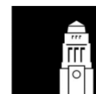

This is a repository copy of *Combining Minds: How to Think about Composite Subjectivity*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/195898/

Version: Accepted Version

Article:

Leidenhag, J orcid.org/0000-0002-1164-7032 (2022) Combining Minds: How to Think about Composite Subjectivity. Journal of Consciousness Studies, 29 (5-6). pp. 207-212. ISSN 1355-8250

https://doi.org/10.53765/20512201.29.5.190

© Imprint Academic. This is an author produced version of an article published in Journal of Consciousness Studies. Uploaded in accordance with the publisher's self-archiving policy.

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

Luke Roelofs
Combining Minds: how to think about composite subjectivity
Oxford: OUP, 2019
ISBN:9780190859053
Reviewed by Joanna Leidenhag, University of Leeds

Luke Roelofs' *Combining Minds* is one of the most ambitious and exciting books on philosophy of mind I have read. Roelof's goal is to challenge an almost universal orthodoxy in Western philosophy of mind; anti-combinationism, the intuition that two subjects cannot become one and one subject cannot be made up of many composite subjectivities. Roelofs tackles anti-combination at multiple levels of scale; he considers electron-sized subjects combining to form a unified larger subject, describes brains as combinations of hemisphere-subjects (or other brain regions), and uses science-fiction thought experiments to imagine how social groups might form a unified single consciousness. Whether for philosophers committed to anti-combination as a premise on which to build arguments, or scholars who desire to articulate some form of mental combination, Roelofs arguments are clear and worth engaging.

The structure of the book works on two levels with four Divisions, each divided into two Chapters, designed to overcome what Roelofs takes to be the two main obstacles facing combinationism. Division 1 (chapters 1 and 2) provides a lucid overview of the current literature, definition of key terms, and an outline of Roelofs' argument, with particular attention to recent claims for and against panpsychism (defined as the "the view that consciousness Is omnipresent among the fundamental things of the universe: all matter Is conscious" (p.14)). These first two chapters lay an invaluable foundation for later constructive developments.

The first obstacle is that the key terms under discussion are slippery, and interlocutors can unintentionally move the goalposts by pivoting between different views on the scope of 'subjects', 'wholes', and 'unities'.

Sometimes we talk about 'subjects' in a metaphysical way, meaning that which supports consciousness, (i.e., the brain, or bodily organism, or the soul), and at other times we talk about 'subjects' in a psychological way, referring to a persona or sense of personal identity (this way of talking about subjects is often associated with the 17th century philosopher John Locke, and so is often referred of as neo-Lockean). 'Wholes' can be mere aggregates, like a pile of sand, and these are division-invariable, which means that no matter how you divide a pile of sand all you get is, well smaller a pile of sand; nothing much changes. On the other hand, some 'wholes' are structure-specific and division-variable like a car; it matters that a car's parts are shaped and connected in a certain way if you want the car (as a whole) to function in a certain way. Unlike a pile of sand, if you cut a car in half, it's not really a car anymore.

Roelofs discusses five types of 'unity' relating to consciousness; (i) "global consistency" (e.g., when a persons beliefs and perceptions are internally consistent), (ii) "causal interdependence" (e.g., when a new perceptive causally impacts a corresponding belief), (iii) "access-unity", (the ability to access or recall beliefs, perceptions, memories, feelings and experience them as 'mine'), (iv) "representational unity", (e.g. that we experience redness and squareness as a single

red square), and (v) "phenomenal unity," (that there is 'something it is like' to have a combined experience, that is not quite captured in what it is like to have either single experience in isolation, e.g. enjoying candles and music whilst soaking the bath) (pp.45-48). The pivotal question how these different forms of unity relate to one another. If two subjects are unified in one way, are they automatically unified in every other way? In particular, Roelofs shows that some views take 'phenomenal unity' to be primitive, which means that it can't be further analysed by breaking it down into parts and it doesn't come in matters of degree - you either have a phenomenally unified experience or you don't. However, others think that phenomenal experience Is not primitive, but is the result of increasing the forms of unity described in (i)-(iv), and so does increase or decrease by matters of degree. The problem for 'subjects', 'wholes' and 'unity' is that sometimes philosophers are assuming one meaning, but their Interlocutor Is assuming another, and so they talk past one another.

Divisions 2-4 tackle this obstacle by stabilizing these terms according to different competing philosophies of mind, resulting in four separate theories of subject combination.

Division 2 develops a theory of combinationism for constitutive panpsychists, who believe that subjects are substrates of experiences which combine to form division-invariable aggregates and that phenomenal unity is primitive. Division 3 develops two theories of combination regarding intelligent subjects, one for panpsychist-functionalists building on the previous chapter and a separate theory for pure functionalists. In both these, functionalist accounts of combined subjects are substrates of experience, wholes are division-variable as structure-specific, and for pure functionalists (but not panpsychist-functionalists) phenomenal unity is reducible to other forms of unity. Division 4 develops a theory of psychological combination; the combination of 'persons' with self-consciousness and agency, where subjects are taken to be (neo-Lockean) personas. One of the most ambitious aspects of this book is Roelofs' suggestion that the different theories of consciousness and combination described in these separate Divisions need not be rivals, "they could be viewed as allies, illuminating different and complementary aspects of reality." (p.296).

By offering readers not one, but four theories of subject combination, Roelofs has quadrupled the work required to overcome the second obstacle. The second obstacle, as Roelofs describes it, is that there are three distinct types of problems that all theories of combination need to deal with. Drawing on categories from Sam Coleman and Thomas Nagel, Roelofs refers to these as "internal problems," with the very possibility of mental combination, "bridging problems," arising from contingent and empirical facts about the kinds of parts and wholes under discussion, and "lack-of-theory problems," which is just the problem that we don't know how to think about composite subjectivity (p.32).[1] Roelofs uses chapters 3, 5, and 7 to dissolve the internal problems. He then uses chapters 4, 6 and 8 to sketch possible ways forward for empirical investigations to make progress on the bridging-

[1] Sam Coleman, "Panpsychism and Neutral Monism: How to Make Up One's Mind," *Panpsychism: Contemporary Perspectives* (eds.) G. Brüntrup and L. Jaskolla (Oxford: Oxford University Press, 2017), 250-254; Thomas Nagel, *The View from Nowhere* (Oxford: Oxford University Press, 1986), 50.

problems. Throughout each division Roelofs responds to the no-theory problem by imagining what it might feel like for the various parts and wholes in these different relationships of combination. Roelofs occasionally makes reference to neuroscientific discoveries or ideas (particularly, in his discussion of Integrated Information Theory, split-brain patients, and his mention of the Hogan twins who are fused at the skull), but the vast majority of his arguments are the result of conceptual analysis and thought experiments with the express purpose of trying to imagine how subject combination could work (but I'm sure Roelofs would jump at the chance to have these ideas empirically tested). To give readers some sense of how this goes, I will sketch the picture of consciousness and combination that Roelofs offers us in *Combining Minds*.

Constitutive panpsychist see matter as conscious, so they want to paint a picture of the universe where the description of the mental closely follows the description of the physical. Therefore, at the fundamental level Roelofs sees the universe as not only a great sea of mass/energy, as physics describes, but also an ocean of phenomenality where there is no clear way to divide and distinguish subjects. Parts and aggregate-wholes overlap and share their phenomenal consciousness wherever there is any kind of physical interaction, because subjectivity only entails weak privacy and weak independence (experience-sharing and grounding does not demand that subjects are identical, only that they overlap).

But what kind of experiences could electrons or quarks possibly have? Obviously, we can't know what it like to be an electron (and we probably will never know this, I'm not even 100% if God know this), but Roelofs suggests that all he needs to do to ward off the implicit objection in this question is show that there is some conceptual space between the 'ceiling' of how far we attribute complex experience to it, and the 'floor' of the explanatory challenge posed by consciousness (p.77). His best guess is that the experiences had by fundamental subject(s) has the quality of contrast: a "something" against the backdrop of "something else", and "an internal motivating power" that, for instance, "motivates the electron to move blindly toward or away from the source of the force it is feeling, in virtue of the way that it feels." (p.78).

Fundamental subject(s) are not, however, *intelligent* subjects. An intelligent subject is "A conscious subject whose consciousness is structured and whose experiences play the functional roles characteristic of intelligence in virtue of their phenomenal character, conscious structure, and representational content." (pp.151-152). Roelofs considers how unintelligent (fundamental) and/or intelligent subjects combine into structure-specific wholes with a measure of control and sensitivity (intelligent functioning) and structured consciousness. Chapter 6 provides discussion of three possible examples of composite intelligent subjects: the 'Nation-Brain' thought experiment, split-brain patients, and ordinary dual-hemisphere brain functioning.

Division 4 explores combinationism for person(a)s, with personal identities, psychological continuity, self-consciousness, and agency. The basic claim is "that when the wills of two subjects are related in the right way, they will each experience the other not as distinct but as an extension of themselves, and the actions of both will be attributable to each." (p.230). Roelofs uses analogies to social groups, the

human experience of inner conflict, and the phenomenon of dissociative identity disorder to motivate the idea that persons might fuse. Hypothetically, two fused persons gradually stop experiencing their interaction as interpersonal and come to experience it like that which holds within a single person's mind. There is no 'third' mind that is created out of their cooperation, but the consciousness of the parts and whole form two sides of the same coin.

Let me conclude with some more general remarks. Roelofs book is aimed at a scholarly audience (I would love to see him attempt a popular level version!). It is fun and exceptionally lucid, but I would only recommend it to postgraduate students and scholars (I found Chapter 5 the hardest to understand, where Roelofs argues that to form a structure-specific intelligent subject, each part's phenomenal field is "superimposed" onto the others through a process called "phenomenal binding", which to the external observer probably looks like integrated information and on the inside feels like co-presentation).

The arguments that will stay with me longest are the various places where Roelofs shows us where philosophy of mind has developed some unhelpful dichotomies between perfectly compatible positions such that we may have been wasting our argumentative efforts; minds as substrates *or* personas, that a philosopher is a panpsychist *or* a functionalist, and a panpsychist is an atomistic panpsychist *or* holistic cosmopsychist (for the latter, see p.168). Often Roelof's reveals the arbitrary nature of such divisions by maintaining a healthy balance between metaphysical and pragmatic concerns; sometimes it is helpful to see a subject as a persona and sometimes as a substrate; sometimes a split-brain patient appears to be one person, and at other times two. In such scenarios, combinationism means we don't have to come down definitely on one side or the other.

Many different types of philosophers and other scholars will find something in these arguments for them, and plenty to disagree with. But Roelofs' aim "has not been to persuade the reader that these theories are true (though I think they are); it has been to use these theories to undermine anti-combination by showing multiple, consistent, plausible ways that minds might combine." (p.296). In this more modest task, at least, this reader finds *Combining Minds* a success.