

Consciousness, Revelation, and Confusion

Are Constitutive Panpsychists Hoist
by their Own Petard?

LUKE ROELOFS

[doi:10.48106/dial.v74.i1.04](https://doi.org/10.48106/dial.v74.i1.04)

Luke Roelofs. 2020. "Consciousness, Revelation, and Confusion: Are Constitutive Panpsychists Hoist by their Own Petard?" *Dialectica* 74(1): 61–93.
[doi:10.48106/dial.v74.i1.04](https://doi.org/10.48106/dial.v74.i1.04).



Consciousness, Revelation, and Confusion

Are Constitutive Panpsychists Hoist by their Own Petard?

LUKE ROELOFS

Critics have charged constitutive panpsychism with inconsistency. Panpsychists reject physicalism for its seeming inability to explain consciousness. In making this argument, they commit themselves to the idea of “revelation”: that we know, in some especially direct way, the nature of consciousness. Yet they then attribute properties to our consciousness—like being constituted out of trillions of simpler experiential parts—that conflict with how it seems introspectively. This seems to pose a dilemma: either revelation is false, and physicalism remains intact, or revelation is true, and constitutive panpsychists are hoist by their own petard. But this is too simplistic. Constitutive panpsychists can say that our minds contain innumerable phenomenal states that are “confused” with one another: immediately present to introspection only en masse, not individually. Accepting revelation does not require ignoring the attentional, conceptual, and interpretive limitations of introspection, and these familiar limitations remove the tension between panpsychism and revelation.

What is the relationship between being conscious and knowing about consciousness? In answering this question, constitutive panpsychists face a delicate balancing act: their own case against physicalism requires that being conscious reveals something of the metaphysics of consciousness, but the stronger they make this claim of revelation, the stronger becomes an objection to their own view sometimes called “the revelation problem”. In this paper I argue that this balancing act, though delicate, is not impossible: there is a plausible, well-motivated “medium-strength” sort of revelation, strong

enough to bring down physicalism but weak enough to leave constitutive panpsychism standing.

In section 1, I lay out the background to the panpsychism-physicalism debate; in section 2, I distinguish six “revelation theses”; in section 3 I analyse the structure and varieties of the revelation problem; and in section 4 and section 5 I outline how to address this problem while retaining as much as possible of the theses discussed in section 2.

1 Are Panpsychists Hoist by their Own Petard?

Panpsychists think all the fundamental physical things are phenomenally conscious, where “fundamental physical things” is a placeholder for whatever fundamental entities feature in the true physical theory (particles, fields, strings, spacetime, etc.). The “constitutive” part of “constitutive panpsychism” describes the relationship between macroexperiences (the experiences of humans and other animals) and the postulated microexperiences of the fundamental physical entities.¹ This relationship should be something like the relationship between the physical features of human bodies (macrophysics) and the physical features of the fundamental entities (microphysics). That relationship (which we might call being constituted, being grounded, or being nothing over and above) generates no “explanatory gap”: even when the details currently elude us, it seems clear that macrophysics is fully accounted for by microphysics. When you have the right particles, arranged in the right pattern, exerting the right forces on one another, and the right laws governing them, there is no further problem about how to get hands, chairs, planets, etc.: those “come for free” when the microphysical foundations are there.

The failure of consciousness to fit into this neat picture is the objection to physicalism that motivates most contemporary panpsychists. Whereas the distribution of and relations among subatomic particles seems to explain everything about my body, it leaves unexplained why there is anything it feels like to be me, and why it feels the particular way it does. In particular, even

1 Some panpsychists would not link “macro” and “micro” (terms conveying size) with “human-like” and “fundamental” in this way. In particular, “cosmopsychists” think that the fundamental physical entity is the cosmos as a whole, which is (obviously) bigger than a human being, not smaller (see Gaudry 2008; Jaskolla and Buck 2012; Shani 2015; Nagasawa and Wager 2017; Goff 2017). Though I am sympathetic to cosmopsychism, I do not believe that it changes the essential contours of the revelation problem, though it requires some re-formulating, as noted in footnotes 11 and 14. For now I will, for convenience, speak as though the fundamental physical entities are very small.

knowing the full story about the particles seems to be compatible with not knowing what the experiences are like (this is the “knowledge argument,” cf. Jackson 1982; Nemirow 1990; Ball 2009), and it seems that a world might have been physically identical and yet differed from ours in respect of consciousness (the “conceivability argument,” cf. Kripke 1980; Chalmers 2009). There is a vast literature on whether these are good reasons to reject physicalism (see, e.g. Chalmers 1996; Dennett 2007; Stoljar 2006; Díaz-León 2011), but here I will assume that they are. What comes next? In particular, is constitutive panpsychism, often offered as an attractive non-physicalist alternative, defensible?

Constitutive panpsychism treats consciousness as a fundamental ingredient of nature, but tries to treat it the same as other fundamental ingredients (mass, charge, spin, force, location, etc.). Just as those other fundamentals are widespread in nature, with human beings as simply one particular arrangement of them, so is consciousness: human experience is not metaphysically special, just a complicated combination of widespread components. Constitutive panpsychism thus retains the monistic spirit of physicalism despite recognising consciousness as fundamental. Importantly, non-constitutive versions of panpsychism, on which human consciousness somehow “emerges from” or is “caused by” microconsciousness but not literally “made up of” it, do not secure this advantage. The macrophysical properties of the brain seem to be wholly constituted by the microphysical properties of its parts, so if its macroscopic consciousness is not similarly constituted by microconsciousness, the hoped-for reconciliation of mind and matter falls apart.

This imposes an explanatory burden: constitutive explanations of human consciousness in terms of microconsciousness have to do better than physicalist explanations. And one major line of criticism has been that they do not: there is just as much difficulty in explaining how many simple minds combine into complex minds as in explaining how mindless things generate minds. This broad objection is often called “the combination problem” (Seager 1995, 280; Chalmers 2017; Roelofs 2019), and has received much discussion from both defenders and critics of panpsychism.

One specific strand of the combination problem is “the revelation problem”: macroexperiences do not *seem* introspectively to be built up out of microexperiences. And constitutive panpsychists can’t just say: “Well they *are*, sometimes things aren’t what they seem.” That would license physicalists to likewise say: “Exactly! Consciousness *seems* distinct from purely physical facts, but it’s actually not.” If being conscious doesn’t reveal the true nature of

consciousness, the case against physicalism is weakened; if it does, then the truth of constitutive panpsychism should be introspectively obvious, which it is not.

This talk of “seeming” and “obviousness” is not the most precise way of presenting things. Authors articulating the sense that there is a problem here say things like:

[...] it is hard to see how smooth, structured macroscopic phenomenology could be derived [from microexperiences isomorphic to microphysics]; we might expect some sort of “jagged,” unstructured phenomenal collection instead. (Chalmers 1996, 306)

It is hard to see how [microexperiences] could somehow add up to the phenomenal properties with which we are familiar—properties with the specific, homogeneous character with which we are all acquainted [...]. (Alter and Nagasawa 2012, 90–91)

[Revelation is] inconsistent [...] with my conscious experience turning out to be, in and of itself, quite different from how it appears to be in introspection: i.e. turning out to be constituted of the experiential being of billions of micro subjects of experience [...]. (Goff 2006, 57; cf. Lee 2019, 290–98)

Similar remarks were made by certain non-reductive mind-brain identity theorists in the last century, writing about a perceived “grain problem”:

[Any experience’s] physiological substrate, presumably, is a highly structured, not to say messy, concatenation of changes in electrical potential within billions of neurons in the auditory cortex [...]. How do all these microstructural discontinuities and inhomogeneities come to be *glossed over* [...]? (Lockwood 1993, 274)

How is it that the occurrence of a smooth, continuous expanse of red in our visual experience can [...] involve particulate, discontinuous affairs such as transfers of or interactions among large numbers of electrons, ions, or the like? (Maxwell 1978, 398)

Indeed, Lewis makes a very similar argument, though he rejects the idea that experience reveals its nature and so presents the argument as a *reductio* of this idea:

If we know exactly what the qualia of our experiences are, they can have no essential hidden structure - no “grain” - of which we remain ignorant. If we didn’t know whether their hidden “grain” ran this way or that, we wouldn’t know exactly what they were. [...] if nothing essential about the qualia is hidden, then if they seem simple, they are simple. (Lewis 1995, 142, fn. 14)

Although I think all the above quotations express a similar sort of concern, they do so with different emphasis and framing, and the exact nature of the problem involved is far from clear. In section 3 I try to identify the problems more precisely, and in section 4 and section 5, I resolve them.

2 The Revelation Problem and the Revelation Thesis

Before examining the revelation problem for panpsychism, we need to examine the background idea of a “revelation thesis” connecting consciousness to knowledge of consciousness. There are actually several different ideas under the broad heading of “revelation”: I will distinguish a total of six distinct revelation theses, resulting from a two-fold distinction permuted with a three-fold distinction.

The two-fold distinction concerns whether the claim says (a) that the full truth about consciousness will always be manifest (a “reality→appearance” direction of implication), or (b) that what is manifest about consciousness is always true (an “appearance→reality” direction of implication).² Claims of the first sort rule out any aspect of consciousness being “hidden” from us, while claims of the second sort rule out any sort of “illusion” about consciousness.

The three-fold distinction is about the topic of a revelation thesis - what kind of reality it connects with what kind of appearance. Putting things for now in reality→appearance terms, we can distinguish the claims:

² Byrne and Hilbert (2007, 77), draw this distinction for colour properties: they “treat Revelation as equivalent to the conjunction of two theses [...] SELF-INTIMATION [and] INFALLIBILITY”, with the former being reality→appearance and the latter appearance→reality.

1. That someone having an experience³ can know that they are presently having that token experience;
2. That someone having an experience can gain a special kind of understanding of that phenomenal property;
3. That this understanding reveals “the complete nature” of a certain type of experience.

The first thesis is sometimes called “self-presentation” or “luminosity”, as distinguished from “revelation” (Stoljar 2006, 223). But in other discussions it is presented as an integral part of a broader idea called “revelation.” (e.g. Goff 2017, 109–10). The second thesis is sometimes put in terms of forming concepts, sometimes of special sorts (e.g. Chalmers 2003b; Goff 2017, 109–10) and sometimes just in terms of “understanding” (e.g. Stoljar 2006, 229). The third thesis is sometimes put in terms of knowing a phenomenal property’s “essence” or “nature”, or knowing all the essential or necessary truths about it.⁴ Sometimes the term “revelation” or “revelation thesis” is used specifically for one of these theses, or for the set of them together, or for the conjunction of the second and third. But they are worth distinguishing because, as I will show, they support quite distinct revelation arguments against constitutive panpsychism, which need to be addressed in quite different ways.

Moreover, we can distinguish reality→appearance and appearance→reality directions of each of the three, yielding a total of six revelation theses (RT1–RT6), as follows:

Topic	Reality → Appearance direction	Appearance → Reality direction
-------	--------------------------------	--------------------------------

-
- 3 Different authors speak variously of qualia, experiences, types of experience, and types of conscious state: for clarity I will in what follows speak of *phenomenal properties* as the things which phenomenal concepts capture, and whose natures they reveal, and of *experiences* as instantiations of phenomenal properties. To have an experience is to instantiate a phenomenal property, i.e. to be conscious.
 - 4 Some example formulations: the special understanding of an experience type we gain from undergoing it “reveals the essence of Q [the experience type]: a property of Q such that, necessarily, Q has it and nothing else does” (Lewis 1995, 141–42); “for every essential truth T about E, [the subject] knows, or is in a position to know, T” (Stoljar 2006, 228); “the complete nature of the type to which [the experience] belongs is apparent to the concept user” (Goff 2017, 110). Cf. also colour-revelation theses: “If it is in the nature of the colors that p, then after careful reflection on color experience it seems to be in the nature of the colors that p” (Byrne and Hilbert 2007, 77); “The intrinsic nature of canary yellow is fully revealed” (Johnston 1992, 223). Cf. Lee (2019, 291–93), Liu (2019, 2020).

Instantiation	Revelation Thesis 1: If someone instantiates a phenomenal property, it will introspectively seem to them that they are instantiating that property. (Call this the “luminosity” thesis.)	Revelation Thesis 2: If it introspectively seems to someone that they are instantiating a phenomenal property, then they really are instantiating that property. (Call this the “no illusions” thesis.)
Understanding	Revelation Thesis 3: If someone instantiates a phenomenal property, they will be in a position to form a pure phenomenal concept of it. (Call this the “understanding from experiencing” thesis.)	Revelation Thesis 4: If someone is in a position to form a pure phenomenal concept of a phenomenal property, they must be instantiating that property. (Call this the “no understanding without experiencing” thesis.)
Knowledge of nature	Revelation Thesis 5: If someone has a pure phenomenal concept, reflection upon it can reveal the whole nature of the corresponding phenomenal property. (Call this the “self-intimation” thesis)	Revelation Thesis 6: If someone’s reflection upon a pure phenomenal concept presents some feature as pertaining to the nature of the corresponding phenomenal property, that feature really does pertain to the nature of that property. (Call this the “infallibility” thesis)

I think these six theses, though logically independent, form a fairly natural package together, and I will refer to this package (i.e. the conjunction RT₁–RT₆) as “the revelation approach”.⁵ This package is particularly important for undergirding modal arguments against physicalism, a role which it is held to have both by its defenders and its critics (e.g. [Stoljar 2009, 2013](#); [Damnjanovic 2012](#); [Liu 2019, 2020](#)). Lewis, for instance, attributes RT₅ and RT₆ to Kripke, as a presupposition of the latter’s inference from the conceivability of pain without any associated brain state to their separate possibility

⁵ The component theses are often connected by the idea that subjects stand in a certain special relation of “acquaintance” to their experiences (see [Chalmers 2003b](#); [Goff 2015](#)): being directly acquainted with our experiences is what lets us know of their occurrence, and understand their properties in a way that fully reveals their nature. Acquaintance is often taken to be one species of a broader category of relations, called “awareness”, which likewise enable knowledge of various kinds, but which include more mediated forms of awareness like visual awareness, auditory awareness, etc. I am very happy to accept these claims about acquaintance and awareness, but they will not be distinctively important in the discussion that follows.

(Lewis 1995, 328, fn. 3). Goff (2017, 74–76, 96–106) likewise argues that the conceivability and knowledge arguments require that phenomenal concepts be “transparent”, effectively meaning that RT5 and RT6 must be true.⁶ And Chalmers’ version of the conceivability and knowledge arguments relies on the premise that the primary and secondary intensions of phenomenal concepts are equivalent (Chalmers 2003a, 2009), which implies RT5 and RT6.⁷

Although RT5 and RT6 have the clearest role, the falsity of the other revelation theses would also leave the anti-physicalist arguments on a shaky footing. For instance, if RT3 were false, we could worry whether we possessed the pure phenomenal concepts whose “transparency” drove the arguments; if RT2 were false, we could worry that the properties these concepts expressed were not even instantiated (as argued by, e.g. Pereboom 2016, 2019); and RT4 is essential to the knowledge argument, which relies on the premise that someone who has never experienced colour cannot know what seeing colour is like.⁸

3 What is the Revelation Problem, Exactly?

So what exactly is the supposed problem for panpsychists? How is it distinct from other aspects of the combination problem? Fundamentally, it concerns a perceived incompatibility between three things:

- the way human consciousness appears in introspection;

6 The arguments might not require going all the way to RT5 and RT6. Stoljar (2006, 229–30) suggests that all that is strictly required is that we have a form of access to the natures of phenomenal properties that allows us to know at least something, if not everything, about these natures. Goff argues against such an intermediate position, saying that for any property whose nature we grasp only part of, we can “split” the property into two components, one with an unknown nature and one with a known nature. The arguments against physicalism can then be run just with respect to “that aspect of phenomenal properties whose nature we know”, and for that sub-property RT5 and RT6 will be true. In this paper I will suppose that Goff is right, and seek to defend RT5 and RT6 in their “whole nature” form.

7 A concept’s primary intension is available to reflection, while its secondary intension is the nature of the property that concept expresses, so the coincidence of these two intensions implies that the natures of the properties expressed by pure phenomenal concepts are available to reflection by those who possess the concepts.

8 The revelation approach also comes up in other places. RT1, the “luminosity” thesis, is sometimes appealed to as a distinguishing feature of consciousness (Rosenthal 1993, 359; Kriegel 2009; Strawson 2015, 9). Other philosophers draw on RT1 and RT2 to develop an epistemology of introspection (Chalmers 1996, 218–19; 2003b; Smithies 2019).

- the way human consciousness would be, if constitutive panpsychism were true;
- revelation: the idea that introspection gives special insight into the reality of consciousness.

The third element makes any discrepancy between the first and second seem fatal. Yet that third element is also something panpsychists cannot readily give up.

How should we spell out these core elements? I think there are actually three slightly different arguments to be made here, and then a fourth argument which engages with the debate on a different combination problem, the “palette problem”. Let us consider the pure revelation arguments first, which differ primarily in whether they rely on the appearance→reality or reality→appearance direction of implication: the first argument says, “Consciousness appears to be X, but panpsychism implies it is not really X,” while the second and third say, “Consciousness fails to appear to be X, but panpsychism implies it really is X.” The first focuses on some positive introspective appearance, and accuses constitutive panpsychists of treating that appearance as an “illusion”. The others focus simply on the *absence* of a certain appearance.

We can call the first argument the “no illusions” argument, since its third premise is RT2, the “no illusions” thesis:

1. If constitutive panpsychism is true, then human consciousness is always “particulate”.
2. Human consciousness (often) appears introspectively to be “smooth”.
3. Consciousness can’t appear a way that it’s not. (RT2)
4. Being “smooth” and being “particulate” are incompatible.
5. Human consciousness is (often) smooth. (from 2 and 3)
6. Human consciousness is (often) not particulate. (from 4 and 5)
7. Constitutive panpsychism is false. (from 1 and 6)

Obviously much turns on the meaning of the terms “particulate” and “smooth”, but despite the frequency with which they (and similar terms like “continuous” and “fragmented”) appear in statements of the problem, it is unclear how to define them, and consequently unclear how plausible premises 1, 2, and 4 are. This definitional question will be central to my discussion in the next section.

The second and third arguments (involving a “reality→appearance” implication) are both suggested in Chalmers’ formulation of what he calls “the revelation argument” (2017, 190). Chalmers notes that although constitutive

panpsychism holds consciousness to be “constituted by a vast array of microexperiences”; this vast array is not revealed to us in introspection. This poses a problem if we think both that introspection reveals the nature of consciousness, and that “whatever constitutes consciousness is part of its nature”.

I distinguish two arguments here because I think talk of “introspection” upon “consciousness” can be taken in two quite different ways. One is that introspection focused *on macroexperiences* doesn’t reveal *that* they are constituted by microexperiences. The other is that introspection focused *on microexperiences* isn’t even possible. The former appears to violate what I above called RT5, the “self-intimation” thesis: reflection upon a pure phenomenal concept reveals the whole nature of a phenomenal property. The latter appears to violate both what I above called RT3, the “understanding-from-experience” thesis, and RT1, the “self-presentation” thesis: having an experience should allow knowledge of its occurrence and a pure phenomenal concept of it.

Focusing on either macroexperiences or microexperiences yields the following two arguments, which I will call the “macroexperience-focused” and “microexperience-focused” argument. The first runs thus, with RT5 as third premise:

1. If constitutive panpsychism is true, each human experience (“macroexperience”) is constituted by a vast array of microexperiences.
2. A vast array of microexperiences is not revealed by reflection on macrophenomenal concepts (i.e. phenomenal concepts based on macroexperiences).
3. The nature of a phenomenal property is revealed by reflection on phenomenal concepts based on experiences of it. (RT5)
4. Whatever constitutes something is part of its nature.
5. The natures of macroexperiences do not involve vast arrays of microexperiences. (from 2 and 3)
6. Macroexperiences are not constituted by vast arrays of microexperiences. (from 4 and 5)
7. Constitutive panpsychism is false. (from 1 and 6)

Clearly, the soundness of this argument depends crucially on what is meant by talk of a property’s “nature”, since that will affect the meaning of premises 3 and 4; this question will be at the heart of my discussion in the next section.

The third (“microexperience-focused”) revelation argument runs thus, with a conjunction of RT₁ and RT₃ as its third premise:

1. If constitutive panpsychism is true, consciousness is constituted by a vast array of microexperiences.
2. We cannot know introspectively about microexperiences, nor form microphenomenal concepts (i.e. phenomenal concepts based on microexperiences).
3. If a subject is having an experience, they can know introspectively that they are, and form phenomenal concepts based on it. (RT₁ and 3)
4. If experiences constitute a subject’s consciousness, that subject undergoes them.
5. We are not undergoing a vast array of microexperiences. (from 2 and 3)
6. Human consciousness is not constituted by a vast array of microexperiences. (from 4 and 5)
7. Constitutive panpsychism is false. (from 1 and 6)

Finally, there is an interaction between a revelation thesis, specifically RT₅, and another aspect of the combination problem, the “palette problem”. How do the huge range of phenomenal qualities that humans experience arise from a fundamental base which appears to involve only a quite small number of fundamental properties? One solution is the “small palette hypothesis”: there are only a few basic phenomenal qualities, corresponding to the fundamental physical properties, which are somehow “blended” to generate a plethora of different qualities for different macroscopic creatures (see [Roelofs 2014](#); [Coleman 2015, 2017](#); [Chalmers 2017, 204–6](#)), whose pattern of similarities and differences are explained by their differing proportions of the basic ingredients. Some critics of the small palette hypothesis object that some of our phenomenal qualities are too heterogeneous to be blended out of a small set of common elements, because they are *completely* dissimilar, with nothing phenomenally in common. Goff ([2017, 195](#)), for instance, claims that, “Minty phenomenology and red phenomenology have nothing in common” (cf. a similar argument in [McGinn 2006, 96](#)). This line of criticism relies on RT₅ to rule out these qualities being similar in a way that we cannot recognise ([Goff 2017, 195–97](#)). Call this the “small-palette revelation argument”, the full structure of which is very similar to that of the macroexperience-focused revelation argument:

1. If the small palette hypothesis is true, then any two phenomenal qualities experienced by humans have something phenomenal in common.
2. Reflection on some pairs of human experiences (e.g. red and minty) does not reveal them to have anything phenomenal in common.
3. The nature of a phenomenal quality is revealed by reflection on phenomenal concepts based on experiences of it. (RT5)
4. The natures of two things determine whether they have anything phenomenal in common.
5. If a pair of phenomenal qualities has something phenomenal in common, reflection on phenomenal concepts based on experiences of them will reveal this. (from 3 and 4)
6. Some pairs of human experiences have nothing phenomenal in common. (from 2 and 5)
7. The small palette hypothesis is false. (from 1 and 6)

All four arguments have a similar four-premise form: first, a supposed implication of constitutive panpsychism (or small-palette forms of it); second, an introspective datum; third, an epistemological thesis about introspection; and fourth, a metaphysical claim, given which the other three premises entail the falsity of constitutive panpsychism (or small-palette forms of it). But despite their common form, I will argue that the arguments go wrong in quite different ways.

4 Ways of Responding to the Revelation Arguments

The challenge for constitutive panpsychists is to rebut the above four arguments without abandoning the revelation approach, components of which underpin all of them. I will show how to rebut each argument in turn, while keeping the relevant revelation theses as strong as I can.

4.1 *The No-Illusions Revelation Argument*

Consider first the “no illusions” argument, which had the following four premises:

1. If constitutive panpsychism is true, then human consciousness is always “particulate”.
2. Human consciousness (often) appears introspectively to be “smooth”.

3. Consciousness can't appear a way that it's not.
4. Being "smooth" and being "particulate" are incompatible.

One option for constitutive panpsychists is to deny premise 1, based on defining "particulate" in such a way that a field-based ontology, or a substance-monist ontology, or some other account of physical reality, renders it false that the material world, and any consciousness isomorphic to it, is particulate (see in particular Nagasawa and Wager 2017, 120–21). If the other three premises (and constitutive panpsychism) are accepted, this implies that the kind of consciousness we enjoy is incompatible with some physical theories (those which make matter "particulate") and that we know introspectively that our world is not any of those ways.

However, I think this approach is a mistake. Even if particles are not ultimately real, Lockwood's point still holds: even the simplest experience involves billions of neurones, ions, and neurotransmitters. Even if the space containing two sodium ions is ultimately just a set of derivative aspects of the one substance, there is still a striking difference in the electrical properties of different regions of that space. To dismiss the problem because particles are not in the fundamental ontology would be too easy. Consequently, I suggest the following definition of "particulate":

X is particulate iff *X* comprises a very large but finite number of parts which differ significantly (in some properties) and discontinuously (on some dimension).

This definition makes the physical brain particulate whatever the fundamental physics turns out to be. Of course this definition will only be as precise as "very large" and "differ significantly and discontinuously". The vagueness of such terms does not stop us from taking "a trillion or more" as a clear case of "very large", and "the mass and charges differences between a water molecule, a potassium ion, and a region of empty space between them" as a clear case of "differ significantly and discontinuously".⁹

⁹ Note also that the definition requires only that the properties of the parts vary discontinuously in *some* dimension, i.e. on some natural way of ordering them, not on all: intuitively, the salient facts about brain parts like potassium ions are things like the abrupt drop in mass from inside the ion's nucleus to outside it, but this abrupt drop might vanish if we instead consider all parts of the brain in a list ordered by mass. But if we want to define "particulate" in a way that does justice to the no-illusions argument, the possibility of finding some dimension on which all variation is continuous should not disqualify the brain from being particulate.

That leaves three remaining options: deny premise 2 (i.e. contradict the supposed introspective observation), deny premise 3 (i.e. reject this particular revelation thesis), or deny premise 4 (i.e. deny that smoothness and particularity are incompatible). But everything depends on what “smooth” means. What is the feature of experience that is being reported by those who feel the pull of this argument?

One option is to define “smooth” by ostension: consider some experiences without discernible internal structure, what Lockwood (1993, 274) calls a “phenomenally flawless” experience, and stipulate that “smooth” means the noteworthy feature of those experiences. That would ensure the truth of premise 2, but would make it hard to adjudicate the truth of premise 4. My preference is to define “smooth” in such a way as to ensure the truth of premise 4, e.g:

X is *smooth* iff it is not particulate.

There are then a few different ways for something to be smooth: since being particular requires parts, for instance, simple things would count as smooth by default. Alternatively, something might be smooth if its parts do not differ significantly in any respect, or do not differ discontinuously along any dimension. The panpsychist must then deny either premise 2 or premise 3: either say that experience does not appear smooth, or say that it does but isn't.¹⁰ At first glance, both options look difficult: premise 3 is, after all, part of the Revelation Approach (RT₂), and if premise 2 is false, why did anyone ever advance the argument in the first place?

The way out lies in scrutinising the word “appears”, and drawing a distinction between illusions, strictly so-called, and easy misinterpretations. Consider some non-mental examples: at first an act appears noble, an argument compelling, a speech beautiful, and yet then I find that upon giving the matter more thought, this appearance vanishes, and I come to think I was mistaken. The act now appears fanatical, the argument sophistical, the speech saccharine; I think myself foolish for being gullible enough for the act, argument, or speech to ever appear otherwise to me. I might say I was subject to an “illusion”, but all this mean is that the act, argument, and speech were such that they could be very readily misjudged.

¹⁰ Using the ostensive definition would just translate denial of premise 2 into denial of premise 4: either way, the claim is that there is no property incompatible with particularity that consciousness introspectively seems to have.

Contrast this with a white object seen under pure red light, or a straight stick seen half in water, or an ambitious Scottish nobleman hallucinating a dagger. The object appears red but isn't, the stick appears bent but isn't, and there appears to be a dagger, but there isn't. Here no reflection on the appearances will change them, and the subject cannot hold themselves rationally accountable for being subject to them (perhaps for forming beliefs based on them, but not for the appearances themselves). Here we have a stronger sense of "illusion": it is not that these perceptions are easy to misjudge, it is that their very content is false. Call this the "quasi-perceptual" sense of "appears", contrasting with the "ready-interpretation" sense (cf. [Stoljar 2013](#); [Kammerer 2018](#)).

Premise 3 (RT₂) is most plausible if read with the "quasi-perceptual" sense of "appears". Plausibly it makes no sense to think that my impression of my own experience is an "illusion" in this stronger sense: surely it would be the "impression" that deserves to be called my experience, since this is what I am immediately aware of. To think that consciousness might appear falsely in this way seems to involve forgetting that consciousness is how things appear to me (cf. [Liu 2020](#)). Or at least, this thought has some appeal, and panpsychists need not disagree with it.

But premise 3 is less plausible if understood in terms of the "ready-interpretation" sense of "appears", saying that if consciousness is readily interpreted as having some property, it must actually have that property. After all, which interpretations come readily depends on the subject's expectations, background assumptions, interpretive style, etc. An absolute principle, that no false interpretation could come readily to *anyone*, would be very close to saying, implausibly, that consciousness was never misinterpreted.

So we should read premise 3 as saying that consciousness cannot appear a way it's not, in the quasi-perceptual sense of "appear". For the argument to remain valid, premise 2 must also be read in terms of the quasi-perceptual sense of "appear", not the "ready-interpretation" sense. But now premise 2 is much more deniable. We can deny premise 2, in this strong sense, by taking the appearance of smoothness to be a matter of what interpretations come readily, and not of how things quasi-perceptually appear.

This is my preferred response to the "no illusions" argument: our consciousness really is particulate, not smooth, but it is readily misinterpreted as smooth. But this misinterpretation demands an explanation - what is it about the way consciousness *does* appear, which makes us judge it "smooth"?

One answer appeals to the difference between represented structure and structured representations: that is, experience represents things as being smooth, rather than itself being smooth (versions of this proposal appear in: Clark 1989; Stoljar 2001). Critics have worried that experience itself really does seem to display the relevant sort of smoothness (e.g. Alter and Nagasawa 2012, 91), and that representing a smooth expanse may be insufficient for introspectively seeming, even in the weak sense, to be smooth (consider the sentence “space is infinitely divisible”). Another answer is to say that many experiences quasi-perceptually appear to have, and thus (by RT₂) actually have, some property similar to, but not identical to, “smoothness”. In section 5 I flesh out this approach.

4.2 *The Macroexperience-Focused Revelation Argument*

Next, consider the macroexperience-focused argument, whose premises are:

1. If constitutive panpsychism is true, each human experience (“macroexperience”) is constituted by a vast array of microexperiences.
2. A vast array of microexperiences is not revealed by reflection on macrophenomenal concepts (i.e. phenomenal concepts based on macroexperiences).
3. The nature of a phenomenal property is revealed by reflection on phenomenal concepts based on experiences of it.
4. Whatever constitutes something is part of its nature.

I see little prospect for denying premises 1 and 2,¹¹ and premise 3 is one of the revelation theses I want to preserve. Chalmers, when he lays out the argument of which this is a variant, advises panpsychists to attack premise 4: to drive a wedge between something’s nature and what constitutes it. I agree that this is the right tack, but everything turns on what kind of “nature” is in question, which in turn depends on how we read premise 3, the self-intimation thesis. I think there is a plausible and well-motivated sense of “knowing a nature”

¹¹ It might look like cosmopsychists can wriggle out of premise 1. But this is illusory: the only way cosmopsychists can deny premise 1 is to commit to an analogous premise that supports a *harder* revelation argument. If they deny that the brain is constituted by neurones, ions, etc., they must instead accept a replacement premise 1*: “If constitutive panpsychism is true, each human experience (‘macroexperience’) constitutes a vast array of microexperiences.” We then run the same argument, with premise 4 replaced by 4*: “Whatever something constitutes is part of its nature.” And I think premise 4* is noticeably *more* plausible than premise 4.

which explains why premise 4 is false, without undermining anti-physicalist arguments.¹²

First consider this common gloss: knowing the nature of a property means being in a position to know a priori every necessary truth about that property.¹³ If I know the nature of squareness, I am in a position to know a priori every necessary truth about squareness (like what squares' internal angles sum to, or what kinds of triangles they can be divided into), though not to know contingent truths about it (like whether it is my sister's favourite shape). Likewise if I know the nature of being water, I can know every necessary truth about being water (like that water is a chemical compound, or its molecular mass), though not every contingent truth about it (like whether it is instantiated on Earth). This suggests that we know the natures of mathematical properties, but do not automatically know the natures of chemical properties, though perhaps we do now, given scientific progress. And those results seem plausible.

But this gloss is inadequate. Consider someone who knew the nature of squareness but not the nature of triangularity (if that were possible). They would not be in a position to know a priori that every square can be divided into four right-angled triangles. This suggests a refinement: knowing the nature of some property means being in a position to know *a priori* all the necessary truths about that property which involve only other properties whose natures you also know. To put it another way, to know a priori a necessary truth involving two properties, you need to know the natures of both: just knowing the nature of one is not enough.¹⁴ This implies, in particular, that knowing the nature of a constituted property is not sufficient to know about its constitution relationships to other properties, without also knowing the natures of those other properties.

-
- 12 The argument discussed in Lee (2019) combines premises 3 and 4 into a single claim, "Structure Luminosity: If a subject introspects an experience, then that subject is in a position to know the phenomenal realizers of that experience" (2019, 292). Lee argues (in my view plausibly) that this is false, but does not clearly identify which elements of it remain true, and whether they are enough for anti-physicalist arguments.
- 13 I am abstracting away from difficulties of memory, attention, and general cognitive skills: in practice, many necessary truths might be just too complicated or subtle for a human mind to entertain, but that should not stop us from saying that someone is in a position to know them if all they would need to do so is an enhancement of their general cognitive skills.
- 14 This is not a retreat from the idea that the phenomenal property's "whole nature" is revealed. There is no part of its nature that is hidden: there are only hidden connections between its nature and other natures, and those connections are hidden for the simple reason that those other natures are hidden.

I think this provides a plausible reading of “knowing a property’s nature”, and thereby of RT5, which does precisely what constitutive panpsychists need it to do: substantiate their arguments against physicalism, without substantiating the revelation argument against their own view. For on this reading of “knowing a nature”, that we know the natures of macrophenomenal properties implies that for any other set of properties whose natures we know, we are in a position to tell a priori whether those properties are sufficient to constitute macrophenomenal properties. And the case against physicalism is that physical properties do not seem a priori to constitute macrophenomenal properties. Of course, this attack only works if we know the natures of physical properties (e.g. if we think of them as exhausted by what physics says about them, as what Stoljar (2001) calls the “t-physical” properties, and what Strawson (2006) calls “physical” properties). It will not work if we think of physical properties as whatever properties physical things have which in fact account for their satisfying the descriptions given by physics (what Stoljar (2001) calls the “o-physical” properties). But that way out is no use to standard physicalism, which needs physical properties to be well-understood: to say that the reason the conceivability argument fails is that there is some mysterious hidden nature of the physical, which plays some crucial role in accounting for consciousness, is to embrace the kind of “non-standard physicalism” (cf. Stoljar 2006) that is no longer incompatible with panpsychism.

But why doesn’t knowing the natures of macrophenomenal properties substantiate a parallel argument against constitutive panpsychism? Because panpsychists do not claim that we know the natures of microphenomenal properties, because we are not the microsubjects who instantiate those properties (though see the next subsection for some complications of this claim). Without knowledge of the candidate constituting properties, we cannot determine a priori their suitability to constitute macrophenomenal properties. All the constitutive panpsychist is committed to is a conditional claim: *if* we were able to grasp the natures of microphenomenal properties, then we could, in principle, see a priori that, when properly arranged, they constitute macrophenomenal properties.

4.3 *The Microexperience-Focused Revelation Argument*

Thirdly, consider the microexperience-focused revelation argument: why can’t we introspect microexperiences like we can macroexperiences? The premises of this argument are:

1. If constitutive panpsychism is true, consciousness is constituted by a vast array of microexperiences.
2. We cannot know introspectively about microexperiences, nor form microphenomenal concepts.
3. If a subject is having an experience, they can know introspectively that they are, and form phenomenal concepts based on it.
4. If experiences constitute a subject's consciousness, that subject undergoes them.

Again, I see little hope in denying premises 1 or 2,¹⁵ which leaves three options: deny premise 3 (“we *are* undergoing microexperiences, but cannot introspect them”), deny premise 4 (“microexperiences constitute our consciousness, but we do not undergo them”), or show the argument to be invalid.

Goff's approach in his (2017, 189ff.) is to deny premise 4, to “loosen” the relation between microexperiences and macroexperiences, so that although microexperiences in some sense constitute (or “ground”, “compose”, or “form”) macroexperiences, the phenomenal character of the latter contains nothing of the former. The cost of this is that the constitution relation between microexperiences and macroexperiences is thereby made more mysterious. If this relation were one in which both constituted and constituter were undergone by the same subject, it could be akin to familiar relations among macroexperiences. For instance, the relation between my total phenomenal field right now and the component experiences that it subsumes (sounds I'm hearing, colours I'm seeing, twinges of physical discomfort, etc.) is plausibly something like constitution. It would be nice if panpsychists could assimilate the microexperience-macroexperience relation to familiar relations like this, where a single subject undergoes all the experiences involved; without that link it is hard to see why microexperiences should really be said to “constitute” a macroexperience, as opposed to somehow giving rise to it as a distinct product.

I think the best approach is to say the argument is invalid *when premise 3 is qualified* in certain ways that are independently necessary to make it plausible. An unqualified form of premise 3 faces easy counterexamples: ferrets

15 Again, though one might think cosmopsychists can deny premise 1, there is no advantage to be gained thereby: the replacement premise 1* - “If constitutive panpsychism is true, human consciousness constitutes a vast array of microexperiences” - will support a revised version of the argument, when paired with 4* - “If experiences are constituted by a subject's consciousness, that subject undergoes them.” And again, 4* seems to me even more plausible than 4.

undergo many experiences, but cannot form phenomenal concepts, or know that they are having experiences. But plausibly this is not a counter-example to what premise 3 was intended to say! The problem is not that ferrets' experiences are somehow hidden from them, but just that they lack the conceptual competence to recognise their experiences as such. A qualified version of premise 3 would allow for this: it would say that certain kinds of knowledge and concept-formation are possible whenever a subject undergoes an experience *and* meets various other conditions. Another plausible requirement is attention: one must focus on an experience in order to introspect it, and if one is unable to direct one's attention, introspection will be impossible.¹⁶

So let us consider a qualified reading of premise 3, that includes these conditions: introspective knowledge is possible whenever a subject undergoes an experience, *and* is capable of conceptualising it, *and* focuses their attention on it. The argument has now become invalid: line 5 ("we are not undergoing a vast array of microexperiences") no longer follows from 2 and 3. There are two reasons why we might be phenomenally undergoing microexperiences but be unable to know them introspectively, compatibly with this weaker reading of premise 3: if humans cannot conceive of experiences as such, or if they are unable to attend to microexperiences. While the first of these options is clearly false, the second is, I think, the best option for the constitutive panpsychist in rebutting the microexperience-focused argument.

This implies that while microexperiences are phenomenally conscious for us, they are not access-conscious for us. That is, microexperiences are presented to us, "right there", characterising the phenomenal character of our consciousness, but they are not presented in such a way that we can cognitively select, access, and identify them. Our relationship to them is rather like our relationship to elements of our experience that are very faint, which require a lot of effort to focus on and distinguish from their surroundings, and which it is correspondingly easier to distract us from. If something in my peripheral vision is roughly the same colour as its surroundings, it would be hard for me to notice it, and if I were distracted, exhausted, or inebriated I might find attending to it all but impossible. Yet it is still part of my phenomenology, not somehow hidden from me. The constitutive panpsychist, I am suggesting, should claim that this near-impossibility of attending to peripheral vision while distracted is intensified to a real practical impossibility

16 Goff's statement of revelation (2017, 109–10) mentions attention explicitly, and Chalmers appeals to inattention as a primary reason for thinking that his principles of "detectability" and "reliability" can only hold for the most part, not absolutely (Chalmers 1995, 326; 1996, 218–19).

with microexperiences. In section 5 I situate this impossibility claim within a broader picture of how the mind is constituted by microexperiences, which will help to motivate this response to the microexperience-focused argument.

4.4 *The Small-Palette Revelation Argument*

Finally, consider the small-palette revelation argument, whose premises are:

1. If the small palette hypothesis is true, then any two phenomenal qualities experienced by humans have something phenomenal in common.
2. Reflection on some pairs of human experiences does not reveal them to have anything phenomenal in common.
3. The nature of a phenomenal quality is revealed by reflection on phenomenal concepts based on experiences of it. (RT5)
4. The natures of two things determine whether they have anything phenomenal in common.

Since this is not an argument against constitutive panpsychism per se, there are technically five options for constitutive panpsychists in responding to it: deny one of the premises, or accept the conclusion. Accepting the conclusion would mean accepting a “large palette” version of constitutive panpsychism, with all human and animal qualities present in the base even though that is more than there are distinct physical roles to play (see, e.g. [Lewtas 2013](#)). The downside is that this sacrifices the appealing parsimony, and isomorphism with physics, that had seemed to set constitutive panpsychism apart from traditional sorts of dualism. Denying premise 3 is also unattractive, since it undermines the case for panpsychism over physicalism.

Denying premise 4 here (as Lee does, [2019, 300–301](#)) is harder than denying premise 4 of the macroexperience-focused argument, that “what constitutes something is part of its nature”. I denied the latter because knowing a property’s nature is not enough to know necessary truths about it which involve the nature of another property; we would have to know that other property’s nature as well. But when it comes to comparing two qualities that we do experience distinctly, it seems to follow that we should be able, in principle, to discern every necessary truth about how those qualities relate, and that should include their resemblance or common constituents.¹⁷

¹⁷ Could we find a more carefully qualified version of RT5, on which knowing the natures of two properties enables us to know whether one suffices to constitute the other, but not whether and

We might deny premise 4 in the same way we might deny premise 4 of the microexperience-focused argument, by saying that although the basic qualities constitute the macroqualities, they do not characterise them - the “blending” leaves no trace of the ingredients at all. But this has the same downsides discussed in the last subsection: if microqualities in no way characterise the macroqualities, the form of constitution involved seems mysterious.

That leaves denying premise 1 or premise 2. Premise 1 might seem undeniable, due to the “interchangeability” of different neurons: experiences of redness and of mintiness involve neurones made of all the same sorts of subatomic particles, so how can one contain any ingredient missing from the other? Any ingredient of the redness experience comes from electrons, quarks, photons, etc., and those same things are all present in the physical basis of a mintiness experience, so how could they not show up in the latter? But this falsely assumes that each macroexperience should contain every ingredient present in its neural basis, as though each one were the independent product of one discrete subset of neurones. It might instead be that several macroexperiences are all grounded in the activity of the same neurones, being just different aspects of the complex, differentiated experience produced by those neurones.

Consider a bar magnet, whose macroscopic behaviour displays a “north pole” and “south pole”. The north pole does not arise from one half of the magnet, and the south pole from the other half: both macroscopic features arise from very same microscopic physical things, because those things are themselves internally differentiated and their different aspects add up to what looks, from a macroscopic perspective, like two different things. It would be a mistake to say “since all the particles generating the magnet’s north pole also have south poles, why don’t their south poles show up in the magnet’s north pole?” Perhaps mintiness and redness are likewise different aspects of the same complex experience, itself arising from the combination of a great many internally differentiated microexperiences, combining in different ways depending on such things as firing rates and degrees of neural synchrony. Then they might have nothing phenomenal in common, despite being constituted by the same things.

However, there are limitations to this response. It might allow for a few completely dissimilar pairs to be compatible with the SPH, but not that many -

how they resemble each other? Maybe, but this feels ad hoc to me; I see no plausible way to motivate it.

if there are a hundred completely dissimilar qualities experienced by humans, saying that they arise from the way that internally differentiated aspects of microexperiences are combined starts to load microexperiences with too much structure for us to retain the SPH. To keep the palette small, there shouldn't be too many completely dissimilar pairs of qualities, which is why this response to the argument works best when combined with another: denying premise 2.

Denying premise 2 means denying that redness and mintiness have absolutely nothing at all in common. After all, our ability to recognise two things as akin to one another is usually enhanced by our ability to recognise and attend to the features they share, and if we never experience their shared features in isolation, we may take them to be entirely unlike even if they are not. Sometimes, of course, two qualities seem inarticulately alike even without an identifiable shared feature; this is why we routinely describe qualities of one modality using terms drawn from another (warm, harsh, sweet, soft, loud, etc.). The SPH and RT5 can both be retained as long as idealised scrutiny of these inchoate likenesses would reveal a system of qualitative connections over our entire experiential range. This view is defended by Coleman:

[...] just as it's possible to move across the colour spectrum in tiny, almost undetectable steps, it must be possible to move from tastes to sounds, sounds to colors, and so on, via equally tiny steps. Tip-toeing between modalities already seems *conceivable* in certain cases, perhaps even actual. We know that what we experience as "taste" is really some kind of fusion of qualia sourced from the nose and from the tongue [...]. To address qualitative incommensurability we must stretch to conceiving of such continuities as the rule rather than the exception. (Coleman 2017, 264, emphasis in original; cf. Coleman 2015; Hartshorne 1934, 35ff.)

This claim does not seem to me obviously false, but it is at least dubitable. Consequently, the revelation approach may be most threatening to constitutive panpsychists not through any of the three pure revelation arguments, but through intensifying the palette problem. Accepting revelation pushes constitutive panpsychists towards either a large-palette solution like Lewtas's, or towards Coleman's very bold and ambitious form of the small-palette hypothesis.

5 Confusion and Revelation

Identifying a premise of an argument that might be false is often not, by itself, an effective way to persuade critics. For all that I have said so far, this “medium-strength” version of revelation, interpreted and qualified so as to undermine arguments against panpsychism while substantiating arguments against physicalism, might be technically consistent but ad hoc and unmotivated, a dingy corner of logical space which panpsychists can awkwardly retreat to. But in fact, these qualified revelation theses flow from a reasonable picture of the limits of human self-knowledge, on which the introspective ignorance that constitutive panpsychism implies differs only in degree from familiar forms of introspective ignorance.

It is commonplace to say that when two experiences become phenomenally unified, they form a composite experience which subsumes them: they still exist, and are still undergone by the subject, but they are now “undergone together”. We easily recognise this when we can discern introspectively not just the composite experience but also its components: but what if the discernibility of the component experiences is not an automatic consequence of the composite experience being composite? We might consider the idea that it depends instead on having the right structure of informational relations among the components.¹⁸ Perhaps if these relations make the subject’s overall dynamics differentially sensitive to multiple distinct features of the experience, the composite experience will be characterised by contrast among those features: they will stand out as distinct things. If not, those features will be present in the composite experience in an undifferentiated way, as a single element whose phenomenal quality is a seamless blend of its components. In short: the component experiences all go in together, but the way they are present in the composite experience depends on how they are organised.

What explains why experiences should compose in this way is a further question, which I cannot here address (though see [Roelofs 2016; 2019, 123–25, 166–70](#)). But suppose some conditional like this were true: when distinct experiences are unified, they can be distinguished by the subject only if they have the right informational structure. Although the human brain is an exquisitely structured processor of information, it has limits. The overall dynamics of the brain can perhaps be sensitive to whether a neurone fires, but not (as far as we know) to which ions in that neurone played which roles in its firing. Since

¹⁸ This is a long-standing idea among panpsychists, though spelling it out in detail is not simple. See [Chalmers \(1996\)](#), 284–292; [Chalmers \(2017\)](#), 209–210; [Gabora \(2002\)](#); [Roelofs \(2019\)](#), 171–176.

individual events at the microscopic level are informationally inaccessible, they will be experienced by the whole in a blended way. They each make a minute difference to the quality of some element of the whole's experience, but they do not stand out as distinct elements of it. To use a term made famous by Leibniz, they are "confused" with one another, the way that the sounds of each bit of water striking the shore are "confused" in the roar of the sea.¹⁹

I have elsewhere elaborated more fully on the idea of confusion as I understand it (2019, 126–29), but the essential idea is captured in the following definition:

Two experiences are *confused* with each other, relative to a subject, iff that subject cannot distinguish them by attending to one without simultaneously attending to the other.²⁰

It is important to emphasise that confusion is not a matter of a subject "perceiving" things outside themselves so poorly that they cannot distinguish the parts of that outside thing. Confusion is a matter of how the subject's own states are related, not a relation between them and something external. For example, someone viewing a pointillist painting, for whom the many dots of paint "blur together", is not thereby subject to confusion, if they simply have a single experience that is the product of many external objects. A better example would be someone with an untrained palate, who drinks coffee and experiences (let us stipulate) the same diversity of taste and flavour experiences as a practiced connoisseur but experiences them together as a single blended flavour, without being able to pick out the bitterness from the aroma, etc.

Confusion may depend on circumstances. When we are tired, distracted, or drunk we often cannot distinguish things which we could under better conditions. Then our experiences are confused only relative to those circumstances. Confusion can also depend on a subject's conceptual repertoire: sometimes we cannot distinguish two things using their present concepts, but would

19 This idea of the mind as comprising a vast number of "little perceptions", most of which cannot be distinguished from one another by the subject, is arguably present in several early modern writers as well as Leibniz, in particular Spinoza, Wolff, and Kant. For discussion see Wilson (1980), Thiel (2011), Liang (2017), and Indregard (2018). To use a more modern phrasing from Andrew Lee (2019), they make up the non-introspectible "microstructure of experience".

20 In the primary instance confusion is defined over tokens, but we can easily define a secondary sense in which two types are confused for a subject when any token of those types onto which a given subject could direct a given operation would be confused with a token of the other type.

be able to if we learnt new ones. Call confusion which can be removed by adjusting the subject's bodily surroundings or condition, or improving their conceptual repertoire, or in some similarly mild way, "shallow confusion", and call confusion which persists even into ideal conditions, "robust confusion".

In between shallow and robust is confusion which persists until the subject becomes distinctly acquainted with a token of the same type as the confused elements. For example, suppose the sensory component of pain is robustly confused with the unpleasant affect pain involves, except for subjects who have experienced "pain asymbolia", the rare condition of feeling pain without finding it at all unpleasant (cf. [Grahek 2007](#); [Klein 2015](#)). If they regain normal pain experiences, they might find themselves newly able to attend to its sensory element in isolation. If this were to happen, we might say that their original confusion was "nearly-robust": removable only by somehow acquainting them with (a token of the same type as) one of the confused elements on its own.²¹

When confusion is shallow, we have an easy way to tell that we suffer from it: we remove it and contrast the resulting distinction with the earlier confusion. With sufficiently robust confusion, we would not have such means of recognising it; we could not tell that we were confused. And if we suffered from confusion that was "nearly-robust", it would be undetectable, except by means of independent acquaintance with elements of the same type as the confused ones. We could, that is, be subject to a lot of confusion without being able to tell, introspectively. And if constitutive panpsychism is true - in particular, if micro-experiences corresponding to all the physical details of our brains were somehow present in our consciousness - then we should expect just that: all the experiences of our microparts would be confused relative to us. Call this the Radical Confusion Hypothesis.

Confusion is defined functionally, but that does not imply that confusion is a purely functional fact that makes no phenomenal difference. My suggestion is that undergoing two confused experiences feels different to undergoing two distinguishable experiences, even if those experiences are the same in all intrinsic respects. When the components of an experience are distinguishable by the subject, they are phenomenally present as discernible, separate, parts - there is an experience of phenomenal contrast, of things standing out against

²¹ In other work ([2019, 128–29](#)), I also distinguish between "strong" and "weak", and "symmetrical" and "asymmetrical" confusion, but this does not substantially affect the argument so I omit it here for simplicity.

other things. But when they are confused, they are present qualitatively, as contributions to the total quality of the experience they blend into.

How would the Radical Confusion Hypothesis help with the four revelation arguments? Recall that in response to the “no illusions” argument, I denied premise 2: that human consciousness positively appears introspectively to be “smooth” (there defined as “not particulate”). I maintained that this is false if “appears introspectively” is read in a strong, quasi-perceptual sense; it is true only if “appears introspectively” is read in a weaker sense, as meaning “it is easy and natural to interpret experience this way”.

Now I can say *why* this misinterpretation is easy and natural: because many human experiences display something close to “smoothness”, namely, all their component experiences are nearly-robustly confused with each other, distinguishable only by a subject who already knows what to look for. A subject who lacks any distinct acquaintance with the ingredients will be unable to distinguish them or discern their internal structure. We might say that experiences all of whose components are confused with one another are “pseudo-smooth”, and it is true (and introspectively obvious!) that many of our experiences are pseudo-smooth. But to infer genuine smoothness from pseudo-smoothness is a metaphysical over-interpretation which goes beyond the introspective deliverances: it is inferring absence of structure from the failure of structure to be manifest in a certain way (it is thus very similar to the “headless woman illusion” discussed by Armstrong (1968), where not seeing someone’s head gives us the vivid but false impression that they have no head). The noticeable quality that some experiences have, which prompted the “no illusions” argument, is just what radical confusion feels like.

Second, in response to the macroexperience-focused argument I denied premise 4, that whatever constitutes something is part of the “nature” that is revealed to us by pure phenomenal concepts. I suggested that a priori reflection tells us only those necessary truths that involve *only* properties whose nature we know - such as whether one could constitute the other. But just knowing the nature of one property does not tell all the things that could constitute it, nor what constitutes a particular instance of it.

I can now elaborate on this distancing of constitution from “nature”. Macroexperiences are composite experiences composed of many microexperiences confused with one another. Their phenomenal character is determined by combining the phenomenal characters of those component experiences, which they subsume in fundamentally the same way that a person’s total experience at any one time subsumes the partial experiences they are having

at that time. But just as two composites might end up sharing certain properties despite being constituted by different sets of parts, and despite their properties being mere combinations of the properties of their parts, two composite experiences might have the same phenomenal character, despite being constituted by different sets of microexperiences. The particular parts might be essential to the particular macroexperience, but not to the property that it is an instance of.

I also said, in response to the small-palette revelation argument, that distinct macroexperiences might arise from the same neural basis: we need not assume that each distinguishable element of our consciousness contains the entire phenomenal nature of one discrete subset of physical entities. The radical confusion hypothesis reinforces this point: it says that which experiences phenomenally contrast or phenomenally blend with one another in human experience reflects the informational structure of the brain, so a single macroexperience might not correspond to any discrete section of the underlying physical substrate. Instead, it will correspond to a set of features of the substrate such that information about them collectively is extracted and used by the brain, but information about them individually is not. Thus different macroexperiences based in the same brain area might have different, even non-overlapping sets of phenomenal ingredients, because they reflect different features of the same microexperiences.

Finally, in response to the microexperience-focused argument I suggested that our ignorance of microexperiences is compatible with our undergoing them, if we cannot attend to them. Now I can add that our inability to attend to microexperiences is part-and-parcel of their being confused for us. Their radical confusion is explained by the limitations discussed above on how much information about microscopic brain events can be extracted by the rest of the brain.²² Because radically confused experiences cannot be distinctly attended to, we cannot know them or their natures, even though the experiences “present themselves” in the sense that if their subject could attend to them they could know them and their natures by introspection.

An opponent might object that even though attending to particular experiences can be harder or easier, depending on, e.g. architectural facts about the brain, it cannot be *strictly impossible* for me to attend to an experience, if it is really is an experience I am undergoing. I reply that distinctly attending

22 This allows for a limited sense in which microexperiences *are* accessible: namely that they can be accessed only by acts which are also accessing many other microexperiences at the same time. They cannot be *individually* accessed, but they can be accessed *collectively*.

to microexperiences is *not* strictly impossible, just impossible in practice (as discussed in Lee 2019, 296–97). They are manifest in our consciousness, but incredibly difficult to pick out. After all, it is very difficult for the large-scale dynamics of our brain to be sensitive to changes in a single particle, but there is no in-principle impossibility in there being such sensitivity, perhaps using advanced technology or strange altered states of consciousness.²³

6 Conclusions

The idea of “revelation”, that having an experience provides a special insight into its nature, is a key weapon in the armoury of anti-physicalists. But for constitutive panpsychists there is a risk it will blow up in their faces. I have argued, however, that a suitably-qualified form of the revelation approach can bring down physicalism while leaving panpsychism standing: a form which reconciles the profound fallibility of the human mind’s self-knowledge with the perfect transparency of its access to its itself. Although nothing does or could “conceal” our own experiences from us, we are nevertheless limited in our ability to attend to their elements, prone to misinterpret them, and consequently unable to tell introspectively just how composite they might really be.*

Luke Roelofs
New York University
luke.mf.roelofs@gmail.com

References

ALTER, Torin, and Yujin NAGASAWA. 2012. “What Is Russellian Monism?” *Journal of Consciousness Studies* 19 (9–10): 67–95.

²³ Note that there need not be any sharp boundary between “the simplest experiential element that we can distinguish” and “the most complex experiential element that is radically confused.” For different people, under different conditions, different distinctions among one’s internal states and processes may be possible. Radically confused experiences are not a qualitatively distinct sort of experience from distinguishable ones, any more than “places I can walk to in ten minutes” are a sharply separate set of places from those I can walk to in ten minutes; my walking ability, like my introspective discernment, waxes and wanes as I change and as conditions change.

* This paper expands on ideas presented over pages 132–137 of Roelofs (2019). Their further development owes a great deal to audiences at the Australian National University and the CEU’s workshop “Russellian Monism: Time for the Details” in Budapest.

- ARMSTRONG, David M. 1968. "The Headless Woman Illusion and the Defence of Materialism." *Analysis* 29 (2): 48–49. doi:10.1093/analys/29.2.48.
- BALL, Derek. 2009. "There Are No Phenomenal Concepts." *Mind* 118 (472): 935–62. doi:10.1093/mind/fzpz134.
- BRADDON-MITCHELL, David, and Robert NOLA, eds. 2009. *Conceptual Analysis and Philosophical Naturalism*. Cambridge, Massachusetts: The MIT Press.
- BYRNE, Alex, and David R. HILBERT. 2007. "Color Primitivism." *Erkenntnis* 66 (1–2): 73–105. doi:10.1007/s10670-006-9028-8.
- CHALMERS, David J. 1995. "Absent Qualia, Fading Qualia, Dancing Qualia." In *Conscious Experience*, edited by Thomas Metzinger, 309–29. Paderborn: Ferdinand Schöningh.
- . 1996. *The Conscious Mind: In Search of a Fundamental Theory*. Oxford: Oxford University Press.
- . 2003a. "Consciousness and Its Place in Nature." In *The Blackwell Guide to the Philosophy of Mind*, edited by Stephen P. Stich and Ted A. Warfield, 102–42. Blackwell Philosophy Guides. Oxford: Basil Blackwell Publishers.
- . 2003b. "The Content and Epistemology of Phenomenal Belief." In *Consciousness: New Philosophical Perspectives*, edited by Aleksandar Jokić and Quentin Smith, 220–72. Oxford: Oxford University Press.
- . 2009. "The Two-Dimensional Argument Against Materialism." In *The Oxford Handbook of Philosophy of Mind*, edited by Brian P. McLaughlin, Ansgar Beckermann, and Sven Walter, 313–37. Oxford Handbooks. Oxford: Oxford University Press.
- . 2017. "The Combination Problem for Panpsychism." In *Panpsychism. Contemporary Perspectives*, edited by Godehard Brüntrup and Ludwig Jaskolla, 179–214. Oxford: Oxford University Press.
- CLARK, Andy. 1989. "The Particulate Instantiation of Homogeneous Pink." *Synthese* 80 (2): 277–304. doi:10.1007/bf00869488.
- COLEMAN, Sam. 2015. "Neuro-Cosmology." In *Phenomenal Qualities. Sense, Perception, and Consciousness*, edited by Paul Coates and Sam Coleman, 66–102. Oxford: Oxford University Press.
- . 2017. "Panpsychism and Neutral Monism: How to Make up One's Mind." In *Panpsychism. Contemporary Perspectives*, edited by Godehard Brüntrup and Ludwig Jaskolla, 249–82. Oxford: Oxford University Press.
- DAMNJANOVIC, Nic. 2012. "Revelation and Physicalism." *Dialectica* 66 (1): 69–91. doi:10.1111/j.1746-8361.2012.01290.x.
- DENNETT, Daniel C. 2007. "What RoboMary Knows." In *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*, edited by Torin Alter and Sven Walter, 15–31. Oxford: Oxford University Press.
- DÍAZ-LEÓN, Esa. 2011. "Reductive Explanation, Concepts, and a Priori Entailment." *Philosophical Studies* 155 (1): 99–116. doi:10.1007/s11098-010-9560-x.

- GABORA, Liane. 2002. "Amplifying Phenomenal Information: Toward a Fundamental Theory of Consciousness." *Journal of Consciousness Studies* 9 (8): 3–29.
- GAUDRY, Justin. 2008. "Does Physicalism Entail Cosmopsychism?" <https://panexperientialism.blogspot.com/2008/05/does-physicalism-entail-cosmopsychism.html>.
- GOFF, Philip. 2006. "Experiences Don't Sum." *Journal of Consciousness Studies* 13 (10–11): 53–61.
- . 2015. "Real Acquaintance and Physicalism." In *Phenomenal Qualities. Sense, Perception, and Consciousness*, edited by Paul Coates and Sam Coleman, 121–45. Oxford: Oxford University Press.
- . 2017. *Consciousness and Fundamental Reality*. Oxford: Oxford University Press.
- GRAHEK, Nikola. 2007. *Feeling Pain and Being in Pain*. 2nd ed. Cambridge, Massachusetts: The MIT Press.
- HARTSHORNE, Charles. 1934. *The Philosophy and Psychology of Sensation*. Chicago, Illinois: University of Chicago Press.
- INDREGARD, Jonas Jervell. 2018. "Consciousness as Inner Sensation: Crusius and Kant." *Ergo* 5 (7): 173–201. doi:10.3998/ergo.12405314.0005.007.
- JACKSON, Frank. 1982. "Epiphenomenal Qualia." *The Philosophical Quarterly* 32 (127): 127–36. doi:10.2307/2960077.
- JASKOLLA, Ludwig, and Alexander J. BUCK. 2012. "Does Panexperiential Holism Solve the Combination Problem?" *Journal of Consciousness Studies* 19 (9–10): 190–99.
- JOHNSTON, Mark. 1992. "How to Speak of the Colors." *Philosophical Studies* 68 (3): 221–63. doi:10.1007/bf00694847.
- KAMMERER, François. 2018. "Can You Believe It? Illusionism and the Illusion Meta-Problem." *Philosophical Psychology* 31 (1): 44–67. doi:10.1080/09515089.2017.1388361.
- KLEIN, Colin. 2015. "What Pain Asymbolia Really Shows." *Mind* 124 (494): 493–516. doi:10.1093/mind/fzu185.
- KRIEGEL, Uriah. 2009. *Subjective Consciousness: A Self-Representational Theory*. New York: Oxford University Press.
- KRIPKE, Saul A. 1980. *Naming and Necessity*. Oxford: Basil Blackwell Publishers.
- LEE, Andrew. 2019. "The Microstructure of Experience." *Journal of the American Philosophical Association* 5 (3): 286–305. doi:10.1017/apa.2019.4.
- LEWIS, David. 1995. "Should a Materialist Believe in Qualia?" *Australasian Journal of Philosophy* 73 (1): 140–44. doi:10.1080/00048409512346451.
- LEWTAS, Patrick Kuehner. 2013. "What Is It Like to Be a Quark?" *Journal of Consciousness Studies* 20 (9–10): 39–64.
- LIANG, Yibin. 2017. "Kant on Consciousness, Obscure Representations and Cognitive Availability." *The Philosophical Forum* 48 (7): 345–68. doi:10.1111/phil.12169.
- LIU, Michelle. 2019. "Phenomenal Experience and the Thesis of Revelation." In *Perception, Cognition and Aesthetics*, edited by Dena Shottenkirk, Manuel Curado,

- and Steven S. Gouveia, 227–51. *Routledge Studies in Contemporary Philosophy* 119. New York: Taylor & Francis.
- . 2020. “Explaining the Intuition of Revelation.” *Journal of Consciousness Studies* 27 (5–6): 99–107.
- LOCKWOOD, Michael. 1993. “The Grain Problem.” In *Objections to Physicalism*, edited by Howard Robinson, 271–92. Oxford: Oxford University Press.
- MAXWELL, Grover. 1978. “Rigid Designators and Mind-Brain Identity.” In *Minnesota Studies in the Philosophy of Science, Volume IX: Perception and Cognition: Issues in the Foundations of Psychology*, edited by C. Wade Savage, 365–404. Minneapolis, Minnesota: University of Minnesota Press.
- MCGINN, Colin. 2006. “Hard Questions: Comments on Galen Strawson.” *Journal of Consciousness Studies* 13 (10–11): 90–99.
- NAGASAWA, Yujin, and Kai WAGER. 2017. “Panpsychism and Priority Cosmopsychism.” In *Panpsychism. Contemporary Perspectives*, edited by Godehard Brüntrup and Ludwig Jaskolla, 113–29. Oxford: Oxford University Press.
- NEMIROV, Lawrence. 1990. “Physicalism and the Cognitive Role of Acquaintance.” In *Mind and Cognition*, edited by William G. Lycan, 490–99. Oxford: Basil Blackwell Publishers.
- PEREBOOM, Derk. 2016. “Illusionism and Anti-Functionalism about Phenomenal Consciousness.” *Journal of Consciousness Studies* 23 (11–12): 172–85.
- . 2019. “Russellian Monism, Introspective Inaccuracy, and the Illusion Meta-Problem of Consciousness.” *Journal of Consciousness Studies* 26 (9–10): 182–93.
- ROELOFS, Luke. 2014. “Phenomenal Blending and the Palette Problem.” *Thought* 3 (1): 59–70. doi:[10.1002/tht3.113](https://doi.org/10.1002/tht3.113).
- . 2016. “The Unity of Consciousness, Within Subjects and Between Subjects.” *Philosophical Studies* 173 (12): 3199–3221. doi:[10.1007/s11098-016-0658-7](https://doi.org/10.1007/s11098-016-0658-7).
- . 2019. *Combining Minds: How to Think about Composite Subjectivity*. Oxford: Oxford University Press.
- ROSENTHAL, David M. 1993. “State Consciousness and Transitive Consciousness.” *Consciousness and Cognition* 2: 355–63. doi:[10.1006/ccog.1993.1029](https://doi.org/10.1006/ccog.1993.1029).
- SEAGER, William E. 1995. “Consciousness, Information and Panpsychism.” *Journal of Consciousness Studies* 2 (3): 272–88.
- SHANI, Itay. 2015. “Cosmopsychism: A Holistic Approach to the Metaphysics of Experience.” *Philosophical Papers* 44 (3): 389–437. doi:[10.1080/05568641.2015.1106709](https://doi.org/10.1080/05568641.2015.1106709).
- SHOTTEKIRK, Dena, Manuel CURADO, and Steven S. GOUVEIA, eds. 2019. *Perception, Cognition and Aesthetics*. *Routledge Studies in Contemporary Philosophy* 119. New York: Taylor & Francis.
- SMITHIES, Declan. 2019. *The Epistemic Role of Consciousness*. Oxford: Oxford University Press.
- STOLJAR, Daniel. 2001. “Two Conceptions of the Physical.” *Philosophy and Phenomenological Research* 62 (2): 253–81. doi:[10.1111/j.1933-1592.2001.tb00056.x](https://doi.org/10.1111/j.1933-1592.2001.tb00056.x).

- . 2006. *Ignorance and Imagination*. Oxford: Oxford University Press.
- . 2009. “The Argument from Revelation.” In *Conceptual Analysis and Philosophical Naturalism*, edited by David Braddon-Mitchell and Robert Nola, 113–38. Cambridge, Massachusetts: The MIT Press.
- . 2013. “Qualitative Inaccuracy and Unconceived Alternatives.” *Philosophy and Phenomenological Research* 86 (3): 745–52. doi:[10.1111/phpr.12030](https://doi.org/10.1111/phpr.12030).
- STRAWSON, Galen. 2006. “Realistic Monism: Why Physicalism Entails Panpsychism.” *Journal of Consciousness Studies* 13 (10–11): 3–31. doi:[10.1093/acprof:oso/9780199267422.003.0003](https://doi.org/10.1093/acprof:oso/9780199267422.003.0003).
- . 2015. “Self-Intimation.” *Phenomenology and the Cognitive Sciences* 14: 1–31. doi:[10.1007/s11097-013-9339-6](https://doi.org/10.1007/s11097-013-9339-6).
- THIEL, Udo. 2011. *The Early Modern Subject. Self-Consciousness and Personal Identity from Descartes to Hume*. Oxford: Oxford University Press.
- WILSON, Margaret Dauler. 1980. “Objects, Ideas, and ‘Minds’: Comments on Spinoza’s Theory of Mind.” In *The Philosophy of Baruch Spinoza*, edited by Richard Kennington, 103–20. *Studies in Philosophy and the History of Philosophy* 7. Washington, District of Columbia: The Catholic University of America Press.