Synthese (2025) 206:252
https://doi.org/10.1007/s11229-025-05338-3

ORIGINAL RESEARCH

How to resist the Fading Qualia Argument

Andreas L. Mogensen[1]

Received: 30 July 2024 / Accepted: 18 October 2025

Abstract
The Fading Qualia Argument is perhaps the strongest argument supporting the view that in order for a system to be conscious, it does not need to be made of anything in particular, so long as its internal parts have the right causal relations to each other and to the system's inputs and outputs. I show how the argument can be resisted given two key assumptions: that consciousness is associated with vagueness at its boundaries and that conscious neural activity has a particular kind of holistic structure.

Keywords Consciousness · Substrate-Independence · Vagueness · Transparency of consciousness · Holism

1 Introduction

Many believe that in order for a system to be conscious, it does not need to be made of anything in particular, so long as its internal parts have the right causal relations to each other and to the system's inputs and outputs. As a result, many also believe that the right software could in principle allow there to be something it is like to inhabit a digital computer, controlled by an integrated circuit etched in silicon. A recent expert report concludes that if consciousness requires only the right causal relations among a system's inputs, internal states, and outputs, then "conscious AI systems could realistically be built in the near term" (Butlin et al., 2023, p. 6).

Is it really plausible that any system with the right functional organization will be conscious - even if it is made of toilet paper and stones (Searle, 1980) or consists of a large assembly of people with walky-talkies (Block, 1978)? My goal in this paper is to raise doubts about what I take to be our strongest argument supporting the view

✉ Andreas L. Mogensen
andreas.mogensen@philosophy.ox.ac.uk

[1] Faculty of Philosophy, University of Oxford, Oxford, UK

Published online: 05 November 2025

that consciousness is *substrate independent* in this sense. The argument I have in mind is Chalmers's *Fading Qualia Argument* (Chalmers, 1996, pp. 253–263). I show how it is possible to resist the argument by appeal to two plausible (but controversial) assumptions: namely, that consciousness is associated with vagueness at its boundaries and that conscious neural activity has a particular kind of holistic structure.

I begin in Sect. 2 by explaining what the Fading Qualia Argument is supposed to show and the broader dialectical context it inhabits. In Sect. 3, I give a detailed presentation of the argument. In Sect. 4, I show how the argument can be answered given the right assumptions about vagueness and the structure of conscious neural activity. At this point, I rely on the assumption that vagueness gives rise to truth-value gaps. In Sect. 5, I explain how the argument can be answered even if we reject that assumption. In Sect. 6, I say more about the particular assumption about the holistic structure of conscious neural activity needed to resist the Fading Qualia Argument in the way I outline.

2 Organizational invariance and two arguments

Chalmers argues for the *principle of organizational invariance*, according to which, given any system that has conscious experiences in any nomologically possible world, any other system with the same fine-grained functional organization will have qualitatively identical experiences in any nomologically possible world, regardless of its material composition.[1]

Two physical systems are said to realize the same *functional organization* if they can be divided into physical components with the same causal dependencies relating their inputs, outputs, and internal states. For sameness of *fine-grained* functional organization, we require that the sameness of causal dependencies relating the systems' inputs, outputs, and internal states holds at a sufficiently fine-grained level of description that the two systems determine exactly the same behavioral capacities. Moreover, we require that the systems realize corresponding states at corresponding times. Any two systems sharing the same fine-grained functional organization, so understood, are said to be *functional isomorphs*.

Chalmers (1996) presents two arguments in support of the principle of organizational invariance: the *Fading Qualia Argument* and the *Dancing Qualia Argument*. In the first instance, the Fading Qualia Argument is intended to rule out the nomological possibility of functional isomorphs that differ in respect of being conscious at all, while the Dancing Qualia Argument is intended to rule out the nomological pos-

[1] Note that the principle of organizational invariance is consistent in principle with the hypothesis that there are no nomologically possible functional isomorphs of conscious neural systems that are not composed of neurons, glia, and the like (Godfrey-Smith, 2016; Cao, 2022). It supports only the conditional claim that if such systems are nomologically possible, then they will have qualitatively identical experiences. Accordingly, my aim is to show that Chalmers fails to mount a convincing argument for accepting this conditional, without necessarily taking a stand on whether inorganic functional isomorphs of conscious neural systems are genuinely nomologically possible. Thanks to an anonymous referee for pressing me to clarify this point.

sibility of functional isomorphs that are both conscious but nonetheless differ in the character of their experiences.

I focus my discussion on the Fading Qualia Argument. Some readers of *The Conscious Mind* may recall that Chalmers (1996, p. 270) argues that the Dancing Qualia Argument can be transposed through minor alterations to buttress the Fading Qualia Argument and that this modified Dancing Qualia Argument is actually a stronger argument. However, Chalmers has since changed his mind and now views the Fading Qualia Argument as the stronger of the two (see Chalmers, 2010, p. 24 n.7). Moreover, Chalmers's transposition of the Dancing Qualia Argument is intended to buttress the conclusion that the spectrum of cases involved in the Fading Qualia Argument does not involve fading qualia. The response to the Fading Qualia Argument that I outline is designed to allow us to answer the argument without committing ourselves to the hypothesis that fading qualia are involved.

3 The Fading Qualia Argument

Suppose that there exists a system, Robot, with the same fine-grained functional organization as Dave during a time when Dave undergoes a vivid conscious experience, such as the experience of watching a loud, colorful basketball game. Suppose for reductio that Robot lacks any conscious experience altogether, because Robot is built from silicon chips, and silicon chips are the wrong sort of components for realizing consciousness, in spite of their ability to duplicate the functional organization of Dave's brain.

We then imagine a spectrum of cases ranging from Dave to Robot, across which more and more of Dave's brain is replaced with functionally isomorphic components made of silicon chips. At the first step, we imagine replacing a single biological neuron with an artificial replacement component made of silicon that performs the same fine-grained functional role. As we move across the spectrum, we replace additional neurons, one by one, always maintaining the same fine-grained functional organization.[2] Eventually, no biological material remains.[3]

By assumption, Dave is conscious and Robot isn't. The question is how this transition from consciousness to unconsciousness plays out as we move along the spectrum. One possibility is that consciousness degrades gradually as we imagine more and more of Dave's brain replaced. Another is that there is no gradual degradation

[2] In principle, we could also imagine more fine-grained replacements that don't involve replacing whole neurons: see Chalmers (1996, p. 255). One might also suppose that more coarse-grained replacements could suffice for maintenance of the same behavioural capacities, thereby preserving the system's fine-grained functional organization, as defined by Chalmers. While a neuron-by-neuron replacement scenario may not be strictly necessary to preserve fine-grained functional organization, this is the scenario on which Chalmers focuses, and I follow suit. I'm grateful to an anonymous referee for pressing me to clarify this point.

[3] Even if we grant the possibility of a functional isomorph of the human brain realized in silicon, we might doubt that it is possible to replace individual neurons with silicon components while retaining the same fine-grained functional organization. Here, I set aside any worries about the feasibility of single-neuron replacement. See Chalmers (1996, pp. 259–260) for suggestions about how the argument can be reformulated or amended to address such concerns.

and instead a point at which vivid conscious experience abruptly switches off completely between two neighboring cases. Each of these possibilities is argued to be extremely implausible, and, with it, the assumption that Robot isn't conscious.

Chalmers makes quick work of the idea that there might be *suddenly disappearing qualia*, i.e., a point at which consciousness abruptly switches from vivid to totally absent. He rejects this possibility because, he says, an abrupt transition from consciousness to unconsciousness of this kind would imply "brute discontinuities in the laws of nature unlike those we find anywhere else" and because any point of discontinuity would be "entirely arbitrary" (Chalmers, 1996, p. 255).[4]

As for the possibility of *fading qualia*, Chalmers thinks we can reject this possibility for the following reason. Imagine a system halfway along the spectrum between Dave and Robot. Call this system Joe. By assumption, Joe's experience is degraded relative to Dave's. Thus, where Dave sees glaring, bright red and yellow uniforms worn by the players, Joe might experience only a tepid pink and murky brown; where Dave experiences the loud noises of the roaring crowd, Joe experiences only a sort of distant rumble. Nonetheless, Joe isn't able to report any of this. Since he is Dave's functional isomorph, he, like Dave, reports having vivid experiences of sound and colour. Nor is Joe in a position to notice that these reports systematically misdescribe what he's experiencing, assuming that noticing requires a particular kind of cognitive processing that supervenes on the fine-grained functional organization of the brain. After all, Dave is Joe's functional isomorph, and he fails to notice any such discrepancy, since none exists for him. On a functional construal of belief, Joe will even count as believing, like Dave, that his experiences really are vivid, although they are really dim and murky. All this, Chalmers (1996, p. 257) claims, is extremely implausible: "For a sentient, rational being that is suffering from no functional pathology to be so systematically out of touch with its experiences would imply a strong dissociation between consciousness and cognition. We have little reason to believe that consciousness is such an ill-behaved phenomenon, and good reason to believe otherwise."[5]

[4] In an earlier presentation of the argument (Chalmers, 1995), an additional consideration against the possibility of suddenly disappearing qualia is suggested: namely, that if they are possible then "we could switch back and forth between a neuron and its silicon replacement, with a field of experience blinking in and out of existence on demand." (315). However, the possibility that conscious experience is normally composed of discrete frames with intervening gaps is an empirical hypothesis worth taking seriously: see, *inter alia*, VanRullen (2016) and Herzog et al. (2020) for evidence supporting this kind of discretization. If there are such gaps in consciousness, it is to be expected that they would go unnoticed, since there is no one 'there' to register the gaps when they occur (Neumann, 1990). By contrast, in the sorts of cases that drive the subsequent Dancing Qualia Argument (Chalmers, 1996, pp. 266–274), we are to imagine that my experience changes markedly while I am conscious throughout.

[5] Block (2023, pp. 454–455) suggests that, on a natural reading, Joe *can* be said to suffer from functional pathology, since the cognitive processes that are recruited in phenomenal introspection are completely misfiring in his case. Arguing for the contrary conclusion, Chalmers (1996, p. 257) tells us that "Joe's processes are *functioning* as well as [Dave's] – by hypothesis, he is functionally isomorphic." (Emphasis in original.) Block (2023, p. 455) understandably objects that it is question-begging to define pathology as supervening on fine-grained functional organization in the present context. I think it's clear that the assumptions about the accessibility of phenomenal character to introspection on which the argument relies need to be made clearer and put on a firmer basis. However, since it's obvious that something

4 Vagueness and the Fading Qualia Argument

The Fading Qualia Argument is obviously similar to the Sorites Paradox. Still, Chalmers (1996, p. 261) insists that his argument isn't soritical. It doesn't say: Dave is conscious; if X_i is conscious, then if X_{i+1} is obtained by replacing a single neuron in X_i with silicon, X_{i+1} is conscious also; therefore, by repeated application of *modus ponens*, Robot is conscious. In conceding, at least for the sake of argument,[6] that consciousness might in principle fade gradually to nothing as we make small changes, Chalmers seems to allow that some small change in the system, like replacement of a single neuron, could take us from a case of faint and badly degraded conscious experience to no conscious experience at all. Therefore, he does not assume the conditional premise stated above.

Nonetheless, it seems to me that we can resist the Fading Qualia Argument by interpreting the spectrum of cases to which it appeals as giving rise to vagueness. Here is how. At the outset, we say that it is determinate that the individual has a vivid experience of the basketball game unfolding. The colors are bright, the sounds loud. Once we've replaced a good portion of the original neural tissue, it's not the case that the individual is determinately conscious and experiencing the world as murky and muffled. Instead, it's indeterminate whether the individual is having an experience that is qualitatively identical to the original experience, and determinate that they aren't having any other experience. At this point, it is therefore indeterminate whether they are having any experiences at all. After enough of the brain has been replaced, it becomes determinate that they aren't having an experience that is qualitatively identical to the original experience, and it remains determinate that they aren't having any other experience. At that point, it's therefore determinate that they aren't having any experiences at all.

Suppose we enumerate the members of the sequence as $X_1, X_2, \ldots, X_n$, where X_1 is Dave and X_n is Robot. Let instances of the sentence schema $\ulcorner X_i$ has $E\urcorner$, obtained by replacing the variable i with the name of an integer between 1 and n, assert that the corresponding element in the sequence has exactly the same vivid experience of the basketball game unfolding that Dave has. Assume we reject an epistemicist theory of vagueness (Williamson, 1994) and treat vagueness as giving rise to truth-value gaps. Assume, furthermore, that we accept the description of the sequence outlined in the previous paragraph. We can then say that the sentence schema $\ulcorner X_i$ has E and X_{i+1} does not have $E\urcorner$ has no true instances. In that sense, there are no suddenly disappearing qualia.

Assume, furthermore, that we accept a theory of vagueness that recognizes degrees of truth (e.g., Goguen, 1969; Lewis, 1970; Kamp, 1975; Machina, 1976; Edgington, 1992, 1996; Smith, 2008). For concreteness, suppose we accept a supervaluationist theory that lets us talk about smaller and larger subsets of the set of permissible sharpenings and treats sentences that are true according to larger subsets of sharpen-

intuitively bizarre is going on in Joe's case, I wouldn't be surprised to learn of a successful patch for this vulnerability in the argument.

[6] Elsewhere, he writes: "There is something odd about the idea that a system with n elements could not be conscious but a system with $n+1$ elements could be" (Chalmers, 1996, p. 297).

ings as having greater degrees of truth (see Williamson, 1994, p. 154–156; Keefe, 2000, pp. 171–172 for discussion). Then we can say that instances of the sentence schema ⌜X_i has E⌝ decrease bit by bit in their degree of truth as we move across the spectrum. This provides some sense in which consciousness 'degrades' gradually, although ⌜X_i does not have E⌝ has no true instances prior to the point where it is also true that the system is completely unconscious.[7] In that sense there is no determinate change in phenomenal character as we move across the spectrum, setting aside the eventual loss of consciousness. But nor are there suddenly disappearing qualia.

Besides allowing us to go between the horns of Chalmers's dilemma, why might someone who rejects the principle of organizational invariance wish to interpret the $X_1, X_2, \ldots, X_n$ spectrum in this way? Suppose that we are attracted to a biological theory of consciousness on which consciousness essentially is a particular kind of functional organization of neural tissue, and so requires a neurophysiological substrate, much as water essentially is the liquid phase of H_2O, and not the liquid phase of just any polar molecule that is potable, transparent, etc. Evidence suggests that conscious experiences in human subjects involve what Dehaene (2014, p. 137)[8] calls "a coherent brain-scale assembly": a state involving synchronized and integrated information processing that binds together distributed populations of neurons and facilitates their shared encoding of a coherent neural representation.[9] Therefore, one natural way to develop a biological theory of consciousness is to propose that E is some particular holistic pattern of brain activity, N, and any experience, E', that is determinately different in character is a holistic pattern of brain activity, N', of a determinately different kind. Arguably, what happens as more and more neurons are replaced isn't that we transition between different holistic patterns of brain activity. At each stage, if there is some holistic pattern of brain activity going on, it can only be N. After all, the overarching pattern of activity in the system remains the same. What changes is the material composition of the system. As we proceed through the sequence, it instead

[7] A somewhat similar idea is proposed by Bostrom (2006), who suggests that there is no difference in *quality* between Dave and Joe's experiences, but a difference in the *quantity* of experience that each has. However, in Bostrom's conception, the sense in which Dave has more experience than Joe is supposed to be the same sense in which there is more pain when more individuals are in pain. This, I take it, is distinct from the idea that ascriptions of phenomenal consciousness admit of degrees of truth and that such ascriptions are less true in respect of Joe than in respect of Dave. Moreover, the hypothesis that there is only a difference in phenomenal quantity between Dave and Joe's experiences does not suffice to rebut the Fading Qualia Argument. We need to know what happens at other points in the spectrum. If Bostrom's view is that decreases in the degree of consciousness always preserve phenomenal character, it follows that Robot is conscious. If his position is that decreases in the degree of consciousness preserve phenomenal character up to the final step, we have suddenly disappearing qualia, whereas Bostrom makes no objection to Chalmers's dismissal of the possibility of suddenly disappearing qualia.

[8] It is worth emphasizing that Dehaene himself does not endorse the view that consciousness requires a neurophysiological substrate. Dehaene (2014, pp. 259–261) argues that simulations of the relevant cortical architecture appear feasible in principle, and he sees "no reason why they would not lead to an artificial consciousness." (p. 261) Thanks to an anonymous referee for encouraging me to clarify this.

[9] Of course, it is hardly uncontroversial that conscious neural activity involves global patterns of activity of this kind: see Lamme (2006, 2010) and Block (2007). Nonetheless, global theories of the neural correlates of consciousness are reasonable to believe given current evidence and it would be a significant liability if the Fading Qualia Argument required us to assume their falsity. Moreover, as I argue at the end of sect. 6, local theories appear to put independent pressure on the epistemological assumptions made by the argument.

becomes less and less clear that the pattern of activity that's in place can be described as a pattern of *brain* activity.[10] Exactly what counts as *brain* activity is surely vague. Just how much active matter needs to be neural matter? There's presumably no sharp boundary. Once we're far enough along the sequence - but not too far - it's simply going to be indeterminate whether the particular assembly of neural cells and silicon parts counts as a brain, as opposed to something else.[11][12]

Assume we are happy to go along with the assumptions I've outlined. Assuming we do not endorse an epistemicist theory of vagueness and instead treat vagueness as giving rise to truth-value gaps, we ought then to say that the schema ⌜X_i has N and X_{i+1} does not have N⌝ has no true instances. Assuming that we accept a theory of vagueness that recognizes degrees of truth, such as the supervaluationist theory that identifies higher degrees of truth with larger subsets of the set of permissible sharpenings, we can also say that instances of the sentence schema ⌜X_i has N⌝ decrease bit by bit in their degree of truth. If we then also count all instances of the biconditional schema ⌜X_i has E iff X_i has N⌝ as determinately true, then, on the standard supervaluationist logic, indeterminacy of any instance of ⌜X_i has N⌝ entails that ⌜X_i has E⌝ must also be indeterminate. In this way, we can end up committed to saying everything I said above.

If we can say all that, we appear to have defeated the Fading Qualia Argument. We don't have suddenly disappearing qualia, in the sense that ⌜X_i has E and X_{i+1} does not have E⌝ has no true instances.[13] Nor do we have fading qualia, in the sense Chalmers imagines. Joe is not determinately having an experience as of muted colors and muffled sounds. Rather, it is indeterminate that Joe is having any experience at all, since it is indeterminate that he has the same vivid experience of the basketball game Dave has, and determinate that he has no other experience. Nothing in the set- up leads to bizarre failures of introspective access to the character of experience on behalf of any subject that can truthfully be described as conscious: for any phenomenal property, P, such that Dave is in a position to know that E has P, the sentence schema ⌜X_i has E and X_i is not in a position to know that E has P⌝ has no true instance.

Should we nonetheless be concerned that Joe fails to register that he isn't determinately conscious and may assert the contrary? I don't see why. After all, we are already granting for the sake of argument that Robot does not register that they are not conscious and will assert the contrary. Since the possibility of a system like Robot

[10] By analogy, suppose a bronze of Aristotle is cast and put atop a pedestal at the Lyceum. His admirers gradually replace imperceptibly small parts of the sculpture with gold, retaining the same form throughout. Eventually, no bronze is left. At some point, it will be indeterminate whether there is a bronze atop the pedestal. *Ipso facto*, it will be indeterminate that there is a bronze of Aristotle atop the pedestal. Nonetheless, it is determinate that there is not a bronze of any other philosopher, nor of any other subject.

[11] The same can be said in respect of whether the system counts as *biological*. Thus, any theory that requires a biological substrate for consciousness can be defended using the conceptual resources outlined here, even if it does not insist on a neuronal basis for consciousness. A neurofunctional theory is relied on purely for illustration.

[12] Compare Carruthers (2019, pp. 151–152), who argues for the possibility of indeterminate instances of phenomenal consciousness by arguing that it may be indeterminate whether a given animal shares the global broadcast architecture with which he identifies phenomenal consciousness.

[13] If we are supervaluationists, we do, however, remain committed to the truth of the sentence, "There is some i such that X_i has E and X_{i+1} does not have E." If that worries you, read on to Sect. 5.

follows trivially from denying the principle of organizational invariance, Chalmers's argument rests on a claim about the introspective powers of *conscious* individuals: namely, that "conscious beings are generally capable of forming accurate judgments about their experience, in the absence of distraction and irrationality" (p. 257). On the present analysis, no point in the sequence constitutes a falsifying instance for a principle of this kind.

Moreover, Chalmers (1996, pp. 257, pp. 269–70) emphasizes that the introspective access of conscious subjects to the character of their experience is to be understood as an empirical regularity that reflects our observations and experiences. The posit that a person who is neither determinately conscious nor determinately unconscious might mistakenly assert or even believe that they are determinately conscious violates no empirically supported regularity of which I am aware, not least because there are no uncontroversial instances of indeterminate consciousness from which to generalize.

Others might be worried about the proposed response to the Fading Qualia Argument simply in virtue of my reliance on the assumption that there are borderline cases of phenomenally consciousness. This is a controversial assumption. Many philosophers, including Chalmers (1996, p. 105, p. 297), have the intuition that there are no indeterminate cases of phenomenal consciousness (Searle, 1992; McGinn, 1997; Tye, 2021).

However, explicit arguments for this claim are few (Antony, 2008; Simon, 2017) and they appear to face significant challenges (Hall, 2023). Given that the evolutionary trajectory from the earliest single-celled prokaryotes to human beings proceeds gradually as a result of the accumulation of selectively advantageous mutations over the course of billions of years, the hypothesis of a sharp boundary that separates conscious from unconscious life seems incredible, leaving extreme views like panpsychism or eliminativism as the only alternatives if we reject the possibility of borderline cases (Godfrey-Smith, 2020; Schwitzgebel, 2023).

Furthermore, the assumption that consciousness is sharp is close to question- begging in the current context. It is obviously vague what counts as brain activity. Therefore, anyone who finds plausible the view that consciousness is essentially a form of brain activity has good reason to demur from the assumption that consciousness cannot have borderline instances, given our current assumptions about vagueness. Indeed, the same can be said given virtually any mainstream theory of the physical basis of phenomenal consciousness (Antony, 2006; Schwitzgebel, 2023).

5 An apology for suddenly disappearing qualia

In the previous section, I relied on the idea that a plausible way to develop a biological theory of consciousness is one on which E is some particular holistic pattern of brain activity, N, and what happens as more and more neurons are replaced isn't that we transition between different holistic patterns of brain activity, but simply that it becomes less and less clear that the pattern of activity that's in place can be described as a pattern of *brain* activity, although it is determinate throughout that no other pattern of brain activity with the holistic character needed for consciousness is involved. In that case, epistemicism entails that $\ulcorner X_i$ has N and X_{i+1} does not have $N \urcorner$ has a true

instance. If we assume that all instances of the biconditional schema ⌜X_i has E iff X_i has N⌝ hold, it follows that ⌜X_i has E and X_{i+1} does not have E⌝ must also have a true instance. Assuming no other experience is had at any point in the spectrum, epistemicism thereby commits us to the possibility of suddenly disappearing qualia.[14]

If we find epistemicism plausible, this might seem like an embarrassment.[15] After all, Chalmers tells us that the possibility of suddenly disappearing qualia is absurd and implies that for it to be plausible that replacement of a single neuron results in the loss of consciousness, the person's experience must already be of severely muted colors and badly muffled sounds. I disagree. If we think that brain activity is necessary for consciousness and take seriously the idea that replacement of *this* particular neuron can make the difference as to whether a system constitutes a brain – as epistemicists surely must – we need feel no embarassment in supposing that cut-off is associated with a sudden loss of vivid conscious experience.

To explain why, I rely on an analogy. Imagine that I use a pencil to write out the sentence, "Teru is wearing brightly colored yellow socks," on a sheet of paper. This sentence has a certain intentional content. It represents Teru as the wearer of bright yellow socks. Imagine now that I write out this sentence using a successively fainter and fainter hand, until I apply so little pressure to the page that no sentence at all is inscribed thereon. As you scan your eyes down the page, the sentence "Teru is wearing brightly colored yellow socks" becomes gradually fainter and fainter, until it disappears altogether. But the same doesn't hold true of the color of Teru's socks, as represented by the sentences I write. Those colors don't get any fainter. So long as there is a determinate sentence there to be read at all, the sentence represents Teru's socks as being exactly as brightly colored as its predecessors, no matter how faint my handwriting. The content of the sentence does not change, insofar as there is any determinate sentence present to have a content at all. It is only the vehicle that gradually degrades.

Consciousness too has intentional content. Some, including Chalmers (2004), go so far as to hold that phenomenal properties are identical to certain representational properties (Harman, 1990; Dretske, 1995; Tye, 1995; Lycan, 1996; Byrne, 2001). As with the case of written sentences, we need to distinguish the properties represented *in* experience from properties *of* the experience. Recall Dave, who is watching a loud and colorful basketball game in action. Suppose part of Dave's experience is of a yellow triangle of fabric, seen as part of a basketball player's jersey. That experience is presumably not itself either yellow or triangular.

[14] As noted previously, it also follows from supervaluationism that the sentence "There is some i such that X_i has E and X_{i+1} does not have E" comes out true, although ⌜X_i has E and X_{i+1} does not have E⌝ has no true instances. The discussion in this section may therefore serve also to reassure readers who have residual doubts about the ability of supervaluationism to avoid an implausible commitment to suddenly disappearing qualia.

[15] Epistemicism does potentially allow us to drop one of the philosophically controversial assumptions outlined at the end of the previous section: namely, the assumption that 'phenomenally conscious' is a vague predicate. Even if 'phenomenally conscious' is coextensive with some neurophysiological predicate that is vague, it need not follow that 'phenomenally conscious' is itself vague (compare Antony, 2006, pp. 522–523). Epistemicism arguably entails that any vague predicate is coextensive with some sharp predicate (if only in an idealized language): namely, a sharp predicate that picks out the location of the borderline between the vague predicate's extension and its complement.

A plausible hypothesis, then, is that there exists a spectrum of possible cases involving conscious experience that is like the sequence of ever-fainter sentences that I described above. In other words, we can imagine a spectrum of cases involving conscious experiences with the same intentional content, but which differ in that the vehicle of content – i.e., the conscious experience – gradually degrades from one case to the next, up to the point at which there is finally nothing there at all.[16] The content remains the same throughout, until there's no content. Moreover, we can imagine that the spectrum appealed to in the Fading Qualia Argument works like that. If so, I think the argument can be defeated while conceding that there is a point at which experience abruptly switches from vivid to totally absent.[17]

A key motivation for identifying phenomenal and intentional properties is a line of argument going back at least to Moore (1903), which maintains, roughly, that our ordinary mode of access to the character of experience is indirect and goes via access to the intentional content of experience. If I try to bring my attention to the character of my experience of the blue of the sky, I find that I end up focusing on the blue of the sky, as represented in my experience. There doesn't seem to be any separate mental blue in my mind. We needn't here assume the *strong transparency hypothesis* – rejected by Moore – on which it is *impossible* to become directly aware via introspection of intrinsic properties of experience distinct from the experience's intentional content (Harman, 1990; Tye, 1995). In what follows, I assume only the

[16] In saying that the pre-existing conscious vehicle for the given content ceases to exist, we obviously need not in principle rule out the possibility that a different unconscious vehicle supporting a similar content comes into being in parallel, assuming we do not think that phenomenal consciousness is a necessary condition for intentionality (Searle, 1990, 1992; Horgan & Tienson, 2002; Loar, 2003; Mendelovici, 2018). Thanks to an anonymous referee for pressing me to clarify this point.

[17] In a recent paper, Sebastián and Martínez (2024) also try to defend the possibility of suddenly disappearing qualia, appealing to an imagined scenario where, H, a measure of the vividness of a person's experience, is governed by the non-linear differential equation $\dot{H} = LH + H^3 - H^5$, where the value of L corresponds to the number of active neurons in the brain. A dynamical system governed by this equation exhibits a bifurcation at which an arbitrarily small decrease in the value of the control parameter L changes a positive value of H from a stable to an unstable equilibrium point. Sebastián and Martínez assert that when this occurs, "[t]he system 'falls off a cliff', and quickly evolves to the only remaining attractor, $H=0$ … [E]xperience is suddenly extinguished." (p. 308) However, Sebastián and Martínez offer no positive reason to suppose that the vividness of experience is in fact governed by a non-linear differential equation exhibiting this kind of behaviour, asserting merely that we should not assume otherwise. By contrast, the analogy that I propose is one we have good reason to take seriously, since it serves only to highlight general properties of intentional contents and their vehicles, and I take it as obvious that conscious experiences are vehicles for intentional contents. Moreover, in Sebastián and Martínez's system, the value of the variable H still evolves continuously from a positive value to a zero value, and we cannot in fact say whether the time evolution is rapid or slow in any intuitive sense without additional assumptions. It is true that there is a discontinuous change in the stability of certain values of H. However, since stability is arguably not a magnitude, this does not conflict with Chalmers's observation that, in all known cases, "the dependence of one magnitude on another continuous magnitude is continuous" (p. 256). One could perhaps suggest that the presence of consciousness depends on a variable of this kind taking a value that represents a stable equilibrium point relative to the control parameter within a system of this kind. However, this can't be true of H, considered as a measure of the vividness of experience: non-zero values for H entail the presence of phenomenal consciousness, regardless of stability. It's not clear what variable could be such that it is necessary for someone to be conscious at a time that the value of that variable be at a stable equilibrium at that time and such that its stability is subject to discontinuous change as a result of changes in the volume of neural matter that preserve fine-grained functional organization.

more plausible *weak transparency hypothesis*, on which it's at the very least rare and/or difficult (Kind, 2003).

Suppose Joe has had a large part of his brain replaced with silicon chips, but is still conscious. Assuming weak transparency, for Joe to be thoroughly and bafflingly out of touch with the character of his experiences, he arguably needs to be mistaken about their intentional content. Even if intentionalism is false and there is some non-intentional difference in the character of Joe's and Dave's experiences, given weak transparency, Joe's failure to notice would hardly be baffling or extraordinary. If Joe is to be mistaken about the intentional content of his experiences, that presumably requires their intentional content to differ from the content of Dave's experiences. But that need not be true, since Joe's experience might be degraded relative to Dave's only in the sense that a faint inscription of "Teru is wearing brightly colored yellow socks" is degraded relative to an inscription made by pressing hard with a newly sharpened pencil. Just as those two sentences nonetheless have the same intentional content, so Dave and Joe would have experiences with the same intentional content, and those experiences would be very difficult, if not impossible, to distinguish introspectively.

More generally, as we traverse the spectrum, the world is not experienced as dimmer and fainter. In that sense, there are no fading qualia. It is the vehicle that gradually decays, not the content. The intentional content of experience remains fixed up to the very end, when the vehicle of content becomes so utterly degraded that it fails to support any content at all. In that sense, we have a case of vivid conscious experience disappearing suddenly as a result of a minor transition. But there needn't be anything arbitrary or mysterious going on here, since the analogy with fainter and fainter inscriptions of "Teru is wearing brightly colored yellow socks" shows us that this sort of change is perfectly intelligible. It's just a fact about the way intentional contents relate to their vehicles that exactly the same intentional content can be supported by a badly degraded vehicle all the way up to the point at which some slight additional degradation of the vehicle means there's actually no vehicle left to support any content at all.

6 Holism

We never said explicitly according to what pattern neurons are replaced with silicon chips as we make the transition from Dave to Robot. It might have been natural to imagine that individual neurons were being replaced at random. To assess the robustness of the proposed response to the Fading Qualia Argument, we should make sure to explicitly consider alternative scenarios that might be used to run the argument.

Imagine, then, that we do not randomly replace individual neurons, but first replace all and only those neurons encoding, say, visual information. Suppose also that Joe occurs at a point along the spectrum at which replacement of all and only those neurons encoding visual information is complete. Someone might imagine that those who treat conscious experience as being essentially a pattern of brain activity will think that in this scenario, Joe's experience will be like Dave's except that it has no visual qualia, since we have preserved those patterns of brain activity that encode

the remaining information of which Dave is conscious (compare Searle, 1992, pp. 66–67; Schneider, 2020, pp. 451–453). Maybe it was indeterminate at some point along the way whether the system still had visual qualia. Even so, we might think, if conscious experience is essentially a pattern of brain activity, then once we've replaced all the neurons encoding visual information, it has to be determinate that visual qualia are gone.

If Joe's experience were like Dave's except that it had no visual qualia, then Joe would exhibit a bizarre failure of introspective access to the character of his current experience if he failed to notice that he lacks any conscious visual experiences.[18] The conclusion that he would fail to notice can easily be derived by repurposing the arguments set out in Sect. 3 for thinking that Joe would not be able to notice if his visual qualia are faded. Thus, if we should expect that Joe's experience will be like Dave's but for the absence of visual qualia when the transition from Dave to Robot involves first replacing all and only those neurons encoding visual information, the Fading Qualia Argument is alive and kicking. To be able to respond to the argument in full generality, we have to say that replacement of all and only those neurons encoding visual information does not result in a case where Joe has an experience much like Dave's but without visual qualia.

As a matter of fact, I did say that. At least, I said something that directly entails it. In Sect. 4, I said that as more and more neurons are replaced with silicon components as we make the transition from Dave to Robot, we do not transition between different holistic patterns of brain activity and so do not transition between qualitatively different conscious experiences. That entails that it is not true that Joe is having an experience qualitatively different from Dave's, as would be the case if Joe were having an experience much like Dave's but without visual qualia.

Still, it might not have been obvious that entailment was supposed to hold even in the case when the transition from Dave to Robot involves first replacing all and only those neurons encoding visual information – a case that might not have been salient to you. Noticing that entailment might lead you to have (new) doubts about the assumption on which I relied previously about the holistic character of conscious neural processing.

I claim that how seriously we should take those doubts depends to a large extent on how seriously we should take the view known as *phenomenal holism* in the philosophy of consciousness (Searle, 2000; Bayne, 2010; Bayne & Chalmers, 2003): very roughly, the view that the conscious whole is prior to its parts. That's because the assumption on which I've relied so far follows from phenomenal holism in conjunction with identity statements it would be natural for adherents of a biological theory of consciousness to endorse. I will also suggest that certain reasons for rejecting holism that might especially appeal to some adherents of biological theories of consciousness (see Block, 2009), independently cast doubt on the Fading Qualia Argument.

[18] Note, however, that cases of *Anton syndrome* exhibit something like this profile. Thus, Chalmers needs to say why a commitment to the nomological possibility of Joe's unnoticed blindness is significantly harder to accept than a commitment to the actuality of Anton syndrome. See Chalmers (1996, pp. 260–261) and Block (2023, pp. 455–456) for discussion. Compare the discussion in footnote 5.

I'll start by doing more to clarify the assumption on which I've relied about the holistic character of conscious neural processing. I'll call this assumption *neural holism*. Consider some overarching pattern of brain activity that binds together activity in different populations of neurons, thereby realizing an integrated and coherent *total phenomenal state*, understood as the overarching experience that subsumes each of the more particular experiences had by a conscious subject at a given point in time. Neural holism says that those of the network's proper parts that encode experiences subsumed by the total phenomenal state do not realize conscious experiences in and of themselves: they do not realize any conscious experience except by virtue of forming part of an overarching pattern of brain activity that similarly integrates information across distributed populations of neurons so as to realize a coherent neural representation.

It follows that replacement of all and only those neurons encoding visual information does not result in Joe having an experience much like Dave's but without visual qualia. Those patterns of brain activity that encode the non-visual information of which Dave is conscious don't suffice on their own for realizing proper parts of the original experience. If the overarching pattern of information processing of which they form part is correctly described as a pattern of brain activity, then they realize the relevant proper parts of the original experience, which is itself realized in its entirety. If it does not, they do not. If it is indeterminate, then it is indeterminate whether any proper part of the original experience occurs.[19]

I claim that neural holism follows from phenomenal holism, given identities that it would be natural for adherents of a biological theory of consciousness to accept. Phenomenal holism is the view that the total phenomenal state is basic and its parts are derivative. The specific version of phenomenal holism I have in mind says that any component of any total phenomenal state is conscious only because there exists a total phenomenal state of which it is part. In this sense, consciousness is inherited by the parts from the whole.

Bayne (2010) formulates holism about consciousness slightly differently, namely as the view that "the components of the phenomenal field are conscious only as the components of *that* field" (Bayne, 2010, pp. 225). (My emphasis.) I take him as denying that one and the same token phenomenal state can occur within the context of different total phenomenal states.[20] We may, however, find it plausible that one and the same pain can persist throughout changes in the overall character of a person's

[19] If that's what neural holism says about Joe, does it also entail, absurdly, that people who are blind as a result of lesions to the visual cortex aren't having total experiences that are like the experiences of sighted people but lacking in visual qualia? No. Here's the crucial difference. In cases of cortical blindness, neural populations encoding sensory information participate in an overarching pattern of brain activity encoding a coherent content from which visual information is missing. In Joe's case, by contrast, if those same patterns of activation occur, they do so as part of a neuron-silicon assembly encoding a coherent content to which visual information contributes. Only in the latter case does the substantial replacement of neural matter by silicon call into question the biological status of the overarching information-processing network responsible for integrating information across sensory channels.

[20] Bayne (2010, pp. 243) writes that "tokens of a single fine-grained phenomenal state type can occur within the context of various total phenomenal state types." Other philosophers such as Sprigge (1983) may be understood as putting forward a form of holism on which tokens of a fine-grained phenomenal state type can only occur as parts of one and the same total phenomenal state type, because the phenomenal

experience and so participate in distinct total phenomenal states. In my formulation, phenomenal holism allows for this possibility, but insists that any component of any phenomenal field depends for its being conscious at a time on its inclusion at that time in the given conscious whole of which it then forms part. In a similar fashion, a sound wave may pass from a body of air to a body of water, but is in each case nothing more than a pattern of disturbance in the given medium. In a roughly similar sense, phenomenal holism understands the components of the phenomenal field as "bumps or forms or features in the unified field of consciousness" (Searle, 2000, pp. 574).

Holism in my formulation remains supported by the same arguments relied on by Bayne. For example, one key argument for holism (Bayne & Chalmers, 2003; Bayne, 2010, pp. 236–238) is that it explains *phenomenal unity*, understood as the idea that for any set of experiences had by any conscious subject at a given point in time, there is an overarching experience that subsumes them. Phenomenal unity is a claim about how things stand for the subject at a fixed point in time and so doesn't touch on the ability of one and the same token experience to participate in different total phenomenal states at other times. It stands to reason that phenomenal unity is explained at least as well by my formulation of phenomenal holism as by Bayne's.

Say that any total phenomenal state is identical to some overarching pattern of brain activity that binds together distributed populations of neurons to realize an integrated and coherent neural representation. Identify its components with those proper parts of the overarching pattern that encode different aspects of the coherent representational state realized across the pattern as a whole. These identities strike me as ones it would be natural for adherents of a biological theory of consciousness to endorse. And with these identities in place, phenomenal holism entails neural holism: that the components of a given total phenomenal state are conscious only because there exists some total phenomenal state of which they are part entails that the proper parts of an overarching pattern of brain activity identified with a given total phenomenal state are conscious only as the parts of an overarching pattern of brain activity of that kind.

Obviously, phenomenal holism is controversial. Still, Chalmers accepts it.[21] (Bayne & Chalmers, 2003), and there are good arguments in its favor. As noted, it explains the unity of consciousness (Bayne & Chalmers, 2003; Bayne, 2010, pp. 236–238). It also gains support from scientific theories of consciousness like the *global neuronal workspace theory* (Dehaene, 2014) (see Bayne, 2010, pp. 228–229). Moreover, some of the best reasons to doubt phenomenal holism independently cast doubt on the Fading Qualia Argument. As noted by Bayne (2010, p. 228), phenomenal holism can be challenged by evidence supporting theories of consciousness like Lamme's *local recurrent processing theory* (Lamme, 2006, 2010). According to Lamme, visual experience can be realized by local activity within circumscribed regions at the back of the brain and does not depend on global patterns of neural activity (see also Zeki & Bartels, 1999; Block, 2007). This suggests an atomistic picture of perceptual consciousness as built up from different streams of information processing, each

character of every part of any total phenomenal state reflects the character of every other. See Dainton (2000, pp. 183–239) for discussion.

[21] Albeit with some caveats, see Chalmers (2014, pp. 792–796).

conscious in itself in virtue of local recurrent processing in circumscribed cortical areas. But local theories like Lamme's also call into doubt the assumptions about the accessibility of phenomenal character to introspection on which the Fading Qualia Argument relies: i.e., that 'conscious beings are generally capable of forming accurate judgments about their experience, in the absence of distraction and irrationality.' Local processing of visual stimuli at the back of the brain need not be available for introspective report. As a result, theories like Lamme's strongly suggest the possibility of dissociations between experience and introspective access to the character of experience even in the absence of any pathology or interference with the character of normal conscious processing (see Block, 2023, pp. 451–459).

7 Conclusion

I've shown how the Fading Qualia Argument can be resisted given assumptions to the effect that consciousness is associated with vague boundaries and that conscious neural activity has a particular kind of holistic structure. I regard each of these assumptions as plausible, although both are obviously controversial.

There is an important asymmetry between these assumptions. While the claim that consciousness is associated with vague boundaries is controversial, it seems very hard to reject without simply begging the question against biological theories of consciousness, for the reasons noted at the end of Sect. 4. By contrast, holism is a logically independent assumption and can be rejected without begging the question.[22]

If my argument in this paper is on track, adherents of biological theories of consciousness need not be forced to reject Chalmers's assumptions about the accessibility of phenomenal character to introspection, provided that they embrace holism.[23] Conversely, philosophers, like Chalmers, who embrace holism cannot rely on the Fading Qualia Argument to challenge biological theories of consciousness. Moreover, as I noted at the end of the previous section, those who reject holism can rely on the argument only if their reasons for rejecting holism do not independently cast doubt on Chalmers's assumptions about introspective access. There may perhaps be ways between the horns of this dilemma.[24] Nonetheless, I take myself to have exposed important weaknesses in the Fading Qualia Argument.

[22] Block (2009, 2023) is arguably the most prominent contemporary philosopher who endorses a biological theory of consciousness, and he accepts an atomist-friendly theory of the neural correlates of consciousness much like Lamme's.

[23] As noted previously, some are willing to bite the bullet and reject Chalmers's assumptions about introspective access. See esp. Block (2023, pp. 451–459). Compare Udell and Schwitzgebel (2021, pp. 136–137) and Birch (2024, p. 69 fn. 55).

[24] For example, Lee (2014) presents an argument against holism that does not rely on a localist theory of the neural correlates of consciousness, and he argues that theories like the global neuronal workspace are in fact consistent with atomism. Although I do not have space to explore the issue in depth, I think the version of holism on which I've focused is immune to Lee's objections by virtue of being weaker than more standard formulations in the literature.

Acknowledgements For helpful comments on previous drafts of this paper, I'd especially like to thank Brad Saad, as well as Eric Schwitzgebel, Derek Shiller, and Teru Thomas, audiences at the 13th Oxford Workshop on Global Priorities Research in December 2023 and at the Mind/Aristotelian Society Joint Session at the University of Birmingham in July 2024, as well as three anonymous referees at the journal.

Funding No funding was received for this work.

Data availability No data was gathered for this work.

Declarations

Conflict of interest No conflicts of interests arose in the writing of this article.

References

Antony, M. V. (2006). Vagueness and the metaphysics of consciousness. *Philosophical Studies*, *128*(3), 515–538. https://doi.org/10.1007/s11098-004-7488-8

Antony, M. V. (2008). Are our concepts CONSCIOUS STATE and CONSCIOUS CREATURE vague? *Erkenntnis*, *68*(2), 239–263. https://doi.org/10.1007/s10670-007-9061-2

Bayne, T. (2010). *The unity of consciousness*. Oxford University Press.

Bayne, T., & Chalmers, D. J. (2003). What is the unity of consciousness? In A. Cleeremans (Ed.), *The unity of consciousness* (pp. 23–58). Oxford University Press.

Birch, J. (2024). *The edge of sentience: Risk and precaution in humans, other animals, and AI*. Oxford University Press

Block, N. (1978). Troubles with functionalism. *Minnesota Studies in the Philosophy of Science*, *9*, 261–325

Block, N. (2007). Consciousness, accessibility, and the mesh between psychology and neuroscience. *Behavioral and Brain Sciences*, *30*(5), 481–548. https://doi.org/10.1017/s0140525x07002786

Block, N. (2009). Comparing the major theories of consciousness. In M. Gazzaniga (Ed.), *The cognitive neurosciences* IV (pp. 1111–1123). MIT Press.

Block, N. (2023). *The border between seeing and thinking*. Oxford University Press.

Bostrom, N. (2006). Quantity of experience: Brain-duplication and degrees of consciousness. *Minds and Machines*, *16*(2), 185–200. https://doi.org/10.1007/s11023-006-9036-0

Butlin, P., Long, R., Elmoznino, E., Bengio, Y., Birch, J., Constant, A., & VanRullen, R. (2023). *Consciousness in artificial intelligence: Insights from the science of consciousness. arXiv*. Retrieved October 13, 2025, from https://arxiv.org/abs/2308.08708

Byrne, A. (2001). Intentionalism defended. *The Philosophical Review*, *110*(2), 199–240. https://doi.org/10.1215/00318108-110-2-199

Cao, R. (2022). Multiple realizability and the spirit of functionalism. *Synthese*, *200*, 506. https://doi.org/10.1007/s11229-022-03524-1

Carruthers, P. (2019). *Human and animal minds: The consciousness questions laid to rest*. Oxford University Press.

Chalmers, D. J. (1995). Absent qualia, fading qualia, dancing qualia. In T. Metzinger (Ed.), *Conscious experience* (pp. 309–328). Ferdinand Schoningh.

Chalmers, D. J. (1996). *The conscious mind: In search of a fundamental theory*. Oxford University Press.

Chalmers, D. J. (2004). The representational character of experience. In B. Leiter (Ed.), *The future for philosophy* (pp. 153–181). Oxford University Press.
Chalmers, D. J. (2010). *The character of consciousness*. Oxford University Press.
Chalmers, D. J. (2014). Strong necessities and the mind-body problem: A reply. *Philosophical Studies*, *167*(3), 785–800. https://doi.org/10.1007/s11098-013-0194-7
Dainton, B. (2000). *Stream of consciousness: Unity and continuity in conscious experience*. Routledge.
Dehaene, S. (2014). *Consciousness and the brain: Deciphering how the brain codes our thoughts*. Viking Press.
Dretske, F. (1995). *Naturalizing the mind*. MIT Press.
Edgington, D. (1992). Validity, uncertainty and vagueness. *Analysis*, *52*(4), 193–204. https://doi.org/10.1093/analys/52.4.193
Edgington, D. (1996). Vagueness by degrees. In R. Keefe & P. Smith (Eds.), *Vagueness: A reader* (pp. 617–630). MIT Press.
Godfrey-Smith, P. (2016). Mind, matter, and metabolism. *The Journal of Philosophy*, *113*(10), 481–506. https://doi.org/10.5840/jphil20161131034
Godfrey-Smith, P. (2020). *Metazoa: Animal minds and the birth of consciousness*. William Collins.
Goguen, J. A. (1969). The logic of inexact concepts. *Synthese*, *19*(3–4), 325–373. https://doi.org/10.1007/bf00485654
Hall, G. (2023). Is consciousness vague? *Australasian Journal of Philosophy*, *101*(3), 670–684. https://doi.org/10.1080/00048402.2022.2036207
Harman, G. (1990). The intrinsic quality of experience. *Philosophical Perspectives*, *4*, 31–52. https://doi.org/10.2307/2214186
Herzog, M. H., Drissi-Daoudi, L., & Doerig, A. (2020). All in good time: Long-lasting postdictive effects reveal discrete perception. *Trends in Cognitive Sciences*, *24*(10), 826–837. https://doi.org/10.1016/j.tics.2020.07.001
Horgan, T., & Tienson, J. (2002). The intentionality of phenomenology and the phenomenology of intentionality. In D. J. Chalmers (Ed.), *Philosophy of mind: Classical and contemporary readings* (pp. 520–533). Oxford University Press.
Kamp, J. A. W. (1975). Two theories about adjectives. In E. L. Keenan (Ed.), *Formal semantics of natural language* (pp. 123–155). Cambridge University Press.
Keefe, R. (2000). *Theories of vagueness*. Cambridge University Press.
Kind, A. (2003). What's so transparent about transparency? *Philosophical Studies*, *115*(3), 225–244. https://doi.org/10.1023/a:1025124607332
Lamme, V. A. F. (2006). Towards a true neural stance on consciousness. *Trends in Cognitive Sciences*, *10*(11), 494–501. https://doi.org/10.1016/j.tics.2006.09.001
Lamme, V. A. F. (2010). How neuroscience will change our view on consciousness. *Cognitive Neuroscience*, *1*(3), 204–240. https://doi.org/10.1080/17588921003731586
Lee, G. (2014). Experiences and their parts. In D. Bennett & C. Hill (Eds.), *Sensory integration and the unity of consciousness* (pp. 287–321). The MIT Press.
Lewis, D. K. (1970). General semantics. *Synthese*, *22*(1–2), 18–67. https://doi.org/10.1007/bf00413598
Loar, B. (2003). Phenomenal intentionality as the basis of mental content. In M. Hahn & B. T. Ramberg (Eds.), *Reflections and replies: Essays on the philosophy of Tyler Burge* (pp. 229–258). MIT Press.
Lycan, W. G. (1996). *Consciousness and experience*. MIT Press.
Machina, K. F. (1976). Truth, belief, and vagueness. *Journal of Philosophical Logic*, *5*(1), 47–78. https://doi.org/10.1007/bf00263657
McGinn, C. (1997). *The character of mind: An introduction to the philosophy of mind* (2nd ed.). Oxford University Press.
Mendelovici, A. (2018). *The phenomenal basis of intentionality*. Oxford University Press.
Moore, G. E. (1903). The refutation of idealism. *Mind*, *12*(48), 433–453. https://doi.org/10.1093/mind/xii.4.433
Neumann, O. (1990). *Some aspects of phenomenal consciousness and their possible functional correlates*. Retrieved October 14, 2025, from https://wwwhomes.uni-bielefeld.de/cgi_psy_ae01/publikat/aspects_of_consciousness.pdf
Schneider, S. (2020). How to catch an ai zombie: Testing for consciousness in machines. In S. M. Liao (Ed.), *Ethics of artificial intelligence* (pp. 439–458). Oxford University Press.
Schwitzgebel, E. (2023). Borderline consciousness: When it's neither determinately true nor determinately false that experience is present. *Philosophical Studies*, *180*(12), 3415–3439. https://doi.org/10.1007/s11098-023-02042-1

Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, *3*(3), 417–457. https://doi.org/10.1017/s0140525x00005756
Searle, J. R. (1990). Consciousness, explanatory inversion and cognitive science. *Behavioral and Brain Sciences*, *13*(4), 585–642. https://doi.org/10.1017/s0140525x00029587
Searle, J. R. (1992). *The rediscovery of the mind*. MIT Press.
Searle, J. R. (2000). Consciousness. *Annual Review of Neuroscience*, *23*(1), 557–578. https://doi.org/10.1146/annurev.neuro.23.1.557
Sebastián, M. Á., & Martínez, M. (2024). Gradualism, bifurcation and fading qualia. *Analysis*, *84*(2), 301–310. https://doi.org/10.1093/analys/anad050
Simon, J. A. (2017). Vagueness and zombies: Why 'phenomenally conscious' has no borderline cases. *Philosophical Studies*, *174*(8), 2105–2123. https://doi.org/10.1007/s11098-016-0790-4
Smith, N. J. J. (2008). *Vagueness and degrees of truth*. Oxford University Press.
Sprigge, T. (1983). *The vindication of absolute idealism*. Edinburgh University Press.
Tye, M. (1995). *Ten Problems of consciousness: A representational theory of the phenomenal mind*. MIT Press.
Tye, M. (2021). *Vagueness and the evolution of consciousness: Through the looking glass*. Oxford University Press.
Udell, D. B., & Schwitzgebel, E. (2021). Susan Schneider's proposed tests for AI consciousness: Promising but flawed. *Journal of Consciousness Studies*, *28*(5–6), 121–144
VanRullen, R. (2016). Perceptual cycles. *Trends in Cognitive Sciences*, *20*(10), 723–735. https://doi.org/10.1016/j.tics.2016.07.006
Williamson, T. (1994). *Vagueness*. Routledge
Zeki, S., & Bartels, A. (1999). Toward a theory of visual consciousness. *Consciousness and Cognition*, *8*(2), 225–259. https://doi.org/10.1006/ccog.1999.0390