Journal of Philosophical Theological Research
Print: 1735-9791
Online: 2538-2500

The Importance of Being Conscious in the Debate over AI

Charles Taliaferro

Professor Emeritus of Philosophy, St. Olaf College, Northfield, Minnesota, USA

taliafer@stolaf.edu

Abstract

This paper defends the evident reality of consciousness, including the existence of conscious, enduring selves, against recent objections raised by Helen Yetter-Chappell and Anil Seth. Yetter-Chappell advances a non-theistic idealism that echoes Hume's bundle theory, denying a substantial self behind experience, while Seth characterizes the self as a controlled hallucination generated by predictive brain processes. I argue that both accounts mischaracterize the common-sense phenomenology of self-awareness: when one reads, thinks, or feels pain, one is directly aware of oneself as a substantial, enduring subject – not of a "lurking" entity behind experience. I further contend that Seth's rejection of appearances is epistemically precarious and that his analogy between consciousness and heat fails to explain the subjective sensation of hotness. The paper then turns to the debate over artificial intelligence. If consciousness is bracketed and intelligence is analyzed behavioristically or functionally, AI may indeed equal or surpass human intelligence; however, this is profoundly different from being intelligent in the sense of conscious self-awareness. Through a critical engagement with Daniel Dennett's thought experiment of the unconscious robot "Zeke," I show that consciousness is both evident and indispensable to what we value in friendship, love, work, and moral life. The paper concludes that while behaviorism and functionalism may grant intelligence to AI, they cannot capture what it means to be a conscious, self-aware being – a reality of paramount importance for how we understand ourselves, other creatures, and God.

Original Research

Keywords:

consciousness, intelligence, Yetter-Chappell, Anil Seth, artificial intelligence, theism.

Received: 2026/02/05; **Received in revised form:** 2026/04/12; **Accepted:** 2026/05/27; **Published online:** 2026/06/10

Taliaferro, C. (2026). The Importance of Being Conscious in the Debate over AI. *Journal of Philosophical Theological Research*, (Special Issue on AI, Consciousness, and Human Agency), 28(3), 21-38. https://doi.org/10.22091/jptr.2026.17505.3695

Copyright © The Author

Introduction

Consciousness plays an essential role in world religions. In some forms of Hinduism, Brahman is pure, infinite consciousness. Buddha-consciousness is a central notion in most forms of Buddhism. In the Abrahamic faiths, God's creation and sustaining of the cosmos is a conscious divine act, and God's omniscience consists in God's awareness of all dimensions of the created order. God's justice and mercy, God's love and providential care are all within the scope of God's boundless conscious life. These traditions also stress the importance of our consciously responding in intentional acts of prayer, worship, repentance, reverence, loving and caring for the creation. We are to live life in light of God's presence (coram Deo), in the faith and hope for life after this life (Taliaferro, 2024).

In the book *Consciousness and the Mind of God,* I argued that the credibility of theism is closely tied to the credibility of believing in the reality of human consciousness (Taliaferro, 1994). If certain forms of eliminative materialism give us compelling reasons to doubt the reality of consciousness in humans and some nonhuman animals, then God's consciousness should also come under suspicion.. Something similar is true in the current debate over artificial intelligence.

If we bracket the reality of consciousness and assume, for example, a behaviorist or functional analysis of intelligence, then AI has already equalled or surpassed human intelligence, or will soon do so. AI has surpassed us in solving mathematical problems, composing texts, creating songs, practicing surgery, offering therapy,planning , and even conducting wars with the use of mechanical flying robots or drones, and so on.

In this paper, I submit there are compelling reasons to recognize the reality of ourselves as self-aware conscious subjects who think, reason, have sensation, experienced emotions, agency, memory, and so on over time. I then contend that the reality of consciousness plays a vital role in our assessment of AI.

Conscious Selves

Along with many other philosophers, including David Lund, Roderick Chisholm, Stewart Goetz, J. P. Moreland, and others, I contend that we are immediately aware of ourselves enduring over time with an array of sensations, thoughts, feelings, and possessing the power of agency as physically embodied beings (Loose, Menuge & Moreland, 2018). In order not to replicate what I have published in other books, in this paper I will defend the

reality of conscious selves against objections raised by two philosophers who present recent challenges to the position I hold: Helen Yetter-Chappell, in *The view from Everywhere: Realist Idealism without God,* and Anil Seth, in *Being You: A New Science of Consciousness.*

Yetter-Chappell articulates and defends the coherence and plausibility of a form of idealism that does not require a theistic basis. Whatever its merits (and there are many), I focus on her view of the conscious self.

In several places, Yetter-Chappell supports David Hume's bundle theory of the self (Yetter-Chappell, 2025, pp. 17, 65). Famously, Hume claimed he never perceived himself, but only perceptions of pleasure or pain, love or hatred, heat or cold, and so on (as cited in Yetter-Chappell, 2025, p. 17). She echoes Hume's claim in her own experience:

> Like [Mark] Johnson (and Hume), I fail to find a self lurking about my mental acts, as something over and above what I experience. I find the computer before me, my hands typing, the thoughts running through my head, the sound of a bird, the ache in my left calf. I do not find the computer in its entirety, but a particular perspective on the computer. All these experiences are unified into the total experience of me at a moment (Yetter-Chappell, 2025, p. 173).

She distinguishes three positions in terms of the self and experience, which she calls the robust mind, the thin mind, and no mind. The first refers to the mind or self as a substantial reality, an individual subject. The second is described as an experiencer (some kind of subject?), but not something that adds to an ontology of experience. The third is the denial that there is any kind of subject.

Claims like Hume's – and, by extension, Yetter-Chappell's – have bewildered many (but not all) philosophers. Arguably, any of Hume's claims about perception (or emotion or sense) seem inextricably to be examples of self-awareness. Who can experience pain without the person feeling herself in pain? What sense can be made of a person claiming to feel hot without her being committed to the view that she herself is feeling hot? Yetter-Chappell's terminology seems skewed. Those who claim that we are aware of ourselves as substantial subjects enduring over time would make no sense of the idea of a lurking self behind Yetter-Chappell's experiences. In her description of her own experiences, who sees the computer if she does not? When she is typing, doesn't she have an awareness that it is not just her hands that are typing, but that she is the one typing? To read her account, cited above, don't you need to experience yourself (a substantial individual person who endures over time) as reading the whole text over time? The so-called "thin mind" seems (at least to

me) preposterous. How might you claim there is a person (a reader) who is reading the text, but there is no ontological significance to the claim there is more than the experience of reading (namely, there is a reader)?

My challenge to Yetter-Chappell is philosophical and based on what seems to me (and many others) common-sense phenomenology. But the neuropsychologist Oliver Sacks has proposed that some severe pathology can so disturb a person's inner life that something like a Humean predicament can seem real.

"For Hume, personal identity is a fiction – we do not exist, we are but a consecution of sensations, or perceptions. This is clearly not the case with a normal human being, because he owns his own perceptions. They are not a mere flux but his own, united by an abiding individuality or self. But what Hume describes may be precisely the case for a being as unstable as a super-Touretter, whose life is, to some extent, a consecution of random or convulsive perceptions and motions, a phantasmagoric fluttering with no center or sense. To this extent, he is a 'Humean' rather than a human being. This is the philosophical, even theological fate which lies in wait, if the ratio of impulse to self is too overwhelming. It has affinities to a 'Freudian' fate, which is also overwhelmed by impulse – but the Freudian fate has sense (albeit tragic), whereas a 'Humean' fate is meaningless and absurd" (Sacks, 1985, p. 146).

Let us now turn to Anil Seth. In *Being You,* he acknowledges that we appear to exist as conscious subjects but raises the question whether this is an illusion: "The notion that there is a single unique conscious self (a soul?) that persists over time may be grossly mistaken" (Seth, 2022, p. 4). In his view, we are "Beast machines," and being you and me emerges "from the way our brain predicts and controls the internal states of the body" (Seth, 2022, p. 6). Like Yetter-Chappell, Seth construes those of us who believe we have immediate self-awareness as positing a lurking self; he charges: "The self is not some immutable entity that lurks behind the shadows of the eyes" (Seth, 2022, p. 154). He rejects common-sense appearances: "the way things seem is often a poor guide to how they actually are" (Seth, 2022, p. 17).

Seth's rejection of common-sense appearances is also reflected in his account of our experience of colors (red is his specific example) and his view that our perception of the world should be thought of as what he calls a controlled hallucination. He claims that what it is like to see red (or see something red) should be largely understood as an experience of what it is not (blue, green, etc.) (Seth, 2022, p. 53). He grants that there is a phenomenological subjective experience of red (Seth, 2022, p. 85), but it is part of a neurological hallucination: "The entirety of perceptual experience is a neuronal fantasy that remains yoked to the world through a continuous making

and remaking of perceptual best guesses, of controlled hallucinations" (Seth, 2022, p. 87).

Seth does not claim that we have established the identity of the mental with the physical, but he claims we are making progress. His example is the scientific discovery of heat being mean kinetic energy.

> The history of thermometry, and its impact on our understanding of heat, offers a vivid example of how the ability to make detailed, quantitative measurements, on a scale defined by fixed points, has the power to transform something mysterious into something comprehensible … If consciousness turns out to be like temperature – which is to say, if there is a single physical process that underlies and is identical to 'being conscious' – the payoff would be spectacular (Seth, 2022, p. 35).

Seth cites the fact that neurological damage can undermine our sense of being a unified self as evidence that there is no unified self. Under normal, healthy conditions, we experience ourselves as unified centers of consciousness (Seth, 2022, p. 54), but this is vulnerable to disintegration (Seth, 2022, pp. 149–150).

Seth ends his book offering some consolation about our death: Your death is the death of "a controlled hallucination" (Seth, 2022, p. 276). We will all pass into oblivion or nothingness, but this is "a return to the eternity that each of us at some time emerged from" (Seth, 2022, p. 276).

I offer five replies.

First, Seth's depiction of those of us who believe in our reliable, direct awareness of ourselves is mistaken. We believe that self-awareness is not an inference to a *lurking* entity behind experience. Instead, we contend that when you read this sentence, you experience yourself as a substantial individual reader. Seth describes those of us who believe in a conscious self as maintaining that the self is immutable (cited above). Not at all. An immutable (non-changing) self would be akin to a statue or frozen self, not a living, changing, conscious being. Instead, we maintain that selves experience and cause change.

Second, Seth's repudiation of trusting appearances is precarious epistemically; it can put one on a slippery slope. Arguably, Seth has to rely on an appearance principle when he claims that our perceptual life is a controlled hallucination. That is, Seth is committed to the view that (at least in his case), based on his scientific view of the brain, *it appears that our perceptual life is a controlled hallucination.* A repudiation of appearances needs to be qualified, as in Smith's claim *that when he has taken LSD*, *the apparent reliability of his*

emotions (feeling paranoia, for example) *is less reliable than when he is not using psychotropic drugs.* In my view, Seth still needs to accept some version of an appearance principle (e.g., if some state of affairs appears to be the case to a subject, she has some reason to believe the state of affairs is true). (For a defense of appearance epistemology, see Taliaferro & Reasoner, 2026.)

Third, on redness and controlled hallucination: unless we defy all experience, we should (in my view) acknowledge the reality of our experience of redness and hallucinations. While I reject the view that our normal, healthy perception of each other and the world is a hallucination, I grant that people sometimes (under the influence of LSD, for example) actually hallucinate. While the hallucination of a butterfly is not a butterfly, it is a visual image. Seth seems to leave us in limbo about the status of visual images and fields. He acknowledges that when a person sees red, there is no redness out there in the world, nor is there any redness in the brain (Seth, 2022, p. 86). Well, where is it? One promising option is to claim that conscious selves perceive the world by means of a visual field or representation of the world.

Fourth, Seth's example of heat does nothing to advance the claim that consciousness is a brain process. The analysis of heat in terms of molecular motion is a fine depiction of non-sentient heat, but it does nothing to account for the actual conscious sensation of hotness.

Fifth, the evidence that our self-awareness is subject to distortion, fragmentation, or utter destruction due to all kinds of harms is no evidence at all that there are no unified, conscious selves. Actually, the embedded, intuitive appearance that we are conscious selves can come to the fore when we contemplate the death of someone we love. Shortly before I started writing this paper, a very dear friend of mine for many years, Michael, whom I love dearly, died. Speaking for myself, I find no consolation in believing Michael never was a conscious self, but a controlled hallucination who (or that?) has irreversibly perished.

Actually, while Seth seems to foreclose the possibility that Michael or any of us might survive death, he does think we should take reports of near-death and out-of-the-body experiences (OBEs) seriously.

> When people report supernatural or otherwise bizarre experiences, like OBEs, we should take seriously their reports. They probably do have the experiences they say they have. People have had real out-of-the-body experiences for millennia, but this does not mean that immaterial selves or immutable souls have ever actually left any physical bodies. What these reports reveal is that first-person perspectives are put together in more complex, provisional, and

> precarious ways than we will ever have direct subjective access to (Seth, 2022, p. 159).

This is not the place for me to philosophically assess OBEs. But I suggest that what these reports might also reveal is the enduring, ostensibly sensible idea that we are not controlled hallucinations but conscious selves who may not share the same fate as our bodies at death. I suggest that taking OBE accounts "seriously," should lead us to accept that persons claim to have direct subjective awareness of themselves some distance from their bodies, even if we contend that such experiences are caused by nero-chemical processes. Moreover, I have yet to come across a report of an OBE in which someone claimed they have experienced themselves as "immaterial selves or immutable souls." The reports available in the literature are far simpler, e.g. I witnessed physicians operating on my body from a distance (see Gary Hbermas 2023, 323-356).

AI withoutConsciousness

I have conceded at the outset that if we deny the reality of consciousness, and settle for some form of illusionism (there are no conscious selves, or the conscious self is an illusion or hallucination), then the status of AI is philosophically formidable. If some machines can outperform humans at most, if not all, tasks, shouldn't we treat them as persons or person-like creations, perhaps entitled to human rights and duties?

To develop this argument, let us consider a proposal advanced by Daniel Dennett.

Dennett asks us to imagine a time (perhaps sooner than we think) when scientists and philosophers will think that the belief in consciousness is akin to believing in geo-centricity, the idea that the sun rotates around an immovable earth. He envisions "a day when philosophers and scientists and lay-people will chuckle over the fossil traces of our earlier bafflement about consciousness. 'It still seems [they will say] as if these mechanistic theories of consciousness leave something out, but of course that's an illusion. They do in fact explain everything about consciousness that needs explanation'" (Dennett, 2001, p. 37).

To enhance his position, he asks us to imagine a robot he calls Zeke. Zeke is an AI who has all the (apparent) cognitive faculties we have; she has reasoning, sensory perception, memory, the power to move, speak, and hear, but all of these (as well as Zeke herself) are not conscious or, in Dennett's terminology, unconscious. While unconscious, Zeke has sensory awareness of

all her states. Dennett then imagines the following conversation with Zeke:

> "Zeke, do you like me?" /
> "Of course I do." /
> "Did you mind me asking?" /
> "Well, yes, it was almost insulting." /
> "How do you know?" /
> "I just recall feeling a bit annoyed or threatened." (Dennett, 2013, p. 291)

Dennett creates a scenario in which it seems impossible to tell whether Zeke is a conscious self or an unconscious machine. He asks us why we would care whether Zeke is conscious or not: "Or, more personably, why would you care whether you are, or became, a zimbo [a non-conscious or unconscious being]? In fact, you'd never know" (Dennett, 2013, p. 292).

If there is no such thing as consciousness, as a real state of self-awareness and awareness of love and the world, Dennett has a point. But I suggest the thought experiment can bring to light the evident reality and value of consciousness.

Consider three points.

First, beginning with a minor point, Dennett's analogy with geo-centricity can be misleading. He seems to assimilate those of us who persist in believing in conscious selves to those who persist in believing in a stationary earth with an orbiting sun. Dennett seems to be saying that while it appears that the earth is stationary, it is not. What is misleading here is that motion is measured in frameworks. In my current environment, a coffee house in North America, the *place on earth I am working is stationary.* This is not *a mere appearance,* something to outgrow with basic science, but true, notwithstanding that from the standpoint of the solar system the earth is rotating at approximately 1,000 miles per hour, orbiting the sun. When I walk at 2 miles an hour to pay for my coffee, it would be bizarre to claim that I am travelling at a rate of 1,002 miles per hour. Back to consciousness: the reality of consciousness (say, my awareness of an extreme toothache) does not depend on spatial frameworks or even conceptual (or philosophical) frameworks. I would have no less certainty in the reality of my toothache if I was assured that such agonizing pain would not be recognized from the standpoint of Dennett's naturalism. So, belief in the appearance of consciousness might be likened to the appearance that the earth is stationary insofar as both are true, while the second depends on the framework of measurement (establishing the compatibility of recognizing

stationary parts of earth and the movement of the earth itself), but the appearance of consciousness is so resilient that any framework that omits it has left out a vital, evident reality.

Second, more significantly, there is something vexing about using the terms "nonconscious" and "unconscious" interchangeably. The term "unconscious" is typically used either to describe persons who have lost consciousness or who have beliefs and desires that are repressed or not directly accessible to conscious awareness. A familiar Freudian claim would be that a patient has an Oedipus complex, an erotic desire for one's mother. In answer to Dennett's question of "why care," I do not care about the feelings of non-conscious objects, but I very much care about the unconscious feelings of my own and those around me. I earlier alluded to my friend Michael. I believe that we both enjoyed a reciprocal, loving friendship, but it would break my heart to learn (based on Michael's extensive psychoanalysis) that unconsciously he regarded me as a vile imposter whom he loathed. If the unconscious exists (and some philosophers are skeptical on the issue), I think we have abundant reasons to care about unconscious beliefs and desires.

Third, and most importantly, if we take up the question of why anyone should care whether they are conscious, I am somewhat baffled. I would rather be anesthetized during a major medical procedure which would otherwise cause me extreme pain. But in life itself, when we enjoy or endure friendships, romance, worthwhile work, show compassion and seek justice, play, make art, drive, vote, compete in sports, write and read, conscious awareness is indispensable.

What I believe Dennett's thought experiment brings to light is that if consciousness is an illusion, whether a machine is conscious is immaterial. But if it is not illusory, then it matters a great deal, both in our own case, in the case of relating to nonhuman animals, extraterrestrials (if there are any), God (if there is one) or other sacred realities (Brahman, Buddhahood, and more), and in relating to the machines we make.

Tentative Conclusion

If I am right about the evident reality and value of our being conscious selves (and thus right about the existence of consciousness), then we have reason to be suspicious of philosophies that eliminate or denigrate consciousness. While we therefore have reason to reject behaviorism and functionalism with respect to humans (and perhaps some nonhuman life), we may grant that behaviorism and functionalism would recognize that AI is intelligent. I would only add that this is profoundly different from recognizing AI as consciously self-aware beings.

The whole point of this paper is to bring to light the evident reality and importance of being conscious. As a Christian theist who has defended at length the reality of God as a supremely conscious living creator and sustainer of the cosmos, it is not my aim to argue that only biological embodied beings can be conscious. Indeed, it is my hope (along with billions of other religious practitioners) that this supremely conscious God will deliver us from what Seth describes as oblivion or nothingness.

I thank Kirk Allison, Jock McGreger, Greg Jessen, and members of the L'Abri Rochester community for fruitful dialogues about the nature of consciousness and AI.

Conflict of Interest

The author declares no competing interests.

References

Dennett, D. (2001). The zombie hunch: Extinction of an intuition? In A. O'Hear (Ed.), *Philosophy at the new millennium*. Cambridge University Press.

Dennett, D. (2013). *Intuition pumps and other tools for thinking*. W. W. Norton.

Habermas, G. (2023). Evidential near-death experiences. In A. Menuge, B. Rickabaugh, & C. Hasker (Eds.), *Minding the brain* (pp. 323–356). Discovery Institute Press.

Loose, J., Menuge, A. J., & Moreland, J. P. (Eds.). (2018). *The Blackwell companion to substance dualism*. Wiley-Blackwell. https://onlinelibrary.wiley.com/doi/book/10.1002/9781119468004

Lund, D. (2005). *The conscious self*. Humanity Books.

Sacks, O. (1985). *The man who mistook his wife for a hat*. Summit Books.

Seth, A. (2022). *Being you: A new science of consciousness*. Faber.

Taliaferro, C. (1994). *Consciousness and the mind of God*. Cambridge University Press.

Taliaferro, C. (2024). Mind and consciousness. In *St. Andrews Encyclopedia of Theology*. https://www.saet.ac.uk/Christianity/MindandConsciousness

Taliaferro, C., & Reasoner, P. (2026). *The meaning of life from sacred and secular perspectives*. Bloomsbury.

Yetter-Chappell, H. (2025). *The view from everywhere: Realist idealism without God*. Oxford University Press.