Chapter 8 – The Combination Problem II: 
Top-Down

Russellian monists postulate narrowly material subjects or proto-subjects, of which we are not familiar, in order to explain ordinary experience, of which we are familiar. In the last chapter we started with micro- (proto) subjects, assuming that the micro-level is the fundamental level (we will question this assumption in the next chapter), and tried to make sense of their potential to combine into 'big' subjects of experience. In this chapter we start with ordinary consciousness itself, with our own conscious experience. Could the conscious experience I am enjoying right now possibly be grounded in, or intelligible caused by, more fundamental kinds of experience? Or do I have conclusive grounds for thinking that my consciousness is primitive, fundamental or irreducible? Reflection on ordinary consciousness – o-consciousness as I have been calling it throughout the book – poses considerable challenges to the Russellian monist. In what follows we will outline some of those challenges and consider how the Russellian monist may respond.

Mental chemistry and the palette problem

The first difficulty we shall consider begins from the observation that o-experience is extraordinarily rich and diverse. Even within a single sense modality, such as vision, there
are a vast range of kinds of experiences. More strikingly, the experiences of different sense modalities seem of wildly different kinds, with no commonality between them; there seems, for example, to be an unbridgeable gulf between an experience of taste and an experience of colour.

This richness and variety is a problem for the Russellian monist because she is committed to o-experience arising from a relatively small number of properties. Recall that according to Russellian monism the explanatorily basic properties of the universe are material properties: the properties picked out by fundamental physics, such as mass, spin and charge. There is more to the nature of these fundamental properties than is revealed by physics: they are the deep nature of matter. Nonetheless, they are the properties named by fundamental physics, and those properties are relatively few in number. Hence, if Russelian monism is true, the vast range of kinds of consciousness – from colours, to tastes, to the experience of echo-location – must emerge from, as it were, a tiny palette of qualities. This seems prima facie an implausible, perhaps even impossible, supposition. I will follow Chalmers in calling this ‘the palette problem’.¹

We may take the palette problem in one of two ways. On a stronger understanding the palette problem takes the form of an argument to the conclusion that Russellian monism is impossible, as the variety of o-experience could not possibly result from a relatively small number of properties. The milder understanding of the palette problem takes the form of a challenge to the Russellian monist, a demand to explain how o-consciousness in all its variety could come from half a dozen or so properties. Let us take each of these in turn.

The strong palette problem

There are certainly similarities between different experiences within a sense modality: the experience of red is similar to the experience of orange, to return to an example we explored in chapter 5. And there is some prima facie plausibility to the idea that these similarities result from common phenomenal elements: intuitively, phenomenal orange is what you get when you mix phenomenal red with phenomenal yellow. Noting such similarities and commonalities may give the constitutive Russelian monist hope for the

prospects of a kind of ‘mental chemistry’, a science of how more complex phenomenal qualities result from combinations of more simple kinds.

Unfortunately, such hope is cruelly disappointed when we compare and contrast phenomenal qualities from distinct sense modalities. There may perhaps be structural or representational commonality between distinct senses: I can represent the shape of a ball with vision and with touch. But the media in which these representations are formed seem entirely unlike. On the face of it, the taste of mint on the one hand, and the experience of red on the other, have nothing whatsoever in common. How then could they be built up from the same elements?

We could press the strong palette problem against the constitutive Russellian monist with the following argument:

**The Palette Argument**

1. If constitutive Russellian monism is true, for any two o-conscious experiences X and Y, X and Y have common constituents.
2. Minty phenomenology and red phenomenology have nothing in common.
3. Therefore, constitutive Russellian monism is false.

The emergentist Russellian monist thinks that o-experience is intelligibly caused by, rather than built up from, micro-experiential elements. One might find baffling the idea that qualities as diverse and dissimilar as those we find in o-experience intelligibly arise from a small number of properties. However, although it is prima facie difficult to see how this could be so, it’s also difficult to see how one could demonstrate that it could not be so. Thus, the emergentist Russellian monist is subject only to the mild form of the palette, which will give her an advantage over the constitutive Russellian monist unless the latter has an effective response to the above argument.

Luke Roelofs has argued that there may be elements of our experience which we are in a sense unaware of, due to what he calls ‘confusion’, a notion he draws from Spinoza and Leibniz:

The perceptions of our senses even when they are clear must necessarily contain certain confused elements... [for] while our senses respond to everything, our soul
cannot pay attention to every particular... It is almost like the confused murmuring which is heard by those who approach the shore of a sea. It comes from the continual beatings of innumerable waves.²

The human body, being limited, is only capable of distinctly forming a certain number of images within itself at the same time; if this number be exceeded, the images will begin to be confused; if this number... be largely exceeded, all will become entirely confused one with another... When the images become quite confused in the body, the mind also imagines all bodies confusedly without any distinction, and will comprehend them, as it were, under one attribute.³

As Roelofs understands the notion of confusion, a subject is confused with respect to two mental elements when she is able to attend or think about those two elements together, but unable to attend or think about them individually.⁴

Let us illustrate the notion of confusion with respect to an imagined alien race, the Kenzars. When a Kenzar smells a rose she has an experience qualitatively just like the experience a human being has when she smells a rose: call that experience the ‘R-experience’. But she also has – when smelling a rose – an experience qualitatively just like the experience a human being has when she hears middle C: call that the ‘C-experience’. Suppose that, as a matter of contingent fact, a Kenzar never has an R-experience without also having a C-experience. Suppose further that Kenzar’s brains are wired such that, although they are able to attend to and think about both the R-experience and the C-experience together, for example, when they smell a rose, they are unable to think about or attend to R-experiences without also thinking about/attending to C-experiences. In Roelof’s terms, the Kenzars are confused with respect to R-experiences and C-experiences.⁵

---

² Leibniz 1686/2012: 96.
⁴ Roelofs 2014: ch. 5. Roelofs does not claim that this notion of confusion captures exactly what either Leibniz or Spinoza understand by the term in the above quotations.
⁵ Roelofs distinguishes between a number of different kinds of confusion, and his central notion of radical confusion is not quite the one I am employing here. My notion is asymmetric, as the Kenzars can think about C-experiences without thinking about R-experiences, e.g. when they think about S-experiences. I do this in order to discuss the possibility of being ignorant of a common element of experiences; I don’t think this affects the argument.
It might be natural to infer from this that the Kenzars would take the experience they have when they smell a rose to be a primitive, irreducible experience—call it an ‘RC-experience’—and would be unaware that it is composed of two distinct elements: R-experiences and C-experiences. To develop the example, we can imagine that when Kanzars taste sugar they have an experience qualitatively just like the experience a human being has when she tastes sugar—call it the ‘S-experience’—but they also have a C-experience. As in the case of smelling roses, although they can attend to and think of S-experiences and C-experiences together, they are unable to attend to/think about S-experiences without also attending to/thinking about C-experiences. It might be natural to infer that the Kenzars take the experience they have when they taste sugar—SC experience—and the experience they have when they smell a rose—RC experience—to both be primitive, irreducible experiences with no elements in common. Contrary to this supposition, these experiences are composite and have a common element: C-experience.

If all this makes sense, we might suppose that something similar is true of human beings. Perhaps whenever we see red we actually have two experiences—an X experience and a Y experience—but are confusedly unable to think about or attend to X and Y individually. Similarly, whenever we taste mint we have two experiences—an X experience and a Z experience—but are confusedly unable think about or attend to X and Y individually. We are thus unable to discern the common element—X-experience—in red experiences and minty experiences.

There is a worry that this ingenious solution to the strong palette problem is inconsistent with Phenomenal Transparency. I am able to attend to and form direct phenomenal concepts of both minty experiences and red experiences. Given Phenomenal Transparency, these concepts reveal the complete essences of those experience. If there is an essential element in common between these experiences, then that common element ought to be a priori accessible to someone who has both concepts. This common element may be difficult to discern, but it is inconsistent with Phenomenal Transparency to suppose it utterly inaccessible to reason.

However, what is being assumed in the above worry is that the constituents of a given experience are essential to that experience, and we need not take this to be so. Compare: being a heart is a functional property which is grounded in certain specific biological
properties in human beings. But the biological nature of a human heart is not essential to the general property of being a heart. The hearts of other organisms have a different biological nature, and indeed we can suppose in the future that some humans may have wholly prosthetic hearts. I can completely understand the essence of the property of being a heart, by understanding the functional role required for there to be a heart, without understanding anything of the biological goings on which ground that property in human beings.

Now red experiences are not functional properties, but it may be that like red experiences they are multiply realisable: in humans they are grounded in experiences X and Y, whilst in other creatures they may be grounded in certain other experiences or proto-experiences. In this case, I can completely understand the essence of red experience, what it is for something to have a red experience, whilst knowing nothing about the phenomenal or protophenomenal properties which ground that property in human beings. In general, grasping the essence of a property does not entail grasping the essence of the properties which ground it.\(^6\)

One might object that if X-phenomenology partly constitutes red phenomenology, then X-phenomenology is part of what it’s like to have red phenomenology, and thus ought to be in principle introspectively discernible to someone having a red experience. However, this plays on an ambiguity already discussed in our consideration of the mereological subject-summation argument in the last chapter. To say that X-phenomenology is ‘part of’ red-phenomenology might mean that X-phenomenology partly characterises red-phenomenology, but it might merely mean that X-phenomenology party constitutes minty-phenomenology, and the latter does not obviously imply the former.

Once we appreciate this, we can see that a commitment to confusion is not necessary in order to make sense of our ignorance concerning the constituents of our experiences, for the constituents of my experiences may not be experienced by me at all. Roelof’s view seems to be that:

1. O-experiences are identical with a very large number of micro-experiences standing in certain relations to each other. In our imaginary example, for a Kenzar to have an

---

\(^6\) Chalmers makes this point in relation to what he calls ‘the revelation argument’ in Chalmers forthcoming b.
RC experience just is for her to have an R-experience at the same time as a C-experience. In the real world Roelof’s view is presumably that to have a red experience just is to instantiate a huge number of micro-experiences at the same time (perhaps standing in certain relations to one another).

2. Confusion renders me unable to think about any of the micro-experiences individually.

This view is indeed inconsistent with Phenomenal Transparency: we have direct phenomenal concepts of many o-experiences, and hence if they were identical with complex micro-experiential states our direct phenomenal concepts of those states would reveal this to us. Of course we might give up on Phenomenal Transparency, but if direct phenomenal concepts do not reveal the essence of o-conscious states then we have no a priori grounds for denying that those states have an entirely physical nature, and hence we lose our motivation for adopting panpsychism (I will not repeat here the arguments I gave in chapter 5 against the halfway house of Phenomenal Translucency).

However, if we suppose that o-phenomenal states are grounded in, rather than identical with, certain micro-experiential states, then we need not suppose that those micro-experiential states are experiences of any human being, and so need not suppose that they are experiences any human being is aware of. Thus we can reject premise 1 of the palette argument. Although we entirely grasp the nature of red phenomenology and minty phenomenology, we may be entirely ignorant of the nature of their experiential or proto-experiential components. Hence, we have no introspective grounds for supposing that minty phenomenology and red phenomenology states are not constructed from common experiential or proto-experiential elements.

The mild palette problem

Even it is fails to present a knock down argument against Russellian monism, the palette problem clearly poses a serious challenge for a Russellian monist, of either constitutive or emergentist stripe, wanting to fill in the details of her view. How might it be that such diversity results from a small number of properties? The ideal Russellian monist theory would postulate a specific explanatoryly basic phenomenal or proto-phenomenal property for each of the properties of fundamental physics, and would then provide, for each of the
qualities we encounter in our experience, an explanation as to how that quality arises from those basic phenomenal/proto-phenomenal properties. A less ambitious way of responding to the palette problem would be to give a broad brush strokes account of how such seemingly radically diverse qualities might result from a small number of elements.

Work has already begun to address the challenge of the palette problem. Sam Coleman, drawing on the work of Charles Hartshorne, argues that the gap between sensory modalities might not be as unbridgeable as we imagine. Perhaps we think there is an unbridgeable gap between colours and tastes simply because we lack the experiences which would bridge that gap. Perhaps there is a certain range of possible experiences not had by humans which lie in between auditory experiences and colours experiences, such that if we instantiated those ‘in between’ experiences we would be able move in imagination from colours to sounds as seamlessly as we move between shades of blue. And maybe future brain science will find a way of inducing in us these ‘in between’ experiences.

There may also be more commonality than strikes us at first glance between the experiences we already have. Coleman suggests the following subtle commonalities between experiences of distinct sensory modalities:

Sometimes, especially when falling asleep, I am uncertain whether I heard or felt something thud somewhere (it doesn’t seem to be both): auditory/tactile overlap thus seems quite feasible. Again, one thinks of the experience of deep bass drumming. This suggests the possibility that tactile sensations are (qualitatively speaking) just ‘strong’, more ‘impactful’, forms of auditory qualia. Additionally, one might take seriously empirical data indicating that what we commonly think of as the ‘flavour’ of things is really a composite of smell and taste qualia (from receptors in the nose and tongue respectively). It’s very easy to separate out the taste element, just hold your nose while eating. Is our failure to notice this fusion experientially because these two species of qualia overlap qualitatively?

Coleman, following Hartshorne, speculates that, contrary to initial appearances, all experiential qualities may lie on a single continuum. If future research could identify such a

---

7 Coleman Forthcoming, Hartshorne 1934.
8 This is an extract from an early draft of Coleman Forthcoming which was later removed.
continuum and the location of our experiences within it, then we could perhaps come to understand how a relatively small number of properties could ‘hone in on’ a certain point on that spectrum and thereby explain the nature of resultant o-experiences.

An alternative approach to the palette problem is to try to reconceive the relationship between micro-experiences and o-experiences. Perhaps the emergence of o-experiences from micro-experiences is not so much a process of composition but of distillation. Keith Turausky speculates that the qualitative nature of each and every possible o-experience may be in some sense already present in the experience of each and every particle, in something like the way that all the colours of the spectrum are present in white light:

...instead of postulating that phenomenology “fades out” as one moves down the scale of complexity, we might consider that the simplest, most fundamental phenomenology is a sort of “white out”. In other words, perhaps experientiality in its most basic form is modally, intentionally, and informationally undifferentiated: not watered-down at all, but rather a super-saturated...experientiality. Consider what it would be like to experience everything at once: hot and cold, rising and falling, quiet and loud, light and dark, wet and dry, happy and sad, etc. ad infinitum. Not, again, that we wish to claim the humble electron actually feels any of those specific things; the point, rather is to get a sense for “experiencing everything at once” actually implies: an utter undifferentiation that is, in a certain sense, the antithesis of consciousness as we know it. To experience experiential state $E$ with every experiential state not-$E$ would be to have no useful content concerning either oneself or one’s environment. It would be a senseless seething stasis of phenomenology – but phenomenology it would be!...

Admittedly these are all extremely speculative stabs in the dark. But this is early days in a wholly new research programme. Neuroscience has come a long way in the last eighty years, but we have also wasted a lot of time on consciousness-denying third-person absolutist approaches to dealing with phenomenology, and precious little time has been devoted to working out the details of the Russelian framework. Darwin’s principle of natural selection was a very general hypothesis concerning the origins of life; filling in the

---

9 Turausky MS.
details has taken a long time and we’ve still got a long way to go. Similarly, Russellian monism is an extremely elegant unification of mind and matter, but it is also a very general hypothesis and it will take time to fill in the details.

The Russelian approach has enjoyed a small revival within contemporary analytic philosophy of mind. But it is still not a well-known view, and it is almost completely unknown in other relevant disciplines. Decades of serious work within a Darwinian framework led to the development of genetics. It will presumably take at least several decades of serious interdisciplinary effort within a Russelian framework before we have any kind of real understanding as to how everyday conscious experience emerges from the fundamental components of matter. We are a long way from even beginning to do what is required. It is ironic that the very people who accuse anti-physicalists of being ‘nay-sayers’ are the quickest to claim they know a priori that a panpsychist or panprotopsychist research programme will never bear fruit.¹⁰

Having said that, it is possible that the details will forever elude us. The great success of the physical sciences has given us great confidence about our powers to uncover the essential nature of space, time and matter. However, as we have discovered in chapters 1 and 6, this is false confidence: the success of the physical sciences is precisely due to the fact that since the scientific revolution they have not been in the business of finding out the essential nature of space, time and matter. From Galileo onwards, physicists have ignored the concrete categorical nature of matter, and just focused on mapping its causal structure. We human beings have no idea how much potential we have for finding out the concrete categorical nature of the world we live in. Assuming our faculties evolved for survival in a state of nature rather than for uncovering the essential nature of matter, it is perhaps more likely than not that we won’t get very far.

It’s well worth trying to develop a better understanding of why we have the kind of consciousness we do, and we may yet surprise ourselves. If theorising within the Russelian framework produces detailed panpsychist/panprotopsychist theories which accurately predict the nature of a given subject’s o-conscious on the basis of more fundamental

¹⁰ For more about ‘nay saying’ see my discussion of Churchland in chapter 1. I am basing my comment here on discussions with Daniel Dennett, Patricia Churchland, and other neuro-fundamentalists.
material goings on, then this would provide powerful evidence for these theories. However, it may be that we end up having to accept some degree of noumenalism about the deep nature of matter: I argued in the last chapter that this is likely with respect to the deep nature of fundamental spatial relations (if there are such things, which we will question in the next chapter). At the very least the business of guessing at the deep nature of matter is likely to involve a great deal more speculation and educated guessing than we are used to in the physical sciences.

Even if this proves to be the case, I would argue that the elegant way in which Russellian monism brings what we know about reality introspectively and what we know about reality empirically together in a unified view of the world, gives us strong reason to believe it to be true. Even if humans prove to be incapable of filling in the details, Russellian monism in its general form remains our best guess as to the nature of reality.

The structural mismatch problem

Like physicalists, most constitutive Russellian monists believe that o-experience results from micro-level goings on in the brain. Indeed, assuming that the macro-level brain is intelligible produced by such micro-level goings on, it is natural for the constitutive Russellian monist to identify the human brain and its macro-level states with the o-conscious subject and its o-conscious states. Of course the mind and the brain seem like very different things, but the Russellian monist puts this down to the fact that from the outside we only get at the causal structure of the brain, whilst the mind is identical with the deep material entity underlying that structure.

A problem remains, however, as whilst the physical sciences don’t reveal the concrete categorical nature of the brain, they do reveal its structure. And on the face of it the structure of the brain seems radically unlike the structure the o-conscious mind as revealed through introspection. But if the mind and brain are identical, or at least grounded in the same micro-level base, one might expect them to have the same structure.

David Chalmers puts the intuitive sense that mind and brain have different structures as follows:
Our phenomenology has a rich and specific structure: it is unified, bounded, differentiated into many different aspects, but with an underlying homogeneity to many of the aspects, and appears to have a single subject of experience. It is not easy to see how a distribution of a large number of individual microphysical systems, each with their own protophenomenal properties, could somehow add up to this rich and specific structure. Should one not expect something more like a disunified, jagged collection of phenomenal spikes?\(^{11}\)

As with the palette problem, the structural mismatch problem might be seen in either a stronger or a milder form.\(^ {12}\) In its milder form, it is challenge to the Russellian monist to explain how something with the structure of the o-conscious mind could possibly arise from something with the structure of the micro-level brain. This is not an argument against Russellian monism, but a target for future research. In its stronger form, the structural mismatch problem takes the form of a deductively valid argument to the conclusion that Russellian monism is false. Chalmers expresses the stronger form as follows:

**The Structural Mismatch Argument**

1. If constitutive Russellian monism is true, micro-(proto)phenomenal structure is isomorphic to micro-material structure.
2. If constitutive Russellian monism is true, micro-proto(phenomenal) (and micro-material) structure constitutes o-phenomenal structure.
3. Micro-material structure constitutes only macro-material structure.
4. If micro-(proto)phenomenal structure is isomorphic to micro-material structure, then any structure constituted by micro-(proto)phenomenal structure (and micro-material structure) is isomorphic to a structure constituted by micro-material structure.
5. O-phenomenal structure is distinct from macro-material structure.
6. Constitutive Russellian monism is false.\(^ {13}\)

---

\(^{11}\) Chalmers 2002: 266.


\(^{13}\) I have framed the argument using my terminology.
The emergentist Russellian monist is entirely untouched by the structural mismatch problem. According to emergentist, o-experience is causally brought about by micro-(proto)experience, and there is no reason to think that an effect need be structurally isomorphic with its cause. Compare: I was causally brought into being by my parents, but my shape and size are quite different to that of my parents. To the extent that the structural mismatch problems raises trouble for the constitutive Russellian monist, to that extent the emergentist Russellian monist has the upper hand.

Premise 1 may be questioned. Physics captures the causal structure of micro-(proto)phenomenal properties, but the categorical nature of these properties may involve structure that goes beyond their causal structure. More importantly, we have good reason to think that premise 5 is false. This is because we have good reason, as Chalmers is well aware, to think that the structure of o-consciousness is captured in information states realised by the brain. Indeed, in the article and then book on which he made his name,\textsuperscript{14} Chalmers put this forward as a principle, the \textit{Principle of Structural Coherence}, which can be used as a constraint when trying to formulate a theory of consciousness:

\begin{quote}
Whenever there is conscious experience, there is some corresponding information in the cognitive system that is available in the control of behaviour, and available for verbal report....the geometrical structure of the visual field [for example] is directly reflected in a structure that can be recovered from visual processing. Every geometrical relation corresponds to something that can be reported and is therefore cognitively represented. If we were given only the story about information-processing in an agent’s visual and cognitive system, we couldn’t \textit{directly} observe that agent’s visual experiences, but we could nevertheless infer those experiences’ structural properties.\textsuperscript{15}
\end{quote}

The basic point is that a person is able to report on the structure of their conscious experience, which seems to imply that that structure is represented in the informational structures of that person’s brain.

\textsuperscript{14} Chalmers 1995/1996.
\textsuperscript{15} Chalmers 1995.
Appreciating this we can see that the ‘structural mismatch’ problem is at best misnamed. There is no structural mismatch, as the relevant structures are there in the brain. But a challenge remains, and that is to explain why phenomenology mirrors abstract and highly specific informational states which don’t seem notable from the perspective of physics. Chalmers is alert to this way of understanding the difficulty:

From the perspective of physics, high-level information structure are derivative aspects of a more encompassing and more basic [macro-material]...structure. It is not easy to see why, on a constitutive Russelian view...[o-phenomenal] properties should have this structure rather than the more encompassing and more basic structure....there should certainly be [o-phenomenal] structure corresponding to [the more basic macro-material]...structure....Why should we expect there also to be...[o-phenomenal] structure corresponding to information structure? 16

We can make the point clearer with an analogy. All green objects have the property of being grue, where something is grue just in case it is either green before 2020 or blue after 2020. All blue objects after 2020 have that property too. That property is there, grounded in the greenness of the objects, but it is a highly abstract, rarefied kind of property. It would be weird if some fundamental or significant truth about the universe depended on the distribution of grueness throughout it. Similarly, although there are informational states corresponding to the structure of o-consciousness, those information states seem, at least from the third-person perspective, to be highly abstract and rarefied. It seems odd that an important property such as o-consciousness should correlate with states of this kind.

How should we respond to this difficulty? Luke Roelofs argues for the coherence of the following line: o-experience really does have the incredibly complex structure we might expect it to have given its grounding in the micro-level structure of the brain, but confusion (in the technical sense explicated in the above discussion of the palette problem) renders us unaware of much of that structure. We are unable to think about the components of our experiences in isolation, and thus we are blinded to their complexity. But this approach is inconsistent with Phenomenal Transparency. In so far as I can form direct phenomenal

16Chalmers Forthcoming b.
concepts of my o-conscious states, those concepts accurately represent the essential nature, and thereby the structure, of those states.

What are we to make then of the fact that our o-conscious states share the structure of highly rarefied information states, rather than the structure of more natural and fundamental macro-level brain states? I think the moral to be drawn is that o-phenomenology is not the unique or privileged kind of property we ordinarily take it to be, but is only one of many forms of consciousness in the macro-level brain, and not an especially interesting one. If we suppose that there is a vast multiplicity of kinds of consciousness corresponding to vast multiplicity of structures in the brain, the mystery as to why there is a form of consciousness mirroring a seemingly quite arbitrary brain-structure disappears; many macro-level brain structures correspond to phenomenology, and the information states corresponding to o-consciousness are just one amongst this many. If we suppose that there is one, or few, kinds of consciousness at the macro-level of the brain, it is extremely odd that there should be consciousness corresponding to information states. This given us strong reason to suppose that there are many forms of consciousness instantiated in the macro-level brain.

One might object that surely our conscious experience is special and privileged. Certainly o-consciousness is special for us. My o-consciousness is the form of consciousness which corresponds to the informational states which govern the overall behaviour of my body. It is not surprising that this is the kind of consciousness which gets to call itself ‘my consciousness’, and which other people think of as such. O-consciousness is special from the perspective of those interested in human society. But from the God’s eye view it may be a highly arbitrary and uninteresting slice from a world teeming with phenomenology. We have got used to the Copernican idea that we are not at the centre of the universe but simply one not particularly exciting planet amongst many. Perhaps it’s time for a Copernican revolution about our own consciousness.

Of course there is always the response: ‘But it’s really weird to think that there is so much consciousness.’ A similar ‘it’s just too weird’ response could be given to the theories of Darwin or Einstein or Heisenberg. The rational metaphysician values theoretical virtue over common sense, and the above considerations demonstrate that the supposition that there
is only one mind in the macro-level brain, although it’s what we ordinarily think, is too improbable to be taken seriously.\footnote{The problem of the many (Unger 1980) also provides us with a reason to think that there are many conscious subjects involved in the brain.}

**The Subject Irreducibility Problem**

In the last chapter we considered the subject-summing problem as a bottom-up problem: from what we know about the nature of micro-subjects, is there any reason to think that they cannot sum to form a composite subject? When considered in this way, I don’t think we find any conclusive arguments against the possibility of subject-summing. However, we will now consider the possibility of mental (or proto-mental) composition starting with the allegedly composite subject itself, i.e. with the human conscious mind. Our question is: Is there anything we know about the human conscious mind which rules out its existence being grounded in micro-level facts? When the issue is posed in this way, I believe we do meet insuperable challenges. The problem is that we have very good reason to think that subjects are irreducible.

I will use the term ‘Subject Irreducibility’ to denote the following thesis: *What it is for there to be a conscious subject $S$ cannot be analysed into facts about some other individual or individuals, none of which need be identical with $S$. More precisely, it is the thesis that there are no deflationary analyses of subject-hood*, i.e. no true analyses of the following form:

*General Form of a Deflationary Analysis of Subject-Hood* – For it to be the case that there is conscious subject $X$ is for it to be the case that there are $Y$s that are $F$, where $X$ need not be one of the $Y$s.

To make this clearer, contrast with the case of parties, which do admit of deflationary analysis:

*Deflationary Analysis of Party-Hood* – For it to be the case that there is a party is for it to be the case that there are people revelling.

A person is not a party; hence in the above analysis we define what it is for a given party to exist in terms of things which are not that party. Subject Irreducibility tells us that the same
cannot be done with subjects: we cannot analyse what it is for a given subject to exist in terms of things which are not that subject.

Physicalists have ways of avoiding Subject-Irreducibility. Analytic functionalists would offer the following a priori analysis:

*Analytic Functionalist Deflationary Analysis* – For it to be the case that there is conscious subject X is for it to be the case that there are Ys that perform such and such a functional role.

The conscious subject itself will not be among the Ys; they will ultimately be certain particles in the body and brain of the conscious subject acting in concert. Thus, what it is for the conscious subject to exist is analysed in terms of things which are not that subject. Type-B physicalists are likely to hold that there is a deflationary analysis of subject-hood, in either physical or functional terms, but that it is only available a posteriori.

Of course these possibilities have been ruled out by options already defended in the first half of the book. The epistemic gap – defended in chapter 3 – rules out analytic functionalism, and Direct Phenomenal Transparency – defended in chapter 5 – rules out a posteriori physicalism. Furthermore, as I explained in the last chapter, Direct Phenomenal Transparency entails that the analysis of subject-hood is a priori accessible. This is because the property of being a conscious subject is a determinable of which any conscious state is a determinate; feeling pain, for example, is a specific way of being a conscious subject. And 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.

According to Direct Phenomenal Transparency, I am able to grasp the essence of 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.

So much for physicalist proposals for deflationary accounts of subject-hood. Can Russellian monists fare any better? It is important to bear in mind that we are not looking merely for a *reductive analysis* of subject-hood, which could be of the following form:
**General Form of a Non-Deflationary but Reductive Analysis of Subject-hood** – For it to be the case that there is a conscious subject x is for it to be the case that x is F, where ‘F’ expresses a property non-identical with the property of being a conscious subject.

If we were merely looking for a reductive analysis of subject-hood, the proposal that consciousness can be analysed into subjectivity and qualitivity would suffice:

**Sample Reductive Analysis of Subject-Hood** – For it to be the case that there is a conscious subject x is for it to be the case that there are certain qualitative contents which are subjectively presented to x.

But such an analysis is not deflationary as the bearer of consciousness has not been analysed away. Recall that we are looking for is an analysis of the following form:

**General Form of a Deflationary Analysis of Subject-Hood** – For it to be the case that there is conscious subject X is for it to be the case that there are Ys that are F, where X need not be one of the Ys.

What kind of analysis could the constitutive Russellian monist offer? It would natural for the panpsychist to offer something like the following:

**Panpsychism Analysis** – For there to be conscious subject X is for certain subjects, S₁, S₂...Sₙ, (none of which is identical with X) to bear relation R to each other.

The first problem with such an analysis is that it’s hard to see how we might complete the proposal by say what relation R is. In the last chapter we speculated that there might be some unknown relation – the deep nature of the spatial relation – which somehow bonds subjects such that they produce a further subject. However, assuming Phenomenal Transparency, we are looking for a relationship which is *available a priori*. There does not seem to be such a relation available upon reflection.

More generally, if the analysis of subject-hood is a priori then it must in some implicit sense be what we *mean* when we judge that there is a conscious subject. What is it for me to judge that there is a conscious subject, call it ‘Jane’? It is simply not plausible that my judgment that Jane is a conscious subject consists in the judgment that there are a large
number of micro-subjects, none of which is identical with Jane, bearing certain relations to each other.

The panpsychist might try to avoid this concern by giving an analysis not in terms of micro-subjects themselves, but in terms of the experiences of micro-subjects. What relation between experiences ties them together into a subject? Barry Dainton has proposed the co-consciousness relation, the relation two experiences bear to each other when they are experienced together.\(^\text{18}\) Following this line of thought, we might be led to the following analysis:

\[
\text{Co-Consciousness Analysis (1st formulation) – For it to be the case that there is conscious subject X is for it to be the case that there are certain experiences which are co-conscious with each other.}
\]

The panpsychist could then hold that a large number of the experiences had by the micro-subjects making up brain bear the co-consciousness relationship to each other, and it is in virtue of this that my conscious mind exists.

This proposal analyses subjects into experiences. But what is an experience? Following Martina Nida-Rümelin, I am inclined to think that phenomenological reflection reveals an experience to be nothing other than an event of a certain subject bearing certain experiential properties.\(^\text{19}\) Consider for example my experience of pain. All it is for that experience of pain to exist is for it to be the case that I feel pain of a certain kind at a certain time. If this is correct, then experiences are analysed in terms of subjects, which rules out the above analysis of subjects into experiences.

Perhaps the proponent of the co-consciousness analysis could try the following:

\[
\text{Co-Consciousness Analysis (2nd formulation) – For it to be the case that there is conscious subject X is for it to be the case that there are certain experiential properties which are co-conscious with each other.}
\]

But what is it for certain experiential properties to be ‘co-conscious’? It is surely nothing other than for those properties to be experienced by a single subject. For the experiential

\(^\text{18}\) Dainton 2011.

\(^\text{19}\) Nida-Rümelin Forthcoming.
property of pain and the experiential property of anxiety to be ‘co-conscious’ is for there to be a single subject which both feels pain and feels anxiety. This turns our analysis into the following:

Co-Consciousness Analysis (3rd formulation) – For it to be the case that there is conscious subject X is for it to be the case that there are certain experiential properties which are had by a single subject.

The problem of course is that the ‘single subject’ can be nothing other than X itself. We have failed to analyse away the bearer of subject-hood, and hence failed to give a deflationary analysis of subject-hood. These panpsychist options seem prima facie plausible, but on deeper reflection turn out to be either incoherent or not really deflationary.

The problem for the Russellian monism is that no other options seem available. Once we have set aside functionalist analyses, there seems to be no other options for an a priori accessible deflationary analysis of subject-hood. What is it for Jane to exist as a conscious individual? An initial answer: it is for Jane to exist and to have experience, for there to be something that it’s like to be Jane. Can we analyse this without quantifying over Jane? A functionalist could say that it for there to be certain particles acting in concert to play such and such a functional role; this is all that is essentially required for there to be conscious Jane. But once we reject functionalism there just doesn’t seem to be any way of a priori analysing what it is for there to be conscious Jane without quantifying over Jane.

Maybe we just haven’t thought of the correct deflationary analysis yet? Even if one has a transparent concept of a property, certain implications of its nature can be difficult to derive. It is pertinent at this point to consider Arnauld’s famous objection to Descartes:

Suppose someone knows for certain that the angle in a semi-circle is a right angle, and hence that the triangle formed by this angle and the diameter of the circle is right-angled. In spite of this, he may doubt, or not yet have grasped for certain, that the square on the hypotenuse is equal to the squares on the other two sides; indeed he may even deny this if he is misled by some fallacy.

---

20 This argument is influenced by criticisms of co-consciousness solutions to the combination problem in Coleman 2016.
Perhaps the deflationary analysis of subject-hood is subtle and difficult to discern, just as Pythagoras’s theorem is a not easily discernible implication of the essence of trianguarity. Perhaps Russellian monists can look forward to the day when the ‘Pythagoras of consciousness’ works it out.

However, what is being demanded here is not some subtle implication of the essence of subject-hood, but simply what is essentially required for there to be a conscious subject (for there to be an x such that x is a conscious subject). A more appropriate analogy would be to suggest that someone could have a transparent concept of a triangle without knowing that for there to be triangle is for there to be an object with three sides.

Of course there are many cases in which we have an implicit grasp of what is required for property P to be instantiated, and yet find it very difficult to fully articulate that requirement as necessary and sufficient conditions for there to be a token of P. Consider for example the ongoing travails of philosophers trying to give a precise definition of the everyday concept of knowledge. If several decades of our greatest conceptual analysts beavering away cannot give us necessary and sufficient conditions for the instantiation of knowledge, something of which we plausibly have a transparent concept, then maybe it’s not so implausible to think that Phenomenal Transparency is consistent with our failure to fully articulate the essence of subject-hood in such a way as to reveal its deflationary essence.

However, whilst it may be unreasonable to demand of the constitutive Russellian monist precise necessary and sufficient conditions for the instantiation of subject-hood, it is not unreasonable to demand some kind a gesture towards what is required. When it comes to those properties of which we have a transparent concept, although we generally find it hard to express what is required for there to be a token of the property as precise necessary and sufficient conditions, we can always formulate some rough and ready approximation of what is required. For there to be knowledge is roughly for there to be true and justified belief. For there to be a party is roughly for there to be people gathered together having a good time. There are of course counterexamples to these rough definitions, but they serve to communicate the idea, and in doing so give us a sense of how the entity might be grounded in more fundamental features of reality. On the basis of such examples, Theodore
Sider argues that a reductionist is obliged to offer a ‘toy’ reductive analysis, to give credence to our belief that a fully spelt out reductive analysis is available to God if not to us.  

However, in the case of subjectivity, we’re not able even to gesture at its supposed deflationary analysis. Perhaps this is not a knock down argument for Subject-Irreducibility, but the fact that sustained and careful consideration yields no sign of a non-functionalist a priori accessible deflationary analysis of subject-hood, make the thesis that there is one seem like an unsupported and implausible leap of faith.

Thus, I think the following argument is a powerful argument for Subject Irreducibility:

Premise 1 – If the analysis of subject-hood is a priori, then it is deflationary only if either analytic functionalism is true.

Premise 2 – Analytic functionalism is false.

Premise 3 – The analysis of subject-hood is a priori (implied by phenomenal transparency).

Conclusion – The analysis of subject-hood is not deflationary, i.e. Subject Irreducibility.

Why does it matter that subjects are irreducible? In chapter 6 we defined constitutive Russellian monism in terms of constitutive grounding, and decided that – unless and until we could find some other way of satisfying the Free Lunch Constraint – we would be understanding constitutive grounding in terms of grounding by analysis. The details of grounding by analysis haven’t been important until now, so let us remind ourselves of them:

Fact X is grounded by analysis in fact Y iff:

- X is grounded in Y, and
- Y logically entails what is essentially required for the entities contained in X to be part of reality.

To take a concrete example, the fact (F1) that there is a party is grounded by analysis in the fact (F2) that Rod, Jane and Freddy are revelling because:

---

22 Sider 2012: 117-18. Sider puts the point in terms of the notion of ‘metaphysical truth conditions’, which is central to his framework, but the substance is the same. I show in Goff 2015 and Goff 2016b how the argument of this section can be put in terms of Sider’s framework.
• F1 is grounded in F2, and
• The fact that Rod, Jenny and Freddy are revelling logically entails what is essentially required for a party to exist, i.e. that there are people revelling.

Now if we cannot analyse what it is for conscious Jane to exist without quantifying over conscious Jane, then the only way the micro-level facts could logically entail what it is essentially required for the reality of conscious Jane, is by directly logically entailing the existence of conscious Jane. In other words, the reality of conscious Jane is grounded by analysis in the micro-level facts only if the non-existence of conscious Jane is logically inconsistent with the obtaining of the micro-level facts.

The trouble is that mereological nihilism – the view that there are no objects with proper parts – is logically coherent. There is no contradiction in the assertion that there are particles arranged table-wise, planet-wise, etc., but there are no tables, planets, etc. And thus, assuming that o-subjects are macro-level entities, we can pose the following argument against constitutive Russellian monism:

**The Subject Irreducibility Argument against Constitutive Russellian Monism**

**Premise 1** – The fact that human conscious subject Jane exists is grounded by analysis in the micro-level facts only if the micro-level facts logically entail what is essentially required for Jane to exist.

**Premise 2** (Subject irreducibility) – There are no true analyses of the following form: For it to be the case that conscious subject X exists is for it to be the case that there are Ys that are F, where X need not be one of the Ys.

**Conclusion 1** – The fact that Jane exists can be grounded by analysis in the micro-level facts only if the micro-level facts logically entail the existence of Jane.

**Premise 3** – Jane is a macro-level entity, and hence her non-existence is logically consistent with the complete micro-level facts.

**Conclusion 2** – The fact that Jane exists is not grounded by analysis in the micro-level facts.

Why doesn’t this form of argument apply quite generally to rule out the grounding of any macro-level entities? Or to put it another way, why does it not follow from the coherence of mereological nihilism that facts about micro-level entities never ground the existence of
macro-level entities? Because in general the essences of macro-level entities are rich enough to account for their grounding in the micro-level facts. It is plausibly in the nature of a table that all that is required for there to be a table is for a (very difficult to precisely specify) functional role to be realised, and it is plausible that this is logically entailed by certain facts concerning micro-level entities. Thus, even though the micro-level facts in and of themselves do not logically entail the existence of tables, the micro-level facts logically entail a certain condition C, such that table-hood is essentially such that C is all that is essentially required for there are tables. Once we ‘unpack’ the nature of the grounded entities, we secure the logical entailment. The problem in the case of conscious subjects is that, if Subject Irreducibility is true, the nature of subject-hood is not rich enough to do ‘reach out’ to the micro-level facts in this way.

I am thus inclined to think that mental combination is indeed impossible, but that the impossibility is apparent only when we approach the combination problem top-down rather than bottom-up. When the focus is on the work stuff at the bottom has to do, the noumenalist panprotopsychists may declare: ‘We have no idea what the nature of micro-level proto-phenomenal properties are. How on earth could we rule out that they have some weird and wonderful nature which somehow produces consciousness?’ Likewise, Russellian panpsychists may announce: ‘We have no clue as to the deep nature of spatial relations, so how could we possible know that their nature is not sufficient to bind micro-subject into o-subjects?’ Fair enough, as far as it goes. But attention to the details of grounding by analysis reminds us that the constituents of the grounded fact must also play their part in cases of constitutive grounds; it takes both sides of the grounding relation to produce an ontological free lunch. And given Phenomenal Transparency, it is not an option to be noumenalist about the grounded reality of subject-hood; we cannot get rid of the problem by saying ‘We have no clue as to the nature of subject-hood, and hence no way of ruling out it that it admits of deflationary analysis.’

None of this threatens emergentist forms of Russellian monism, as the emergentist takes conscious subjects to be fundamental and unanalysable entities which causally arise from

---

23 Consciousness may be an aspect of a more expansive property, a view we will discuss at length in the next chapter. But this does not remove the need to account for the grounding of the experiential aspect of that more expansive property, and the bearer of that experiential aspect qua bearers of experience. And the non-deflationary of subject-hood renders it impossible to do this via grounding by analysis.
the narrowly material. However, this is an extremely powerful challenge against constitutive Russellian monism, and I can’t see how the forms we have so far considered can survive it. We need a metaphysical picture which can accept that conscious subjects are irreducible whilst at the same time make sense of their being non-fundamental, grounded in more fundamental features of reality. Fortunately, I think there is metaphysical view which is able to square this circle. This will be the subject of the next chapter.