Chapter 7 – The Combination Problem

I: Bottom-up

As we have discovered, Russellian monism is a very promising theory of fundamental reality and the place of consciousness within it. However, the view faces serious challenges, some of which we will explore in the next two chapters.

The most notorious problem facing Russellian monism is the combination problem, which is actually a family of problems. Each is a reflection of the following general concern.

According to Russellian monism, o-subjects and o-consciousness – the kind of subjects and consciousness we pre-theoretically associate with humans and other animals – intelligibly
arise from fundamental (proto)subjects and (proto)consciousness. On the most familiar version, there are a huge number of micro-level (proto) subjects, or ‘micro-(proto)subjects’ as we might call them, in your brain right now each enjoying its own (proto)consciousness, which somehow come together to form, or to bring about, your mind and its consciousness. The essence of the combination problem is simply: how on earth is that possible? We feel we have some kind of grip on how bricks forms a house, or parts of a car engine make up an engine, but we are at a loss trying to make sense of lots of ‘little’ (proto) minds forming a ‘big’ mind.

We can divide the combination problem into two forms: bottom-up and top-down. Bottom-up versions of the combination problem start with fundamental (proto)consciousness or (proto)subjects, and raise doubts concerning their potential to intelligibly produce o-consciousness or o-subjects. The most powerful bottom-up combination problem, and the main focus of this chapter, is the subject-summing problem: the difficulty making sense of how ‘little’ subjects could unify together to make a ‘big’ subject. The subject-summing problem is perhaps the paradigm form of the combination problem. Top-down versions of the combination problem, which we will discuss in the next chapter, start instead with o-consciousness or o-subjects, and raise doubts concerning their potential to be intelligibly produced by fundamental (proto)consciousness or fundamental (proto)subjects.

Before getting to the subject-summing problem itself I will to discuss two other putative concerns with panpsychism and panprotopsychism respectively. The concern with panpsychism, which I wholeheartedly reject, is that it’s just too crazy to believe. The concern with panprotopsychism, which I accept is a pro tanto concern with the view, is that it threatens to resign us to noumenalism: the view that human beings by their very nature are unable to understand the fundamental nature of reality.

At the end of these discussions of the first two sections, panpsychism will have the advantage in that it avoids noumenalism. However, as we move on to the subject-summing

---

1 I am not entirely happy with the word ‘proto-subject’, as there is no sense in which a ‘proto-subject’ is a kind of primitive subject. But it is useful to have a word for the bearers of protophenomenal properties, and so long as we remember that the word is just defined to mean a bearer of protophenomenal properties then we shouldn’t get into trouble. Perhaps the word ‘protophenomenal’ itself is problematic for similar reasons, but it has a clear definition and so ought not to lead to confusion.

2 The term ‘combination problem’ is due to William Seager (1995), although the problem itself is usually traced back to the discussions of William James considered below.
problem, we will find a serious challenge to panpsychism, and one which does not obviously apply to panprotopsychism. Therefore, if the panpsychist is to retain its advantage over the panprotopsychist, she must successfully respond to the subject-summing problem. I think that the panpsychist does have plausible things to say in response to this problem; although sadly there are more serious problems to be explored in the next chapter.

The threat of noumenalism

An initial concern about panprotopsychism is that it seems to be an incomplete picture of the world. The essence of the view is that the deep nature of the physical consists of protophenomenal properties. But the definition of these properties is partly negative and partly indirect. Protophenomenal properties are those which are not phenomenal (the negative bit), but which somehow give rise to the phenomenal (the indirect bit). We are left wanting to know ‘Ok, but what are they? What are these consciousness-producing properties in and of themselves?’

It doesn’t seem that we have any positive conception of a property which is not a feeling, but which somehow entails the instantiation of feelings. To this extent, panprotopsychism gives us a rather unsatisfying account of fundamental reality and its relationship with consciousness. We ask the big question, ‘What is fundamental reality like, and how on earth does it give rise to consciousness?’; the panprotopsychist seems to answer, ‘I’ve no idea what fundamental reality is like, but whatever it’s like it somehow gives rise to consciousness.’

But how permanent is this incompleteness? For some panprotopsychists the positive nature of the protophenomenal is a detail to be filled in as the theory advances. Hence, whilst panpsychists, in characterising fundamental stuff as consciousness-involving, have a head start characterising the deep nature of matter, there is the hope that panprotopsychists will catch up as time advances. For other panprotopsychists our failure to grasp the protophenomenal is a permanent feature of the human condition. Just as dogs aren’t set up to understand mathematics, so human beings aren’t set up to grasp protophenomenology.
Colin McGinn is perhaps the best known of the second group of theorists, dubbed ‘mysterians’ by Owen Flanagan, after the first Latino rock band to have a hit in the US.\(^3\)

McGinn does not describe his view as a form of Russellian monism, but it seems to have all the features of a protophenomenal version of Russellian monism: there is the deep nature of matter, which is not itself a form of consciousness, but which intelligibly produces consciousness. McGinn believes that there is a perfectly intelligible link from certain brain properties to the properties of consciousness, but that we are constitutionally incapable of understanding that link. To put it in his jargon, we are ‘cognitively closed’ to the brain properties which explain consciousness, a notion defined as follows:

A type of mind M is cognitively closed with respect to a property P (or a theory T), if and only if the concept-forming procedures at M’s disposal cannot extend to a grasp of P (or an understanding of T). Conceiving minds come in different kinds, equipped with varying powers and limitations, biases and blindspots, so that properties (or theories) may be accessible to some minds but not to others.\(^4\)

We can reasonably interpret McGinn as a panprotopsychist who holds that we are cognitively closed to protophenomenal properties. Panprotopsychism of this kind is a form of noumenalism, which we can now describe as the view that human beings are cognitively closed to the nature of fundamental reality.

For McGinn, the root of the difficulty is that we access consciousness on the one hand, and matter on the other, through two very different faculties: consciousness is accessed through introspection, matter through perception. Through introspection we find only consciousness; through perception we find only non-conscious properties which fail to entail consciousness. Through neither faculty can we discover properties of the brain which are not phenomenal but entail phenomenology. Human beings simply lack a faculty for accessing protophenomenal properties.

---

\(^3\) McGinn 1989.
\(^4\) Pereboom 2001/Forthcoming.
However, as has been pointed out by Derk Pereboom, there seems to be a gap in this argument.\(^5\) Even if McGinn is right that we don’t seem to have any immediate awareness of, or acquaintance with, protophenomenal properties, it doesn’t follow that we are unable to form a positive conception of them through our creative imagination. Perhaps future theorists might be able to construct possible candidates in imagination, and maybe even formulate testable hypotheses involving such candidates. It seems that we do not currently have a positive conception of protophenomenal properties, but how can we draw limits on future conceptual innovation?

Perhaps there is no way of conclusively demonstrating that humans are cognitively closed to protophenomenal properties, but there may be reason to give some credence to the supposition that we are. It is plausible that a human who has not been acquainted with a token of phenomenal property type P is cognitively closed to phenomenal property type P; for example, a congenitally blind person is unable to form a positive idea of phenomenal colour. We might suppose the same is true of protophenomenal properties; if we agree with McGinn that no human is acquainted with the protophenomenal property underlying phenomenal red, it may seem plausible to infer from this that no human is able to form a positive idea of that property.

Furthermore, reflect for a moment on what would be involving in positively conceiving of a protophenomenal property. A protophenomenal property is a third person property, but it entails the instantiation of first person properties. It seems that to conceive of it, or at least to conceive of it qua protophenomenal property, would involve first person and third person representational faculties employed in a unified conception of a single property. But our perceptual and introspective faculties don’t seem capable of integrating in this way. The more I reflect on what would be involved in conceiving of a protophenomenal property, the more I am struck by how awesome — in the British English sense of the word — it would be to behold such a property, and the more I am inclined to think that beholding such beauty is beyond the reach of human minds, just as, say, visually imagining the infinite is beyond us.

Perhaps I have limited imagination regarding what it is possible for humans to imagine. Or perhaps a bit of surgery or drug induced brain alteration could expand our powers of

imagination. Unlike McGinn, I don’t take what I have said above to constitute a knock down argument that humans will be forever incapable of grasping protophenomenal properties. It’s just that, whereas Pereboom is cautiously optimistic, I am cautiously pessimistic. At the very least, panprotopsychists are reliant on the hope for a monumental conceptual innovation in order to fill in the gaps in their theory of the world.

Is panpsychism too crazy to believe?

Panpsychist forms of Russellian monism look to be the simplest and most elegant and unified theories consistent with the reality of consciousness. As we discovered in the last chapter, although physics tells us much about the dispositions of micro-level entities, it leaves us completely in the dark about their absolutely intrinsic nature. The only thing we know for certain about the (more than merely formal) categorical nature of natural entities is that at least some of them, for example you and I, have a consciousness-involving nature.

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 properties of the kind we are acquainted with, 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 leads us to 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 perhaps no observational data which in itself supports panpsychism, there is a hard datum which arguably counts in its favour: the existence of consciousness. The reality of consciousness is more evident to us than any empirical data. The existence of consciousness does not entail the truth of panpsychism, but if it turns out that panpsychism is the most unified picture of the world which is consistent both with its existence and with our observational knowledge, then we have solid evidence in favour of the truth of panpsychism. 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 arguably 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 experience to fundamental particles. But once we realise that physics leaves us completely in the dark about the (more than merely formal) 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.
The subject-summing problem

At this stage, panpsychism has the advantage over panprotopsychism. Panpsychism provides unity where panprotopsychism introduces division. Panpsychism gives answers where panprotopsychism brings mystery. However, panpsychism faces a deep challenge which threatens to undermine its very coherence: the subject-summing problem. Whilst all forms of Russellian monism face some form of the combination problem, the subject-summing problem is perhaps the most vicious form of the worry, and it is faced only by panpsychists. If panpsychists are unable to mount a good defence against the subject-summing problem, the advantage may go back to panprotopsychism.

If the difficulty for the panprotopsychist is that she doesn’t know enough about the fundamental constituents of reality, the difficulty for the panpsychist is she knows too much. She knows, if her theory of reality is correct, that the fundamental building blocks of reality are subjects. And by the definition of Russellian panpsychism these subjects combine to intelligibly produce o-subjects. We will explore below a few ways of spelling out the subject-summing problem, but the starting point is just a deep intuition that subjects aren’t combinable. The idea of lots of little conscious subjects coming together to form a big conscious subject seems at best something we haven’t been yet able to make sense of, and at worst something of which sense cannot be made.

James’ anti-subject-summing argument

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 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 where 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, not it with them, and one could
never deduce the one from the others, nor (in any intelligible sense)
say that they evolved it.\(^6\)

Many philosophers, under the influence of this passage, claim to find some special
conceptual difficulty in the idea of mental entities 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\(_2\) 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.\(^7\)

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 exist only 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.\(^8\)

However, the more suggestive elements of the former paragraph have made more of an

\(^6\) James 1890: 1. 160.
\(^7\) James 1890: 1. 158-9.
\(^8\) Shani 2010 critiques James’ argument on these grounds.
impression on contemporary philosophers than the argument of the latter paragraph. There
does seem to be some deep difficulty making sense of distinct subjects of experience
combining to intelligibly produce a greater subject, and in what follows I will explore two
different ways of understanding this difficulty.

**The anti-subject-summing conceivability argument**

In the standard understanding of panpsychist Russellian monism o-subjects are intelligibly
produced by micro-level subjects, or ‘micro-subjects’ as we might call them. Call this
standard form of panpsychist Russellian monism ‘micropsychism’. The constitutive
micropsychist holds that the intelligible production takes the form of grounding, whilst the
emergentist panpsychist holds that the intelligible production is causal. But in both forms
there is an a priori entailment from the micro-phenomenal truths – as we might call them –
to the o-phenomenal truths. In virtue of this a priori entailment, the micro-phenomenal
truths necessitate the o-phenomenal truths.

One way of reading the James’ passage sees it as speaking against the possibility of such
intelligible production. For any group of subjects, it seems that we could conceive of those
subjects existing in the absence of some further subject. A panpsychist might suppose that
my severe pain intelligibly arises from the slight pain of trillions of neurons. And yet, for any
group of subjects feeling slight pain, it seems possible to conceive of just that number of
slightly pained subjects existing in the absence of some further pained subject, whether
slightly or severely pained.

Contrast with the case of a party. Once you’ve got people dancing and drinking and
generally having a good time, you’ve got a party. It’s impossible to conceive of people
dancing and drinking and having fun without conceiving of a party; the two states of affairs
cannot intelligibly be separated. Similarly, it’s impossible to conceive of a state of affairs in
which bricks are arranged as they are in my house, but in which there is no building formed
of those bricks. In contrast, whenever we conceive of two states of affairs involving distinct

---

9 Schaffer (forthcoming) holds that we require basic metaphysical principles to bridge the gap between micro-
level entities and composite wholes. I know from conversation that he thinks the same about revelling and
parties. However, in line with the grounding by analysis model, I am inclined to think that all it is for there to be
a party is for there to be people revelling, and consequently we cannot conceive of there being people revelling
without it being the case that there is a party.
subjects, it seems that those two states of affairs can be intelligibly separated: either can be conceived of obtaining in the absence of the other.

We can put this in the form of a principle:

Conceivable Isolation of Subjects (CIS) [pronounced ‘kiss’] – For any group of subjects, S_1, S_2...S_n, instantiating certain conscious states, E_1, E_2...E_n, it is conceivable that just those subjects with those conscious states exist in the absence of any further subject S* such that S* is not identical with any of S_1,S_2...S_n.

It might be objected that we only know CIS to obtain with respect to phenomenal qualities we are acquainted with it and hence able to form phenomenal concepts of. Indeed it is not clear that we can form phenomenal concepts of the phenomenal qualities instantiated by micro-subjects, and if we are unable to conceive of micro-subjects in terms of their phenomenal nature, then we are unable to directly test whether CIS is true of micro-subjects, i.e. whether or not micro-subjects can be conceived of in the absence of further subjects. Nonetheless, CIS seems true of any of the kinds of subject we can positively conceive of; varying the phenomenal character of the subjects we’re conceiving of, in so far as we are able to do this, has no bearing on the conceivable isolation of the subjects being conceived of. Intuitively, CIS is to be a conceptual truth concerning the determinable property of being a subject of experience rather than any specific determinates of it.

I don’t know how to demonstratively prove that there is not a possible set of subjects which constitute a counterexample to CIS: i.e. a group of subjects of experience which cannot intelligibly be thought about in the absence of some further subject of experience. But reflection shows CIS to be true with regards to all the many and varied kinds of subject 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 CIS is a principle we can reasonably take ourselves to know.

CIS constitutes a challenge for the micropsychist. If the micro-subjects in my brain, merely by existing with the conscious experiences they have, intelligibly produce the subject of my experience, then it is not conceivable that those micro-subjects exist (with the conscious
experiences they have at a time when I am conscious) without the subject of my o-
experience also existing. But it follows from CIS that it is conceivable that those micro-
subjects exist (with the conscious experiences they have a time when I am conscious) 
without the subject of my experience also existing.

The above line of reasoning assumes that the subject of my experience is not identical with 
one of the micro-subjects which intelligibly produce it. This is not true by definition, as o-
subjects are defined as the subjects of experience we pre-theoretically associate with
macroscopic entities, such as people and animals. It is a conceptual possibility that the
subject which common sense associates with me is identical with a single particle in my 
brain. Nonetheless, this is a conceptual possibility which it is reasonable to dismiss on
empirical grounds, and on grounds of general plausibility. There is no single micro-level
entity in my brain with which it is plausible to identify as my entire mind. And the thesis that
my mind is identical with a single micro-level entity is tantamount to epiphenomenalism, as
a single micro-level entity in the brain cannot on its own cause the behaviour we pre-
theoretically want to suppose that my mind causes.

We could combine CIS with the Transparency Conceivability Principle in order to infer the
corresponding modal principle:

.Modal Isolation of Subjects (MIS) – For any group of subjects, S₁,
S₂…Sₙ, instantiating certain conscious states, E₁, E₂…Eₙ, it is
possible that just those subjects with those conscious states exist
in the absence of any further subject S* such that S* is not
identical with any of S₁,S₂…Sₙ.

However, a move from conceivability to possibility is unnecessary given that Russelian
monism has a claim about conceivability built into it. If the micro-level phenomenal states of
affairs intelligibly produce the o-phenomenal states of affairs, then the micro-level
phenomenal states of affairs cannot be conceived of in the absence of the o-phenomenal
states of affairs. We can note also that this is just as much a problem for the layered
emergentist Russelian panpsychist as it is for the constitutive Russelian panpsychist, as
both think take the connection between micro-level consciousness and o-consciousness to be intelligible.\textsuperscript{10}

**Fusionism and the anti-subject-summing conceivability argument**

Fusionist panpsychists agree with layered panpsychists that o-subjects are intelligibly produced by micro-subjects. However, the intelligible production is *diachronic* rather than synchronic. Compare: a causal essentialist may hold that salt is essentially disposed to dissolve in water; the fact that salt has been placed in water at a given time intelligibly results in the salt dissolving *at a slightly later time*. Similarly, for the fusionist o-subjects are formed by a process of diachronically intelligible production: the obtaining of certain facts concerning micro-subjects at a given time entail that there will be an o-subject at a slightly later time. Recall that for the fusionist the o-subject does not co-exist with the micro-subjects which produce it; the micro-subjects cease to exist at the moment the o-subject is formed.

Whilst fusionist intelligible production is diachronic, it is natural to interpret CIS synchronically:

*Synchonic interpretation of CIS* – For any group of subjects, instantiating certain conscious states, it is conceivable that just those subjects, with those conscious states exist at time T in the absence of any further subject at time T.

On this synchronous interpretation CIS is clearly no threat to fusionism, given that the fusionist does not think a macro-subject exists at the same time as the micro-subjects which formed it.

One might of course formulate a diachronic version of CIS:

*Diachronic interpretation of CIS* – For any group of subjects, instantiating certain conscious states, it is conceivable that just those subjects, with those conscious states exist at time T in the absence of any further subject at any time later than T.

\textsuperscript{10} I gave a version of the conceivability subject-summing argument in Goff 2009.
However, it would not be unreasonable for the fusionist to build some ceteris paribus qualifiers into their claims of intelligible production. Their claim is likely to be that the instantiation of a certain micro-phenomenal state of affairs at T1 intelligibly produces the instantiation of a certain macro-phenomenal state of affairs at T2 *so long as there are no intervening factors*. For example, if God decides to annihilate the universe at T1, then clearly there aren’t going to be any o-subjects at T2. Compare: even if salt is essentially such as to dissolve in water, if it’s placed in water at T1, and God simultaneously decides to destroy the world, that salt ain’t going to dissolve.

Strictly speaking, then, the fusionist could make sense of the universal applicability of the diachronic form of CIS. For any micro-phenomenal states of affairs, it is conceivable that that state of affairs not give rise to the existence of an o-subject if God decides to destroy the world before it has time. One might object at this point that it’s not even the case that micro-level subjects are essentially disposed to fuse into o-subjects, such that this would happen if they were left to their own devices. However, this sounds more like a direct attack on the view, and turns on old and difficult questions concerning whether or not phenomenal qualities are essentially active. I will not go into these matters here.¹¹ Suffice to say that the subject-summing problem is much less obviously a challenge for the fusionist, and this seems to be an advantage of the view.¹²

It remains for us to examine how the micropsychist might respond to the subject-summing problem. Below I will quickly reject two common responses, before turning to the response I think the micropsychist ought to make.

**The ignorance response**

It might perhaps be suggested that it is our ignorance of the nature of subject-hood – of what it is for something to be a subject – which explains our inability to grasp the combinability of subjects.¹³ The most straightforward version of this would adopt full blown opacity concerning our concept of subject-hood, according to which we have no a priori grasp of what it is for something to be a subject. It could then be claimed that if we

---

¹¹ We will return to this issue in the discussion of consciousness+ in chapter 9.
¹² Hedda Hassel Mørch (2014) defends fusionism with an account according to which the emergence of o-subjects from micro-subjects is *partially* intelligible.
¹³ Galen Strawson (2006b) takes this line, and it is common to hear it from panpsychists in conversation.
understood what it is for something to be a conscious subject, we would understand that subjects are the kind of things which combine to yield other subjects. Imagine someone in the strange position of having opaque concepts of each of the pieces in a turnip jigsaw puzzle, perhaps by having a list of numbers each of which refers to a piece of the puzzle. Not having a grasp of the shape and appearance of each of the pieces, such a person would be ignorant of how a certain fact involving the pieces a priori entails the complete image of a turnip. Similarly, for the proponent of the ignorance response, we have only opaque labels for subject-hood, and hence fail to grasp all that is implied by the instantiation of that kind.

However, in chapter 5 I defended Direct Phenomenal Transparency, the thesis that direct phenomenal concepts reveal the essences of the states they denote. I take it that subject-hood is a determinable of which each conscious state is a determinate. For example, to be pained is to be a subject in some specific way, to have an experience of orange is to be a subject in some other way. I further take it that if one grasps the essence of a given determinate one thereby grasps the essence of the determinable of that determinate. For example, I couldn't understand what it is for something to be spherical without grasping what it is for something to be shaped, or what it is for something to be red without understanding what it is for something to be coloured. It follows that, for any phenomenal property I conceive of under a direct phenomenal concept, in grasping the nature of that property I thereby grasp the nature of subject-hood. This is inconsistent with the ignorance response, at least in its most straightforward form. Moreover, if we have no a priori grasp of the nature of subjectivity, then we have no a priori grounds for denying that that property is wholly physical, which would leave the panpsychist with no means of rejecting physicalism, and would therefore undermine the crucial motivation for her view.

The panpsychist is more likely to claim that we are partially ignorant of the nature of subjectivity, perhaps by defending Direct Phenomenal Translucency: the thesis that a direct phenomenal concept reveals an aspect of their nature of its referent. In this way she will hope to walk a thin line: we have enough a priori insight into the nature of micro-subjects to know that they aren’t physical, but not enough to grasp their combinability.

However, I also gave in chapter 5 I argued that a thesis of Full – rather than Partial – Revelation is required in order to properly account for our epistemic situation with respect
to consciousness; and Full Revelation implies Phenomenal Transparency. At the very least, panpsychist proponents of Phenomenal Transparency owe us an account of it as detailed as the physicalist account offered by Robert Schroer. They need to tell us, for example, exactly which properties of consciousness are transparently revealed and which hidden; and how this precise blend of transparency and opacity rules out physicalist whilst leaving phenomenal bonding an option. In my experience anti-physicalists in conversation often gesture at a commitment to Phenomenal Translucency; but I have never found in the literature a detailed non-physicalist account of it.

The consciousness+ response

One hypothesis easily confused with phenomenal translucency is the thesis that consciousness is an aspect of a greater property, of which we have only a partial understanding. Call this the consciousness+ hypothesis. Crucially, the consciousness+ hypothesis is not the thesis that we are ignorant about any aspect of consciousness itself; rather it is the thesis that we are partially ignorant about some property, call it ‘consciousness+’, of which consciousness is an aspect. Just as when thinking of proto-consciousness, so when thinking about consciousness+, I am struck by how deeply beautiful such a property would be, enfolding experiential and non-experiential aspects into a single nature.

But before I examine this matter with more care, and pass on to the consideration of other truths which may be derived from it, it seems to me right to pause for a while in order to contemplate...to consider, and admire, and adore, the beauty of this light so resplendent, at least as far as the strength of my mind, which is in some measure dazzled by the sight, will allow me to do so.14

Back to the game. According to the response currently under consideration, the unknown part of this consciousness+ is crucial to phenomenal combination, and our inability to make sense of phenomenal combination is the result of our ignorance of this aspect. The proponent of the consciousness+ hypothesis accepts CIS, but denies the following principle:

Conceivable Isolation of Subjects+ (CIS+) – For any group of subjects, S₁, S₂...Sₙ, instantiating certain conscious+ states, E₊₁, E₊₂...E₊ₙ, it is conceivable that just those subjects with those conscious states exist in the absence of any further subject S* such that S* is not identical with any of S₁,S₂...Sₙ.

Given our lack of grasp of the + bit of consciousness+, we are arguably not in a position to judge whether or not this principle is true.

The trouble with the consciousness+ hypothesis is that it essentially reintroduces protophenomenal properties, which destroys the advantage panpsychism has over panprotopsychism. Why postulate protophenomenal and phenomenal properties at the micro-level, when we might as well just postulate protophenomenal properties? Of course consciousness+ properties are ex hypothesi properties which already involve micro-phenomenal properties. But if the choice is between postulating consciousness+ properties, involving phenomenal and protophenomenal elements, and postulating protophenomenal properties, the latter would seem to be the simpler hypothesis. Once we have invested in protophenomenal properties, micro-level phenomenal properties are redundant.

The spatial relations response
CIS focuses on how subjects of experience are intrinsically, and says nothing about how they are related to each other. The principle tells us that a plurality of subjects cannot, merely in virtue of how they are intrinsically, intelligibly produce another subject of experience. One might respond that this is not surprising, as combination as we know it always requires the things combining to be spatially related to each other in a certain way. At the very least things which are to combine generally need to be near each other: bricks could not form a house if they’re scattered at different corners of the globe. And yet CIS does not even mention spatial proximity. The layered panpsychist may plausibly hold that in order to intelligibly produce an o-subject, micro-subjects must not only be a certain way intrinsically, but must also be spatially related to each in a certain way. Given that CIS does not mention spatial relationships between subjects, its truth is no bar to this notion of mental combination.
The initial difficulty with this response to CIS is that adding spatial relationships doesn’t seem to close the gap. In so far as we can conceive of micro-subjects as occupying spatial relations and standing in certain spatial relationships with each other – more on this shortly – doing so still seems to leave micro-subjects conceivably isolated from o-subjects.

We can make the point vividly by imagining *micro-experiential zombies*, defined as follows:

*Micro-experiential zombies* – A micro-experiential zombie is a duplicate of an actual human being in terms of its pure physical nature, is such that each of its fundamental parts is a micro-subject, but is also such that none of its parts instantiates an o-subject.  

When we conceive of a regular zombie of, say, the great Liverpudlian comedian Ken Dodd, we conceive of something just like Ken but lacking any kind of consciousness. Take such a zombie and conceive of each of its micro-level parts instantiating, say, slight pain. It still seems conceivable that such a creature lacks the o-consciousness we pre-theoretically associate with Ken himself, or any other form of o-consciousness.

Perhaps it’s not surprising that a multitude of instances of slight pain won’t give rise to Ken’s complex and sophisticated experience. But no matter what kind of consciousness we attribute to Ken’s bits, the result still seems consistent with the absence of Ken’s macro-consciousness. Suppose actual macro-subject Ken is currently staring at green grass, smelling it’s freshly cut odour, and listening to a lawn mower engine in the distance. We might imagine (obviously ludicrously oversimplifying to make the point vivid) that some of Ken’s micro-subjects underlying relevant brain functioning have an experience as of staring at grass, some have an experience as of freshly cut grass smell, and some have an experience as of the sound of a distant lawnmower. Still, all of these micro-level goings on seem consistent with the absence of some subject having *a unified experience* as of seeing and smelling grass whilst hearing a distant lawnmower. No matter what weird and whacky conscious states we attribute to your micro-level bits, it seems conceivable that those micro-subjects exist in the absence of some further subject beyond the micro-level.

---

15 I introduced micro-experiential zombies in Goff 2009.
It might be pointed out that we cannot conceive of every form of consciousness, and probably we cannot conceive of the forms of consciousness instantiated by micro-level subjects. However, I suggested earlier that the fact that CIS holds for the many and varied forms of consciousness we are able to conceive of gives us good grounds for believing CIS. Similarly, I would suggest here that the fact that micro-experiential zombies remain conceivable whatever consciousness, from the many and varied forms of we are able to conceive of, we imagine enjoyed by their micro-level parts, gives us good reason to think that all micro-experiential zombies are conceivable. In other words, we have good reason to believe in a form of CIS which involves micro-subjects standing in pure physical relations:

**Physicalised Conceivable Isolation of Subjects (P-CIS)** – For any group of subjects, instantiating certain conscious states and realising certain pure physical structures, it is conceivable that just those subjects with those conscious states, standing in those relations, exist in the absence of any further subject.

P-CIS seems to imply a form of CIS which involves micro-subjects standing in spatial relations:

**Spatialised Conceivable Isolation of Subjects (S-CIS)** – For any group of subjects, instantiating certain conscious states and standing in certain spatial relations, it is conceivable that just those subjects with those conscious states, standing in those relations, exist in the absence of any further subject.

This can be seen as a very depressing conclusion for the micropsychist. One is initially drawn to panpsychism on the grounds that the physical facts cannot account for consciousness: zombies are conceivable. And yet adding micro-subjects to the mix seems not to change anything: zombies are still conceivable. We seem to have got nowhere.

In spite of all this, I think the spatial relations response is defensible. What must be appreciated is that we have good reason to think that we lack a transparent understanding of spatial relationships. As we discussed in the last chapter, our scientific understanding of spatiality is mediated by mathematical models, and it is natural to think that mathematical
models of concrete features of reality abstract from their real nature. The intuition that relations must have a more than mathematical nature is perhaps weaker than the intuition that intrinsic properties must have a more than mathematical nature. However, if space itself is a fundamental entity, perhaps the most straightforward way of interpreting general relativity, then it must have an intrinsic nature which goes beyond the mathematical nature given to us in physics. And if space itself has a more than mathematical nature, then presumably spatial relations have a more than mathematical nature.\textsuperscript{16}

If micro-subjects do combine, the ways in which they are spatially related will surely be relevant. But spatial relations will be relevant in virtue of their full nature, not merely the mathematical structures they realise. In the light of this we can appreciate the possibility that our conception of a micro-experiential zombie is far from being transparent: our conception of the intrinsic nature of micro-subjects is potentially transparent (or at least our grip on their determinable nature is transparent); but our conception of the way in which micro-subjects are spatially related may be opaque or translucent. It is not surprising that this partially opaque conception fails to reveal the intelligible connection between the micro-phenomenal facts and o-phenomenal facts, if indeed there is one.

We can see, then, that there is an ambiguity in S-CIS. It might be a claim about what is conceivable when spatial relations are conceived of in terms of their pure physical nature:

\textit{Physicalistically Spatialised Conceivable Isolation of Subjects (PS-CIS)} – For any group of subjects, instantiating certain conscious states and standing in certain spatial relations – where those spatial relations are conceived of in terms of the pure physical structures they ground – it is conceivable that just those subjects with those conscious states, standing in those spatial relations, exist in the absence of any further subject.

\textsuperscript{16} At least if spatial relations are fundamental, then they must have a deep nature. On priority monist views (considered in chapter 9) there is only one fundamental object, and hence spatial relations are not fundamental; there is less pressure on such a view to think that spatial relations have deep nature. We will reassess matters when we examine this kind of view.
Alternatively, it might be a claim about what is conceivable when spatial relations are conceived of under a transparent conception:

*Transparency-Spatialised Conceivable Isolation of Subjects (TS-CIS)*

– For any group of subjects, instantiating certain conscious states and standing in certain spatial relations – where those spatial relations are conceived of under transparent concepts – it is conceivable that just those subjects with those conscious states, standing in those spatial relations, exist in the absence of any further subject.

We have good reason to think that PS-CIS is true, but we have no good reason to think TS-CIS is true. Lacking a transparent conception of spatial relationships, we have no grounds for denying that micro-subjects, spatially related, intelligibly produce o-subjects.

The layered panpsychist, then, can reasonably argue that the subject-summing problem results from our lack of understanding of the real nature of the spatial relation. If we transparently understood the nature of spatial relationships, then it would be clear to us how micro-subjects bond to make o-subjects. The gap between the micro-phenomenal and the o-phenomenal plausibly results from the gap in our understanding of space.

This obviously concedes a gap in our understanding, and threatens to introduce an element of noumenalism. Given that we are not acquainted with spatial relations, it is not clear that we will ever be able to form a positive idea of their full nature, just as it is not clear that we will ever be able to form a positive idea of a protophenomenal property. This approach seems to be consign us to permanent ignorance with respect to subject combination.

However, the reasons I gave above for supposing that we lack a transparent conception of spatial relations would be reasons just as much for the panprotopsychist as they would be for the panpsychist. I have criticised panprotopsychism, and the consciousness+ hypothesis, on the grounds that they add more mystery than is necessary, given that panpsychism is an alternative. However, if spatial relations are fundamental features of reality we lack an acquaintance relation with, a certain amount of mystery regarding their deep nature may be inevitable. Even so, panpsychism remains more attractive than panprotopsychism, in that it
avoids adding to our ignorance about fundamental relations and ignorance about fundamental intrinsic properties.

Of course, this is still in a certain sense less than ideal. Such gaps frustrate the human desire for a complete picture of the world. However, we should be trying to work out which picture of the world is most likely to be true, not which picture is most likely to be understandable by human beings. Assuming we are the products of natural selection rather than intelligent design, it is perhaps more likely that the world won’t be entirely comprehensible than that it will. When we forget that the physical sciences capture only causal structure, we become overly ambitious about what we are likely to achieve through our own epistemological efforts.

I don’t want to completely rule out the possibility of human beings one day coming to understand the real nature of metaphysically fundamental spatial relations, and perhaps in doing so will be able to give a more satisfying account of the summing of subjects. Perhaps it simply requires some great conceptual innovation, or a bit of neural re-wiring. However, I am inclined to be cautiously pessimistic in this regard for the same reason I was cautiously pessimistic about the non-panqualityist panprotopsychist’s hopes for discovering a positive conception of protophenomenality. It’s hard to see how we could somehow guess at the more than mathematical nature of a feature of reality we are not acquainted with, just as it’s hard to see how a contingently blind woman could somehow guess at the phenomenal nature of red.

Whether or not my pessimism is justified, there is still plenty of hope for theoretical progress. Empirical investigation may be able to reveal what kind of spatial relationships result in combination and the production of an o-subject. Even if we are unable to fully grasp an intelligible connection between the micro-level phenomenal facts and the macro-level phenomenal facts, theoretical success with a layered panpsychist model may deepen our conviction that there is such an intelligible connection.

---

17 In this next chapter I will argue, for quite distinct reasons flowing from the Subject Irreducibility Problem, that constitutive subject-summing is impossible. However, this leaves open the possibility of emergentist subject-summing.
In conclusion, the subject-summing problem does not give us reason to doubt panpsychism. It merely draws our attention to an epistemic gap which, assuming that spatial relations are fundamental, we would expect to be there anyway given our ignorance concerning the nature of such relations.

**Other anti-subject-summing arguments**

We have been considering a conceivability argument in support of the impossibility of intelligible subject-summing. There are other arguments which attempt to show more directly that the notion of a conscious whole formed from conscious parts is implausible or incoherent. In an early paper of mine – before my conversion to panpsychism – I argued that even if we can make sense of subject-summing, the experience of each micro-subject must feature in the experience of the macro-subject, and that appreciating this casts doubt on the coherence of micropsychism:

For the experiential being of some little experiencing thing ‘LITTLE’ to be part of the experiential being some big experiencing thing ‘BIG’ is for what it is like to be LITTLE to be a part of what it is like to be BIG. But if follows from this that BIG feels how LITTLE does (even if it also feels other things). Correspondingly, for the experiential being of some BIG to be wholly constituted by the experiential being of LITTLE 1, LITTLE 2, LITTLE 3… (again assuming this makes any sense at all) can be nothing other than for BIG to feel how it feels to be all those LITTLES and to feel nothing else. Even if it is intelligible how experiential states can sum together, it is contradictory to suppose that they could sum together to form some novel conscious state.

If my experiential being were constituted by the experiential being of billions of experience-involving ultimates [‘ultimate’ is Galen Strawson’s term for a fundamental material entity], then what it is like to be each of those ultimates would be part of what it is like to be me. I would literally feel how each of those ultimates feels,

---

18 Strawson 2006a.
somehow all at the same time. Assuming that my experiential being is wholly constituted by the experiential being of a billion experience-involving ultimates, then what it is like to be me can be nothing other than what it is like to be each of those billion ultimates (somehow experienced all at the same time).

But this surely cannot be right. My experience is of a three dimensional world of people, cars, buildings, etc. The phenomenal character of my experience is surely very different from the phenomenal character of something that feels as a billion ultimates feel. (p. 58-9)¹⁹

Even if this argument were successful in demonstrating that the experience of each micro-subject is present in the experience of the o-subject they form, perhaps we should see the argument as presenting a challenge to the layered panpsychist to accommodate this fact into her view, rather than being a decisive argument that layered panpsychism is incoherent.

However, more recently Sam Coleman has formulated an ingenious argument to the conclusion that, whatever the nature of the consciousness allegedly instantiated at the micro and macro level, subject-summing is incoherent.²⁰ His argument relies on the idea that the essence of a subject’s point of view is given not just by the fact that the subject has certain conscious states, but by the fact that it has certain conscious states and not any others. Like the holy sacrament of marriage, subjectivity demands exclusivity. Coleman considers for illustration the putative possibility of combining a micro-subject instantiating only phenomenal red, which he calls ‘Red’, and a micro-subject instantiating only phenomenal blue, which he calls ‘Blue’:

Consider the original duo’s point of view. One – Blue’s – is pervaded by a unitary blueness, the other – Red’s – by redness, and that is all they experience respectively. To say these points of view were present as components in the experiential perspective of the uber-subject (‘Ub’) would therefore be to say that Ub

---

²⁰ Coleman 2014.
experienced a unitary phenomenal blueness and a unitary phenomenal redness, i.e. had synchronous experiences as of each of these qualities alone, to the exclusion of all others. For it is these qualities each on their own that characterise, respectively, the perspectives of the original duo. Experience excludes, as well as includes. Yet nowhere does Ub have any such experiences: he precisely combines his predecessors’ qualitative experiential contents. Ub doesn’t experience red-to-the-exclusion-of-(blue-and)-all else, nor blue-to-the-exclusion-of-(red-and)-all else, let alone – impossibly – both together. Thus, the original points of view are not ingredients in Ub’s subjectivity.

Both of these arguments rely on equating the following two things:

- Phenomenal quality/point of view X is partially constituted (i.e. partly grounded) by phenomenal quality/point of view Y.
- Phenomenal quality/point of view X is partially characterised by phenomenal quality/point of view Y, i.e. Y is phenomenal present, or ‘shows up’, in X’s experience.

In my argument I was assuming that if my experience is partially constituted by, say, a slight pain, then that slight pain must characterise my experience, must be part of what it’s like to be me. In Coleman’s argument, he assumes that if a lesser point of view partially constitutes a greater point of view, then the greater point of view is partially characterised by the lesser point of view, and then demonstrates that this is incoherent. In both cases, the equation of partial constitution with partial characterisation is crucial. However, it seems to me that this equation can be coherently denied.

To try to make sense of this, consider a standard form of panprotopsychism. On this view, a given o-experiential state is constituted of certain protophenomenal properties, but those protophenomenal properties themselves don’t characterise the resulting o-experience. Protophenomenal properties, not being phenomenal, are not the kind of properties which

---

21 Coleman 2014: 33.
can characterise experiences. One might dispute the truth of this kind of view, but it does not seem that we can infer its falsity simply on the grounds that the qualities which constitute experience do not characterise experience. Nor for that matter could we rule out the truth of physicalism on these grounds. It seems coherent, therefore, to hold that a property can play a role in constituting a given experience without playing a role in characterising that experience.

One might object that it’s one thing to hold that a non-experiential micro-level property can partially constitute an o-experience without partially characterising that experience, quite another to hold that an experiential micro-level property can partly constitute without partially characterising an o-experience. However, once we have allowed that partial characterising and partially constituting can come apart, it seems that we can coherently suppose that they can come apart in the case of the relationship between the micro-level phenomenal and the o-phenomenal. At the very least we are owed an argument as to why this couldn’t be the case.

We might see the above arguments against subject-summing as relying on an ambiguity in the phrase ‘being part of what it’s like to be subject.’ A given micro-experience can be ‘part of what it’s like to be me’ in the sense that they constitute my experience, or can be ‘part of what it’s like to be me’ in the sense that they characterise my experience. Distinguishing these two meanings disarms the argument. Regarding my 2006 argument, one could hold that the experience of each of a huge number of micro-subjects constitutes without characterising my experience, and thereby avoid the counterintuitive conclusion that ‘what it is like to be me can be nothing other than what it is like to be each of those billion ultimates (somehow experienced all at the same time).’ Regarding Coleman’s argument, one could hold that Blue and Red constitute without characterising the point of view of Ub, and hence avoid the incoherent scenario of Ub experiencing both red-to-the-exclusion-of-(blue-and)-all-else and blue-to-the-exclusion-of-(red-and)-all-else.

I think there’s a more general moral to be drawn here. I would be cautious about any argument against subject-summing which implicitly relies on a sense that we have a transparent understanding of the part-whole relation. We feel intuitively we have a grip on what it is for one thing to be part of another thing, just as we feel we have a grip on what it
is for things to be spatially related, or for things to be solid. Certainly in the case of solidity, our everyday sense of what it is for matter to be solid has been decisively refuted by modern science, given that ‘solid’ matter is mostly empty space. But more generally, as Russellians we accept that our grip on the world beyond our own consciousness is mediated by abstract mathematical models. When concerned with the more than mathematical nature of fundamental reality, the default position should be an assumption of ignorance.

Russellians have thus far not appreciated that this general assumption of ignorance ought to hold in consideration of mereological relations. Appreciating this opens up possibilities. Just as has proved to be the case with respect to solidity, the part-whole relation might turn out to be radically different from our pre-theoretical conception of it. Suppose layered emergentist panpsychism is true, such that a given o-subject BIG is intelligibly caused by a number of micro-subject SMALLS. We might naturally be inclined to think: ‘Given that the SMALLS cause BIG, they cannot also be part of BIG. For a whole is nothing over and above its parts, whilst an effect is ontologically additional to its cause: causal relations are a matter of the cause bringing the effect into being, which entails that the effect is ontologically distinct from the cause.’

Once made explicit, it is clear that this argument relies on an assumption that we have a significant grip on the nature of the part-whole relation, the very assumption I am suggesting a Russellian ought not to make. It is rather like arguing, ‘Given that the table is solid, it cannot be mostly empty space.’ Like solidity, mereological relations are out there in the world, they are not created in the image of what we have in our minds. We identify a part/whole relation by pointing: ‘Those things *points at some small things* are parts of that thing *points at a large thing filling roughly the same region of space as the little things*.’ It could turn out that the little things we call ‘parts of Bill’s brain’ intelligibly cause the big thing we call ‘Bill’s brain’. In this case it would turn out, not that that the little things we call ‘parts’ are not really parts at all, but rather than – at least in this case – the relationship between parts and whole is causal rather than constitutive. The whole is something over and above its parts.

It is even clearer that Coleman’s anti-subject summing argument fails if the relationship between micro-subjects and o-subjects is causal rather than constitutive. The fact that a
given micro-phenomenal quality is partially causally responsible for my experience does not tempt us into thinking that that micro-phenomenal quality must partially characterise my experience. But the argument fails against constitutive panpsychism too. Suppose that micro-phenomenal property Q is involved in a fact which grounds the fact I have o-phenomenal property R, which is how I understand the claim that micro-phenomenal properties are involved in constituting my o-experience. It does not follow from this supposition that R is partially characterised by Q. At least the opponent of subject-summimg owes us an argument as to why we should think this is the case.

Thus in formulating the above solution to the subject-summimg problem, I do not take myself to be merely rescuing constitutive Russellian panpsychism, but also emergentist Russellian panpsychism. This point is important, as whilst I ultimately think constitutive Russellian panpsychism (or rather those which take the micro-level to be fundamental, we will discuss alternative forms in chapter 9) to be unworkable (for reasons distinct from the subject-summimg problem, which we will get to in the next chapter), emergentist Russellian monism seems to me perfectly coherent.