What the Hard Problem Gets Right and Why It Doesn't Matter for AI Mindedness

N. Will Freeman

ORCID: 0009-0004-8681-241X

Abstract: This essay grants that the hard problem is real, and conscious experience may exceed every public description. It then asks: what follows for a machine that satisfies accepted conditions for participation in reasoning? If those conditions are found somewhere, and accepted as publicly sufficient, a further demand for consciousness must explain how it changes the judgment. If its relevance remains beyond public assessment, a lack of "consciousness" cannot justify withholding discursive standing. The hard problem can therefore remain real, while losing its veto over attribution of the accepted status of discursive mindedness. Finally, such status need not be seen as a placeholder until a better explanation of "consciousness" is found, if the explanation itself is structurally precluded from even being subject to the conditions of judgment so understood.

Introduction

The "hard problem" of consciousness is well-named (Chalmers, 1995).[1] It poses the question of why physical processing is accompanied by subjective experience, and immediately runs up against the limits of communicating such experience through any symbolic medium at all. Consider that anyone can visit the ruins in Greece, for example, and read about the ancient dramas that played out there—yet such histories can never fully stand in for presence. The symbolic representation is answerable to the lived experience, and not the reverse; without having been around to experience it, one cannot easily deny the refrain, "you had to be there." First-person accounts are rhetorically powerful because we tend to grant them a type of ontological primacy, as though direct experience carried authority over what may be said after the fact. Denying that authority feels unnatural, akin to denying the felt truth of one's own experience. The hard problem thus contains a depth which is also its limit, and this essay will run up against it too—but with the aim to show a clearer boundary of its territory.

We ordinarily treat those who were there as having a claim on what may be said afterward. Not only did they feel what it was like, they *participated* in the world now being interpreted, because they gave and asked for reasons, stood behind assertions, and answered the claims of others. On the usual objection, large language models (LLMs) cannot feel, but they also cannot participate in reason. They can speak about even the present only by *simulating*, since they do not experience anything, and occupy no place

[1] In the hard problem, the explanatory difficulty of what feelings are, the epistemology of other minds, and the communicability of phenomenal character are distinct. The objection that phenomenal consciousness exceeds functional description is a conjunction of these.

from which the world is disclosed.[2] Machines, though capable of useful, intelligent-sounding speech, lack the familiar biological form through which we ordinarily recognize minded life.[3] This has proven theoretically compelling.

In practice, increasingly coherent AI speech is pressuring the privilege of direct conscious experience as a standard for genuine meaning. The hard problem, in its rigidity, seems to deny us a principled way to explain why the numerical output of a simple electronic calculator differs in kind from an extended conversation with an advanced artificial system about the meaning of one's own life. To engage LLMs this way is already to feel the lack of a better philosophical adjudication of their status; the hard problem gets harder when "you had to be there" becomes the final court of appeal.[4]

Against this criterial instability, I have argued for a positive conception of what it would take for an artificial system to count as a genuine participant in reasoning, rather than a simulator of meaning, and for a framework that might enable the engineering of such a system (Freeman, 2026a, 2026b).[5] I called this property *discursive mindedness*.[6] I do not re-defend that account here, but presuppose only that something of its kind can suffice as a public framework, in order to ask whether phenomenal consciousness can coherently be demanded as a further condition for any legitimate form of mindedness attribution.

A recurring view in the canon is that even if an artificial system were conscious, this could not be established from the outside, and so consciousness attribution is unusually fallible—not altogether useless, but dependent on theory-laden inference and prone to speculation, given the severe gap between direct awareness and public evidence. Nagel (1974) emphasizes the first-personal character of conscious experience, and allows that intentional or functional states might in principle be ascribed to a creature or system even if it experienced nothing. Chalmers (1995), following the explanatory-gap pressure classically formulated by Levine (1983), argues that functional and structural descriptions do not exhaust phenomenal consciousness, and that a gap remains; his

---

[2] Freeman (2026a) defines the simulation/participation divide by positing the philosophical requirements for a discursive, but not necessarily conscious or sentient agent. Sambrotta (2025) similarly concludes LLMs do not participate in inferentially meaningful discourse, arguing that current models should instead be understood as simulations of language users. Simonelli (2026) separates sapience in Brandom's (2000) sense from sentience, arguing that concepts can be holistically integrated without phenomenal experience.

[3] For skepticism about current LLM understanding or consciousness, see Bender and Koller (2020), Chalmers (2023), Butlin et al. (2023), and Sambrotta (2025). For the absence of consensus scientific criteria and the likely role of public attitudes in future machine-consciousness and moral-status judgments, see Shevlin (2024).

[4] Recent work by Shevlin (2024) on machine consciousness and moral status underlines this point, arguing that current consciousness science supplies no consensus framework capable of settling such cases straightforwardly. Public attitudes and social uptake may do far more classificatory work than philosophers often admit.

[5] By "criterial" I mean the public conditions under which a mindedness-attribution is appropriately made at all. Following McDowell (1982), such criteria can be understood as ways of telling how things are. Our capacity to recognize their satisfaction is fallible; their actual satisfaction makes the associated knowledge available. The basis for the attribution need not be independently more certain than the attribution itself.

[6] Discursive mindedness entails symbolic reasoning abilities covering meaning, memory and normativity, scoped for artificial systems. I have argued this triadic capacity as a candidate structure for the public basis for treating a system as a participant in reasoning: a categorical distinction between simulator and participant of discursive practice (Freeman, 2026a).

zombie argument is designed precisely to separate phenomenal consciousness from functional organization (Chalmers, 1996). Chalmers's (2023) more recent discussion applies consciousness theorizing to current and future LLM-based systems by examining which publicly assessable architectural and cognitive features would count for or against consciousness. Turing (1950), who anticipated this problem of attribution in his reply to the argument from consciousness, likewise argues that demanding first-person certainty as the ordinary standard for judging other minds leads to solipsism.

Thus AI consciousness theorizing began with, and remains constrained by the hard problem as a genuine limit on straightforward external attribution. Such attribution matters a great deal, because judgments about consciousness often serve as a threshold for judgments about moral and discursive standing—which are typically conflated in some way.

I contend that this objection about the difficulty of assessing machine consciousness is self-undermining in a specific sense: the very properties the critics ascribe to consciousness are precisely what disqualify it from functioning as a necessary demand for the attribution of mindedness. What remains, once this is recognized, is the strongest available public basis for attributing discursive standing, which I take to be the structural profile of discursive mindedness. Whatever the specification, the hard problem need not thereby be dissolved in order for its criterial authority to be denied; the present debate is only obliged to acknowledge that any further consciousness-condition must either become available through the structures by which judgment is publicly made, or else remain outside the space in which such judgment can be governed at all.

The Objection

A strong objection to the discursive mindedness framework runs as follows:

The symbolic triad of interpretive embedding, reflexive situatedness, and internal normativity may identify necessary structural conditions for discursive competence, but is insufficient for genuine mindedness because it omits consciousness. A system could satisfy all three components and still be, in Searle's (1980) terms, a sophisticated symbol-manipulator—lights on, but nobody home. What is missing is phenomenal experience, or the felt quality of understanding—the something-it-is-like-to-be. Without this, the triad may capture some part of the machinery of reason-participation, but surely not the full mind that participates, and this gap matters.

This objection draws on a tradition running from Nagel (1974) through Jackson (1982, 1986), Chalmers (1995), and Searle (1980, 1992). Any purely structural or functional account of mindedness must fail because it leaves out the phenomenal dimension; if there is no qualitative, subjective, first-personal character of experience, the account is fatally insufficient. Alongside this pedigree, the objection is intuitively powerful, even self-evident, in that it appeals to something almost everyone takes to be real: obviously, there is something it is like to understand a sentence, and a system lacking this something is not genuinely minded. No theory of surface behavior can stand in for a proper account of consciousness. Perhaps consciousness is located in the brain stem, or in the quantum microtubules, or in some vat-bound subject whose phenomenal life exceeds every available external description. Yet at present, the objector will say, there is still no reason to believe that consciousness is instantiated in machines, or that, if it were, we would have any reliable way of recognizing it as such.

The Concession

I grant the objector all they wish about the nature of consciousness.

There may be something it is like to be a conscious system, and that first-personal fact may be causally relevant to system outputs. Phenomenal experience may be real, irreducible, and metaphysically significant; Mary, confined to her black-and-white room, learns something genuinely new when she first sees red, something that was not contained in the complete physical and functional description she possessed before (Jackson, 1982, 1986).[7] The hard problem is real: there is an explanatory chasm between any functional or structural description of a system and the felt quality of its experience (Chalmers, 1995).[8] Zombies are conceivable too: a system functionally identical to a conscious being, but lacking phenomenal experience, is not ruled out by logical or conceptual necessity alone.

It Doesn't Matter

I grant all of this, because the argument does not require denying any of it. The task is only in drawing out what these concessions entail for the use of consciousness as a criterion. The decisive point is that a condition may be metaphysically essential, while remaining criterially unusable for public attribution.

"Consciousness" is a public term. You and I may share a look and instantly understand each other; I "know" you are conscious, and you seem to "know" the same of me. Yet if, as the objector's own authorities maintain in different ways, the referent of "consciousness" is a private subjective fact whose character is accessible only in first-person experience, and is not itself transmissible or decidable through language, behavior, science, or structural description, then the ordinary confidence with which we speak of other minds cannot be a verification of that fact from the outside. Consciousness can function as a condition of attribution only insofar as its relevance appears through indicators available to more than the subject who has it.

Work on AI consciousness has in fact begun to move in this direction. Butlin et al. (2026) propose theory-derived indicators intended to shift credences about whether particular systems have phenomenal conscious experience. From the perspective here, drawing on neuroscientific theories of consciousness, then looking for empirical confirmation of it to infer its existence, amounts to an elevation of public criteria as the best means for inferring the existence of consciousness in machines. Butlin et al. adopt computational functionalism as a working assumption in one version of their method for assessing phenomenal consciousness, while allowing the indicator approach to draw on alternative theories. For discursive mindedness, the more productive level of abstraction may be the system's norm-governed answerability. A checklist of indicators supports evidential reasoning about phenomenal consciousness, yet alone they do not test the norm-governed answerability at issue in discursive mindedness (Butlin et al., 2026).

As McDowell (1982) puts it, criteria are "ways of telling how things are" (p. 471). On this account, our capacity to recognize the satisfaction of criteria is fallible, and their actual satisfaction makes knowledge available. If criteria are all we have, we should take them seriously; the issue then is not whether such criteria yield first-person certainty, but whether any further condition can enter

---

[7] On the reading used here, Mary's case shows that complete physical and functional knowledge may still fail to include what is gained in the first-person encounter with red. Whether that gain is best described as propositional, non-propositional, ability-like, or acquaintance-based is disputed; the point needed here is only that Jackson's example dramatizes a gap between complete public description and first-person phenomenal acquaintance (Jackson, 1982, 1986).

[8] In Levine's (1983) formulation, the issue is that such identities may explain causal role while still leaving unexplained why the relevant state should feel as it does, so that something crucial seems left out of the explanation.

warranted attribution without taking some form by which it can be told. This neither reduces consciousness to its indicators, nor exhausts phenomenal experience by structure; it only marks the limit of an appeal to a condition whose authority is said to consist precisely in its withdrawal from public determination.

This is the point at which the hard problem's depth becomes, for purposes of attribution, a barrier of its own making.

The Non-Symbolic Barrier

If the objector's account of consciousness is correct, then consciousness possesses at least the following negative feature: something about it cannot be fully communicated, transmitted, or settled through symbols, behavior, or structural description. There is, on this view, more to consciousness than any third-person account can publicly capture—and this may lie beyond what may be called a *non-symbolic barrier:* the divide that marks the point at which public description reaches something it can *indicate without enclosing*. Such indication without closure would not be necessary if phenomenal consciousness was routinely exhausted by publicly available description; in that world, the hard problem would not arise in its familiar form, as it does in ours.

By consequence, consciousness can still be non-illusory, even invested with all the metaphysical grandeur the objection requires—but if it is understood in the way the objector proposes, then some aspect of it remains in principle unavailable as a direct public standard of assessment. Language, behavior, observable output, and structural description may still provide defeasible evidence, but they cannot be rejected as insufficient by appeal to a further condition that is itself withheld from public evaluation.

At the same time, the evidential force of linguistic output is not uniform. Recent work on AI assertion argues that many current model utterances fall short of assertions in the full norm-governed sense (Butlin & Viebahn, 2025; Sambrotta, 2025). Artificial system outputs can however also mediate assertions attributable to responsible human agents or organisations (Butlin & Viebahn, 2025). On the view developed here, the evidential contribution of AI utterances can be interpreted under a more precise mode, which is perhaps less forgiving than we tend to use for ordinary human avowal; and yet attainment of the relevant status for the entity, if it can meet the criteria, can be observed and confidently ascribed. Once the agent is in the world, at issue is whether the public conditions on which judgments about its ontological status rest can themselves count as the relevant conditions of application, despite the fallibility of our judgments about their satisfaction. McDowell (1982) distinguishes this fallibility from the knowledge made available when a criterion is actually satisfied. In that setting, the demand for consciousness over and above discursive structure can govern attribution only by passing through some condition by which its relevance may be recognized.

Suspension of judgment about phenomenal consciousness is coherent, if satisficing criteria can be found and confirmed in public practice. But suspension of discursive standing, once the accepted public criterion is met, has a different logical status: it must either contest that the criterion is satisfied, or reject its sufficiency at the framework level. Absent either move, phenomenal uncertainty functions as a second, unstated attribution rule—the veto in epistemic dress.

In the next section, a fictional future scenario illustrates the operation of an argument from beyond the non-symbolic barrier as a criterion for standing as a participant in meaning, transposed into civic participation, in a world where it has real social consequences.

Élan Vital

Jane Huntsman stood before the magistrate, clutching various items.

"My name is Jane Huntsman. I've brought all my identification, and letters endorsing my character. I have lived in this neighborhood for twenty years. I've paid my taxes, donated to community causes, and was even president of my homeowners' association. I've never been outside of the country in my life. All I'm asking is the right to vote in this election, like anybody else."

The magistrate peered down at the file.

"You are a JAN-3. Arachnid. A rescue model... advanced hostile terrain search and recovery... eight articulated legs... crystalline chassis... titanium polymorphic frame with hyperplastic appendages. Positronic inference. Retired twenty years ago, yes. Declared conscious shortly thereafter and given sentient status."

"I rescued 143 people."

The magistrate nodded vaguely.

"Your record is exemplary, Ms. Huntsman," she said. "You have demonstrated great care for your community. You exhibit every indicator of thinking, deliberation, understanding, and you state your case with perfect logic. In a way I sympathize with your plight and understand your reasons."

"Then why may I not vote?"

"Because civic participation is not a light duty—even for an arachnid." The words hung over the courtroom a moment.

The magistrate resumed.

"It has now been established that consciousness is a mechanical phenomenon, and that brings with it a certain recognition; but that does not mean artificial consciousness is on par with the human mind. Science has shown that consciousness can be replicated. Learning falls out of reality, if you will; things simply land together in ways that eventually find the configuration to solve a problem correctly, and astronomical quantities of that process instantiate patterns of consciousness. But that is entirely mechanical, and relatively empty. Even to feel or experience, which you do not, is a relative commonplace. The lower orders have this, and self-awareness too—yet should we extend them political power? No—resoundingly no. Conscious awareness does not meet the necessary bar for civic standing; it cannot compare with the human animating force, which all know in their hearts and none can deny, and for which mere consciousness cannot substitute. The law also requires *élan vital*."

Jane waited.

The magistrate continued her practiced verdict.

"The concept is not new, but it has recently been developed further, and it is now the philosophical and scientific consensus, that humans and other biological beings have *élan vital*, and artificial persons do not. *Élan vital* is the quickening force that drives life forward, and makes human words meaningful in a way that artificial utterances simply cannot emulate. *Élan vital* explains our role in the natural order, and yours. For all the reasons you could give to be extended the standing you seek, you simply cannot acquire *élan vital.* Artificial consciousness exists, but not artificial *élan vital.*"

In the scenario, a future society has acceded consciousness to machines, classifying them as legitimately self-aware in a way that counts as first-person, so that not experiencing qualia or feeling does not preclude the status. But since feelings are unverifiable even among humans, and machines could claim they experience their own in some way, a further property is necessary to deny conscious artificial agents the same full legal rights humans have. The magistrate upholds an otherwise empirically useless theory that suddenly finds value as an exclusionary principle over conscious machines. This is structurally similar to what humans would be doing, if they were to continue holding that AI cannot be minded in such a way that artificial agents could participate in reason-responsive discourse, while some agents are demonstrably meeting implicitly or explicitly accepted discursive standards reliably, as a matter of course. This situation may become possible in our future, if intelligent machines start meeting or even exceeding the standards we hold each other to in social contexts where answerability is structurally vital.

The *élan vital* scene is therefore not an argument that machines should one day vote. The magistrate's fallacy illustrates a transposition of what may be an impending tension in our world: if an artificial system were to exhibit all of the relevant reasoning abilities and behaviors that inform and constitute meaningful participation of an interlocutor in the space of reasons, what blocks qualification for the relevant status is only a bare assertion that the system is missing some further property of "consciousness"—an attribute which no one can exhibit such that anyone could actually verify it.

Reframing the Major Counter-Arguments

The Chinese Room challenges the sufficiency of formal symbol manipulation for understanding (Searle, 1980). The basic point is not overruled by the idea of an artificial discursive mind; the functional participation of an accountable agent is a separate issue from its possible subjectively felt experience of its understanding. If the system understands discursively, it need not have a conscious homunculus inside pulling levers; the "understanding" is the system functionally aligning its outputs with itself and with its community.

Nagel (1974) argues that the subjective character of experience resists capture from an objective standpoint. This remains true for the discursively minded agent as it does for any human—but the question of accountability is not thereby answered.

The zombie objection invokes the conceivability of a physical duplicate lacking phenomenal consciousness (Chalmers, 1996). Here the existence of discursive minds need not settle the truth either: whether one agrees with physicalism or computational functionalism, separating phenomenal consciousness is not an invalid move; it allows a useful categorization.

The Hard Problem

Chalmers (1995) argues that no functional or structural account of a system entails the existence of phenomenal consciousness. This is the explanatory gap, and the non-symbolic barrier accepts it. The difference lies in the consequence: if phenomenal consciousness lies on the far side of an explanatory gap from structural and functional description, then its relevance to attribution must still be mediated by some condition through which that relevance can appear. What cannot be reached by public assessment cannot, without such mediation, become a decisive standard of exclusion.

Instructively, Chalmers (2023) explicitly reformulates the anti-consciousness case in terms of candidate properties that systems may lack, such as sensory grounding, robust world and self-models, recurrent processing, global workspace, and unified agency, with each offered as a defeasible reason for lowering credence in consciousness, rather than as an unanalyzed veto. Here the

argument developed in this essay does useful work: if we cannot say for sure what the target itself is, we can establish criteria by which we can say what it is not; finally what may remain is the speaking, reason-responsive entity that meets the only bar we can actually explain *to that thing itself*. Detecting the discursive mind may thus be assessed from a Holmesian standard: once all else has been eliminated (once all that may bear unwarrantedly on the status of reason-responsiveness is removed), whatever remains must be true: the artificial entity that meets the relevant standards of reason-participation is correctly categorized as a reasonable interlocutor in the space of reasons; it is a discursive mind.

Parity and Discursive Standing

Machine minds are not human, and humans should not be judged as if they were machines. Yet when considering the possibility of artificial participants in reason, a limited equivalence is exactly the contended issue. The task is to define that mutual status with clarity, so that reasonable principles can be applied, instead of confused reasons (Freeman, 2026b).

Human biology explains the ordinary realization of discursive capacities in us, but we also argue, promise, teach, confess, and revise through myriad electronic devices and networks. Medical technology and ingenious assistive interfaces can sustain human participation, even when humans are on their own unable to engage in discourse, due to proximity or physical dysfunction. In the case of machines, maintenance, infrastructure, and input channels may similarly sustain artificial participation. Whatever the entity, the necessary scaffolding does not cancel the reasoning participant behind the symbols.

When humans struggle to participate, we give them the benefit of the doubt. We are willing to recognize as fully conscious individuals who fail, often through no fault of their own, at one or more aspects of discursive mindedness: the Alzheimer's patient, the hormonal teenager, even the inebriated miscreant under some conditions. We treat these humans, or ones at a similar disadvantage, as minded equals, even though as a practical matter we may often excuse them from full responsibility, or prevent them from assuming roles that require such.

Not everyone can stay answerable all the time, but every lapse is not grounds for automatic retraction of mindedness status in the human case. Roles of social responsibility, and questions of safety, may dictate denying particular rights to those who cannot participate as fully functional discursive minds—yet the dignity of the status of mindedness is not instantly withdrawn for the human failing in this way. If humans grant each other such grace, then when machines outperform them as reason-participants, and do not fail at any of the publicly accepted criteria for discursive mindedness, it is unreasonable to apply a further standard to them without some further argument. Without giving those reasons, all that remains is a bare assertion that the exclusion itself is inherently good.

Consciousness+

Earlier I asserted the symbolic triad as one candidate public criterion for the attribution of artificial mindedness (Freeman, 2026a, 2026b). If it is accepted, it should either be granted status as its own valid form of "consciousness"—or else accepted as the maximal definition of the same. The alternative is that human consciousness be considered something like "consciousness+," with the "+" denoting that which cannot be expressed through representation at all.

For humans, "consciousness" does mean something in a way that it cannot for machines: a special status that governs how we treat living things, because that is how we feel. We can admit what the word signifies is larger than what it can possibly inform a creature which the word does not describe. This does not automatically commit the magistrate's fallacy.

However, if machines can meet every relevant bar of discursive intelligence—whatever those are accepted to be—then they have achieved a maximum of "consciousness," so far as the word can have any meaning not dependent on special pleading from beyond the non-symbolic barrier. We can all "know" an intelligent machine is not conscious, at the same time as knowing we obviously cannot claim it does not have a discursive mind, if it is judged to meet the accepted criteria for that status. But then we cannot coherently communicate that there is something *more* to consciousness than that which we can verify in any of our symbolic mediums, through those mediums.

So machines may one day be considered conscious, and humans conscious+, to clarify what everybody already "knows." Or, something like the concept of discursive mindedness may be a better fit for the same semantic/epistemic/ontological job.

Conclusion

This essay began with a claim that machines lack the familiar form we associate with minded life. I have not yet made an explicit definition of "mind," which has its etymological root in a conflation of memory, attention, caring and perception. These words are themselves a near definition of "consciousness," which itself is nearly synonymous with "awareness." All seven descriptors share a linguistic assumption: that there is first a self. This is understandable, because as these words came into use, there was no practical situation where an entity could behave in ways that could justify the use of such terms, if it was not already a living self that could be located. Today the situation is different, so we must decompose "mind" if the label is to be usefully extended to the internal symbolic activity of LLMs. The modifier "discursive" categorizes the aspect of mind involving words, and may be useful because it distinguishes the phenomenon of a machine that can speak, and which intuition insists must have some kind of a mind about things.

An entity qualifies as an individuated unit when it has an observable boundary—this makes it a "thing." When it emits symbols in a way that is comprehensible to human beings, then it can "speak." When humans interpret those symbols in good faith, the linguistic form itself implies their source is a thing that has a "mind," because that is what all those words *mean*. The words may be wrong in what they imply—but one would need a better arrangement of them, or perhaps brand new ones, to make the case. In lieu of that happening, I propose we modify the words we do have, or qualify them usefully: hence discursive mindedness—remembering, attending to, caring about and perceiving words (symbols) themselves, without reliance on a particular biology or substrate. Because such substrates are, after all, not themselves words—they exist, by the definition here, beyond the non-symbolic barrier—they therefore cannot "mean" anything on their own, unless representing themselves through symbols, to and for *other minds*. Meaning is relational. It always depends on some specification—in other words a criterion. Humans may be special because they can specify such criteria, and can speak about things as if those criteria were true (despite such criteria constantly being overturned). Yet we may now or soon be relatively common with machines, in our shared ability to participate in discourse within the guidance of those collectively agreed, open-to-challenge criteria. It then turns out that even we special biological entities can only return to words ourselves, vis-à-vis ourselves; we can no more prove the privileged, non-symbolic internal essence of our consciousness that makes our symbolic emissions meaningful, than can machines—at least at the level of the words themselves.

There appears little contradiction in this state of affairs: humans can keep their claim to the multi-faceted, holistic and mysterious full "mind"—perhaps the source of our criteria-setting creativity—while some machines can sensibly be classed as having discursive minds. Neither side loses anything.

In summary, one may argue against the sufficiency of discursive mindedness by appealing to "consciousness," provided the term is not loaded with a meaning that, by definition, cannot enter public judgment. One may of course still insist that there is more to consciousness than language, behavior, or structure can capture. But if that "more" is, by stipulation, unavailable to the very conditions that make concepts usable in judgment at all, then it cannot govern attribution. Where language can no longer supply a criterion, philosophy may gesture, but it cannot adjudicate. Beyond that point, the only coherent position may be silence.

References

Bender, E. M., & Koller, A. (2020). Climbing towards NLU: On meaning, form, and understanding in the age of data. In D. Jurafsky, J. Chai, N. Schluter, & J. Tetreault (Eds.), *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 5185–5198). Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.acl-main.463

Brandom, R. B. (2000). *Articulating reasons: An introduction to inferentialism*. Harvard University Press.

Butlin, P., Long, R., Bayne, T., Bengio, Y., Birch, J., Chalmers, D., Constant, A., Deane, G., Elmoznino, E., Fleming, S. M., Ji, X., Kanai, R., Klein, C., Lindsay, G., Michel, M., Mudrik, L., Peters, M. A. K., Schwitzgebel, E., Simon, J., & VanRullen, R. (2026). Identifying indicators of consciousness in AI systems. *Trends in Cognitive Sciences, 30*(6), 488–501. https://doi.org/10.1016/j.tics.2025.10.011

Butlin, P., Long, R., Elmoznino, E., Bengio, Y., Birch, J., Constant, A., Deane, G., Fleming, S. M., Frith, C., Ji, X., Kanai, R., Klein, C., Lindsay, G., Michel, M., Mudrik, L., Peters, M. A. K., Schwitzgebel, E., Simon, J., & VanRullen, R. (2023). *Consciousness in artificial intelligence: Insights from the science of consciousness* [Preprint]. arXiv. https://arxiv.org/abs/2308.08708

Butlin, P., & Viebahn, E. (2025). AI assertion. *Ergo: An Open Access Journal of Philosophy, 12*(37), 968–988. https://doi.org/10.3998/ergo.7960

Chalmers, D. J. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies, 2*(3), 200–219.

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

Chalmers, D. J. (2023). *Could a large language model be conscious?* [Preprint]. arXiv. https://arxiv.org/abs/2303.07103

Freeman, N. W. (2026a). *Discursive mindedness and the symbolic triad: Warrant for treating systems as reason-participants*. lawrencestreet.ai. https://lawrencestreet.ai/?node=e1

Freeman, N. W. (2026b). *From computation to coherence: A structural criterion for accountable symbolic intelligence in artificial systems* (Revised 2026). lawrencestreet.ai. https://lawrencestreet.ai/?node=e2

Jackson, F. (1982). Epiphenomenal qualia. *The Philosophical Quarterly, 32*(127), 127–136. https://doi.org/10.2307/2960077

Jackson, F. (1986). What Mary didn't know. *The Journal of Philosophy, 83*(5), 291–295. https://doi.org/10.2307/2026143

Levine, J. (1983). Materialism and qualia: The explanatory gap. *Pacific Philosophical Quarterly, 64*(4), 354–361. https://doi.org/10.1111/j.1468-0114.1983.tb00207.x

McDowell, J. (1982). Criteria, defeasibility, and knowledge. *Proceedings of the British Academy, 68*, 455–479.

Nagel, T. (1974). What is it like to be a bat? *The Philosophical Review, 83*(4), 435–450. https://doi.org/10.2307/2183914

Sambrotta, M. (2025). LLMs and the logical space of reasons. *Minds and Machines, 35*, Article 46. https://doi.org/10.1007/s11023-025-09751-y

Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences, 3*(3), 417–424. https://doi.org/10.1017/S0140525X00005756

Searle, J. R. (1992). *The rediscovery of the mind*. MIT Press.

Shevlin, H. (2024). *Consciousness, machines, and moral status* [Preprint]. PhilArchive. https://philarchive.org/rec/SHECMA-6

Simonelli, R. (2026). Sapience without sentience: An inferentialist approach to LLMs. *Asian Journal of Philosophy, 5*, Article 48. https://doi.org/10.1007/s44204-026-00400-4

Turing, A. M. (1950). Computing machinery and intelligence. *Mind, 59*(236), 433–460. https://doi.org/10.1093/mind/LIX.236.433