

Panpsychism, intuitions, and the great chain of being

Luke Roelofs¹ · Jed Buchanan² 

Published online: 12 September 2018
© Springer Nature B.V. 2018

Abstract Some philosophical theories of consciousness imply consciousness in things we would never intuitively think are conscious—most notably, panpsychism implies that consciousness is pervasive, even outside complex brains. Is this a *reductio ab absurdo* for such theories, or does it show that we should reject our original intuitions? To understand the stakes of this question as clearly as possible, we analyse the structured pattern of intuitions that panpsychism conflicts with (what we call the ‘Great Chain of Being’ intuition). We consider a variety of ways that the tension between this intuition and panpsychism (or other counter-intuitive theories) could be resolved, ranging from complete rejection of the theory to complete dismissal of the intuition, but argue in favour of more nuanced approaches which try to reconcile the two.

Keywords Panpsychism · Intuitions · Consciousness · Other minds · Animal ethics

Philosophers often find themselves led by theoretical reasoning into views that are sharply—some might say wildly—counter-intuitive. What is the most sensible response to such situations? We discuss a particular instance of this general issue, involving the distribution of consciousness in nature: which beings are, and which are not, phenomenally conscious? How should we react to theories which imply that

✉ Luke Roelofs
luke.mf.roelofs@gmail.com
Jed Buchanan
buchaje@gmail.com

¹ Institut für Philosophie II, Ruhr-Universität Bochum, Universitätsstraße 150, 44780 Bochum, Germany

² Canberra, Australia

consciousness is present much more widely than we might have thought, in beings which it seems intuitively obvious are not conscious? In particular, how should we react to panpsychism, which states that all the fundamental physical entities are conscious? And what practical implications might such views have, particularly for those who think that moral status is importantly tied to consciousness? The sense of breathtaking implausibility that often greets panpsychists is rather like the “incredulous stare” which David Lewis famously reported from many colleagues in response to his views of possible worlds (Counterfactuals, Blackwell, Oxford, 1973, p. 86). Lewis famously added that he could not refute an incredulous stare; while we admit we cannot refute stares, we can perform a charitable exegesis of an incredulous stare, to identify the tacit ideas lying behind it and hopefully refute, re-interpret, or accommodate them. The paper is laid out as follows: in Sect. 1 we outline panpsychism and some motivations for it; in Sect. 2 we attempt to articulate the intuitions with which the theory conflicts; in Sects. 3 and 4 we explore the conflict between these two, and the different approaches that might be taken to resolving it. In Sect. 5 we consider but reject the option of dismissing the intuition outright. Finally, in Sects. 6 and 7 we outline our favoured approaches: Sect. 6 considers how versions of panpsychism that attributes consciousness to as few macroscopic things as possible can mitigate the conflict, while Sect. 7 argues that taking the relevant intuitions to track certain forms of consciousness, not consciousness itself, resolves it entirely. Section 8 closes by considering what these approaches might imply practically about veganism and vegetarianism.

1 Panpsychism

This paper takes panpsychism as a starting point rather than a conclusion, aiming to defuse an objection rather than explain or motivate the view itself. Nevertheless we will briefly sketch the main claims associated with panpsychist theories of consciousness, and the main motivations for accepting them.

We take the defining commitments of contemporary panpsychism (from Greek ‘pan-’ and ‘psuche’, meaning ‘all-’ and ‘mind’) to be the following: all fundamental physical entities are conscious, and complex physical things like humans have whatever consciousness they have in virtue of their fundamental parts being conscious (possibly together with other factors). In sum: matter is inherently conscious, and we are conscious because we are made of conscious matter. Determining what the ‘fundamental physical things’ are, is left to physics: if the final physical theory deals in particles then particles are the fundamental physical entities, and panpsychism says that every particle is conscious. If physics instead turns out to deal in waves, fields, strings, spacetime itself, or something else, then panpsychism says correspondingly that those are conscious. ‘Consciousness’ here is the property a creature has when ‘there is something it is like’ to be that creature (Nagel 1974; cf. Block 1995).

Panpsychists need not think that ‘everything’ is conscious, for many things are non-fundamental, and whether such things are conscious depends on how consciousness combines: on some versions of panpsychism (call them

‘universalist’), all composites of conscious parts inherit consciousness from those parts, while on other versions (call them ‘restrictive’) only a restricted subset of composites would.¹ The aim in either case is a theory which, though it posits many more conscious entities than rival theories, nevertheless aims to derive their existence from the simplest possible set of fundamental rules.

The most influential contemporary argument for panpsychism is the ‘explanatory’ argument. If a non-mental understanding of physical matter leaves an ‘explanatory gap’ between physics and consciousness, and all physicalistic attempts to ‘close the gap’ are abandoned, we should think of consciousness as a fundamental property additional to the fundamental physical properties. But then how is this fundamental property distributed? It seems more parsimonious for it to be systematically correlated with the fundamental physical properties, and thus omnipresent or nearly so, than for it to be correlated only with the specific kind of complexity that characterises particular organs of particular organisms. Panpsychists suggest that we should prefer fundamental laws that are simple and general, generating a variety of forms out of the same few basic elements, over ones that link one basic element to a specific complex structure.

There are other arguments in support of panpsychism, which we can only sketch here. Some authors see such theories as being better-placed to account for mental causation than any other non-physicalist view (e.g. Chalmers 2015); some see physical descriptions as failing to say anything about the ‘intrinsic nature’ of physical objects, and suggest that the most parsimonious view of this intrinsic nature is that it is the same intrinsic nature that we are aware of in our own brain states (e.g. Strawson 2006; Seager 2006).

An argument for panpsychism that is especially pertinent to this paper, is the ‘continuity’ argument (James 1890, pp. 147–148; Clifford 1897, pp. 60–61; Chalmers 1996, p. 297; Goff 2013; Mørch 2014, pp. 153–154). Scientific progress seems to show that there is nothing supernatural about humanity, no point in our development where we break with the rest of nature. If panpsychism were false, there should be a moment when the ‘most advanced’ non-conscious thing was succeeded by the ‘least advanced’ conscious thing, a moment when ‘the lights turn on’. Yet human evolution and foetal growth are so gradual that no candidate for such a discontinuous moment seems plausible.

Panpsychism is our focus, but the issues raised are not specific to it. Non-panpsychist theories can also yield counterintuitive implications about the distribution of consciousness. Integrated Information Theory, or ‘IIT’ (Tononi 2012; Oizumi et al. 2014), for instance, implies that almost all free-floating atoms

¹ Things are complicated here by the existence of views on which many or most sets of microscopic entities do not even compose wholes at all, so that there are in strictness no rocks, tables, cars, rivers, and perhaps no macroscopic objects at all except (some or all) living organisms, or not even those (Unger 1979; Van Inwagen 1990; Merricks 2001). When we speak of ‘panpsychists who attribute consciousness to all macroscopic composites’, we do not have in mind those who endorse a highly restricted view of composition, but attribute consciousness to all of the relatively few composites they do countenance: for our purposes such panpsychists can be treated together with those who accept lots of composites but attribute consciousness to only a few. Rather we have in mind panpsychists who regard all the macroscopic composites recognised by everyday language as both really existing and conscious.

and molecules are conscious, but that component parts of the human brain are not. Moreover, it implies that certain very simple-to-construct grids of logic gates are ‘more’ conscious, in some sense, than human beings (Aaronson 2014a, b). The view that Strawson (2006, p. 7) calls ‘micropsychism’, on which consciousness belongs to some but not all of the fundamental physical things, might imply that everything made out of normal matter is conscious but everything made out of anti-matter, even structural copies of us, is not (cf. Block 1992, p. 116). If some aggressively ‘chauvinistic’ version of the mind-brain identity theory is true, it might follow that aliens or androids with different physiology to ours would not be conscious, even those that are equally behaviourally complex—it may imply this for animals with very different brains, like cephalopods or even birds.² In essence, any theory which does not make functional structure wholly or partly constitutive of consciousness will either imply, or at least make a live possibility, that consciousness is present in things we would normally never think conscious, or absent in things which we would normally assume were conscious.³ But before deciding how to react to such implications, let us clarify what exactly the intuition(s) at work here are: what intuition makes panpsychism so unpalatable to so many?

2 The ‘great chain of being’ intuition

The intuitions that panpsychism conflicts with are most vividly illustrated when philosophers deny consciousness to some physical entities right as they introduce the term ‘consciousness’:

- “There is something it is like for me as I look at the bricks in the wall; there is nothing it is like to be a brick.” (Kirk 2003, p. 75)
- “An organism, such as a bat, is conscious if it is able to experience the outer world... There is... something it is like to be a conscious creature whereas there is nothing it is like to be, for example, a table or tree.” (Gennaro 2012)
- “the defining mark of a conscious organism is that ‘there is something that it is like to be that organism...’ There is something it is like to be a human being; you are experiencing it right now... By contrast, rocks, tables, and chairs lack

² The problem facing panpsychists and integrated information theorists is interestingly complementary to what Block calls the ‘harder problem of consciousness’ (Block 2002): for physicalists who think that both functional structure and material constitution are relevant to whether an entity is conscious, what possible evidence could be adduced to decide whether a being functionally just like us, but made of a very different material, is conscious? While Block worries about the possibility of beings whose behaviour suggests consciousness, who are not conscious in virtue of their constitution, the panpsychist is committed to the opposite: beings whose behaviour suggests non-consciousness, but who are conscious in virtue of their constitution.

³ Of course, theories which *do* make functional or behavioural structure wholly or partly constitutive of consciousness can also be threatened by counter-examples where intuitively non-conscious beings are implied to be conscious (the ‘Nation-Brain’ from Block 1992, being a well-known example). But these beings are generally hypothetical (and often implausibly contrived), whereas the problems that interest us in this paper involve real beings which we all have regular interactions with.

consciousness. There is nothing it is like to be such an entity.” (List 2018, p. 295)

- “What I mean by ‘consciousness’ can best be illustrated by examples. When I wake up from a dreamless sleep, I enter a state of consciousness, a state that continues so long as I am awake. When I go to sleep or am put under general anaesthetic or die, my conscious states cease.” (Searle 1992, p. 83)
- “Everybody knows what consciousness is: it is what vanishes every night when we fall into dreamless sleep and reappears when we wake up or when we dream.” (Tononi 2012, p. 290)

We take these five quotes as articulating a single intuition, though we should say exactly what we mean by ‘intuition’. We think of a *primary* intuitive judgement as one which appears plausible even before any further evidence or argument has been adduced in its support; this does not require that primary intuitive judgements *lack* supporting evidence or argument, only that they seem plausible—they ‘sound right’, they ‘strike us as true’—even when we cannot produce such evidence or argument. Perhaps primary intuitive judgements get their plausibility from some evidence or argument which we understand implicitly, so that when that reasoning is laid out we recognise it as ‘what we meant all along’. The point of calling the judgement intuitive is that we did not have to wait for the articulation of the reasons before finding it plausible.⁴ Thus we do not think of ‘intuitions’ as defined either by a special epistemology or by a special phenomenology(cf. Bealer 1998, p. 307 ff, Cappelen 2012, p. 7 ff).

It’s not clear that the above claims about trees, rocks, or walls are primary intuitive judgements in this sense. Thinking that rocks are not conscious is not mere prejudice: we can readily identify some reasons for it. Rocks don’t move on their own, they don’t respond to their environment in a goal-directed way, they don’t learn from the past, and so on. Humans, cats, bats, rats, and so on, do things which we take as evidence for consciousness, and when they don’t, we don’t attribute consciousness to them. So these judgements do come with articulable reasons. But if we took one of those reasons—e.g. ‘cats are probably conscious because they do X’—and asked ‘why is doing X good evidence for consciousness?’, the typical answer would simply be ‘the things I think are conscious tend to do X, and the things I think aren’t don’t.’ That is, the principles implicit in the reasons for our attributions are themselves dependent on the attributions, in that they are rendered plausible by their fitting those attributions. Perhaps some more fundamental justification is actually at work, and perhaps neurological research could provide independent evidence, but we can reasonably accept the general principle in advance of articulating that justification or discovering that evidence, just because it fits with the attributions we are inclined to make. We will call these particular

⁴ This definition is close kin to Chalmers’ (2014b) definition of intuitions as claims which are taken to be “dialectically justified... in a way that does not depend on an inferential, perceptual, memorial, introspective, or testimonial dialectical justification.” (2014a, b, p. 538). It differs in not being tied to a dialectical context where one person tries to support something to someone else.

attributions ‘secondary intuitive judgements’, meaning that they seem plausible just in light of reasons that which are themselves plausible just in light of fitting enough of the secondary intuitive judgements. Each judgement by itself can be backed up by compelling reasons, but the overall set of of judgements, though it hangs together, has the characteristic unsupported-plausibility of an intuition. When we speak of ‘the intuition’ that conflicts with panpsychism, then, we mean not any specific attribution or denial of consciousness, but the pattern of such judgements, each of which can be supported by reference to the others. This pattern of evidence seems to be challenged by panpsychism, among other theories, in a way that must be addressed by supporters of those theories.

Characterising that pattern precisely is tricky. It’s not simply that living things are conscious, for that would not explain the remark about trees. It’s also not the claim that non-animals are not conscious, since denying the *possibility* of a conscious plant or computer seems to go beyond what’s being said here. Capturing the intuition requires more than just a list of things that are definitely conscious or definitely unconscious: it requires capturing the structure of our uncertainty.

This structured uncertainty is nicely expressed by Nagel: “[although] we all believe that bats have experience... if one travels too far down the phylogenetic tree, people gradually shed their faith that there is experience there at all.” (1974, pp. 438) Between mature humans, who everyone is generally sure are conscious, and ‘rocks, tables, and chairs’, which everyone is generally sure are not conscious, there is a large region where we ‘gradually shed our faith’—that is, our subjective probability of a given creature being conscious steadily declines.⁵

One particularly explicit articulation of this pattern comes from Aaronson (2014b), who argues that the counter-intuitive implications of Integrated Information Theory (noted above) are sufficient to reject it. He identifies a number of ‘paradigm-cases’ associated with the term ‘consciousness’:

- You are conscious (though not when anesthetized).
- (Most) other people appear to be conscious, judging from their behavior.
- Many animals appear to be conscious, though probably to a lesser degree than humans (and the degree of consciousness in each particular species is far from obvious).
- A rock is not conscious. A wall is not conscious. A Reed-Solomon code is not conscious. Microsoft Word is not conscious (though a Word macro that passed the Turing test conceivably would be).

⁵ We are here taking for granted accurate understanding about the causal structure of the world—about what causes what. It has been suggested that humans have a systematic tendency to overestimate how many things act in an agentive way (sometimes called a ‘Hyperactive Agency Detection Device’, cf. Guthrie 1993; Barrett 2000)—to think, for instance, that a tree perceives our actions and does things in response to them, based on its private goals. The GCOB intuition is unrelated to this: it concerns whether we think a tree is conscious, once we know that its behaviour is non-agentive but can be explained as relatively simple reactions to specific environmental factors.

Fetuses, coma patients, fish, and hypothetical AIs are... the tougher cases, the ones where we might actually need a formal definition to adjudicate the truth. (Aaronson 2014b)

Nagel's and Aaronson's rankings of entities by the probability of their being conscious are both offered as merely stating what 'we' all tend to assume. If this pattern of intuitions really is so widespread, it would be good to have a term for it; we like 'the Great Chain of Being (GCOB) intuition'. This label evokes the traditional European idea of the 'scala naturae', often traced to Aristotle's rankings of different creatures by their 'degrees of vitality' (Aristotle c. 350 BCE, Book 8 Part 1, cf. Thorp 2016), which received its fullest analysis in Lovejoy's (1964) 'The Great Chain of Being: A Study of the History of an Idea.' Of course the traditional idea comes with baggage we're not taking on: the GCOB intuition says nothing about the metaphysics of form, matter, perfection, or divinity, nor need it cleave to the specific rankings given by Aristotle or others.⁶ It is simply a convenient label for the widely-shared pattern of judgements which (putting it roughly), ranks human above animals, animals above plants, and plants above inanimate things, in their likelihood of being conscious.

At the top of the ranking (we might say at 'level 1') are awake, adult, human beings like you the reader: if a philosopher denied the consciousness of such creatures they would likely be thought either deeply bizarre or insincere.⁷ Just below this point (at 'level 2') are beings which most of us would think probably conscious: this includes a lot of animals (birds and mammals at least, perhaps reptiles too), as well as humans in their first year of life, and adult humans during dreams. Here we can imagine a philosopher arguing (perhaps even with some persuasiveness) that the beings in question are not really conscious,⁸ but that conclusion would be very surprising.

Around the middle of the scale ('level 3') are beings most of us are unsure about: the majority of animals, from fish to flies to feather-stars; human foetuses at various stages, and human beings in neurologically impaired states like epileptic seizures, sleepwalking, or the 'minimally conscious state' (see Owen et al. 2006; Shea and Bayne 2010; Monti et al. 2010; Bayne et al. 2016; Peterson and Bayne 2017). We think that most people, asked whether there is anything it is like to be one of these

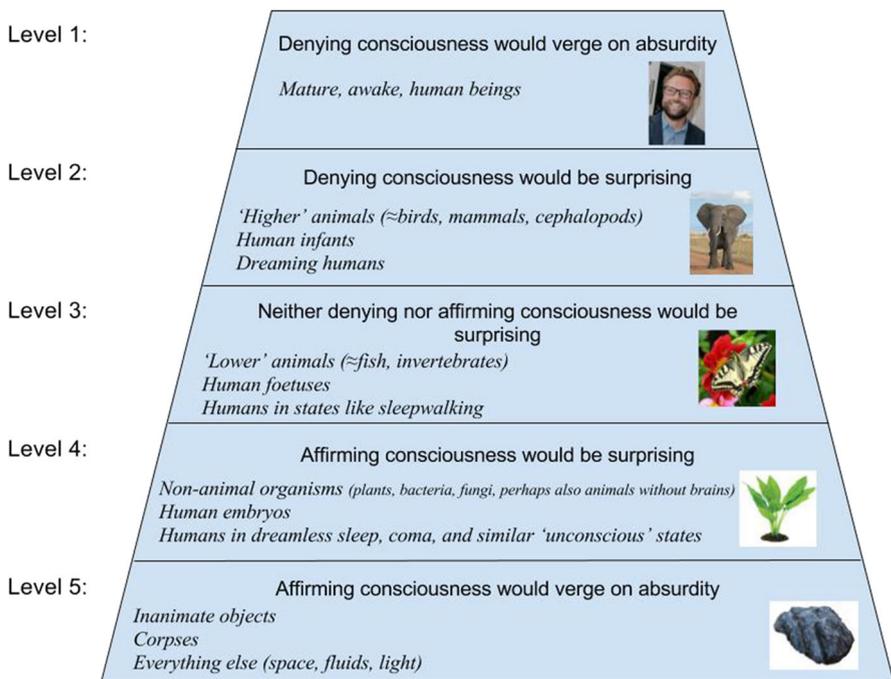
⁶ Ideas similar to the *scala naturae* can be found in other traditions, such as the Jain idea that all beings can be classified as one-sensed (the elements), two-sensed (molluscs et al.), three-sensed (certain insects and arachnids), four-sensed (other insects and arachnids), or five-sensed (vertebrates) (Donaldson 2015, p. 68). This particular ranking does not give rise to the same conflict with panpsychism, however, because insofar as all things have sense and soul (*jiva*), it is itself plausibly interpreted as panpsychist (Donaldson 2015, p. 61).

⁷ When the human being in question is oneself, of course, there is plausibly a special kind of certainty that is absent with others—though I may be sure that other people are conscious, it seems conceivable that I be wrong about them in a way that it is not conceivable that I be wrong about myself. (For a contrary view see, e.g., Carruthers 2011.)

⁸ For examples, consider accounts of consciousness which tie it closely to linguistic abilities, suggesting its absence in non-human animals (Jaynes 2000 [1976]; Dennett 1986). Consider also Dennett's suggestion that memories of dream-experiences might be systematically illusory (1976).

entities, would be genuinely unsure what to say: it is a live possibility that they are conscious, or that they are not.

Below this (at ‘level 4’) are beings most people assume are not conscious, but about which there might be a smidgen of doubt. This includes things which are alive but brainless, like trees, jellyfish, and bacteria, as well as human embryos and people in a deep, dreamless, sleep. Just like someone maintaining that animals or dreaming humans aren’t actually conscious, someone maintaining that plants or dreamlessly asleep humans *were* would be making a surprising claim, but not one that seemed beyond belief.⁹ Finally, right at the bottom (‘level 5’) are completely lifeless things (at least with the exception of hypothetical artificial intelligences; we take it that common-sense intuition yields no definite consensus on such beings’ consciousness). Rocks, raindrops, and radiators are all, intuitively, non-conscious: to ascribe consciousness to them would be about as outlandish as denying it to a human interlocutor.



⁹ For examples, consider Maher’s (2017) defence of plant minds, and more generally proponents of what Thompson (2007, p. 218) calls the ‘deep continuity thesis’, who hold that all living organisms have the germ of something akin to subjectivity (cf. Jonas 1966). Consider also the ongoing research into whether behaviourally unmanifested consciousness remains in patients in a ‘persistent vegetative state’ (e.g. Naci et al. 2017).

Although people vary on where they draw the line of consciousness on this 1–5 scale, the ordering is usually preserved.¹⁰ For example, it would be surprising for someone to hold that plants are conscious, yet fish are not. Equally, it would be surprising for someone who thought cats were unconscious, to also think fish were conscious. This is hardly a universal rule: Descartes is (in)famous for thinking both that nonhuman animals were not conscious (e.g. 1987, p. 140; cf. Cottingham 1978), and that sleeping humans were (1987, p. 247). But we suspect most people maintain the ordering outlined above: anyone who believes that things at level X are conscious, will tend to also believe that things at levels above X are also conscious.

3 Analysing the tension

The GCOB intuition seems to conflict with panpsychism on three points, one general and two specific. On a general level, many forms of panpsychism will conflict with the basic structure of the GCOB intuition, to the extent that they see consciousness as omnipresent while the GCOB intuition is about distinguishing conscious things from other things. Secondly and more specifically, any form of panpsychism entails that at least some things (e.g. electrons) at ‘level 5’ are conscious. And, thirdly, some forms (‘universalist panpsychism’) will add that all other macroscopic things, including rocks, chairs, tables, and others intuitively at ‘level 5’, are also conscious.

This conflict is more than purely theoretical; it is also practical. Consciousness is relevant to a being’s moral status: a conscious being is, just in virtue of being conscious, a being whose interests we must consider, a being who limits our moral freedom of action. They may not have equal moral status with us, particularly if they lack rationality, self-consciousness, or a capacity to recognise our moral status in return. But they count, for moral purposes, and if we can avoid harming them without great cost to ourselves, we should (Regan 1975, 1979; Singer 1990; cf. Mullin 2011).

This idea is particularly prominent as a rationale for ethical vegetarianism or veganism, doctrines which draw an important distinction between animals, on the one hand, and plants, algae, and fungi, on the other. The most straightforward and plausible basis for this distinction is that (many) animals are conscious (or as it is sometimes put, ‘sentient’), while members of other kingdoms are not. This clearly relies on the GCOB intuition. Insofar as panpsychism conflicts with the GCOB intuition, it seems to undermine one major rationale for ethical vegetarianism or veganism. If everything is conscious and conscious beings have moral status, everything has moral status: as well as not eating animals, we should not eat, or otherwise consume, plants, water, air, or anything else. This result seems unworkable in the extreme. So one task for panpsychists, as part of resolving the conflict between their view and the GCOB intuition, is to show that it does not have ruinous ethical implications.

¹⁰ We believe that this is true even if the boundaries of the levels that we have sketched out are defined differently: the point is that our intuitions about consciousness involve an ordinal ranking of confidence.

This tension can also be read as either epistemic or semantic. On the epistemic reading it is a matter of conflicting sources of evidence regarding a thesis which is at least *internally* consistent. The panpsychist hypothesis is supported by some theoretical argument, but pre-theoretical intuition counts against it, with however much epistemic weight that carries. On the semantic reading, the tension arises *within* the panpsychist thesis itself, because the GCOB intuition is inherent in the meaning of the term ‘consciousness’. Consequently, applying the term as widely as panpsychists do risks undermining its meaningfulness, or rendering panpsychism internally inconsistent. We think the tension can be successfully addressed on both construals, but that the semantic construal is easier to answer *for panpsychists*. Here we discuss how panpsychists can resolve the tension, construed semantically; in subsequent sections we consider their options for resolving it on the alternative epistemic reading.

The choice between the semantic and epistemic readings determines what is relevant to resolving the tension. On the semantic reading, what matters is how the term ‘consciousness’ has its reference fixed. On the epistemic reading, what matters is the trustworthiness of the two conflicting sources of evidence. We will first consider the semantic reading, which makes particular sense on the view that intuitions, in general, are always expressions of conceptual competence.¹¹ Aaronson provides a nice articulation of this reading, saying that:

When we consider whether to accept [some theory of] consciousness, we don’t start with any agreed-upon, independent notion of consciousness against which the new notion can be compared. The main things we start with, in my view, are certain paradigm-cases that gesture toward what we mean...

If, for example, our definition told us that [DVD players are conscious], that wouldn’t be a “surprise”; it would just be evidence that we’d picked a bad definition. The definition failed at the only task for which it could have succeeded: namely, that of capturing what we meant. (Aaronson 2014b)

Aaronson’s draws a parallel with the behaviour of scientific terms like ‘heat’: an account of consciousness which ascribes it to DVD players is like an account of heat which “delivered the shocking result that boiling water is actually colder than ice” (Aaronson 2014b). If ‘consciousness’ had first been introduced to cover a cluster of cases that people had encountered before knowing their exact nature, then this critique would make sense. ‘Panpsychism’ would then be on a par with ‘panaillurism’, the doctrine that all the fundamental physical entities are cats. If someone maintained that electrons, light, spacetime, and so on were all in fact cats, it would be fair to say: ‘the word ‘cat’ is only meaningful if it refers to a specific set of middle-sized entities, and distinguishes them from other middle-sized entities. If everything is said to be a cat, then ‘cat’ loses its meaning.’

¹¹ E.g. Ludwig (2010, p. 432) argues that “[philosophical] intuitions...are to be conceived of as judgements or beliefs which are the product of our competence in the deployment of the concepts involved.” Cf. Cappelen 2012, p. 9ff.

However, panpsychists can reply that this misunderstands the semantics of the term ‘consciousness’. We are directly acquainted with conscious states in our own case, and can straightforwardly define the term ‘consciousness’ as meaning ‘having states like *this*’ (cf. Chalmers 2014a, b; Goff 2015). This gives us a grip on what ‘conscious’ means that is independent of who or what else is conscious. If this is how we fix the meaning of the term ‘consciousness’, then the semantic reading of the tension between panpsychism and the GCOB intuition fails.

Of course, we saw in the last section that many authors do use contrasts (with trees, walls, etc.) to introduce the term ‘consciousness’ into discussion. But this does not show that they intend these contrasts to be *part of* the meaning of the term: they might just mean the contrasts to focus readers’ attention. By analogy, I might introduce the term ‘square’ by pointing at the one place on a screen where a square is visible, while saying ‘not here’, ‘not here’, when pointing at other places on the screen. But clearly it is no part of the *meaning* of ‘squareness’ that squares do not appear at those other places—indeed, I may know that there *are* squares there, but that the person I am teaching the word to cannot discern them. My aim is merely to direct their attention.

So is ‘consciousness’ defined in the manner of ‘cat’ or in the introspective manner? We take it that there is room for both views, for after all ‘consciousness’ is famously a ‘mongrel concept’, used by different people in different ways (Block 1995). But this is to the advantage of panpsychists, for all they need is *a* sense of ‘consciousness’ that is not semantically dependent on the structure of contrasts expressed by the GCOB intuition. And if there is any such sense, it is very plausible that ‘phenomenal consciousness’, the ‘raw feel’ of having a subjective point of view at all, would be it. Since it is ‘phenomenal consciousness’ that panpsychists generally talk about, they are well-placed to resist the semantic reading of their problem.¹²

4 Four ways to resolve the tension

Suppose we adopt the epistemic reading of the tension between panpsychism and the GCOB intuition. How should we resolve this tension? Obviously one question is whether the theoretical arguments for panpsychism are sound. But there are enough philosophy papers already on that topic (e.g. Van Cleve 1990; Chalmers 1995; Seager 1995; Nagel 2004; Strawson 2006; Carruthers and Schechter 2006). Another question is how trustworthy the GCOB intuition is, which depends on where it comes from. What psychological mechanisms explain our finding it intuitive to ascribe consciousness to some beings and deny it to others, and are those mechanisms generally a good guide to the truth?

¹² Indeed, Aaronson recognises this point when he distinguishes ‘consciousness’ from ‘apparent consciousness’, defining the latter as “the type of intelligent behavior that ought to lead reasonable people to infer the presence of consciousness.” (Aaronson 2014b, comment 37). A semantic objection could be mounted against the view that all things have ‘apparent consciousness’, but panpsychists do not claim that—they infer the presence of consciousness in the fundamental physical entities from theoretical considerations, not from their intelligent behaviour.

We may distinguish four ‘ideal types’ of resolution in conflicts between theory and intuition. At one extreme, we might uphold the intuition in question, based on what we will call a ‘vindicating’ explanation of its origin, and reject the theory. At the opposite extreme, we might disregard the intuition in question, based on what we will call a ‘debunking’ explanation of its origin, and re-affirm our theoretical conclusions. In between these extremes are options which adhere somewhat to both sides while partially amending one or the other, based on an explanation of the relevant intuitions which is intermediate between full vindication and full debunking.

Our distinction between ‘vindicating’ and ‘debunking’ explanations of an intuition turns on whether the explanation makes essential reference to what the intuition represents. Scientific accounts of vision are largely vindicating: in explaining how vision works we have to at some point mention the light that enters the eye from external objects, and the way that this light correlates with the boundaries, shape, and motion of external objects. Since these features of external objects are a major part of the content of visual experience, these explanations are vindicating: when we come to accept them as true, we should keep trusting visual perception.

By contrast, the explanation of someone hallucinating a pink elephant makes no reference to elephants, but only to states of sleep deprivation, brain chemicals, bright lights, etc., then the explanation is debunking: after accepting it as true we ought to downgrade, or entirely disregard, the pink-elephant-experience as evidence for pink elephants. (Note that things like brain chemicals will also be referred to in explanations of normal vision—it is the irrelevance of elephants, not the relevance of other things, that matters here.¹³)

In between fully vindicating and fully debunking explanations are ‘mixed’ explanations. One sort is what we will call ‘circumscribed’ explanations, which explain why a given faculty is trustworthy in one set of cases but not in another—i.e., why it is explained by what it represents in the former but not in the latter. For example, in the course of explaining human visual perception, we often find that various sorts of optical illusions reveal the conditions under which the perceptual system works well, and the conditions under which it ‘misfires’. Under white light, colour perception is (in some sense) veridical, while under monochromatic red light things will look the same colour which are actually different colours (e.g. red and white surfaces, green and black surfaces). When seeing things in a single medium, our perceptions of spatial relationships are good, but when seeing through a boundary between two media (e.g. water and air) those perceptions mislead.¹⁴ We

¹³ Compare Chalmers 2018, p. 29: “if there is an explanation of our beliefs about X that is independent of X, those beliefs are not justified.” What qualifies as a debunking explanation, and how far such explanations count against trusting an intuition, has received particular examination recently in the literature over whether evolutionary explanations of our moral intuitions undermine moral realism (see, e.g. Street 2006; Joyce 2006; Vavova 2015).

¹⁴ As these examples show, the categories of explanation being considered here are not strictly exclusive, and the same explanatory theory (the psychology and physiology of human vision) may count as vindicating, debunking, or somewhere in between depending on the context we focus on and on how

frequently encounter circumscribed explanations when investigating how human evolution has fitted us well or ill for modern life: our sense of taste, for instance, is generally a good detector of what is good to eat in the sort of environment we evolved in, where finding sufficient calories was a bigger challenge than finding sufficient micronutrients, but often misleading in a modern environment where calories are plentiful.

Another sort of mixed explanation is a ‘paraphrasing’ explanation, which shows a faculty to be trustworthy only when its content is reinterpreted as something other than what it at first seems. We appeal to something that does not match the *prima facie* content exactly, but is close enough that the faculty is still tracking something in the world (more resemblance, at least, than there is between an elephant and sleep deprivation). Examples from metaphysics are numerous (see O’Leary-Hawthorne and Michael 1996), such as Van Inwagen’s ‘paraphrase strategy’, explaining our intuitive acceptance of the existence of composite material objects like tables as really expressing no more than acceptance of “simples arranged tablewise” (Van Inwagen 1990, pp. 98–114). Or consider Haslanger’s account of ‘cool’ (1995, pp. 100–101): many of us can discern which people are ‘cool’ and which are not, and this seems at an intrinsic feature of those people. But when properly understood, our judgements of coolness are really tracking a social relation: whether a given individual’s self-presentation matches or violates the standards which permeate our social context. More significantly, Haslanger argues that our perceptions of race, for instance, track social relations of subordination rather than underlying genetic sameness (Haslanger 2000). The point is not that ‘nobody is really cool’, or that ‘nobody really belongs to any race’, or that our intuitions on these topics are worthless: they are very useful for navigating our social context, and in some cases insightful about social dynamics. They just aren’t tracking what they initially seem to be tracking.

Proponents of paraphrase explanations differ sometimes in how far they regard ordinary intuition to be strictly false, as opposed to ambiguous, and in how far they offer their explanation as a surprising ‘revision’ to ordinary beliefs. Van Inwagen, for instance, avoids saying that people are *wrong* when they talk about chairs and tables; this is why O’Leary-Hawthorne and Michael (1996) describe his approach as ‘semantic compatibilism’. On the other hand, social constructionists often emphasise the error involved in our usual beliefs about social and individual attributes, error that deserves searching and morally impactful critique. Generally, proponents of a paraphrase explanation must work harder, interpretively, the more they want to avoid convicting everyday intuition of error; Van Inwagen, for instance, has been criticised for the implausibility of the interpretations he puts on everyday statements (O’Leary-Hawthorne and Michael 1996; Merricks 2001,

Footnote 14 continued

finely we individuate things to be explained (e.g. ‘visual perception’ vs. ‘the seeing of this particular colour on this occasion’).

pp. 162–190). An approach which freely admits that everyday intuition is getting things wrong would be less vulnerable to this sort of critique.¹⁵

So when an attractive theory conflicts with intuitions, there are four ideal types of response:

Endorse the intuition and reject the theory for being counterintuitive, justifying this by offering a vindicating explanation of the intuition;

Endorse the theory and reject the intuition as mistaken or misleading, justifying this by offering a debunking explanation of the intuition;

Endorse the intuition and amend the theory in such a way that it no longer conflicts with the intuition (or conflicts as little as possible), justifying this by offering a mixed explanation of the intuition (which might be a circumscribed explanation or a paraphrase explanation); or,

Endorse the theory and re-interpret the intuition so that it no longer conflicts with the theory (or conflicts as little as possible), justifying this again by offering a mixed explanation, either circumscribed or paraphrase.

The most obvious response to the conflict between panpsychism and the GCOB intuition is category 1: reject panpsychism for conflicting with the GCOB intuition, and conclude that there must be something wrong with the theoretical arguments that motivate it. Since most people are not panpsychists, this could be called the most popular option. But for the sake of this paper, we set it aside, since our interest is in whether panpsychism can be successfully defended from the present challenge. In the next section we consider, and reject, a category 2 response; then in Sects. 6 and 7 we outline our respective preferred responses, which fall into categories 3 (understanding panpsychism to ascribe consciousness only to fundamental physical things and to those macroscopic things which fit the GCOB intuition) and 4 (taking the GCOB intuition to be a good guide to which things have consciousness *similar to ours*, not to which things have consciousness *per se*).

5 Dismissing the GCOB intuition entirely

Perhaps we should just forget the GCOB intuition, trusting in theoretical argument over everyday intuition and refusing to grant any significant epistemic weight to the fact that ‘we’ (or ‘people’, or just ‘oneself’) find it intuitive that a given being does or does not have consciousness. Giulio Tononi articulates this position in response to Aaronson’s criticism (discussed above):

...it can be dangerous to rely too much on one’s pre-theoretical intuitions, however strong they may seem. Examples in science are numerous, starting

¹⁵ Thus after criticising Van Inwagen, O’Leary-Hawthorne and Michael note that, “So long as it is understood that the project of answering the tracking question is importantly different to that of doing justice to the content of ordinary talk, we fully endorse the deployment of ‘paraphrase’ strategies with regard to the former.” (1996, p. 134).

with the strong intuitions people once had that the earth must be still and the sun must revolve around it...

Concerning consciousness, the reliability of pre-theoretical intuitions is even worse, because different people often hold radically different ones. Most of us agree that, since we ourselves are undeniably conscious, people who are built and behave like us are likely conscious, which is a good start. But from there intuitions diverge. Faced with an unresponsive patient, some people's intuition suggests that the patient is conscious... while others are convinced that there is nobody home... (Tononi 2014, p. 3, cf. also Goff Forthcoming XX)

This response has some initial appeal. Why think that the truth about consciousness, or any other abstruse topic, conforms neatly to the inherited gut feelings of this particular ape species? And why not expect that the full explanation of where the GCOB intuition comes from will mention only the quirks of our evolution, not the true distribution of consciousness? All sorts of 'intuitive' ideas have fallen by the wayside of scientific progress, so why expect our current intuitions to hold up?

But this is a much harder line to take with intuitions about consciousness than it is with intuitions about, say, which celestial bodies orbit which. The GCOB intuition is not simply our gut response to something unfamiliar and arcane: it is a deliverance of our 'mindreading' faculty, a faculty which we use every day and rely on to navigate our lives.

That is, the same faculty that allows us to track, detect, anticipate, and share the mental states of other people, is also what generates the GCOB intuition. We direct our attention onto a walking, talking, person, ask ourselves what they are feeling and thinking, and various answers spring immediately ('intuitively') to mind: they're happy, they haven't seen me yet, they're thinking carefully, etc. We direct our attention in the same way to a rock, asking ourselves what it is thinking and feeling, and get the immediate, 'intuitive', answer: 'nothing'.

This means that we cannot dismiss attitude toward the GCOB intuition without this dismissal extending to everyday mindreading. A straightforwardly debunking explanation of the GCOB intuition would undermine our basis for ascribing consciousness to anyone but ourselves.¹⁶ It is hard to see how we could simultaneously dismiss our intuitions about the extension of consciousness from informing our theories about it, yet continue to rely on those intuitions to guide our

¹⁶ It is important here that the GCOB intuition arises from mature mindreading capacities, employed thoughtfully. It is quite plausible that infant mindreading capacities, or mature ones employed quickly or thoughtlessly, might be very fallible—in particular, evidence from developmental psychology suggests that we are naturally disposed to attribute something like intentionality to anything that displays a human face, or spontaneous motion (see Carey 2009, pp. 157–215). But the very fact that we see such attributions are fallible shows that, with mature faculties and a better grasp of how the world works, our everyday capacities can identify and correct these mistakes.

interactions with other people, with animals, and with the various ‘inanimate objects’ (as we intuitively think of them) that we may ethically impinge upon to go about our lives. Suppose my friend falls down, their leg is bleeding, and they are crying; it seems obvious to me that they are in pain. After all, I am very confident that people often cry when they are in pain, that falling down tends to cause pain, and so on. But am I more confident in these utterly commonplace generalisations than I am in ‘toasters feel nothing’, or ‘there is nothing it is like to be a table’, or ‘rocks are not conscious’? All of these generalisations seem similar in their initial plausibility, and if I regard the latter as untrustworthy why should I trust the former?¹⁷

So it seems to the authors that, even if many intuitions are untrustworthy, our intuitive understanding of which things have which mental states must, on pain of ruinous scepticism, be generally trustworthy. But then, if the GCOB intuition is part of that intuitive understanding, we cannot simply dismiss it out of hand: we must accommodate it somehow. One author believes that option 3 (changing theory to fit intuition) will provide a useful partial mitigation of this tension: the GCOB intuition is right about macroscopic things, and wrong primarily about fundamental physical entities. Moreover, both agree that option 4 (reinterpreting intuition to fit theory) provides a substantive resolution: the GCOB intuition is right about certain *sorts* of consciousness, but not about consciousness per se. The next two sections outline these two options.

6 Restrictive panpsychism and a circumscribed vindication of the GCOB intuition

A third option for panpsychists is to vindicate as much of the GCOB intuition as possible by accepting restrictive panpsychism, where only some collections of physical simples exhibit mental combination, and thus macro-level consciousness. One author believes that the GCOB intuition is a good guide to which of these are conscious: (some) animals and humans yes, rocks and tables no, insects maybe, and so on.

As Mørch puts it, “panpsychism [ascribes consciousness] to all things that are... either fundamental or otherwise properly unified according to a principle yet to be determined” (2014, p. 3), later elaborating that macro-level consciousness belongs to “humans and animals and any other properly unified complex physical things.”

¹⁷ Someone might try to get around this practical skepticism by appealing to a sort of ‘precautionary principle’ (cf. Matheson 2016): the potential negative consequences of ignoring a friend’s pain are worse than the negative consequences of helping an automaton, and this asymmetry might be enough to justify certain sorts of response. This is a reasonable line of thought (it may, in particular, be a good reason not to eat, e.g., insects, when one is genuinely in doubt about their consciousness and associated moral status). But it cannot work by itself. After all, consider my table: if it suffers agonising pain as long as it is not painted red, the negative consequences of not painting it red are worse than the negative consequences of unnecessarily painting it red. But if we are not to be talked into doing infinitely many bizarre actions ‘just in case’, we need to be able to say at some point ‘the probability of that being true is too low for us to act on it, even in a precautionary spirit’.

(Mørch 2014, p. 39). This leaves open that there might be no other ‘properly unified’ things than humans and animals, or at least that the only other examples might be things like intelligent androids, which the GCOB intuition says nothing about. This would yield a version of panpsychism on which:

- The fundamental physical entities are conscious;
- Waking humans and animals are conscious composites of conscious parts;
- Inanimate things are non-conscious composites of conscious parts;
- Plants, sleeping humans, bacteria, and some animals may be either conscious composites of conscious parts, or non-conscious composites of conscious parts.

Thus with the exception of the fundamental physical entities, the GCOB intuition is vindicated. Our intuition that rocks aren’t conscious is correct, just with the caveat that they are nevertheless composed entirely of conscious parts.

The big question for this kind of panpsychist is: under what conditions do conscious parts compose a conscious whole? Unfortunately, this question is hard to answer with any precision. Shani, for instance, says that a system is conscious when it is “a cohesive system with a characteristic organization, or form, maintained through dynamic balance between opposing forces and tendencies.” (2015, p. 416) But this does not really answer the question: any system can be said to have some sort of ‘characteristic organisation’ so what distinguishes the organisation found in a crystalline lattice from that found in a vertebrate brain? And any answer which is clear and precise risks violating the GCOB intuition in its own ways. For instance, consider the following three possible principles of unity:

First, maybe biology plays the unifying role (cf. Van Inwagen 1990): rocks and tables and planets and the oceans and the United States of America are not conscious, while people and dogs are. Unfortunately, this account tells us that *all* biological organisms are conscious, including sea sponges and bacteria, so the desired vindication of the GCOB intuition is quite limited. Perhaps worse, it implies that no inorganic system could be conscious, a result that strikes many as counter-intuitive (e.g. Gray and Wegner 2012, cf. Turing 1950.).

Second, maybe ‘information integration’, as modelled mathematically by Tononi’s IIT, plays the unifying role. But, as we have already noted, IIT has its own deviations from the GCOB intuition: it implies, for instance, that a sufficiently large grid of logic-gates is more conscious than a human being, despite not exhibiting anything like intelligent behaviour (cf. Schwitzgebel 2014).

Third, maybe conscious wholes are marked out by distinctive, strongly emergent, causal powers (rather like the account of composition defended by Merricks 2001, 2005). There are a number of candidates for what these emergent causal powers could be, though all of them are controversial. Merricks believes that free will is a strongly emergent causal power present in humans, but free will skeptics could still endorse other options. Intentionality might count as a strongly emergent causal power, if one had an account of how intentional directedness goes beyond the stimulus-responses of unconscious systems. This might allow us to ascribe consciousness to systems which seem like us in their causal powers, regardless of how they are constructed. But without knowing which systems in fact have strongly

emergent causal powers, we do not know what this view would imply about the distribution of consciousness (indeed, human beings may well have no strongly emergent causal powers at all). So we have no idea in advance how far this would vindicate the GCOB intuition.

A further problem for any way of specifying the conditions for mental combination is the problem of discontinuity. Part of the appeal of panpsychism is that it cements the continuity of nature, by saying that there is no hard delineation between conscious things like humans and the rest of nature. This continuity is attractive because it leaves the panpsychist less vulnerable to accusations of arbitrariness and epiphenomenalism (cf. Goff 2013). Restrictive panpsychism appears arbitrary because it is hard to see how there could be any satisfying reason why the boundary should lie wherever it lies, and not fractionally above or below it. Take Shani's criterion of being "a cohesive system" which maintains "a characteristic organization": how cohesive is cohesive? The most cohesive things still sometimes disintegrate, and even a heap of sand has some degree of cohesion (from friction, gravity, etc.). Wherever we place the threshold between 'not cohesive enough' and 'just cohesive enough', it will seem equally plausible that it could have been a smidge to one side or the other of that point.¹⁸

The worry about epiphenomenalism builds off this: for any precise threshold in (say) cohesiveness, or functional complexity, or integration of information, the difference in overall behaviour between a system just below it (supposedly non-conscious) and a system just above it (supposedly conscious) will be miniscule—at least if we make the threshold genuinely precise.

Restrictive panpsychism rests on treating two aspects of common-sense intuition very differently: on the one hand, it endorses and relies upon our intuitive sense that only 'appropriately unified' beings can be conscious, but on the other it rejects our intuitive sense that only 'intelligently behaving' beings can be conscious. Both factors contribute to our intuition that rocks are not conscious (they are neither unified nor intelligent), but the latter alone seems responsible for our sense that particles are not conscious (they are unified but not intelligent). A full defence of restrictive panpsychism would need to explain why the mechanisms producing the one set of intuitions are trustworthy but the others are not.

The authors differ on whether the theoretical costs of restrictive panpsychism are worth paying, for the circumscribed vindication of everyday intuition they may buy. But whether or not panpsychists restrict mental combination, they will be denying some parts of the GCOB intuition—the restrictive panpsychist still denies that unintelligent things like particles always lack consciousness. Since it is unsatisfying to simply dismiss these intuitions, panpsychists should look for a way to understand

¹⁸ Of course there are boundaries in the world, but for most macro-level features, these boundaries are vague: between what is determinately X and what is determinately not-X, there are things which could be called X or not-X with roughly equal justification. But the most plausible analyses of vagueness tend to trace it to the concept being applied having many slightly different ranges of application, which must themselves be specified as different precise descriptions in some more basic terms. But if consciousness is a fundamental property, this sort of analysis is not available.

these intuitions that explains both their fallibility in some cases and their trustworthiness in others.

7 Paraphrasing the GCOB intuition to better fit panpsychism

The core idea of our final proposal is simple: our intuitions about consciousness are not tracking consciousness per se, but rather a particular type of consciousness. Our sense that consciousness belongs definitely to humans, probably to various animals, and definitely not to plants and rocks accurately reflects something, namely that a certain type of consciousness belongs definitely to humans, probably to various animals, and definitely not to plants and rocks.

This is not the blanket anti-intuition proposal discussed in Sect. 4: because the GCOB intuition tracks a certain type of consciousness, the attributions it generates are reliable *when they are positive*. Having some particular type of consciousness entails being conscious, so if some being seems conscious to us (due, by hypothesis, to its having a particular type of consciousness), it is reasonable to think it really is conscious.¹⁹ But when something intuitively seems non-conscious to us, that intuition should not be trusted: it might be non-conscious, or it might simply have consciousness of a sort we are not equipped to detect. We may express this more explicitly thus:

Asymmetry principle Our ‘positive’ intuitions (that something is definitely or probably conscious) are generally reliable, but our ‘negative’ intuitions (that something is definitely or probably not conscious) are generally unreliable.

This proposal needs to be fleshed out. It is easy to say that our intuitions track only a particular type of consciousness, but we need to say what type that is, and why our intuitions track it alone. We believe that the best thing to say is that our intuitions are not tied to any pre-set type of consciousness, but rather to experiences that are, to a greater or lesser extent, *familiar* to us from our own case. We can intuitively recognize something as experiencing things like pain, attention, contentment, anxiety, desire, and so on, because we have ourselves undergone such experiences. But the outward signs of a completely unfamiliar experience will not be registered by our intuitions. Call this the ‘Invisibility of Unfamiliar States’:

Invisibility of unfamiliar states (IUS) Our intuitions about consciousness in other beings are sensitive only to conscious states of types we are personally familiar with.

We think that IUS is an independently plausible idea: it dovetails with Nagel’s point that we cannot understand experiences too alien to our own (Nagel 1974), and with the long tradition of understanding the basis for our belief in other minds as being

¹⁹ There is thus a sense in which this section’s proposal is ‘circumscribing’ in character, saying a certain subclass of the relevant intuitions are trustworthy; we present the proposal as a ‘paraphrasing explanation’ because this circumscription follows from the claim that the whole set of intuitions is tracking something other than what it presents itself as tracking.

some sort of analogy with or extrapolation from ourselves (e.g. Mill 1889, cf. Hyslop and Jackson 1972; Heal 2003; Steuber 2006). It fits neatly with empirical evidence suggesting that people routinely over-estimate how similar others are to themselves (e.g. Camerer et al. 1989), and with the familiar observation that without undergoing certain experiences ourselves, we have difficulty recognising and understanding others who have.

Of course IUS needs to be carefully qualified. For one thing, experiences differ from and resemble each other in many ways, and someone's experience can be unfamiliar in some ways but familiar in others. As Nagel points out, even if the distinctive qualities of bat sonar are beyond our understanding, we can still understand that the experience of sonar is likely to be a *perceptual* experience, rather than an emotional or cognitive one, because we ourselves enjoy perceptual experiences (Nagel 1974, pp. 339–340). Moreover, an experience's familiarity makes detectable only *together* with factors like how much effort we make, or how much skill we deploy. Experiences that are very familiar to ours are the easiest for us to discern; experiences that are less familiar but still within the human range may remain invisible if we are lazy or unskilled in trying to discern them. Experiences very different from our own might require sustained effort to recognise, perhaps involving their subject actively explaining to us, step by step, exactly how their experiences resemble ours and how they differ. And someone who has spent years observing a particular animal may become able to pick up on subtle cues that reveal something of the richness and variety of its inner life, even if that person remains unable to fully understand them.²⁰

Let us call experiences that are completely unfamiliar to us, in every respect but their being conscious at all, 'alien' experiences.²¹ IUS then suggests that our intuitive faculties for detecting conscious states will not register alien experiences at all, even when something's behaviour is manifesting them right in front of us. Let us call a conscious being all of whose experiences are 'alien', in this sense, an 'alien being'. IUS says our intuitions are no use for identifying what an alien being is experiencing.

IUS is a first step towards explaining and justifying the Asymmetry Principle, but a second step is needed. For all that IUS says, our minds might have two entirely independent mechanisms, one for detecting particular conscious states and one for detecting consciousness itself. IUS says something about the first, but leaves open that the second might be independent of familiarity, and thus still able to detect consciousness in an alien being. Then panpsychism would still face the original objection: if consciousness were widespread, we would discern this intuitively, but

²⁰ Although IUS fits quite neatly with the 'simulation theory' of how mindreading works, and with Nagelian worries about reductive physicalism, it does not by itself imply either, and is compatible with the denial of both.

²¹ Perhaps all possible experiences have certain features in common, beyond just being conscious. In that case, our talk of 'alien experiences' should be read as 'maximally unfamiliar experiences'. For this to undermine our argument, there would have to be a universal feature of experience which is always detectable by our mind reading faculties, and the burden is on our opponents to explain what this feature could be.

we do not. So the second step in panpsychism's defence is to deny a faculty for detecting consciousness independently of particular experiences; call this denial 'epistemic atomism'²²:

Epistemic atomism (EA) Our intuitions about consciousness in other beings arise primarily from detecting particular experiences, and only on that basis inferring the presence or absence of consciousness per se.

EA might at first seem false. Don't we sometimes interact with someone and feel that, though we have no idea what they are feeling, we are sure they are feeling *something*? And don't we sometimes think, observing a strange non-human creature, that there is 'something going on' inside them, though we have no idea what? But it is an overstatement to say that in such cases we have *literally* no idea what the other is experiencing. We distinguish such cases, after all, from cases where a creature is dead or asleep, and we do this by recognising their movements as intentional, and as responsive to their surroundings in a way that suggests perceptual awareness. Having intuitively recognized sensing and willing, we suppose (by analogy with our own case) that something mediates between them—plans, desires, beliefs, etc. While we have no direct evidence of these mediating states, we feel there *are* some because we have intuitively recognised sensation and will, which we know from our own case need something to connect them. This still relies on our capacity to intuitively recognize particular sorts of conscious states; if the being in question did nothing that struck us as perceptual response or intentional action, we would not feel that there was an inner life here whose other contents were mysterious.

If IUS and EA are accepted, then the Asymmetry Principle follows. An alien being will (by IUS) not intuitively seem to us to have any particular conscious states, and will therefore (by EA) not intuitively seem to us to be conscious at all. Since both non-conscious beings and alien conscious beings will fail to register as intuitively conscious, our negative intuitions about a creature's consciousness are not a reliable sign of its lack of consciousness.²³

On this proposal, the GCOB intuition is a decent guide to *something*, but not the presence and absence of consciousness: rather, it tracks how far a being's consciousness resembles our own, and how likely a being is to have consciousness like ours. It is trustworthy, but only if we paraphrase it into saying something other than what it seems to say. Thus understood, it is no threat to panpsychism: panpsychists can agree that a fly's consciousness is probably more different from ours than a cat's, but more similar to ours than a plant's, and that for any given familiar mental state we undergo, it is (other things being equal²⁴) more likely that

²² This sort of 'atomism' about the epistemology of other minds is independent of any sort of phenomenological or metaphysical atomism: the fact that we identify component states, not total states, first does not imply that they in any sense 'come first' in reality.

²³ Moreover, IUS and EA imply that when it comes to familiar sorts of conscious state, our intuitions are reliable, both positive (that a being is in a state of a that sort) and negative (that it is not).

²⁴ Obviously, the specifics of the state and the creature will make a difference: it is more plausible that social animals share our specifically social emotions than that solitary ones do, more likely that animals with colour vision share something like our visual experiences than that blind ones do, and so on.

cats share it with us than that flies do, but more likely that flies share it with us than that plants do. This is compatible with believing, on theoretical grounds, that cats, flies, plants, and everything else is conscious in some fashion.

8 What are the moral implications?

In the preceding sections we have considered the apparent clash between the CGOB Intuition and panpsychism, or any other theory which ascribes consciousness very widely in the physical world. We considered four ways this conflict could be resolved: abandoning panpsychism, abandoning the GCOB Intuition, amending panpsychism to no longer clash with the intuition, and interpreting the intuition as a trustworthy guide to something other than its apparent content. Having defended the latter two, we should finally consider their ethical implications.

Section 7's proposal, that everything is conscious but our intuitions track only 'familiar' sorts of consciousness, might seem to leave vegan panpsychists in an awkward position. If plants are conscious, doesn't the principle of respect for conscious beings demand avoiding harm to them just as much as to animals? Section 6's proposal, that everything macroscopic is made of conscious parts but mostly does not inherit their consciousness, might seem to dodge that particular bullet, since it does not imply consciousness in plants. But it still posits consciousness in fundamental particles, which raises its own puzzling ethical issues (is it wrong to cause some particles to be absorbed, destroyed, or turned into others?). In fact, we believe that panpsychism need have only subtle, rather than catastrophic, ethical implications, and the last section's argument for the Asymmetry Principle shows how.

The significance of consciousness in debates about moral status is more closely tied to particular sorts of conscious state rather than to consciousness per se (cf. Lee 2018). For classical utilitarians, the relevant states are pleasure and displeasure (hence Bentham's famous line, "the question is not, Can they reason? nor, Can they talk? but, Can they suffer?" 1996, p. 283); for preference utilitarians, the relevant states are desire and aversion. For rights theorists, the relevant states are more complicated, but plausibly they involve the structure of hedonic, cognitive, and epistemic experiences that Regan calls 'being the subject-of-a-life' (1983, pp. 242–243).

States like this are ones that we know, and appreciate the significance of, from our own case. Even though there are many unfamiliar pleasures, desires, etc., the very fact that they are pleasures, desires, etc. makes them to that extent familiar. And IUS and EA imply that when it comes to familiar sorts of conscious state, our intuitions are reliable, both positive (that a being is in a state of that sort) and negative (that it is not). If a being is undergoing pain, pleasure, fear, or striving, the fact that it has those states is something that our intuitive mindreading capacities should, with enough effort, detect. To the extent that intuition judges a given being completely unconscious, that is evidence that, even if it is in fact conscious, it does not undergo anything recognizable as suffering, seeking, or other morally significant states.

Of course, what is detectable need not be *easily* detectable: morally relevant conscious states might still require much careful work (imaginative, physiological, ethological, phenomenological, or social) to recognize and understand. But we knew that already: the point is that panpsychism by itself neither performs this work nor dooms it to failure. And if someone judged that the preponderance of evidence supported consciousness in most animals but not in plants, and chose to eat the latter but not the former, their conversion to panpsychism would not undermine the reasonableness of that decision.

Some worries remain. Couldn't there be other sorts of conscious states, alien to us, but morally important in the same way as hedonic and cognitive states? The authors see no way to rule this out, but they equally see no way to rule out a similar possibility about morally important sorts of non-conscious state that are completely unrecognisable to us. Maybe ultimately all that matters is the number of salt crystals in existence. But clearly there is nothing sensible we can do about such possibilities, and so we act reasonably in proceeding as if they were false: likewise panpsychists act reasonably in proceeding as if the only morally relevant sorts of consciousness are those which we can, at least potentially, intuitively recognise.

But what if it is wrong to destroy a conscious being whatever kind of states it has? Here vegan panpsychists have a few options. They might just say 'no': consciousness that is not organized around the pursuit of goals, or the perception of some things as pleasant and others as unpleasant, is morally neutral to destroy, disrupt, or otherwise interfere with. This would be the most conservative option, requiring the least revising of moral judgements.

Alternatively, panpsychists might accept that destroying consciousness is always bad, but point out that since corpses and ashes and piles of rubble are all themselves conscious, the destruction of a conscious being need not mean that consciousness ends and passes into non-consciousness. In a panpsychist world, all that can happen is a change in the form of consciousness, and one might think that that is morally neutral as long as no hedonic, conative, or epistemic states are involved. What would normally be regarded as a destruction of consciousness (e.g. the death of an animal), is to a panpsychist a transformation from one form of consciousness to another. The practical implications of this option, far from constraining our actions more tightly, actually free us from the thought that biological death is intrinsically an evil, because it is *not* an absolute cessation of consciousness. Death is only bad, this option suggests, because of the pleasure it deprives us of, the preferences it frustrates, and the grief it causes to others. Readers must decide for themselves whether this result is welcome or worrying.

Finally, panpsychists might grant that all conscious beings have some moral status (even if that includes plants, rocks, rivers, and everything else), which counts against destructive interference with them, but insist that the moral reasons thus provided are usually outweighed by those provided by the hedonic, conative, and epistemic experiences of animals. That is, they might say that it is morally better not to kill a plant than to kill one, but that animals' need to sustain their richer sort of life justifies killing plants for food. The main practical implication of this option would be that large complexes of plant life, and other natural phenomena, have their own intrinsic moral value—a view already countenanced by some environmentalists

(e.g. Leopold 1949, Callicott 1984, 1989), including some panpsychists (Matthews 2003).

The authors take no stand on which of these three options is best: rather, they point out that all three are already live options, already familiar positions in moral debate. Panpsychism can influence how we think about these questions, but by itself it neither answers them nor renders them pointless. Just as panpsychists still need to evaluate the evidence for (morally relevant states of) consciousness in non-human organisms, they still need to evaluate the arguments for and against these different views about what, ultimately, matters.

References

- Aaronson, S. (2014a). *Why I am not an integrated information theorist*. <http://www.scottaaronson.com/blog/?p=1799>. Accessed 6 Sept 2018.
- Aaronson, S. (2014b). *Giulio Tononi and me: A Phi-nal exchange*. <http://www.scottaaronson.com/blog/?p=1823>. Accessed 6 Sept 2018.
- Aristotle. (c. 350 BCE). *Historia animalium*.
- Barrett, J. L. (2000). Exploring the natural foundations of religion. *Trends in Cognitive Sciences*, 4(1), 29–34.
- Bayne, T., Hohwy, J., & Owen, A. (2016). Are there levels of consciousness? *Trends Cogn Sci.*, 20(6), 405–413.
- Bealer, G. (1998). Intuition and the autonomy of philosophy. In M. DePaul & W. Ramsey (Eds.), *Rethinking intuition: The psychology of intuition and its role in philosophical inquiry* (pp. 201–240). New York: Rowman & Littlefield.
- Bentham, J. (1996). In J. H. Burns & H. L. A. Hart (Eds.), *An introduction to the principles of morals and legislation*. Oxford: Clarendon Press. (originally published 1789).
- Block, N. (1992). Troubles with functionalism. In C. W. Savage (Ed.), *Perception and cognition: Issues in the foundations of psychology*. Minneapolis: University of Minneapolis Press ((1978). Reprinted with revisions In B. Beakley and P. Ludlow (Eds.), *The philosophy of mind: Classical problems/contemporary issues*. MIT Press).
- Block, N. (1995). On a confusion about a function of consciousness. *Behavioral and Brain Sciences*, 18(2), 227–287.
- Block, N. (2002). The harder problem of consciousness. *The Journal of Philosophy*, 49, 1–35.
- Callicott, J. Baird. (1984). Non-anthropocentric value theory and environmental ethics. *American Philosophical Quarterly*, 21(4), 299–309.
- Callicott, J. Baird. (1989). *In defense of the land ethic*. Albany: SUNY Press.
- Camerer, C., Loewenstein, G., & Weber, M. (1989). The curse of knowledge in economic settings: An experimental analysis. *Journal of Political Economy*, 97, 1232–1254.
- Cappelen, Herman. (2012). *Philosophy without intuitions*. New York: Oxford University Press.
- Carey, Susan. (2009). *The origin of concepts*. Oxford: Oxford University Press.
- Carruthers, P. (2011). *The opacity of mind: An integrative theory of self-knowledge*. Oxford: Oxford University Press.
- Carruthers, P., & Schechter, E. (2006). Can panpsychism bridge the explanatory gap? *Journal of Consciousness Studies*, 13(10–11), 32–39.
- Chalmers, D. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies*, 2(3), 200–219.
- Chalmers, D. (1996). *The conscious mind: In search of a fundamental theory*. Oxford: Oxford University Press.
- Chalmers, D. (2014). Comment 36 on Aaronson 2014-b. <http://www.scottaaronson.com/blog/?p=1823#comment-108216>. Accessed 6 Sept 2018.
- Chalmers, D. (2014b). Intuitions in philosophy: A minimal defence. *Philosophical Studies*, 171(3), 535–544.

- Chalmers, D. (2015). Panpsychism and panprotopsyism. In Y. Nagasawa & T. Alter (Eds.), *Consciousness in the physical world: Essays on Russellian monism*. Oxford: Oxford University Press.
- Chalmers, D. (2018). The meta-problem of consciousness. *Journal of Consciousness Studies*, 25(9–10), 1–41.
- Christian List. (2018). What is it like to be a group agent? *Noûs*, 52(2), 295–319.
- Clifford, W. K. (1897). Body and mind. *Lectures and essays by William Kingdon Clifford*, VI, by Clifford, William Kingdon, 1845–1879; Stephen, Leslie, Sir, 1832–1904; Pollock (pp. 1845–1937). Sir: Frederick.
- Cottingham, J. (1978). A brute to the brutes? *Philosophy*, 53(206), 551–559.
- Dennett, D. (1976). Are dreams experiences? *The Philosophical Review*, 85(2), 151–171.
- Dennett, D. (1986). Julian Jaynes's software archeology. *Canadian Psychology*, 27(2), 149–154.
- Descartes, Letter To Plempius, 3 October 1637 (AT I 415; K 37); and Letter to More of 5 February 1649: AT V 276; K 243 ('K' stands for A. Kenny, Descartes' Letters Philosophical (Oxford: Clarendon, 1970)).
- Descartes, The Second Meditation, CSM II 18; AT VII 27. My references here are to The Philosophical Writings of Descartes, translated and edited by Cottingham, Stoothoff and Murdoch (CSM), Cambridge, 1984 and to the Oeuvres de Descartes, edited by Ch. Adam and P. Tannery (AT), revised edition, Vrin, 1964–76.
- Donaldson, B. (2015). *Creaturely cosmologies: Why metaphysics matters for animal and planetary liberation*. Lanham: Lexington Books.
- Gennaro, R. (2012). Consciousness. *Internet Encyclopedia of Philosophy*. <http://www.iep.utm.edu/consciou/>.
- Goff, P. (2013). Orthodox property dualism + linguistic theory of vagueness = panpsychism. In R. Brown (Ed.), *Consciousness inside and out: Phenomenology, neuroscience, and the nature of experience* (pp. 75–91). Berlin: Springer.
- Goff, P. (2015). Real acquaintance and physicalism. In P. Coates & S. Coleman (Eds.), *Phenomenal qualities: Sense, perception, and consciousness*. Oxford: Oxford University Press.
- Goff, P. (Forthcoming). Galileo's error.
- Gray, K., & Wegner, D. (2012). Feeling robots and human zombies: Mind perception and the uncanny valley. *Cognition*, 125(1), 125–130.
- Guthrie, S. (1993). *Faces in the clouds: A new theory of religion*. Oxford: Oxford University Press.
- Haslanger, (1995). Ontology and social construction. *Philosophical Topics*, 23(2), 95–125.
- Haslanger, S. (2000). Gender and race: (What) are they? (What) Do we want them to be? *Nous*, 34(1), 31–55.
- Heal, J. (2003). *Mind, reason and imagination*. New York: Cambridge University Press.
- Hyslop, A., & Jackson, F. (1972). The analogical inference to other minds. *American Philosophical Quarterly*, 9(2), 168–176.
- James, W. (1890). *The principles of psychology*. Cambridge: Harvard University Press.
- Jaynes, J. (2000) [1976]. *The origin of consciousness in the breakdown of the bicameral mind (PDF)*. Boston: Houghton Mifflin.
- Jonas, H. (1966). *The phenomenon of life: Toward a philosophical biology*. Chicago: University of Chicago Press.
- Joyce, R. (2006). *The evolution of morality*. Cambridge: MIT Press.
- Kirk, R. (2003). *Mind and body*. Kingston: McGill-Queen's University Press.
- Lee, A. (2018). Is consciousness inherently valuable? *Philosophical Studies*, 1–17.
- Leopold, Aldo. (1949). *A sand county Almanac*. New York: Oxford University Press.
- Lewis, D. (1973). *Counterfactuals*. Oxford: Blackwell.
- Lovejoy, A. (1964) [1936]. *The great chain of being: A study of the history of an idea*. Cambridge: Harvard University Press.
- Ludwig, K. (2010). Intuitions and relativity. *Philosophical Psychology*, 23(4), 427–445.
- Maher, C. (2017). *Plant minds: A philosophical defence*. New York: Routledge.
- Matheson, Jonathan. (2016). Moral caution and the epistemology of disagreement. *Journal of Social Philosophy*, 47(2), 120–141.
- Matthews, F. (2003). *For love of matter: A contemporary panpsychism*. Albany: State University of New York Press.
- Merricks, T. (2001). *Objects and persons*. Oxford: Oxford University Press.
- Merricks, T. (2005). Composition and vagueness. *Mind*, 114, 615–637.

- Mill, J. S. (1889). *An examination of sir William Hamilton's philosophy and of the principal philosophical questions discussed in his writings* (6th ed.). London: Longman, Green, Longman, Roberts, and Green.
- Monti, M., Vanhauzenhuysse, A., Coleman, M., Boly, M., Pickard, J., Tshibanda, L., et al. (2010). Willful modulation of brain activity in disorders of consciousness. *The New England Journal of Medicine*, 2010(362), 579–589.
- Mørch, H. H. (2014). Panpsychism and causation: A new argument and a solution to the combination problem. *Doctoral Dissertation*. University of Oslo.
- Mullin, A. (2011). Children and the argument from 'marginal cases'. *Ethical Theory and Moral Practice*, 14(3), 291–305.
- Naci, L., Graham, M., Owen, A., & Weijer, C. (2017). Covert narrative capacity: Mental life in patients thought to lack consciousness. *Annals of Clinical and Translational Neurology*, 4(1), 61–70.
- Nagel, T. (1974). What is it like to be a bat? *Philosophical Review*, V83, 435–450.
- Nagel, T. (2004). Panpsychism. In D. S. Clarke (Ed.), *Panpsychism: Past and recent selected readings*. Albany: State University of New York Press. (originally published Cambridge University Press, 1979).
- Oizumi, M., Albantakis, L., & Tononi, G. (2014). From the phenomenology to the mechanisms of consciousness: Integrated information theory 3.0. *PLoS Computational Biology*, 10(5), e1003588.
- O'Leary-Hawthorne, J., & Michael, M. (1996). Compatibilist semantics in metaphysics: A case study. *Australasian Journal of Philosophy*, 72, 117–134.
- Owen, A., Coleman, M., Boly, M., Davis, M., Laureys, S., & Pickard, J. (2006). Detecting awareness in the vegetative state. *Science*, 313(5792), 1402.
- Peterson, A., & Bayne, T. (2018). Post-comatose disorders of consciousness. In R. J. Gennaro (Ed.), *Routledge handbook of consciousness* (pp. 351–365). New York: Routledge.
- Regan, T. (1975). The moral basis of vegetarianism. *Canadian Journal of Philosophy*, 5(2), 181–214.
- Regan, T. (1979). An examination and defense of one argument concerning animal rights. *Inquiry*, 22(1–4), 189–219.
- Regan, T. (1983). *The case for animal rights*. Berkeley: University of California Press.
- Schwitzgebel, E. (2014). Tononi's exclusion postulate would make consciousness (nearly) irrelevant. Blog post at *The splintered mind*. <http://schwitzsplinters.blogspot.com/2014/07/tononis-exclusion-postulate-would-make.html>. Accessed 6 Sept 2018.
- Seager, W. (1995). Consciousness, information and panpsychism. *Journal of Consciousness Studies*, V2–3, 272–288.
- Seager, W. (2006). The 'Intrinsic Nature' argument for Panpsychism. *Journal of Consciousness Studies*, 13(10–11), 129–145.
- Searle, J. (1992). *The rediscovery of the mind*. Cambridge: The MIT Press.
- Shani, I. (2015). Cosmopsychism: A holistic approach to the metaphysics of experience. *Philosophical Papers*, 44(3), 389–437.
- Shea, N., & Bayne, T. (2010). The vegetative state and the science of consciousness. *British Journal for the Philosophy of Science*, 61, 459–484.
- Singer, P. (1990). *Animal liberation*. New York: Avon.
- Steuber, K. (2006). *Rediscovering empathy: Agency, folk psychology, and the human sciences*. Cambridge: The MIT Press.
- Stawson, G. (2006). Realistic monism: Why physicalism entails panpsychism. In A. Freeman (Ed.), *Consciousness and its place in nature: Does physicalism entail panpsychism?*. Exeter: Imprint Academic.
- Street, S. (2006). A Darwinian dilemma for realist theories of value. *Philosophical Studies*, 127, 109–166.
- Thompson, E. (2007). *Mind in life: Biology, phenomenology, and the sciences of mind*. Cambridge: Harvard University Press.
- Thorp, J. (2016). *Aristotle's clivus naturae*. Presentation at University of Western Ontario Arts & Humanities Research Day. <https://ir.lib.uwo.ca/philosophypres/12/>. Accessed 6 Sept 2018.
- Tononi, G. (2012). Integrated information theory of consciousness: An updated account. *Archive of Italian Biology*, 150(2–3), 56–90.
- Tononi, G. (2014). *Why Scott should stare at a blank wall and reconsider (or, the conscious grid)*. <http://www.scottaaronson.com/tononi.docx>. Accessed 6 Sept 2018.
- Turing, A. (1950). Computing machinery and intelligence. *Mind*, 49, 433–460.
- Unger, P. (1979). There are no ordinary things. *Synthese*, 41(2), 117–154.

-
- Van Cleve, J. (1990). Mind-dust or magic? Panpsychism versus emergence. *Philosophical Perspectives*, *V4*, 215–226.
- Van Inwagen, P. (1990). *Material beings*. Ithaca: Cornell University Press.
- Vavova, K. (2015). Evolutionary debunking of moral realism. *Philosophy Compass*, *10*(2), 104–116.