as actual or counterfactual). Given that the primary intension can be evaluated a priori, and the secondary intension reflects the essence of the referent, this is more or less equivalent to the concept's being transparent.

## V – Phenomenal Bonding – A response to the *No Summing of Subjects Argument*

I believe the *No Summing of Subjects Argument* to be sound right up to 4, which is to say that there is a sound argument for *Metaphysical Isolation of Subjects* (MIS). But let us think more carefully about what implications MIS has for the summing of experiences. It follows from MIS that certain 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 accounts for 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 certain subjects of experience *being related in some specific way* which necessitates the existence of some distinct subject of experience.

To put it another way, MIS implies that there is no state of affairs of the form <subject of experience S1 exists with phenomenal character x, and subject of experience S2 exists with phenomenal character y> which necessitates <subject of experience S3 exists with phenomenal character z>. But it does not imply that there is not some state of affairs of the form <subject of experience S1 with phenomenal character x bears relationship R to subject of experience S2 with phenomenal character y> which necessitates <subject of experience S3 exists with phenomenal character z>. Such a sense of experiences summing is not ruled out by MIS.<sup>18</sup>

I don't think we have a transparent conception of such a relation, call it 'phenomenal bonding,' which bonds subjects together to produce other subjects of experience. If we did have such a conception, then the solution to the combination problem would be obvious. Indeed, the problem would never have occurred to us.

However, it is not surprising that we lack a transparent grasp of the phenomenal bonding relation – if such a thing there be – given the nature of our epistemic situation. Our most basic empirical science, physics, yields understanding only of the world's mathematico-causal structure, and the phenomenal bonding relation is not a mathematico-causal relation: conceiving of subjects standing in mathematico-causal relations does not remove their conceptual isolation, and hence does not remove their metaphysical isolation. Apart from its mathematico-causal structure, arguably the only feature of the world we transparently understand is consciousness. And consciousness is a monadic property. Our unfortunate epistemic situation does not afford us a transparent understanding of the (non-mathematico-causal) relations which conscious things bear to each other.

Here is another way of pressing the point. Human beings are able to 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. And it follows from the fact that we can introspect only one subject of experience that we cannot introspect how subjects of

---

<sup>18</sup> Goff 2009a suggests this solution to the combination problem.

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.

Just because we are unable to form a transparent conception of the phenomenal bonding relation does not mean we cannot form a conception of it. We can think of it as ‘the relation such that when subjects stand in it they produce a further subject’ and we can suppose that there is such a thing. We may even be able to identify it with some relation we can observe in the world, or some relation that features in physics. None of the relations that appear in perception or in physics are conceived of *as* phenomenal bonding relations. In the same way, the brain does not appear from the outside *as* a subject of experience, and the properties of physics or neuroscience are not conceived of in those sciences *as* phenomenal qualities. But just as the panpsychist might identify charge with a form of consciousness, so the proponent of phenomenal bonding might identify some empirically known relation as the phenomenal bonding relation.

I can see no principled reason to think the phenomenal bonding relation is not a real relation that certain subjects bear to each other, and I think therefore we have a way of making sense of subjects summing, and hence a way of making sense of panpsychism. The theoretical attractions of panpsychism give us good reason to take this route to saving the view, and hence to believe that there is a phenomenal bonding relation.

The flaw of the *No Summing of Subjects Argument*, then, is in the final stages. We can only move from the *Metaphysical Isolation of Subjects* to

5. For any group of subjects, instantiating certain conscious states, those subjects with those conscious states cannot account for the existence of a further subject.

if we construe ‘account for the existence of’ as meaning ‘wholly explain the existence of just by their mere existence and intrinsic conscious nature.’ And we can only infer the falsity of panpsychism from this construal of 5 if we construe panpsychism as requiring that subjects ‘wholly explain just by their mere existence and intrinsic conscious nature’ the existence and nature of other subjects. But there is no reason to construe panpsychism in this way. The nature of organisms and car engines are accounted for in terms of their parts, but those parts constitute the organism/engine only when related in the right way. The same is surely true of the explicability of subjects in terms of other subjects.

There is a sense in which embracing this solution to the combination problem leads to a kind of mysterianism. In so far as we don’t have a transparent grasp of the phenomenal bonding relation, there is clear sense in which we don’t understand how subjects combine. This lack of knowledge is frustrating, which may cause us to yearn for a different theory. However, as hardnosed metaphysicians, we should be asking not what view we’d like to be true, but what view is most likely to be true. And the great elegance with which panpsychism unifies the existence of consciousness with the facts of observation renders it highly likely to be true. Probably the success of science, and the collective forgetting that all natural science gives us is structure, has given us rather high

expectations for what we can achieve concerning our understanding of the universe. But once we get our epistemic situation right in the way I described earlier we are more humble in our aspirations.

Indeed, independently of the desire to make sense of panpsychism, we have good reason to think that there are relations we lack a transparent conception of. We know that things stand in spatio-temporal relations, and yet physics provides us only with a mathematical conception of such relations. Mathematical descriptions abstract from the concrete nature of things. We abstract from the concrete reality of phenomenology when we describe it merely in terms of its mathematical structure, e.g. we abstract from the real concrete nature of five subjects when we describe them as merely 'five things'. Intuitively, there must be some real nature to spatiotemporal relations which lies behind the mathematical understanding of those relations we get from physics. If that intuition is correct, then it seems we are completely in the dark about the real nature of spatiotemporal relationships.

## **VI – Intelligent emergentist panpsychism versus constitutive panpsychism**

The kind of consciousness we want a theory of consciousness to explain is the kind of consciousness we pre-theoretically associate with humans and other animals. I call this kind of consciousness 'o-consciousness'. I have tried in the above to make sense of the idea that o-conscious subjects are intelligibly produced by the consciousness and the phenomenal bonding of micro-level entities. However, I do not take intelligible production to be the same thing as constitution. I take cases of constitution to be a subset of cases of intelligible production, the defining characteristic of constitution being that constituted states of affairs are nothing over and above the states of affairs which constitute them.

Consider the following examples of intelligible production which do *not* involve constitution. We can move a priori from *God's willing that there be light* to *there being light*, and the latter state of affairs obtains because the former obtains; in this sense the state of affairs of there being light is intelligibly produced by the state of affairs of God willing that there be light. Nonetheless, there being light is not constituted of God's willing that there be light; in willing that there be light God creates new being. Similarly, if the conjunction of dispositional essentialism and determinism is true we can move a priori from facts about the past to facts about the future, but it does not follow that facts about the future are constituted of facts about the past.

Thus, we can distinguish two forms of panpsychism, both of which involve o-phenomenal facts being intelligibly produced by the micro-phenomenal facts:

*Constitutive panpsychism* – O-phenomenal facts are constituted by, and hence are nothing over above, the micro-phenomenal facts.

*Intelligent emergentist panpsychism* – O-phenomenal facts are intelligibly produced by, but are something over and above, the micro-level facts.

David Chalmers has argued that constitutive forms of panpsychism enjoy an important advantage over non-constitutive forms, in so far as they are able to reconcile the causal efficacy of o-

consciousness with the causal closure of the micro-physical.<sup>19</sup> In ‘Against constitutive Russellian monism’, I argue against the constitutive view.<sup>20</sup> It is important in this context to note that the case I made for the panpsychism in the first section of this paper, grounded in considerations of theoretical elegance, is neutral between constitutive and emergentist forms of panpsychism.

## VII – When do subjects sum?

Peter van Inwagen encouraged metaphysicians to ask ‘the special composition question’: under what conditions do objects combine to form a further object?<sup>21</sup> Panpsychists can ask ‘the special *phenomenal* composition question’: Under what conditions do subjects combine to produce a further subject? For believers in phenomenal bonding, this will be equivalent to the question: which subjects bear the phenomenal bonding relation to each other?

One popular answer to the special composition question is ‘always’.<sup>22</sup> For proponents of *unrestricted composition* any objects, no matter how disparate and seemingly unrelated, compose an object: your nose, my teeth and the planet Venus form an object. Other philosophers defend some form of *restricted composition*: some sets of objects are such that their members compose an object, other sets of objects are such that their members do not compose an object.<sup>23</sup> We can distinguish analogous answers with respect to the special phenomenal composition question. According to unrestricted phenomenal composition, for any group of subjects, say, the particles forming your nose, my teeth and the planet Venus, those subjects are related by the phenomenal bonding relation and hence produce a further subject. Obviously, some form of restricted phenomenal composition, according to which some but not all subjects are such that they bear the phenomenal bonding relation to each other, will be more in keeping with pre-theoretical common sense.

One’s approach to answering the special phenomenal composition question will vary radically depending on whether one is a constitutive or an emergentist panpsychism. If emergentism is true, and if emergent entities have distinctive causal powers, then there will be an empirically discernible distinction between those systems which have and those systems which lack such emergent causal powers. The behaviour of the latter but not the former will be predictable on the basis of the behaviour of the system’s parts. Hunting the phenomenal bonding relation, for the emergentist, will be a matter of looking for an empirically distinguished relation which relates the parts of systems with emergent causal powers, but doesn’t relate the parts of systems which lack emergent causal powers. It is likely, then, that the emergentist will support some form of restricted phenomenal composition, looking to the empirical facts for the boundary between systems which are mere aggregates of micro-subjects and systems which are conscious in their own right.

---

<sup>19</sup> Chalmers This volume/Forthcoming.

<sup>20</sup> Goff Forthcoming b. The argument I present in this paper might be seen as a way of pressing a version of the combination problem against constitutive Russellian monism. Interpreted in this way, as a challenge to constitutive panpsychism, I take the combination problem to be insoluble. Interpreted as a challenge to the more general view that higher-level phenomenal facts can be intelligibly determined by more basic phenomenal facts, I take the combination problem to be soluble.

<sup>21</sup> van Inwagen 1990.

<sup>22</sup> Lewis 1986, Sider 2001.

<sup>23</sup> Van Inwagen 1990, Merricks 2001.

For the constitutive panpsychist, in contrast, higher-level subjects are nothing over and above micro-subjects, and their causal powers are entirely derived from the causal powers of their parts. It is less clear in this case that there will be an empirically discernible distinction between systems which have and systems which lack consciousness in their own right. We cannot directly observe either the presence or absence of consciousness in a system. Sciences such as psycho-physics are reliant on the anti-sceptical assumption that what a subject says about her consciousness is a reliable guide to the facts about her consciousness. Such anti-sceptical assumptions give us reason to believe in the presence of consciousness in certain material systems, e.g. systems that talk to us, but they give us no reason to believe in the *absence* of consciousness in any material systems. To avoid sceptical scenarios we must attribute consciousness to organic systems, but we are not required to refrain from attributing consciousness to non-organic systems.

What can we turn to if observation can't help us to answer the special phenomenal composition question? It is probably the case that most people researching into consciousness draw their answer to the special phenomenal composition question from pre-theoretical common sense, taking it to be something like the following:

*Commonsense answer* – Particles form a conscious subject when and only when they form organisms (or a subset of organisms, or the brains/central nervous systems of organisms; I will ignore these subtleties for the sake of simplicity).

Unfortunately, there is a difficulty, arising from considerations of vagueness, with accepting this answer. In what following I will outline this difficulty.

The boundary between the organic and the non-organic is vague. There are what we can call 'organic borderline cases' – cases where there is no fact of the matter as to whether or not we have a human organism – at the beginning and end of an organism's existence. In any particular case there is no utterly precise point in time, after which we have a zygote, and before which we had only sperm and egg. Similarly, in each particular case there is no utterly precise point in time after which we have a corpse and before which we have a living body. Given our macroscopic concerns, this vague boundary is barely discernible. But if we were looking at a complete description of the fundamental particles composing a human organism during, and slightly before and after, its existence, there would be no precise arrangement of particles which constituted the beginning and end of the organism's existence; there would be borderline cases. There would presumably be similar borderline cases at the coming to be/passing away of brains, or coming to be/passing away of brains of the level of complexity required for consciousness.

If the commonsense answer to the special phenomenal composition question is correct, it follows that there are 'phenomenal borderline cases', cases where there is no fact of the matter whether or not we have a conscious subject. If the existence of an organism is necessary and sufficient for the existence of a conscious subject, and if it's sometimes vague whether or not we have an organism, it follows that it's sometimes vague whether or not we have a conscious subject.

Why should this be thought to be a problem? Why should a vague boundary between the conscious and the non-conscious be any more problematic than a vague boundary between the tall and the

non-tall, or the heap and the non-heap? The trouble arises if we want to give a semantic treatment of vagueness. Making sense of vagueness as a semantic phenomenon requires associating each vague predicate with a spectrum of sharpenings, and (given a couple of fairly plausible assumptions) it is hard to make sense of the predicate 'is conscious' being associated with a spectrum of sharpenings.<sup>24</sup>

Let us take this more slowly. According to semantic theories of vagueness, vagueness is the result of *semantic indecision*: for any vague predicate there are multiple 'sharpenings' of the predicate, such that the meaning of the predicate does not settle on any of these sharpenings. Consider the vague predicate 'is tall'. We could stipulate, somewhat arbitrarily, that anything that is exactly 6 feet or taller counts as 'tall', and anything shorter is not tall. This is one 'sharpening' of the predicate 'is tall', that is, one way of making the predicate precise. Alternately, we could stipulate that anything that is exactly six feet and one inch or taller counts as tall, and anything shorter is not tall. This is an alternative sharpening of 'is tall', that is, an alternative possible way of making the predicate precise. The predicate 'is tall' is thus associated with a *spectrum of sharpenings*: a range of possible ways of making the predicate precise.<sup>25</sup>

Semantic theories of vagueness tell us that a vague predicate is vague because no one has bothered to single out one of its sharpenings as the unique meaning of the predicate. To put it metaphorically, the predicate hasn't made up its mind which of those precise meanings it wants to plump for. Suppose John is a borderline case of tallness. According to semantic theories of vagueness, it's not that in reality there is some fuzzy, indeterminate state of affairs of John's neither having nor lacking a certain quality. In the world there's just John with some utterly precise height. It's *the predicate* that is indeterminate such that there's no fact of the matter as to whether or not it applies to things with John's exact height. The indeterminacy is in language rather than the world.

There seems to be two ways of making sense of the predicate 'is conscious' being associated with a spectrum of sharpenings. The first is to adopt analytic functionalism, according to which it is a priori that for something to be conscious is for it to instantiate a certain functional state. On such views, it is natural to take the predicate 'is conscious' to be indeterminate with respect to a range of slightly different functional states, each of which could be captured with a fine grained enough description. The second strategy is to hold that the meaning of the predicate 'is conscious' is determined by facts outside of what is a priori accessible to the concept user, and so, although the meaning of the predicate involves a spectrum of sharpenings, that spectrum of sharpenings is not *a priori* accessible.<sup>26</sup>

The panpsychist is unlikely to be able to adopt either of these strategies. The first strategy entails a deflationary account of consciousness, whilst panpsychism is usually grounded in a robust metaphysical seriousness about consciousness. The second strategy involves the rejection of phenomenal transparency, and as I have already argued above the panpsychist has good reason to accept *phenomenal transparency*.

---

<sup>24</sup> See Dummett 1978: 260; Fine 1975; Russell 1923; Lewis 1986: 212.

<sup>25</sup> With some vague predicates, as with 'is tall', the sharpenings are determinates of a single determinable. In the case of other vague predicates, e.g. 'is a religion', there is a weighted cluster of properties, *involves belief in a supernatural being, involves ritual, involves a moral code*, such that each sharpening involves some of those properties but it is not the case that each sharpening involves all of those properties.

<sup>26</sup> See Papineau 2002.

If we are unable to give a semantic treatment of vagueness, then adopting the common sense answer to the special phenomenal composition question will require thinking of phenomenal borderline cases in metaphysical or epistemic terms. I haven't the courage to wade too much into the overwhelming literature on vagueness, but having dipped in my toe I find myself strongly inclined towards a semantic account of vagueness,<sup>27</sup> and thereby inclined against the common sense answer to the special phenomenal composition question. Others must weigh up these considerations for themselves.<sup>28</sup>

There is a strong possibility, then, that neither the facts of observation, nor the facts of common sense will be able to help the constitutive panpsychist answer the special phenomenal composition question. In such a case, she must turn to theoretical virtue. On account of their elegant simplicity, the two most theoretically satisfying answers to the special phenomenal composition question are: *Nihilism* – subjects never combine to make a further subjects, and *Universalism (unrestricted phenomenal composition)* – subjects always combine to make a further subject. Therefore, on the assumption that nihilism is a non-starter on the grounds that the subjects we are pre-theoretically committed to are composite objects of some sort, universalism looks to be the default position. In a similar way, before we had empirical evidence to the contrary, the default assumption was that the speed light was infinite (zero and infinity being the simplest values). Arbitrary constants are to be avoided if it all possible. Universalism is wildly at odds with common sense, but we have shown that there are deep problems with what common sense has to tell us about phenomenal composition.

If we are prepared to accept universalism, there is an obvious candidate for identification with the phenomenal bonding relation: the spatial relation. If we identify the phenomenal bonding relation with the spatial relation it follows that, for any group of material objects, the members of that group, being spatially related, determine a conscious subject. A nice consequence of identifying the phenomenal bonding relation with the spatial relation is that we end up having some positive conception of the spatial relation. As noted at the end of the section V, the spatial relation must have some real nature that goes beyond the mathematical conception of it we get from physics. Supposing that the real nature of the spatial relation is the phenomenal bonding relation is not entirely satisfying, as we lack a transparent conception of that relation – we don't completely understand what it is for objects to be related in that way. Still, it is better than nothing; at least we know that the real nature of the spatial relation is such as to bond subjects together to constitute further subjects. On the view currently under consideration, we have a reasonable grip on the nature of the world: the only intrinsic determinable is consciousness, the only relational determinable is phenomenal bonding.

## Conclusion

Some form of panpsychism is highly likely to be the true theory of our universe. It's high time we started working out the details.<sup>29</sup>

---

<sup>27</sup> I find persuasive Horgan and Potrc's 2008 argument against the coherence of metaphysical vagueness. For a contemporary account of metaphysical vagueness, see Barnes and Williams 2011.

<sup>28</sup> My formulation of these considerations is influenced by Lewis's (1986) and Sider's (2001) arguments against unrestricted composition, and is similar to arguments I gave in Goff 2011b and Goff 2013.

<sup>29</sup> I am grateful to Luke Roefls, Hedda Hassel and David Chalmers for comments.

## References

- Barnes, E. J. and Williams, J. R. G. 2011. 'A theory of metaphysical indeterminacy,' in K. Bennett and D. Zimmerman (Eds.) *Oxford Studies in Metaphysics*, Oxford University Press.
- Chalmers, D. J. 1996. *The Conscious Mind: Towards a Fundamental Theory*, Oxford University Press.
- Chalmers, D. J. 2002. Consciousness and its Place in Nature, in D. J. Chalmers (Ed.) *Philosophy of Mind: Classical and Contemporary Readings*, Oxford, New York: Oxford University Press, 247-72.
- Chalmers, D. J. 2003. 'The content and epistemology of phenomenal belief,' in Q. Smith & A. Jokic (Eds.) *Consciousness: New Philosophical Perspectives*, Oxford University Press.
- Chalmers, D. J. 2009. 'The Two-Dimensional Argument Against Materialism,' in B. McLaughlin (Ed.) *Oxford Handbook of the Philosophy of Mind*, Oxford University Press, 313–39.
- Chalmers, D. this volume. 'The combination problem for panpsychism'.
- Chalmers, D. Forthcoming, 'Panpsychism and panprotopsychism,' in T. Alter & Y. Nagasawa (Eds.) *Russellian Monism*.
- Dummett, M. 1978. *Truth and Other Enigmas*, London: Duckworth.
- Fine, K. 1975. *Vagueness, Truth and Logic*, Synthese 30, 265-300.
- Goff, P. 2009a. 'Can the panpsychist get round the combination problem?' in D. Skrbina (Ed.) *The Mind that Abides: Panpsychism in the New Millennium*, Benjamins Publishing.
- Goff, P. 2009b. 'Why pansychism doesn't help explain consciousness,' *Dialectica*, 63.
- Goff, P. 2011a. 'A posteriori physicalists get our phenomenal concepts wrong,' *Australasian Journal of Philosophy* 89.
- Goff, P. 2011b. 'There is no combination problem,' *The Mental as Fundamental*, Blaumauer, M. (Ed.) Ontos Publishing House.
- Goff, P. 2013. 'Orthodox property dualism + linguistic theory of vagueness = panpsychism,' in R. Brown (Ed.) *Consciousness Inside and Out: Phenomenology, Neuroscience, and the Nature of Experience*, Springer.
- Goff, P. Forthcoming a. 'Real acquaintance and the phenomenal concepts strategy,' in P. Coates & S. Coleman (Eds.) *The Nature of Phenomenal Qualities*, Oxford University Press.
- Goff, P. Forthcoming b. 'Against constitutive Russellian monism,' in T. Alter and Y. Nagasawa *Russellian Monism*, Oxford University Press.
- Goff, P. MS. *Consciousness and Fundamental Reality*.

Goff, P. and Papineau, D. Forthcoming. 'What's wrong with strong necessities?', *Philosophical Studies*.

James, W. 1890/1950. *The Principles of Psychology* (2 volumes) Dover.

Kripke, S. 1972. 'Naming and necessity,' in G. Harman and D. Davidson (Eds) *The Semantics of Natural Language*; reprinted as *Naming and Necessity* 1980, Harvard University Press.

Lewis, D. 1986. *On The Plurality of Worlds*, Blackwell.

McLaughlin, B. P. 2001. A defense of new wave materialism: A response to Horgan and Tienson, in C. Gillett & B. Loewer *Physicalism and Its Discontents*, Cambridge University Press.

Merricks, T. 2001. *Objects and Persons*, Oxford University Press.

Russell, B. 1923. 'Vagueness,' *Australasian Journal of Philosophy and Psychology*, 1: 84-92.

Sider, T. 2001. *Four-Dimensionalism: An Ontology of Persistence and Time*, Oxford University Press.

Sider, 2013. 'Against parthood,' in K. Bennett & D. W. Zimmerman (Eds.) *Oxford Studies in Metaphysics*, 8.

Papineau, D. 2002. *Thinking about Consciousness*, Oxford University Press.

Papineau, D. 2006. 'Comments on Galen Strawson,' *Journal of Consciousness Studies* 13.

van Inwagen, P. 1990. *Material Beings*, Cornell University Press.