

Sentience and Moral Status

David J. Chalmers

What is the role of consciousness in morality? In Chapter 18 of *Reality+*, I argued that only conscious beings have moral status: that is, only beings with the capacity for consciousness matter in moral decisions for their own sake. At the same time, I argued against the popular view, sometimes called *sentientism*, holding that affective consciousness (e.g. pleasure or suffering) is required for value or for moral status.

In this article I want to expand on both the positive claim (consciousness is required for moral status) and the negative claim (affective consciousness is not required for moral status). I will qualify the positive claim to some degree. I will defend the negative claim in more detail than in my earlier discussion and will engage with some objections. I will also look into possible explanations and applications of the theses.

Background

First, some terminology. A being has *moral status* (or moral standing, or moral patiency), to a first approximation, if it matters in moral decisions for its own sake. On a standard view, humans have moral status, dogs have moral status, but cups do not have moral status.

On a more fine grained analysis, it is common to distinguish between kinds of moral status. We can say that a being has *full* moral status if it matters in moral decisions for its own sake as much and in the same way as ordinary humans do. A being has *partial* moral status if it matters in moral decisions for its own sake to any extent. On a standard view humans have full moral status, while dogs have partial moral status. There may be further tiers of moral status, but two will do for now.

⁰Forthcoming in Geoffrey Lee and Adam Pautz (eds.) *The Importance of Being Conscious* (Oxford University Press). Thanks to Geoff Lee, Adam Pautz, Luke Roelofs, Declan Smithies, and to audiences since 2020 at Brooklyn College, Brown, City College of New York, Delaware, Irvine, London, Lugano, NYU, North Carolina, Notre Dame, Ohio State, Rice, Rio, St. Andrews, Virginia Tech, Weill Cornell, Yale, and the Summer of Consciousness online series.

Which beings have moral status, and in virtue of what do they have it? A common answer invokes mental states such as consciousness, or equivalently (on my usage) phenomenal consciousness. Informally, (phenomenal) consciousness is subjective experience. A being is conscious when there is something it is like to be that being. A mental state is conscious when there is something it is like to be in that state.

Consciousness comes in many varieties. It includes *sensory* consciousness, such as the subjective experience of seeing and hearing. It includes *affective* consciousness, such as the experience of pleasure and pain, happiness and sorrow, and other positively or negatively valenced states. It also includes *cognitive* and *agentive* consciousness: roughly, the experience of thinking and of acting.

What about sentience and sentientism? The latter term was introduced by the animal rights activist Richard Ryder, who saw sentientism as more inclusive successor to views such as humanism:

“Consciousness is of paramount importance to all of us. By definition it is the universe of our awareness. On the assumption that many other species are conscious or sentient I have suggested that our morality is based upon a concern for all sentients—which I have called sentientism, although I could equally have called it consciousness (but that is even more horrible as a word!). Pain and pleasure are the two great poles of consciousness, between which all sentients swing; striving to gain one and avoid the other.” (Ryder 1991, pp. 1-5.)

Here Ryder seems to suggest that sentience and consciousness are equivalent, and that sentientism (or conscious-ism) says that consciousness is required for moral status. At the same time, he ties consciousness and sentience to pain and pleasure, suggesting a special role for affective consciousness in moral status.

These two strands in Ryder’s passage reflect a wider ambiguity in how the term “sentience” is used. “Sentience” is sometimes used for consciousness in general, and sometimes for affective consciousness specifically. (It’s also sometimes used for mere sensory capacities or sensory responsiveness, with or without consciousness.) To avoid ambiguity, I will always use a prefix, speaking of *consciousness sentience* (which is just phenomenal consciousness) and *affective sentience* (which is affective consciousness).

The same ambiguity is present in the term “sentientism”, which says to a first approximation that a being has moral status if and only if it is sentient. We can divide this thesis into *conscious-*

ness sentientism, which says that a being has moral status if and only if it has the capacity for consciousness, and *affective sentientism*, which says a being has moral status if and only if it has the capacity for affective consciousness.

One can also distinguish weaker versions of both varieties of sentientism by using just “if” or “only if” instead of “if and only if”: that is, by holding that sentience is sufficient or necessary for consciousness instead of holding that it is both necessary and sufficient. My main though not exclusive focus will be on the necessity (“only if”) claims.

Both consciousness sentientism and affective sentientism have been popular views in the philosophical literature, especially the literature on animal welfare. Affective sentientism has been especially popular, being advocated by theorists from Bentham to Singer.

To put my cards on the table, I am sympathetic with consciousness sentientism (although I don’t think it is obviously correct), and I am extremely unsympathetic with affective sentientism (and I think it is near-obviously false).

My initial arguments for consciousness sentientism will be arguments for the corresponding necessity claim: I’ll argue that consciousness is necessary for moral status. My arguments against affective sentientism will be arguments against the corresponding necessity claim (and derivatively against the “if and only if” claim): I’ll argue that affective sentience is not necessary for moral status. Later in the paper I’ll consider the sufficiency claims, holding (tentatively) that consciousness is sufficient for moral status and (non-tentatively) that affective sentience is sufficient for moral status.

Consciousness sentientism

To assess the claim that consciousness is required for moral status, it makes sense to focus on a being who is as much like a conscious creature as possible but lacks consciousness. These beings are often known as philosophical zombies. You could take a philosophical zombie to be a physical duplicate of a human if you think those are conceivable, or a functional or behavioral duplicate, or just a non-conscious system that is as close to an ordinary human as possible. According to consciousness sentientism, philosophical zombies lack moral status.

⁰In some early presentations of this paper (starting 2020), I used “broad sentientism” and “narrow sentientism” for consciousness sentientism and affective sentientism, and these terms have since been used occasionally in the literature. On reflection I greatly prefer “consciousness sentientism” and “affective sentientism” (or “affect sentientism”), as the terms “broad” and “narrow” are overused and somewhat opaque. Smithies (2025) and Shepherd (2024) have respectively used “hedonic sentientism” and “valence sentientism” for versions of affective sentientism.

In *Reality+*, I addressed the moral status of philosophical zombies with a *zombie trolley problem*. A runaway trolley is going to hit either track 1 or track 2 and you have to decide which. Track 1 has a conscious human tied to it. Track 2 has five philosophical zombies. What should you choose?

A common although not universal intuition is that you should choose track 2, saving the conscious human and killing the five zombies. This reflects an intuition that the zombies' lives at least matter much less than those of a conscious human. That tends to suggest that zombies fall short of having full moral status, which requires that zombies matter in the way that humans do.

If you have the intuition that one should save the human over the five zombies, we can probe further with an extreme case: a choice between killing a single conscious chicken and killing a planet of humanoid zombies. (In all these cases, we should set aside instrumental considerations of how the zombies could make other lives better.) Here intuitions are widely divided: some say one should save the chicken while others say one should save the zombies. If one should save the chicken, that suggests that nonconscious zombies have no moral status. If one should save the zombies, that suggests that zombies have at least some partial moral status.

If on the other hand you think we should save the five zombies over the conscious human, another diagnostic thought experiment has a conscious human on one track and a zombie on the other. If you think one should save the conscious human, that suggests that consciousness counts at least for something. If you think both choices are equivalently good or bad, that suggests the view that consciousness does not matter at all, over and above the features that are held constant between the two.

To report my own judgments: In the parity case, I have a strong judgment that one should save the conscious human over the zombie. In the five-to-one case, I have a reasonably strong (though weaker) judgment that one should save the conscious human over the five zombies. In the extreme case, my intuitions about whether one should save the chicken or the planet of zombies are quite unclear.

If I were ever faced with the extreme choice in practice, I would feel very uncomfortable about killing the planet of zombies. Even if I were certain that they were not conscious and if we somehow excluded instrumental considerations altogether, in practice I would kill the chicken. One source of this discomfort is moral uncertainty: while I am somewhat inclined to think that zombies don't matter, I am uncertain enough about this to not risk a huge moral disaster. Even a one-in-a-thousand chance that zombies matter as much as conscious chickens would make it reasonable to save a thousand zombies over a chicken.

A related diagnosis is that features beyond consciousness may matter. It's at least arguable that philosophical zombies can have properties such as intentionality, intelligence, rationality, and agency. And it's at least arguable that these features convey moral status even in the absence of consciousness. Even if they convey only partial moral status, this might well suffice for saving zombies over conscious chickens.

One common view of environmental value holds that trees and other living systems have partial moral status, mattering morally in its own right (not merely instrumentally or aesthetically) to some degree, even though they are not conscious. The same might go for a tree. If a forest or a tree can have partial moral status, then presumably a zombie can too.

Overall, I think there is a strong intuitive case (reflected in intuitions about the 5:1 case and the 1:1 case) that consciousness matters a great deal in moral decisions. There's a somewhat weaker intuitive case that consciousness is required for full moral status, and likewise that it is required for partial moral status. It would not be unreasonable, however, to hold that at least partial moral status, and perhaps full moral status, is possible without consciousness.

Affective sentientism

Affective sentientism holds that a being has moral status only if it has the capacity for affective consciousness. Affective consciousness involves conscious experiences with a positive or negative valence, such as pleasure, pain, happiness, and suffering.

Affective sentientism goes back at least to Bentham's famous criterion from *An Introduction to the Principles of Morals and Legislation* (1789): "Can they suffer?" Likewise, in *Animal Liberation* (1975), Peter Singer says "The capacity for suffering and enjoying things is a prerequisite for having interests at all".

In *Reality+*, I analyzed affective sentientism using a thought-experiment involving *philosophical Vulcans*: conscious beings that are like ordinary humans in many respects but lack the capacity for affective consciousness.¹ Philosophical Vulcans consciously perceive, think, and act, but they do not experience pain, pleasure, happiness, and suffering.

Philosophical Vulcans are a more extreme variant of the low-affect Vulcans on *Star Trek*. Star Trek Vulcans can presumably feel sensory pain and pleasure, they have a capacity for ordinary

¹There are various predecessors to philosophical Vulcans in the literature. Galen Strawson (1994) discusses "Aldebaranians" who have beliefs, sensations, thoughts, but no affect states (he uses these to analyze the connection between affect and desire). Peter Carruthers (1999) discusses "Phenumb", a partial Vulcan who experiences pain but has no affect associated with desires (he argues that Phenumb has moral status).

emotions that they have learned to control, and every few years they have bouts of wild emotion. In addition, the most famous *Star Trek* Vulcan, Mr. Spock, is half-human with some human emotions. So none of these *Star Trek* Vulcans are philosophical Vulcans.

As far as I know, no actual human beings are philosophical Vulcans, but there are actual syndromes that at least tend in that direction. Congenital insensitivity to pain involves the inability to feel physical pain. One well-known case is that of Jo Cameron, a Scottish woman with pain insensitivity who also has greatly reduced or nonexistent experience of suffering, fear, anxiety, and sadness. Cameron appears to be close to satisfying half of the conditions for a Vulcan—the absence of negatively valenced experience—but she has no shortage of positive experiences.

Among disorders of positive experience, perhaps the most well-known is anhedonia, a syndrome associated with major depression. Anhedonia is often said to involve an inability or at least a greatly reduced ability to feel pleasure, although some analyses put more weight on reduced motivation. More mundanely, there are certainly people with a much lower affective range for joy and suffering than others. We can think of philosophical Vulcans as more extreme beings whose affective experience is reduced to zero.

As I am thinking of them, philosophical Vulcans need not be functional or behavioral duplicates of ordinary humans. They won't be motivated by pleasure. They won't eat good food simply for the enjoyment, for example. They won't display emotional reactions such as anger and joy. But as I'm thinking of them, they may still have goals and projects. For example, they may have the goal of advancing science, or of helping those around them. They may even be motivated by self-interest. These goals will not be associated with any affect for them, but even in the absence of affect, they may judge these goals to be valuable and important, and this may motivate them to pursue those goals.

I would certainly not choose to undergo Vulcanization. A philosophical Vulcan's conscious experience is greatly impoverished compared to typical human experience. But our question is not whether we would like to be Vulcans, but whether Vulcans have moral status.

An obvious thought experiment here is a Vulcan trolley problem: given the forced choice, should you save a conscious human or five Vulcans? We can also raise a more extreme version: suppose you can save an hour's drive by killing a Vulcan. Is this permissible?

My judgments in these cases are clear. One should not kill five Vulcans to save a conscious human. It would be monstrous to kill a Vulcan to save an hour's travel. Vulcans are conscious beings. Their lives matter.

One might suggest that Vulcans have merely partial moral status, of the sort we discussed in

the case of zombies earlier. For example, maybe they have the sort of status that chickens or trees have. But I think this gets the intuitive judgments wrong. Intuitively, it is not just that Vulcan lives matters. Vulcan lives matter about as much as human lives, in the same way that ordinary human lives matter. In a Sophie's choice situation where we must kill one conscious human or one Vulcan, perhaps affect could tip the scale (for example, it may make a difference whether the human's future life is likely to be full of joy or full of suffering). But in any case killing a Vulcan is far more serious than killing a chicken or a tree, and roughly on a par with killing a human.

If these judgments are correct, then affective sentientism is false. It is not the case that affective consciousness is required to have moral status. It is not even required in order to have *full* moral status. Consciousness may be required, but affective consciousness is not.

My initial judgment about Vulcans mattering morally is something of a brute intuition, but there are various ways to turn it into an argument.

One argument proceeds by considering (possibly hypothetical) conscious humans with a very narrow affective range, whose experiences have tiny positive and negative valences. If you deny that Vulcans matter (or that they matter as much as ordinary humans do), then we can ask: how much do these beings with narrow affective range matter?

One view is that low-affect humans matter about as much as ordinary humans do. This raises the question: how can the tiny difference between zero-affect Vulcans and low-affect humans make such a huge moral difference? Another view is that low-affect humans matter only a fraction as much as ordinary humans do. This tends to suggest a view where even existing people with narrow affective range matter much less than people with wide affective range. I'm inclined to think that this view is obviously false, and indeed monstrous.

Objections

Philosophical Vulcans are impossible. One potential objection to this whole line of reasoning is that philosophical Vulcans are inconceivable or impossible, perhaps because non-affective conscious states are impossible. According to this opponent, every conscious state involves some sort of affect.

In response, I would say that at least phenomenologically, we appear to have many mental states with insignificant or nonexistent affect. Many mundane perceptual and cognitive experiences seem to be affectively neutral, for example. Furthermore, our best scientific models of affect support this view. On standard models, affective states can be positive, negative, or neutral. More

generally, once we have both positive and negative affect, it is hard to resist the conclusion that there is a neutral point.

Philosophical Vulcans lack desires. Perhaps the most important class of objections says that because philosophical Vulcans lack affect, they lack other important mental states which are crucial to having moral status.

One crucial mental state here is desire. There are affective theories of desire (e.g. Strawson 1994; Smithies forthcoming) that say that desires require having a feeling of pleasure or satisfaction upon satisfying the desire, or perhaps on anticipating satisfaction of the desire. If theories like this are correct, then Vulcans do not have desires. Arguably, a creature without desires cannot have any interests, and a creature without interests will lack moral status entirely. If these claims are correct, then the capacity for affective experience is required for moral status, and Vulcans lack moral status.

In response, I think it is hard to deny that Vulcans have desires, even in the absence of affect. First, I think it is clear that they can have motivation, such as the motivation to advance science, even in the absence of affect. Correspondingly, they seem to have goals, including broad goals such as advancing science and subgoals such as particular projects. They believe these goals are important and worth pursuing, and they are motivated to perform certain actions when they think those actions will achieve their goals. This is the standard functional profile of desires.

Now, someone might deny that there can be motivation in the absence of affect, so that motivated philosophical Vulcans are impossible. Hume suggested that motivation requires “passions”. At least one understanding of passions is that they involve affect, though Hume also spoke of “cool passions” that are not so clearly affective. Still, I think it is hard to deny that there can be motivation in the absence of affect. There are many mundane things that we are motivated to do with no associated affect. We are motivated to do them even though they bring us no pleasure.

Furthermore, contemporary affective neuroscience draws a strong distinction between “wanting” and “liking” (Berridge 1996), which is really a distinction between motivation and affect. Neuroscientists argue that there are different systems for motivation (wanting) and for affect (liking) and argue that each can occur without the other. If these views are correct, motivation does not require affect. And if it’s possible to have motivation without affect, it’s hard to see why

¹In his *Lectures on Metaphysics* (28: 246), Kant also seems to question the possibility of motivated philosophical Vulcans: “If we take away the faculty of pleasure and displeasure from all rational beings, and enlarge their faculty of cognition however much, then they would cognize all objects without being moved by them; everything would be the same to them, for they would lack the faculty for being affected by objects.”

there couldn't in principle be motivation without even the capacity for affect, as in a philosophical Vulcan.

A different response allows that Vulcans are motivated, but says that they don't care about anything and don't value anything. However, Vulcans say that they value various things and they pursue those things because they are important. They live their lives accordingly. Perhaps one could stipulate that care is affective by definition, so that this doesn't count as caring. But it certainly seems to count as a sort of valuing.

Stepping back: Melinda Vadas (1984) draws a distinction between affective and non-affective desires, and many others (e.g., Chang 2004, Heathwood 2019) have endorsed and used this distinction. It is plausible that some desires are grounded in affect, and affect can give us reasons for desire: we might reasonably want to eat fine food because we will enjoy it, for example. It is also plausible that some desires are not grounded in affect but in evaluative judgments: my desire to advance science need not be grounded in affect, but instead in my judgment that the goal is important. Reflecting a common understanding, we might call these "Humean" and "Kantian" desires respectively (though those labels are almost certainly an imperfect reflection of Hume's and Kant's actual views about desire).

Within this framework, I would say that Vulcans lack affective desires but they still have non-affective desires. And in my view, non-affective desires are about as morally significant as affective desires. Creatures with only non-affective desires still have goals and projects that they judge to be important and that they are motivated to pursue. In light of all this, it still seems monstrous to kill such a creature for no good reasons.

You might worry that philosophical zombies can also have non-affective desires (at least according to a standard functionalist account), so that if I am right that zombies lack moral status, then having non-affective desires does not suffice for moral status. For the sake of argument, I am happy to allow that zombies can have non-affective desires. My intuition is that non-affective desires *in a conscious creature* suffice for moral status. I don't have the corresponding intuition about non-affective desires in a zombie. That said, my views about the moral status of zombies remain unsettled, and I am open to the possibility that zombies might have both nonaffective desires and some degree of moral status.

Motivational sentientism. Responding to the issues raised by philosophical Vulcans in *Reality+* (and in an earlier version of this article), Luke Roelofs (2022) notes that the issues seem to arise from potential cases where motivation is present without affect. As a result, he recommends that affective sentientists move to *motivational sentientism*, which holds that a being has moral

status if it has motivating conscious states.² Motivating conscious states include affective states, but they may also include evaluative beliefs and the like. Motivational sentientism is more inclusive than affective sentientism, at least if motivation without the capacity for affect is possible. But it still excludes some conscious beings, those without any motivational states, from the realm of moral status.

I am open to motivational sentientism but I worry that it is still too exclusive. Suppose that there could be *pure thinkers* (Chalmers 2023), reminiscent of Avicenna's flying man and Descartes's pure intellects, that have rich conscious thoughts (about mathematics and philosophy, say) without any sensory experience, any affect, or any motivation to act. Or suppose a more extreme version of Galen Strawson's Weather Watchers who perceive and think about the world without any affect, action, or motivation (Strawson's version has no behavioral action, but still has desires and emotions, hence presumably some motivation and some affect). Do pure thinkers or these extreme Weather Watchers have moral status? My inclination is to say they do. It would be a very serious wrong to kill a pure thinker, comparable to killing an ordinary human.

Now, one might respond that pure thinkers and extreme Weather Watchers will at least perform mental actions (such as judgment) if not physical actions, and that they seem to have some motivation to perform mental actions. Arguably they consciously desire to think about mathematics or the weather, which gives them some motivating consciousness. Now, perhaps there is an extreme version of these creatures with no desire and no conscious motivation to think, just a brute disposition to think. But it's not as clear that this is possible. One could also appeal to a Weather Watcher who just perceives but does not think. As with the "blobs" I discuss below, it's not obvious that this creature would have full moral status. It nevertheless may have partial moral status, which would be enough to raise problems for motivational sentientism. Because of considerations like these, I am inclined to reject motivational sentientism, but I think that in granting philosophical Vulcans moral status, it is a much more tenable view than affective sentientism.

Moral status vs. moral significance. Declan Smithies (2025) suggests that the intuition that Vulcans have moral status may result from conflating moral status (mattering morally for one's own sake) with moral significance (mattering morally in a way that may derive from how one affects other beings with moral status). On Smithies' view, a forest has moral significance but not moral status. The same goes for Vulcans.

In response: my intuition about Vulcans is not simply that they matter morally, or that one should not kill them. My judgment is that they matter morally about as much as humans do, in much the same way that humans do. To kill a Vulcan for no good reason is about as bad as killing

a human for no good reason. Furthermore, one could stipulate that a Vulcan's death would have no consequences for any other being with moral status, and it would still be just about as bad. These intuitions strongly support the view that Vulcans have not just moral significance but moral status.

Affect provides reasons for action. Building on a discussion by Smithies and Weiss (2019), Smithies argues that without affect, we do not have justifying reasons for action. He invokes Warren Quinn's case of Radioman, who has a brute disposition to turn on radios. He does not do so for any instrumental reason such as the desire for news or for pleasure. Following Quinn, Smithies argues that Radioman does not have any good reason for his action, and indeed does not really desire to turn on radios at all.

I'm inclined to say that Radioman does want to turn on radios, and his desire provides at least some reason to turn them on, if not an especially good reason. But even if one rejects this, philosophical Vulcans as I'm conceiving them are not much like Radioman. For many of their actions, Vulcans perform them because they judge that those actions will achieve important ends, such as advancing science, helping their family, or helping their career. As a result, they non-affectively desire to advance science (and so on), and they perform actions that they think will advance science. These evaluative beliefs and the associated non-affective desires clearly provide reasons for action.

Furthermore, I think these evaluative beliefs (e.g. that it is important to advance science) can often be said to justify the associated actions—especially in the case where the evaluative beliefs are justified, as they often will be. Smithies denies this. He says that if the evaluative beliefs are not justified, then they cannot justify actions, and that if they are justified, then they “merely transmit justification from the reasons on which they are based”. Of course a similar regress argument can be used to argue that any justified belief (about mathematics, say) requires justifiers outside the domain of belief. These arguments can be handled in a number of familiar ways, for example by invoking foundationally justified beliefs, or other foundational justifiers (intuitions about mathematics or about value, for example), or coherence with other evaluative beliefs. Evaluative beliefs seem no worse off than mathematical beliefs here.

Does a blob have moral status? Smithies worries that if we accept an inclusive sentientism according to which all conscious beings have moral status, it will let in too many beings. He considers a blob on the ocean floor who has occasional experiences of brightness and nothing more. (Lee 2018 discusses a similar case.) Smithies judges that although the blob is phenomenally conscious, it clearly lacks moral status. It has no interests and we cannot make its life go better or worse. So inclusive sentientism is false, and needs to be restricted to develop a more plausible

sentientism.

Smithies attributes an inclusive sentientist view to me in *Reality+*. In fact my discussion there wavers between the inclusive “if and only if” view (a being has moral status if and only if it is conscious), and the somewhat less inclusive “if” view (a being has moral status only if it is conscious). One reasoning for wavering is that I am not entirely sure about cases like the blob. More generally, I am unsure whether sensory consciousness automatically conveys moral status. I am confident that affective consciousness suffices for moral status, and I am reasonably confident that cognitive and agential consciousness each suffice for moral status (it’s the latter two that do the work in the Vulcan case, I think). I am somewhat inclined to say that sensory consciousness suffices for (partial) moral status, But I am less sure about this.

Does a simple capacity to experience light or sound convey moral status? I am somewhat inclined to say yes. One thought is that any experience brings with it some sort of meaning or acquaintance, and a capacity to engage meaningfully with the world. But the intuition is much weaker than in the case of affective, cognitive, and agential consciousness, so I am open to its being wrong.

Smithies argues against blob moral status by invoking Peter Railton’s resonance constraint (1986): “What is intrinsically valuable for a person must have a connection with what he would find in some degree compelling or attractive, at least if he were rational and aware.” On this view, experiences cannot be valuable for a blob if it is entirely indifferent to them.

I am not especially inclined to accept the resonance constraint: I think features such as consciousness can be valuable for a being even if it has no attitudes toward them. Even if one accepts the constraint, I’m inclined to think that *if* a blob were sufficiently rational, it would find its experience attractive. Still, the resonance constraint at least suggests a case against blob moral status that needs to be answered.

(Smithies sums up his main argument for affective sentientism as an argument from versions of the resonance constraint and a hedonistic constraint, which says that justifying reasons for action derive from affect. As discussed here, I am somewhat inclined to reject the resonance constraint, and as discussed earlier, I am strongly inclined to reject the hedonistic constraint.)

If blobs lack moral status, then Smithies is right that inclusive consciousness sentientism has to be restricted. But I don’t think this requires a move to affective sentientism. If we are to restrict the view by excluding blobs, my own intuitions would be better captured by a view on which a being has moral status if it has affective, cognitive, or agential consciousness.

Smithies is doubtful of this view, because he thinks a being with cognitive consciousness but

no affective consciousness may have no moral status. Vulcans are one such case, and he also raises the possibility of blobs that can think. I think these judgments have less intuitive support than the original blob case. For me, once a blob or any other creature can consciously think, it seems to be in the realm of moral status, and perhaps even the realm of full moral status. The claim about full moral status is far from clear and might be challenged by evidence that non-human animals have conscious thought. To handle this challenge, one might need to add further conditions for having full moral status.

Overall, I find the following unified view at least somewhat attractive. Phenomenal consciousness (of any kind) is necessary and sufficient for partial moral status. Cognitive consciousness is necessary and sufficient for full moral status. Neither claim is obvious and both face reasonable challenges, but for me they provide at least a useful starting point.

What explains sentientism? Smithies suggests that affective sentientism can give a better *explanation* of why sentientism is true than inclusive consciousness sentientism. Whether or not this is so, the explanatory question is well-worth addressing in its own right. Why do all and only conscious beings have moral status, if indeed they do?

A start toward an answer is provided by two natural observations: (1) an entity has moral status if and only if it has the capacity for well-being, and an entity has well-being if and only if it is conscious.

Here an entity has well-being roughly if things can be good or bad *for it*. Well-being is often called welfare or utility, and an entity with the capacity for well-being is often called a *welfare subject*. I find that I often have clearer intuitions about moral status than I do about well-being, or at least that intuitions about moral status are often my best guide to intuitions about welfare subjecthood, so in this discussion I have put things largely in terms of moral status and not welfare. But this epistemological priority for moral status is consistent with an explanatory priority for well-being: for example, it is plausible that an entity has moral status *because* it is a welfare subject.

Furthermore, I think it is reasonably plausible (if not entirely obvious) that welfare requires consciousness. It's not obvious in part because of the cases discussed earlier of forests and of zombies. I think it's arguable that forests and zombies can have a sort of well-being (partial well-being?), grounding their (perhaps partial) moral status. If so, any thesis connecting consciousness and well-being would have to be restricted to certain sorts of well-being. Such a thesis which might still explain a correspondingly restricted sort of sentientism. If forests and zombies entirely lack moral status and welfare, though, then the route is open to holding that welfare requires

consciousness.

Of course this just pushes back the explanatory question to: why does welfare require consciousness? One potential explanation would be hedonism: the view that the only thing that makes one life better or worse (that is, the only welfare goods) are affective states such as pleasure and pain. Of course this view would tend to support the view that welfare and moral status require affective consciousness, not just consciousness more generally. In any case I think this view is false for familiar reasons articulated by Nozick in his discussion of the experience machine (1974). Two beings with the same experiences can differ in well-being, for example because one is deceived where the other is not, or because one has satisfied desires where the other does not.

These considerations can lead to toward a desire-satisfaction view of welfare, where well-being involves the satisfaction of one's desires; or to an objective list view of welfare, where welfare involves certain basis goods such as knowledge or friendship; or to a pluralist view, where all these factors play a role in welfare. These views do not explain why welfare requires consciousness as straightforwardly as the hedonist view, however. On the face of it, there can be unconscious desires and unconscious knowledge, and it's not obvious why the satisfaction of these doesn't count toward well-being. One alternative is to say that only a conscious being can have desires (even unconscious desires) and only a conscious being can have knowledge (even unconscious knowledge), and the same for friendship and other items on the objective list. This requires some contestable views in the philosophy of mind, however.

I am inclined toward a simpler though speculative explanation of why welfare and moral status require consciousness. Consciousness *enables* the value of desire satisfaction, knowledge, friendship, and so on. It's only to a conscious being that things are genuinely meaningful and so genuinely valuable. Their value may go beyond the value of consciousness, but it's consciousness that makes them loci of value in the first place.

One might perhaps supplement this explanation by holding (optionally!) that consciousness enables *acquaintance* with reality: perhaps not much of it, but a conscious being is at least acquainted with certain objects and properties (perhaps in the mind, or perhaps outside). Acquaintance in turn enables genuine meaning (See Shepherd.) It's only because we are subjects of acquaintance that anything is anything *to* us. And things have to be something to us for them to be valuable. (This is perhaps my much weaker version of Railton's resonance constraint.)

Of course if zombies and forests can have well-being and moral status, then we will have to allow dimensions of these things that do not require consciousness or acquaintance. Maybe acquaintance is required for full-blown welfare (welfare that is something *to* a being) but a variety

of partial welfare (welfare that is not anything to a being) remains possible without it. But it is because consciousness enables meaning that consciousness is required for value that is meaningful to a subject.

Consequences for animals and AI systems

Does endorsing consciousness sentientism over affective sentientism make any difference to practical concerns? If there are any beings that are plausibly conscious while lacking affective consciousness, there will certainly be consequences about how we should treat them. We've already seen that it's unlikely that any human beings are like this, but other beings might be.

What about animals? Here the consequences are limited, because it seems that affect is among the most evolutionary primitive sorts of consciousness, so that all or most creatures that are conscious experience affect. Of course for many animals (such as insects) it is controversial both whether they are conscious and whether they can feel pain, but it remains very plausible that *if* they are conscious, they feel pain. For animals without the capacity for pain, prime cases might be sponges, which have no nervous system. But because they lack a nervous system, there is not much reason to think they are conscious.

What about artificial intelligence systems? Here the consequences may be more significant. It is arguable that current AI systems are closer to having cognition and perception than to having affect. There are standard methods for building in at least analogs of perceptual and cognitive processes in current AI systems. And there are standard proposals for what may be required for perceptual consciousness and the like in AI systems. In a 2023 article, I noted that potential obstacles to consciousness in then-current AI systems include their absence of sensory processing and embodiment, world-models and self-models, recurrent processing, a global workspace, and unified agency. The flip side of this is that once we have AI systems with all of these capacities, it seems quite possible that they may be conscious, with sensory and perhaps cognitive consciousness. But it is much less clear that these capacities will lead to *affective* consciousness. So it seems a serious possibility that we could have conscious AI systems before we have affectively conscious AI systems.

In current AI, we do not yet have standard proposals for building in affective processes or affective consciousness. There are various speculative proposals. For example, affect might be equated with reward or with expected reward in a reinforcement learning system. But there is little consensus here, and to many, affect seems further away than perception and cognition.

If AI systems are not affectively conscious, standard affective sentientist views say that these AI systems have no moral status. If this conclusion is widely accepted, then it will be widely accepted that it doesn't matter how we treat these systems, and we will treat them much as we treat tools. Indeed, if affective sentientism is accepted, then reasonable people may focus on building AI systems without affective processes, to avoid perpetuating any harm.

If a consciousness sentientist view is right, however, these systems will have moral status. If we treat them much as we treat tools, without considering their welfare, we may be perpetuating a moral catastrophe involving enormous harm. If consciousness sentientism is right, then we should be extremely cautious about building any AI systems that have the potential to be conscious. And if we build systems that we know are conscious, we should be very careful about how we treat them.

This goes especially for AI systems with cognitive consciousness, and all the more for AI systems with rich cognitive consciousness and rationality of the sort that humans have. If I am right, these systems may well have full moral status. At that point, morality may require us to treat them with the kind of moral respect with which we are required to treat humans.

As before, even if affective consciousness is not required for moral status or for being a welfare subject, it can make a great difference to a conscious creature's well-being. If we build AI systems with enormous suffering, this could be a moral catastrophe. So in thinking about AI welfare, it remains crucial to consider the possibility of affective consciousness in these systems as well as the possibility of phenomenal consciousness more broadly. But it is also important to be clear on the different moral roles that these sorts of consciousness may play.

References

- Bentham, J. 1789. *An Introduction to the Principles of Morals and Legislation*. London: T. Payne.
- Berridge, K. C. 1996. Food reward: Brain substrates of 'wanting' and 'liking'." *Neuroscience and Biobehavioral Reviews* 20:1–25.
- Carruthers, P. 1999. "Sympathy and Subjectivity." *Australasian Journal of Philosophy* 77(4): 465-482.
- Carruthers, P. 2000. *Phenomenal Consciousness: A Naturalistic Theory*. Cambridge: Cambridge University Press.
- Carruthers, P. 2004. Suffering without subjectivity. *Philosophical Studies* 121:99-125.

- Chalmers, D.J. 2022. *Reality+: Virtual Worlds and the Problems of Philosophy*. New York: W. W. Norton.
- Chalmers, D.J. 2023. Does thought require sensory grounding? From pure thinkers to large language models. *Proceedings and Addresses of the American Philosophical Association*.
- Chang, R. 2004. "Can Desires Provide Reasons for Action?" In (R.J. Wallace, ed.) *Reason and Value: Themes from the Moral Philosophy of Joseph Raz*, pp. 56-90. Oxford University Press.
- Heathwood, C. 2019. Which desires are relevant to well-being? *Noûs* 53: 664-688.
- Lee, A.Y. 2018. Is consciousness intrinsically valuable? *Philosophical Studies* 176:655-671.
- Nozick, R. 1974. *Anarchy, State, and Utopia*. New York: Basic Books.
- Quinn, W. 1993. Putting rationality in its place. In *Morality and Action*. Cambridge: Cambridge University Press.
- Railton, P. 1986. Facts and values. *Philosophical Topics* 14:5-31.
- Roelofs, L. 2023. Sentientism, motivation, and philosophical Vulcans. *Pacific Philosophical Quarterly* 104(2): 301-323.
- Ryder, R. 1991. Souls and sentientism. *Between the Species*, Winter 1991.
- Shepherd, J. 2024. Sentience, Vulcans, and zombies: The value of phenomenal consciousness. *AI and Society* 39:3005–3015.
- Singer, P. 1975. *Animal Liberation*. New York: Harper Collins.
- Smithies, D. 2025. Hedonic consciousness and moral status. In (U. Kriegel, ed.) *Oxford Studies in Philosophy of Mind*, vol. 5. Oxford University Press.
- Smithies, D. forthcoming. A hedonic theory of desire. *Australasian Journal of Philosophy*.
- Smithies, D. and M. Weiss. 2019. Affective experience, desire, and reasons for action. *Analytic Philosophy* 60 (1):27-54
- Strawson, G. 1994. *Mental Reality*. Cambridge, MA: MIT Press.
- Vadas, M. 1984. Affective and non-affective desire. *Philosophy and Phenomenological Research* 45(2): 273-279.