Luke Roelofs - Nicolas Kuske

If Panpsychism Is True, Then What?
Part 1: Ethical Implications

Panpsychism is a striking metaphysical claim: every part of the physical world has some form of consciousness. Does this striking claim have equally striking ethical implications? Does it change what duties we owe to which beings, or how we should understand the relation between self-interest and altruism? Some defenders of panpsychism have suggested it does; some critics have suggested the same. Others have disagreed. In this paper, we attempt to survey and organize these existing discussions.

We suggest that panpsychism is likely to have significant implications, but that they 1) stop short of some of the most radical implications sometimes drawn (by both critics and proponents), and 2) often depend heavily on the exact sort of panpsychism that is endorsed, as well as how one answers various independent questions about value and identity.

We think this is good news for the plausibility of panpsychism. It might seem like a problem if adopting panpsychism, or any other particular view of the metaphysics of consciousness, could at a stroke overturn and transform major aspects of our practical lives. Our concern for each other, and our practical relations to the world, should have a degree of stability in the face of metaphysical debate. But it might also seem like a problem if adopting panpsychism had no impact at all. It would be odd if someone could come to think the entire world was suffused with consciousness, and yet go on entirely as before.

François Kammerer puts this point nicely, discussing the mirror-image question of what practical implications might follow from illusionism – the equally breathtaking claim that consciousness does not exist, and our sense of being conscious is an illusion:

> «I am a bit suspicious of views that would manage to avoid revisionary consequences altogether. Revisionary normative consequences may often appear undesirable; but it would be very strange, after all, that a radical revision

> of our conception of what reality (in general) and the mind (in particular) are [...] should leave our [ethical lives] completely untouched»[1].

Adapting terminology from Kammerer, we distinguish «conservative», «moderate» and «radical» results in any particular area: the conservative result is that panpsychism makes no difference at all, the moderate result is that it makes some significant difference but within the space of somewhat-familiar options, and the radical result is that it overturns our everyday ways of thinking altogether. In these terms, we will argue that panpsychism's practical implications are mostly moderate, rather than either conservative or radical.

1. *Sentientism and the Distribution of Consciousness*

Does panpsychism imply rights for trees, rights for rocks, rights for electrons? This question arises primarily because of the widespread appeal of an idea often termed sentientism:

> **Sentientism:** Some form of phenomenal consciousness is necessary and sufficient for moral status.

To have moral status means to matter, for moral decisions – in particular, to be the sort of thing that can be benefitted (and which we ought to benefit, where possible) and can be harmed (and which we ought not to harm, unless necessary). To have moral status is to be an individual locus of concern, a being to whom we can have obligations.

Since panpsychism holds that phenomenal consciousness is incredibly widespread, the conjunction of panpsychism and sentientism might seem to suggest that moral status is also incredibly widespread. This would take the commonly-recognised implication of sentientism, that many non-human animals have moral status, and «turn it up to eleven». This implication is sometimes welcomed: Skrbina advocates thinking that «if all things are enminded, then, I claim, they all possess intrinsic value»[2]. But ironically, this radical extension threatens to undermine some of the chief practical implications usually associated with sentientism, such as veganism:

[1] F. Kammerer, *The Normative Challenge for Illusionist Views of Consciousness*, in «Ergo: An Open Access Journal of Philosophy» 6, 32 (2019), p. 919.

[2] D. Skrbina, *Ethics, Eco-Philosophy, and Universal Sympathy*, in «Dialogue and Universalism» 23, 4 (2013), p. 65.

> «If panpsychism is tre [...] trees have value in their own right [...] Admittedly, this does have some difficult implications for the ethics of vegetarianism and veganism. Many vegans and vegetarians feel that it is wrong to kill or to exploit sentient creatures. But if plants also have sentience, what is there left to eat?»[3].

But even assuming the truth of sentientism, the implications of panpsychism depend on at least three questions:

- Which things are conscious?
- What sort of consciousness matters for moral status?
- What sort of consciousness do different things have?

Let us begin with the first question: which things are conscious, according to panpsychism? As it is usually defined, panpsychism is the thesis that the most fundamental physical things – whether those are very small, like subatomic particles, or very large, like the universe as a whole – are phenomenally conscious: there is something it is like to be such entities. This something-it-is-like need not be complex or subtle, and need not be accompanied by any sort of cognition or thought, but it is subjective, and so, fundamentally, it differs only in degree from our own subjective experiences. Panpsychism says that phenomenal consciousness, understood in this way, is inherent to all physical matter. Panpsychists agree that all fundamental things are conscious, and that we human beings are conscious. However, they disagree widely on which other non-fundamental things (like rocks, plants and animals) are conscious. Mørch characterizes panpsychism as holding that «all fundamental or otherwise properly unified things»[4] are conscious: the question then is which things count as properly unified.

At one extreme, universalists say that literally everything – every concrete object and every collection of concrete objects – has some sort of consciousness[5]. This view is defended by Roelofs[6], at one point by Goff[7], and possibly by

[3] P. Goff, *Galileo's Error: Foundations for a New Science of Consciousnes*, Pantheon Books, New York 2019, p. 191. See also J. Buchanan - L. Roelofs, *Panpsychism, intuitions, and the great chain of being*, in «Philosophical Studies» 176, 11 (2019), pp. 312-314, and K. Frankish, *Panpsychism and the depsychologization of consciousness*, in «Aristotelian Society Supplementary Volume» 95 (2021), p. 65.

[4] H.H. Mørch, *Does Panpsychism Mean That "We Are All One"?*, in «Journal of Consciousness Studies» (Forthcoming).

[5] Some universalists may qualify this claim slightly: Roelofs suggests that a collection so widely-scattered that its parts do not physically interact in any way might not have *unified* consciousness, and by some standards that might be tantamount to not really being a subject of consciousness at all. But for all practical purposes, «every collection with any physical interactions among its parts» is so inclusive, given the pervasiveness of gravitational interactions, as to be equivalent to «every collection».

[6] J. Buchanan - L. Roelofs, *Panpychism*, cit., pp. 3006-3009.

[7] P. Goff, *Orthodox property dualism + linguistic theory of vagueness = panpsychism*, in R. Brown

Margaret Cavendish[8]. It is often motivated by the absence of any sufficiently sharp, non-arbitrary dividing line to draw between systems that seem sufficiently unified and those that don't[9].

Other forms of panpsychism impose more restrictive standards. The potential range of restrictions mirrors the variety of non-panpsychist views on the minimal basis for consciousness: there could be «global neuronal workspace panpsychism», on which macroscopic consciousness is present only in structures that work like the global neuronal workspace supposedly enabled by the human neocortex[10], or «unlimited associative learning panpsychism», on which macroscopic consciousness is present only in systems capable of the kind of flexible and complex learning supposedly found in vertebrates, cephalopods, and arthropods but not in sponges, jellyfish, and nematodes[11]. There could be «higher-order thought panpsychism», on which macroscopic consciousness is present only in systems which can represent their own inner states in the right way[12].

Two particularly notable versions of restricted panpsychism are biopsychist and IIT-inspired. Biopsychist panpsychism is the idea that living organisms, understood as a certain sort of self-realizing system, are the only non-fundamental subjects of experience[13]. Freya Mathews has defended a biopsychist theory of

(ed.), *Consciousness inside and out: Phenomenology, neuroscience, and the nature of experience*, Springer, Berlin 2013, pp. 75-91.

[8] M. Cavendish, *Observations upon experimental philosophy*, Cambridge University Press, Cambridge 2001; see discussion in J. McWeeny, *Which Bodies Have Minds? Feminism, Panpsychism, and the Attribution Question*, in K. Maitra - J. McWeeny (eds.), *Feminist Philosophy of Mind*, Oxford University Press, Oxford 2022, pp. 275-279.

[9] See especialy P. Goff, *Orthodox property*, cit.; J. Buchanan - L. Roelofs, *Panpychism*, cit., pp. 91-94; D. Skrbina, *Ethics, Eco-Philosophy*, cit., pp. 64-65.

[10] For the global workspace theory, see B.J. Baars, *A Cognitive Theory of Consciousness*, Cambridge University Press, Cambridge 1993 and S. Dehaene - M. Kerszberg - J.P. Changeux, *A neuronal model of a global workspace in effortful cognitive tasks*, in «Proceedings of the National Academy of Sciences» 95, 24 (1998), pp. 14529-14534.

[11] For the unlimited associative learning theory, see S. Ginsburg - E. Jablonka, *The Transition to Experiencing: I. Limited Learning and Limited Experiencing*, in Biological Theory 2, 3 (2007), pp. 218-230, and S. Ginsburg - E. Jablonka, *The Transition to Experiencing: II. The Evolution of Associative Learning Based on Feelings*, in «Biological Theory» 2, 3 (2007), pp. 231-243.

[12] For an example of a panpsychism-adjacent view combined with higher-order thought theory, see S. Coleman, *The Real Combination Problem: Panpsychism, Micro- Subjects, and Emergence*, in «Erkenntnis» 79, 1 (2014), pp. 19-44, and S. Coleman, *Panpsychism and Neutral Monism: How to Make Up One's Mind*, in G. Brüntrup - L. Jaskolla (eds.), *Panpsychism: Contemporary Perspectives*, Oxford University Press, Oxford 2017.

[13] Compare this with theories that are biopsychist full stop, rather than biopsychist-panpsychist: on such theories, consciousness itself appears only in living systems, rather than existing in all matter

panpsychism at length, with the notable feature that as well as single-celled and multicellular organisms, she takes the cosmos as a whole to also be a self-realizing system and thus a «self»[14]. Goff does not explicitly commit to such a view, but does suggest that plants as well as animals are conscious subjects, suggesting a view which likely corresponds with biopsychism at least in outline.

IIT-inspired panpsychism blends the panpsychist account of fundamental consciousness with the Integrated Information Theory, which associates consciousness with any local maximum of integrated information[15]. This theory already implies a very widespread distribution of consciousness, since even a very simple system might be a local maximum if it is not part of a more integrated larger system. But it is not strictly panpsychist, because simpler systems within a complex system lack consciousness (e.g. water molecules are conscious while in the air, but lose individual consciousness when absorbed into my brain). IIT-inspired panpsychism, by contrast, takes matter to be intrinsically conscious, and appeals to maxima of integrated information only to explain when consciousness combines.

One notable difference between these two views is that biopsychist panpsychism implies consciousness in plants as whole organisms, while IIT-inspired panpsychism does not. Individual plant cells are likely to be more integrated, informationally, than the whole plant, and so «in terms of consciousness, then, a plant would be a society, not an individual»[16]. Other restricted forms of panpsychism might land somewhere between the two: Itay Shani argues that particular forms of consciousness are only instantiated in «esonectic» systems which construct and maintain an inner domain, rather than in «exonectic» systems[17]. Shani identifies animals with brains as paradigmatic esonectic systems (conscious), and mineral structures as paradigmatic exonectic systems (non-conscious), leaving the status of plants undecided.

Can we say anything synoptic about which non-fundamental systems would be conscious, if panpsychism were true? It looks hard: Panpsychists are

and only combining in living systems, see F. Baluška - A. Reber, *Sentience and Consciousness in Single Cells: How the First Minds Emerged in Unicellular Species*, in «BioEssays» 41, 3 (2019), pp. 180-229.

[14] See F. Mathews, *For love of matter: A contemporary panpsychism*, State university of New York Press, NewYork 2003 and F. Mathews, *Living Cosmos Panpsychism*, in *The Routledge Handbook of Panpsychism*, Routledge, London 2020, pp. 131-143.

[15] See especially H.H. Mørch, *Is the Integrated Information*, cit.

[16] H.H. Mørch, *The Integrated Information Theory of Consciousness*, in «Philosophy Now» (2017) https://philosophynow.org/issues/121/The_Integrated_Information_Theory_of_Consciousness .

[17] I. Shani, *Cosmopsychism a holistic approach to the metaphysics of experience*, in «Philosophical Papers» 44, 3 (2015), p. 419.

all over the place. But we can observe that panpsychist theories do tend to be far more expansive in their ascriptions of consciousness, even to non-fundamental things, than non-panpsychist theories. Standard debates about the distribution of consciousness tend to be pitched within the domain of nervous systems: which animals' nervous systems enable consciousness, which don't? But universalist, biopsychist, and IIT-inspired panpsychisms all ascribe consciousness to all things with nervous systems as well as a wide range of things without.

The obvious explanation for this difference is that what panpsychists are looking for is less demanding. They need a system to hang together sufficiently that we can, without arbitrariness, view it as «one thing»[18]. Non-panpsychists, by contrast, are looking for something complex and special enough to effect the seemingly magical appearance of consciousness *de novo*. Given this difference, it is not surprising that panpsychist theories, though they differ on which particular non-fundamental systems are conscious, tend to ascribe consciousness very widely.

2. *Varieties of Consciousness*

Determining which beings are conscious does not necessarily determine which beings have moral status, even given sentientism: We also need to ask what sort of consciousness confers moral status. Here there are two basic answers: either all consciousness matters or only some specific sort of consciousness matters. Chalmers terms the former position "Broad Sentientism":

> **Broad Sentientism:** A being has moral status if and only if it has the capacity for some form of phenomenal consciousness[19].

The latter position can take various forms; the most common is what Chalmers terms Narrow Sentientism, which assigns weight specifically to valenced consciousness, consciousness that feels good and or bad in an affective sense.

[18] Friston offers a mathematical definition of 'thinghood' appealing to self-organization in a non-equilibrium steady-state, purportedly applying all the way down to fundamental physics (K. Friston, *A free energy principle for a particular physics*, in «arXiv» [2019], https://doi.org/10.48550/arXiv.1906.10184). Friston, Wiese, and Hobson deny that this account favors panpsychism generally, which they interpret as property dualism, but advocate a dual-aspect monism which they admit can on certain interpretations collapse into panpsychism (K. Friston - W. Wiese - J.A. Hobson, *Sentience and the origins of consciousness: From Cartesian duality to Markovian monism*, in «Entropy» 22, 5 [2020], p. 516).

[19] D. Chalmers, *Reality+: Virtual Worlds and the Problems of Philosophy*, Penguin Books, New York 2022, cfr. L. Roelofs, *Sentientism, Motivation, and Philosophical Vulcans*, in «Pacific Philosophical Quarterly» 104, 2 (2023), p. 324.

Narrow Sentientism: A being has moral status if and only if it has the capacity for affectively valenced phenomenal consciousness.

Intuitively, the ability to suffer or be happy makes sense as a criterion for moral status, if morally significant harms and benefits primarily consist in suffering and happiness. However, it is not obvious that this exhausts the morally relevant sorts of harm and benefit. The capacity for some degree of desire, choice, and agency might also enable a being to be harmed (by frustration of desire) or benefitted (by satisfaction). And it may be possible for a conscious being incapable of pleasant or unpleasant experiences to still have desires, goals, and thus reasons for action[20]. There is some intuitive pull to seeing such beings as having moral status – even being our moral equals. Given the possibility of such exotic forms of consciousness we might consider a third form of sentientism, «Motivational Sentientism».

Motivational Sentientism: A creature has moral status if and only if it has motivating consciousness.

«Motivating consciousness» means any conscious state which provides or presents reasons for action, including valenced states at a minimum but leaving open the possibility of non-valenced motivating experiences[21]. Other versions of sentientism are entirely possible: e.g., we might think that epistemic consciousness also confers moral status, or we might think that any consciousness which meets some standard of unity or «selfhood» confers moral status[22].

For our purposes here the most important distinction is between Broad Sentientism and its various narrowings. On Broad Sentientism, panpsychism implies moral status, at least to some degree, in all the fundamental physical things. On any narrower form of sentientism, panpsychism might not have this implication, since these simplest conscious beings might lack the sort of consciousness that matters. Thus we should ask: if panpsychism is true, what sorts of beings have which sorts of consciousness?

We can start with the question of «micro-valence» of whether the simplest physical things are capable of valenced experiences – can electrons suffer?

[20] D. Chalmers *Reality*, cit., pp. 335-337, and L. Roelofs, *Sentientism, Motivation*, cit., discuss such dispassionate beings under the label "Philosophical Vulcans", cfr. D. Smithies - J. Weiss, *Affective Experience, Desire, and Reasons for Action*, in «Analytic Philosophy» 60, 1 (2019), p. 30, and J. Shepherd, *Sentience, Vulcans, and Zombies: The Value of Phenomenal Consciousness*, in «AI and Society» (Forthcoming), pp. 1-11.

[21] L. Roelofs, *Sentientism, Motivation*, cit.

[22] See, for instance, F. Mathews, *For love of matter*, cit, p. 58 ff.

This will determine whether panpsychists who are also Narrow Sentientists should extend moral status to them.

One reason to posit micro-valence might be that *we* have valenced experiences, which must come from somewhere. If we thought that valence was strictly irreducible to any sort of non-valenced experience, this would show that some germinal valence must be present at the fundamental level. Call this the Genetic Argument for micro-valence. Gottlieb and Fischer present a version of this argument:

> «What kind of micro-experience would correspond to [a painful] macro-experience? It's unclear. But we do know that whatever this micro-experience is like, it must be of the kind from which the existence of macro-experience is a priori derivable, perhaps in conjunction with some further structural-functional facts. We aren't sure how this would be possible without saying that micro-level experiences have forms of valence themselves»[23].

Their key premise, that valence is irreducible, can be questioned, however. It seems conceivable that valence might arise from a particular arrangement or blending of elements which are not themselves valenced, though they might be in some ways akin to valence. For instance, they might be something like the raw «instinctive excitations» that Freud, in his later work, suggested the psyche has to first «bind» in «an act of preparation» before they could come under the sway of «the pleasure principle», i.e., the drive to maximize pleasure and minimize pain[24]. At least, the question of the ultimate nature of pleasure is not simple.

A different way to approach the question of micro-valence is by looking for functional profiles that might provide evidence of valence, and whose absence might, therefore, be (defeasible) evidence for its absence. At first such a methodology might seem unavailable to those panpsychists who deny that consciousness is explained by function, and who moreover make ascriptions of consciousness that wildly diverge from our normal, intuitive, judgements[25]. But in fact, contemporary panpsychists have often been «nonreductive functionalists», holding that functional structure correlates systematically with

[23] J. Gottlieb - B. Fischer, *The ethical implications of panpsychism*, in «Australasian Journal of Philosophy» (Forthcoming), p. 11.

[24] S. Freud, *Beyond the Pleasure Principle,* International psycho-analytical Press, Austria 1922, p. 80.

[25] K. Frankish, *Panpsychism and the depsychologization*, cit., p. 63, n. 9.

the structure and content of particular conscious states[26]. Moreover, Buchanan and Roelofs argue that panpsychists can maintain the qualified reliability of our intuitive, function-driven, ascriptions of any sort of consciousness that we have personal acquaintance with[27]. Since we are all well-acquainted with suffering and pleasure, this implies that if something seems to us not to «act as though» it has valenced experiences, it probably doesn't. So the question becomes, do electrons act like they have valenced experiences? Brian Tomasik suggests a positive answer:

> «If we take "nociception" to be "noticing and avoiding some stimulus", then electrons very crudely engage in nociception, when they avoid other electrons. Expressing the same point using more suggestive language: an electron receives input information (conveyed via force carriers) about the presence of another electron, and that information «motivates it» to change its state by moving away»[28].

However, this is clearly stretching the notion of nociception pretty far, perhaps too far to be persuasive. And one might observe here a sort of argument from symmetry: nothing in an electron's behavior seems to support thinking that they move away from each other to reduce the pain of proximity over the opposite hypothesis, that they move away from each other to increase the pleasure of distance, or to indulge the pleasure or movement-under-force, or something else. Equilibrium of forces might be blissful contentment that particles strive towards, or a torturous state of frustration. Nothing in the behavior of the particle seems to make a pleasure-based explanation more reasonable than one based in displeasure.

This symmetry could be taken to mean either of two things. If we were convinced by the genetic argument, we might take it to have strictly epistemic importance: electrons are at any given moment either suffering or enjoying, but we have no way to tell which. But if we thought the genetic argument was inconclusive, we might think that the Symmetry argument provides some positive reason to deny that electrons either act like they're feeling pleasure or like they're feeling displeasure. Their behavior is too simple to fit either the pleasure-pattern

[26] D. Chalmers, *The Conscious Mind: In Search of a Fundamental Theory*, Oxford University Press, Oxford 1996, p. 249.

[27] *Ibidem*.

[28] B. Tomasik, *Is There Suffering in Fundamental Physics?*, in *Essays on Reducing Suffering*, 2020 https://reducing-suffering.org/is-there-suffering-in-fundamental-physics/#Panpsychisteliminativist_theories_of_consciousness

or the displeasure-pattern, providing evidence that their consciousness is of neither sort. This would allow panpsychists who endorse Narrow Sentientism to resist extending moral status to electrons and other fundamental physical things.

If valence, or other sorts of ethically important consciousness, are not present at the fundamental level, where might they first appear? Mathews suggests that things can consciously matter for biological organisms, but not for other systems, because only organisms posit and maintain a self:

> «[A] guided missile, for instance, may be heat sensitive, and programmed to detonate when it comes into contact with a certain kind of heat source; but the missile does not 'reach out' for heat sources. It doesn't matter to the missile whether it makes contact with a heat source or not, because nothing matters to a missile; it does not have ends of its own; it is not a self»[29].

Similarly, Baluška and Reber suggest that even a single cell can «subjectively evaluate the beneficial or injurious nature of objects it encounters»[30]. And Goff suggests that we «see the movement of a plant toward the light as expressing its own desire and conscious drive for life»[31]. Certainly, biological life marks out a domain within which we can see a fairly clear way to assign interests to things: plants, bacteria, animals alike, can all be healthy or diseased, injured or intact, flourishing or struggling.

However, a case could also be made for a more neurocentric view (for more detail, see Roelofs Ms.). First, observe that although our human bodies, and many parts of them, exhibit health, illness, and a sort of goal-directedness based on their evolved biological functions, this sort of goal-directedness is distinct from, and often comes apart from, our conscious feelings of pleasure, displeasure, and motivation. Tissue damage doesn't hurt; the brain activity that it usually elicits does (and if nerve damage or anesthetics block that brain activity, there is no pain). Deliberately blocking reproduction is ruinous from the perspective of evolutionary biological goals, but often wonderful from the perspective of the conscious individual. Moreover, a version of the «symmetry argument» against micro-valence sketched above might be repurposed here: simple biological systems can respond to stimuli, but it's not clear simple responses provide any indication of whether they are reducing distress started by the stimulus, or fulfilling pleasure started by the stimulus. By contrast, in a system for integrating multiple

[29] F. Mathews, *For love of matter*, cit., pp. 59-60.
[30] F. Baluška - A. Reber, *Sentience and Consciousness*, cit., p. 5.
[31] P. Goff, *Galileo's Error*, cit., p. 191.

drives and trading them off against each other, we might have more basis for distinguishing pleasure and displeasure, and brains seem well-suited.

A final plank of the case for neurocentrism would be to note that talk of organisms having a «self» might not, in fact, be entirely apt for plants, fungi, and algae, which do not have a fixed body plan nor (consequently) any internal representation of themselves. A tree branch is part of the tree, but if removed and planted or grafted it might become a separate tree, or a part of some other tree. A leaf at its tip does not, as best we can tell, have any inkling of this change: its leafy activities are indifferent to its inclusion in one larger whole or another. Animals with brains, however, have a central organ which collects inputs from the rest of the body and, among its other functions, distinguishes between self and outside world. Such animals seem to be 'selves' in a much richer sense than other organisms.

So just as there are a range of panpsychist views on when consciousness combines, there are a range of views on what kinds of consciousness appear where – often drawing similar dividing lines (e.g. biological vs. non-biological, or nervous-system vs. no-nervous-system). Broadly speaking, genetic reasoning – attributing irreducible properties to fundamental thing to explain their presence in other things – can be used to motivate ascribing more forms of consciousness more widely, at least if those forms can be shown to be irreducible. By contrast, looking at functional profiles as evidence of particular types of consciousness can be used to motivate more limited ascriptions, especially through the symmetry argument sketched above. But relatively little has been written on this topic, and many of these arguments have not seen sustained development or scrutiny: thus this is an area where new work by panpsychists would be especially useful.

3. *Self and Other*

The last two sections both asked about ways panpsychism might change the content of morality. But there is also the possibility that it might impact moral motivation – might help to strengthen, or to justify, our willingness to actually act as morality demands. This is because panpsychism, according to some authors, removes or lessens the boundary between self and other, or between self and world.

How exactly might panpsychism have this result? One observation is that we normally assume that the boundary between self and world is also a boundary between consciousness and the utterly non-conscious. Panpsychism denies this: there is no utterly non-conscious stuff, so the boundary between me and

my surroundings is not a fundamental difference. Things out there are, in this specific regard, the same as me. For instance, Goff claims that dualism «creates a sense of separation [and] implies that, as an immaterial mind, I am a radically different kind of thing from the mechanistic world I inhabit», whereas panpsychism removes this sense of separation[32].

Of course, the self/other boundary is not shown to be unreal or illusory just because it doesn't line up with a conscious/non-conscious boundary. But some versions of panpsychism might push us in that direction. Constitutive panpsychism is often felt to face an especially urgent set of combination problems, in which it may struggle to explain things like the unity of consciousness, the boundaries of consciousness, and the conscious subject or self. One appealing option for solving such problems is to deflate, or even deny, those phenomena: to say that there is no self, or that many selves overlap, or that consciousness is not bounded in the way it seems to be.[33] Such claims may or may not help constitutive panpsychists to address the combination problems, but if they were true they would further weaken the self-other boundary. For example, some versions of panpsychism imply that the relation sometimes called «phenomenal unity» or «phenomenal co-consciousness» obtains universally or near-universally, and not merely within individual minds (though other versions, notably including Mørch's, reject this as self-evidently false, on the basis that we are immediately aware of the lack of phenomenal unity between our experiences and other people's).

But do any of the metaphysical claims made by panpsychists in general, or constitutive panpsychists in particular, actually translate into a clear support for moral behavior? Does it become irrational to sacrifice others to my own interests just because my consciousness is surrounded by, or is continuous with, other consciousnesses? This is the driving question of a recent paper by Mørch, who distinguishes four different sorts of unity that might be claimed among people:

> «(1) that we are all a part or aspect of the same whole, (2) that we all have the same nature, (3) that we all have the same transcendental self, or – most strongly – (4) that we are all the very same person»[34].

[32] *Ibidem*; cfr. D. Skrbina, *Ethic*, cit. p. 69; F. Mathews, *For love of matter*, cit.

[33] For discussion, see G. Rosenberg, *The Boundary Problem for Phenomenal Individuals*, in S. Hameroff - A. Kaszniak - A. Scott (eds.), *Toward a Science of Consciousness: The First Tucson Discussions and Debates (Complex Adaptive Systems)*, MIT Press, Cambridge 1998; L. Roelofs, *The Unity of Consciousness, within and between Subjects*, in «Philosophical Studies» 173, 12 (2016), pp. 3199-3221; L. Roelofs, *The Varieties of (Un)Boundedness*, in «Journal of Consciousness Studies» (Forthcoming); G. Miller, *Can Subjects Be Proper Parts of Subjects? The De-combination Problem*, in «Ratio» 30, 2 (2017), pp. 1-18.

[34] H.H. Mørch *Does Panpsychism*, cit., p. 2.

For each thesis, Mørch considers whether panpsychism implies this kind of unity, whether it makes coherent sense, and whether it actually supports moral motivation «by implying the irrationality of egoism and rationality of compassion for all»[35]. She concludes that although the first two clams are likely consequences of many versions of panpsychism, they do not directly support the rationality of compassion; the third is not obviously coherent; and the fourth, though it makes sense and does support the rationality of compassion, is only moderately supported by panpsychism. Let us consider these four sorts of unity in turn.

First, the same whole thesis is just the claim that we are each part of something larger and more encompassing. Mørch claims that this depends on cosmopsychism, i.e. on holistic forms of panpsychism which make the whole cosmos the fundamental conscious thing.[36] But universalist panpsychists can also agree that the universe as a whole *is* conscious, regardless of whether it is more or less fundamental than its parts. Indeed, even physicalists could likewise claim that we are all part of something, namely the physical universe. But we might think that panpsychist views make this encompassing whole seem more significant, by ascribing it consciousness, and that cosmopsychist views go further in the same direction, by making it fundamental.

The same nature thesis is the claim that, fundamentally, all things «have the same nature in virtue of being forms or modifications (or determinates) of consciousness as such»[37]. Panpsychism does seem to have such an implication, but Mørch suggests that the same whole thesis, and the same nature thesis, both fail to clearly establish the rationality of compassion: while panpsychism might imply that we are «all one» in one or both of these senses, that still leaves us as many individuals in the ways that matter. There is nothing obviously irrational about one part of a whole pursuing its interests at the expense of other parts of the same whole, or about a being pursuing its own interests at the expense of another being with the same fundamental nature. Indeed, even other people's experiences being connected to mine in a certain phenomenal way does not obviously make it irrational for me to care about the difference: perhaps the self-other boundary is a boundary drawn within a continuous phenomenal field, but we might still draw that boundary, and care only about what falls on this side of it.

Mørch next considers the «Same Transcendental Self» thesis – essentially the claim that although we are all in some sense distinct beings, we are also all

[35] *Ibi*, p. 18.
[36] *Ibi*, p. 6.
[37] *Ibi*, p. 7.

in some sense the same being. Many «empirical selves» but only one «transcendental self». Mørch's worry is that «almost by hypothesis, the view does not make complete sense, at least not to non-mystics and non-meditators»[38]. For example, Albahari writes of «our abiding conscious nature as identical with a wider unitive consciousness that is unstructured by any spatio-temporal or subject/object parameters»[39]. My consciousness, of course, *is* structured by spatio-temporal and subject/object parameters, and so even if it depends on, occurs within, or is in some other way *related to* the underlying formless consciousness, it seems it must be recognized as, in some sense, a distinct thing after all.

Insofar as it is paradoxical, the practical implications of this sort of unity claim will be unclear. What becomes of egoism and altruism, privacy and publicity, consent and coercion, if we are all simultaneously one and many? This view seems to amount to breaking down our everyday notion of an individual and reconceiving it in some way, so its practical implications will depend entirely on how we are supposed to reconceive the practical aspects of selfhood.

Finally, Mørch considers the most radical unity claim, that we are all literally the same person. This seems to resolve the paradox in the Same-Transcendental-Self claim by flatly denying the apparent fact of multiple individual minds. However, it still requires some account of how there can *seem* to be multiple minds. A coherent world that accommodates distinct synchronic conscious experiences necessitates localized variations within its causal structure. That means when I brush my teeth, you can keep sleeping. And while you dream about the hiking trip you did last year, I could not indulge in such memories even if I were sleeping because I have never been on this trip. Rationally, within the causal framework of the world, the everyday «I» corresponds to a specific set of experiences associated with a particular body and its needs.

So both the same-person and same-transcendental-self views say that, in some sense, there is just one mind, but that, in another sense, there are many. But in the same-person view, the oneness has become determinately more real and fundamental than the many-ness. Oneness is now supposed to encompass a diversity of particular times, places, and qualities, rather than being formless and beyond all such diversity. This is important for the views' practical implications: by insisting that what is important in our everyday sense of self really falls on the side of the one, rather than being split between the one and many, it seems to

[38] *Ibi*, p. 8.

[39] M. Albahari, *The Mystic and the Metaphysician*: *Clarifying the Role of Meditation in the Search for Ultimate Reality*, in «Journal of Consciousness Studies» 26, 7, 8 (2019), p. 25.

support the conclusion that each of the many ought rationally to care for the one, and by extension to care for all of the many, not just for «themselves».

So the Same-Person thesis has serious practical implications. But does it follow from panpsychism? There are two ways that it could, which we might call metaphysics-heavy and metaphysics-light. The metaphysics-heavy approach goes by way of a radically monistic ontology on which there is just nothing else for us to be: the only real entity in existence is the whole. This approach is sometimes called «existence cosmopsychism», and is defended by Benovsky as the best way to address the combination problem[40]. On his account, there exists only one object, the world, and by virtue of being the only object, it is also the only subject. And while this subject might seem locally tree-ish or human-ish to itself, the moment it starts to believe some human is a real independently existing object or subject, it is in fact lapsing into delusion.

Mørch outlines a different, metaphysics-light, approach to establishing a Same-Person thesis. This begins from the worries about personal identity raised particularly by Derek Parfit[41]. Humans, as described by science, are metaphysically superficial, not fundamental, and in consequence their identity is sometimes indeterminate, determined by extrinsic factors, and potentially a matter of convention. Yet it seems absurd that my existence or non-existence, in the sense that matters to me, should be indeterminate, extrinsic, or conventional. This tension, Mørch suggests, could be resolved in three basic ways:

- By accepting that nothing really exists and matters in the way that we seem to ourselves to exist and matter (No self)
- By supposing that we as individuals are actually more objectively real and fundamental than the human being as described by physical science (Individual Self)
- By identifying ourselves, in some important sense, with something universal and metaphysically fundamental (Shared Self).

Mørch suggests that physicalism is most friendly to the No Self option, substance dualism to the Individual Self option, and panpsychism to the Shared Self option, because in a panpsychist ontology there may be a plausible basis for a shared self (fundamental consciousness), but no good candidate basis for a deep individual self. Although we cannot here evaluate this argument fully, we suggest that it does at least show that adopting panpsychism can shift the balance of plausibility to some extent in favor of some sort of Same-Person view.

[40] J. Benovsky, *Mind and matter: Panpsychism, dual-aspect monism, and the combination problem*, Springer, Berlin 2018.

[41] See, in particular, D. Parfit, *Personal identity*, in «The Philosophical Review» 80, 1 (1971).

In sum, panpsychism and physicalism can both be seen as undermining the reality of the human individual, showing its boundaries to be unreal or superficial. If moral motivation just required undermining the self, panpsychism would not have much to add. Its distinctive contribution is not on the negative side, but on the positive. It attributes consciousness to the more fundamental reality into which the self may turn out to dissolve. In doing so, it might help to make that which encompasses both me and other individuals seem worth caring about.

4. *Take home*

Given such a wide range of views on what things are conscious, what sort of consciousness gives moral status, which things have that sort of consciousness, and how real the self-other boundary turns out to be, panpsychism could have a wide range of ethical implications.

4.1. *Radical Results*

First, there are a few ways panpsychism could yield a «Radical Result». Most radically, combining universalism with Broad Sentientism would imply that literally everything has moral status – the sort of «everything ethic» sometimes discussed (and usually rejected as absurd and unworkable) in environmental ethics[42]. The same result would follow if we combined universalism with both Narrow Sentientism (or a similar view) and a liberal view about the distribution of affective valence (or whatever sort of consciousness is considered important).

Non-universalists might face an almost as radical result: moral status for fundamental particles. This flows from endorsing either Broad Sentientism or Narrow Sentientism together with the genetic argument that macro-valence can only be explained by micro-valence (or some other combination of views where the type of consciousness that confers moral status is fundamental).

[42] See R. Elliot, *Environmental ethics*, in P. Singer (ed.), *A Companion to Ethics*, Blackwell, Oxford 1993, p. 288; C. Belshaw, *Environmental Philosophy: reason, nature and human concern*, Acumen, Teddington 2001, p. 144; S.P. James, *For the Sake of a Stone? Inanimate Things and the Demands of Morality*, in «Inquiry» 54, 4 (2011), pp. 384-397.

The radicalism of these results might be tempered by the claim that «we just don't know what's good and bad for electrons; so, we can't harm or benefit them in expectation (even if we can harm or benefit them in fact)»[43]. This is a major point made by Gottlieb and Fischer, who argue that panpsychism likely does imply that electrons have moral status, but that this does not render panpsychism morally absurd because:

> «Even if we know that pain is bad for electrons, we have no idea what pains them [...] if that's right, then we don't have any reasons to modify our behavior because of potential impacts on them»[44].

We might say that this turns a Radical Result into a merely «disquieting result»: there is a vast realm of things that matter morally, right under our noses, but we cannot tell whether our actions are having wonderful or awful effects on them, so all we can do is carry on and maybe from time to time respectfully remind ourselves of the vast potentially affective sentience of which everything is made. Indeed, Shupe and Giberman have argued that, given certain formal assumptions about parthood and value, a world where panpsychism is true might be the «best of all possible worlds», a conclusion reminiscient of the famous historical panpsychist Leibniz[45].

It's possible that a similar epistemic point can temper the radicalism of the sort of «everything ethic» that universalism and Broad Sentientism together imply. James writes of:

> «[T]he tricky problem of explaining what it could mean to damage an inanimate thing. [...] a china vase, for example, appears to have an optimal state, and certain events will constitute a deviation from that state. A chip, for instance, will generally be regarded as damage. [...] However, adopting the disengaged, objective viewpoint mentioned above, one might wonder whether this description can be upheld. After all (one might ask) is it not rather the case that the vase has simply changed from one state to another? [...] Any judgment to the effect that an inanimate thing is vulnerable to damage will reflect certain human views on the integrity of things»[46].

[43] J. Gottlieb - B. Fischer, *The ethical implications*, cit., p. 21.
[44] *Ibi*, pp. 21-23.
[45] E. Shupe - D. Giberman, *Value in a Limitless World*, Ms.
[46] S.P. James, *For the Sake of a Stone?*, cit., p. 390.

Moreover, even among beings with moral status, there is room for wide variation in its form and degree, which may depend on things like welfare capacity, diachronic identity conditions, intensity of experience, or degree of consciousness[47]. So although moral status for all physical things, or all fundamental physical things, is a radical result, it might be tempered both by radical uncertainty about their interests, or radical differences in their moral weight.

A different radical implication follows from forms of panpsychism that endorse the same-person claim, that what is most truly and properly me is the underlying sea of physical consciousness that makes up the universe. This implies that I share my true self with my greatest idol and with the lion in the savannah as well as with a molecule light years away. But any sense of vanity on this account should be tempered by the recognition that I am also the sewage in the drain and the mosquito I squashed yesterday night. And if the chief practical implication of this view is that it is irrational for me to be selfish, and rational for me to work for the good of all things, then that result is perhaps not so radical at all, and maybe even common-sensical.

4.2. *Conservative Results*

So much for radical results: could panpsychism have a thoroughly conservative result, leaving our ethical outlook essentially unchanged? If Narrow Sentientism, or an analogous view of moral status, was combined with a very restrictive view about the distribution of valenced consciousness (or whatever type of consciousness matters), then panpsychism's expansion of consciousness might not translate into any expansion of moral status.

Likewise, nothing stops a panpsychist, especially of an emergentist bent, from insisting that reality comprises a diversity of separate, disconnected, selves, and that consequently compassion and rational self-interest are often in genuine opposition.

And of course we might reject sentientism entirely, cutting the link between consciousness and ethics. Frankish suggests that something like this is the natural upshot of panpsychism: its expansion of the domain of consciousness should actually push us away from Broad Sentientism, and away from linking consciousness closely to ethics.

[47] See A.Y. Lee, *Degrees of Consciousness*, in «Noûs» 57, 3 (2023), pp. 553-575; cfr. J. Gottlieb - B. Fischer, *The ethical implications*, cit., pp. 14-19.

> «A property so ubiquitous cannot in itself have any ethical significance. Whatever it is that makes us and other animals objects of ethical value, it cannot be that we are conscious subjects in the depsychologized sense»[48].

There is some plausibility in the thought that consciousness being everywhere might reasonably adjust our views on the significance of consciousness. But we suspect few panpsychists will want to sever the consciousness-ethics link entirely. Arguments for panpsychism rely on the idea that consciousness is a deep and important property, intrinsic and irreducible, and this makes it natural to see it as having important connections to other important matters.

4.3. *Moderate Results*

Finally, panpsychism might yield real but moderate ethical implications. If panpsychists adopt Narrow Sentientism, or something else short of Broad Sentientism, then not all conscious beings will be bearers of moral status. If panpsychists moreover adopt either a biocentric or a neurocentric criterion of either when consciousness combines or when these morally important forms of consciousness appear (which many of them seem to do), then the domains of moral consideration will remain anchored to living things. And the idea of moral consideration for all living is already, if not widely accepted, at least familiar.

In particular, panpsychism might do either or both of the following: solidify our belief in the moral status of animals, and expand moral status to plants and other organisms. Intuitively, the threshold for evidence required to conclude that a being in a panpsychist universe exhibits enough unity to have subjectivity, agency, or valenced experiences is substantially lower than the threshold required to conclude that a being in a fundamentally non-conscious world has consciousness. Where a non-panpsychist might be confident in the moral status of mammals and birds but uncertain about that of insects, the average panpsychist (if we may appeal to such a construct) seems to be someone confident in the moral status of insects but uncertain about that of plants.

Finally, panpsychism might have a moderate result regarding the unity of all minds: although individual minds are real and distinct, and might be worth caring about, the oneness which they all exist within also not only exists but is conscious, and to that extent has a claim on our identification. The one and the many are both conscious, and are not sharply separate but exist in some kind

[48] K. Frankish, *Does Panpsychism*, cit., p. 65.

of overlap or synthesis with each other. And depending on how we think our sympathies and identifications should track these two parts of ourself, this result might show selfishness to be incomplete or one-sided, without showing it to be strictly irrational.

In sum, the ethical implications of panpsychism depend on a number of things. They depend, firstly, on what version of panpsychism we are considering: restrictive or universalist, constitutive or emergentist, cosmopsychist or micropsychist, with any of a range of views on what sort of consciousness is present at the fundamental level. The ethical implications depend, secondly on how we answer various independent philosophical questions: what properties confer moral status? What makes some subject me? And they might also depend, thirdly, on personal values and dispositions: do I identify more with the aspect of me that is distinctive and sets me apart, or with the aspect which is more fundamental and (supposedly) is one and the same in all people?

Luke Roelofs - Nicolas Kuske

University of Texas at Airlington - luke.mf.roelofs@gmail.com
Technical University of Chemnitz - nicolas.kuske@gmail.com

Abstract

Panpsychism is a striking metaphysical claim: every part of the physical world has some form of consciousness. Does this striking claim have equally striking ethical implications? Does it change what duties we owe to which beings, or how we should understand the relation between self-interest and altruism? Some defenders as well as critics of panpsychism have suggested it does. Others have disagreed. In this paper, we attempt to survey and organize these existing discussions. We suggest that panpsychism is likely to have significant implications, but that they 1) stop short of some of the most radical implications sometimes associated with the view (by both critics and proponents), and 2) often depend heavily on the exact sort of panpsychism that is endorsed, as well as how one settles various independent questions about value and identity.

Keywords: *Panpsychism*, *Ethics*, *Consciousness*