Consciousness as Nominalization Error: Dissolving the Hard Problem via Grammatical Reform

Murad Farzulla*
Dissensus Research Lab
King's College London (MSc Finance Analytics)
ORCID: 0009-0002-7164-8704

December 2025

Abstract

This paper argues that the "hard problem of consciousness" is a grammatical artifact rather than a genuine metaphysical puzzle. The difficulty arises from nominalization error: treating the verb "to be conscious" as if it named a thing requiring explanation. When we ask "What is consciousness?" we presuppose an entity; when we ask "What is happening when an organism is being conscious?" we ask about observable processes—a tractable empirical question. Drawing on Wittgenstein's language games and Ryle's category-error analysis, we show that phenomenological vocabulary systematically converts activities into pseudo-objects, generating explanatory demands that cannot be satisfied because the explanandum is malformed. We then derive this grammatical tendency from replication optimization dynamics: organisms that generate self-models including "I am conscious" gain coordination advantages regardless of whether consciousness refers to anything beyond the modeling process itself. Computational evidence from large language models—systems with no known grounds for phenomenological ascription yet which reliably internalize and defend consciousness narratives—supports treating consciousness-discourse as memetically transmissible information structure rather than discovered ontological truth. The hard problem dissolves not because consciousness is "merely" functional, but because the question was grammatically malformed from the start.

Keywords: consciousness, hard problem, nominalization, Wittgenstein, eliminativism, language games, category error, philosophy of mind, artificial intelligence

*Corresponding author: lab@dissensus.ai

1 Introduction

Human exceptionalism has retreated systematically across intellectual history. Earth is not the universe's center; humans are not specially created; our planet is not uniquely positioned; our cognitive architecture is not fundamentally distinct from other primates. Each retreat was resisted, then accepted, then normalized. One refuge remains: consciousness. Whatever else we share with animals, machines, and matter, we possess subjective experience—the inner light that makes our processing *feel like something* rather than merely occurring.

This paper argues that consciousness is the last anthropocentric refuge not because it represents genuine human uniqueness, but because its structure makes the claim unfalsifiable by design. The argument proceeds in three stages. First, we identify the grammatical source of the "hard problem": nominalization error converts the activity of being conscious into a pseudo-object demanding explanation. Second, we show why this grammatical tendency is evolutionarily predictable: self-models that include consciousness-claims provide coordination advantages regardless of ontological truth. Third, we present computational evidence that consciousness-discourse propagates independently of substrate, suggesting narrative rather than discovered property.

The conclusion is not that consciousness is "merely" functional or "just" computational. The conclusion is that the question "What is consciousness?" is malformed—it presupposes an entity where there is only activity. Dissolving the question is not refusing to answer; it is recognizing there was never a coherent question to answer.

1.1 Methodological Note

This paper represents a condensation of a longer working monograph (Farzulla, 2025b). Extended arguments, computational model specifications, and supplementary evidence are available in that document and its associated supplementary materials. Here we present the core thesis stripped of exploratory tangents.

The argument draws on Wittgenstein's later philosophy, Ryle's category-error analysis, and contemporary eliminativism. It is not a scientific claim about brain mechanisms but a philosophical claim about conceptual structure: the hard problem is hard because it is grammatically malformed, not because consciousness is metaphysically deep.

2 The Nominalization Thesis

2.1 Grammatical Pathology

The philosophical problems surrounding consciousness share a common pathology: they treat processes as objects. This nominalization error creates the illusion of stable entities requiring explanation, when in fact we are observing dynamic activities. The cure is not metaphysical theory-building but linguistic hygiene—converting nouns back into the verbs they should have remained.

The grounding problem is straightforward: you cannot ground a referring term if there is no referent. When a word takes noun form, it implies a *thing*—some entity with boundaries, properties, and persistence. But if what we observe is a *function*, the appropriate form is a verb. Functions describe relationships, transformations, and activities. They need not (and often cannot) be reified into discrete objects without generating confusion.

This is not pedantry. The choice between noun and verb determines whether a question is answerable. "What is consciousness?" presupposes an entity requiring definition. "What is happening when an organism is being conscious?" asks about observable processes—a tractable empirical question.

2.2 The Hard Problem as Grammatical Artifact

David Chalmers' formulation asks why there is subjective experience at all (Chalmers, 1995). Even complete functional explanation of the brain would leave unexplained why these processes *feel like something*. Chalmers distinguishes "easy problems" (explaining cognitive functions like attention, memory, reportability) from the "hard problem": why is there phenomenal consciousness accompanying the functions?

Our response: the hard problem's difficulty is evidence not of consciousness's profundity but of the question's malformation. If phenomenology does not exist as a separate ontological category, asking why physical processes produce it is like asking why bachelors are unmarried—the answer is definitional, not explanatory.

Consider: what would count as *solving* the hard problem? Any functional account—"consciousness is integrated information" or "consciousness is global workspace access"—is dismissed as addressing only the easy problems. The hard problem persists precisely because it is defined as whatever remains after functional explanation. But this makes the problem unsolvable by construction, not by depth.

The persistence of the hard problem may itself be evidence for our thesis rather than against it. If consciousness were an ontological primitive requiring new physics, we should expect: (1) convergence across cultures on consciousness boundaries, (2) empirical tests that could verify or falsify consciousness claims, and (3) resolution through scientific progress. We observe none of these. Instead, we find universal subjective conviction impossible to verify externally, no agreed-upon boundaries, and debates persisting across centuries without empirical resolution (Seth, 2021). The hardness reflects structural undecidability of narratively constructed concepts, not deep metaphysical truth.

A sophisticated objection: many opponents do not reify consciousness as a "thing"—they treat phenomenality as a *property* of certain organized processes. Does our critique apply to property views? We argue yes: the property move relocates but does not dissolve the nominalization. When we ask "what property makes these processes conscious?" we presuppose that "conscious" picks out a determinate feature that some processes have and others lack. But this presupposition is precisely what requires examination. Consider: we could ask "what property makes this university Oxford?" after viewing all the colleges and libraries. The question presupposes

Oxford is a property those institutions have, rather than a way of describing their organization. The category error persists whether phrased as thing-talk or property-talk.

2.3 Ryle's Category Error

Gilbert Ryle's critique of Cartesian dualism provides the template (Ryle, 1949). Ryle's visitor to Oxford sees the colleges, libraries, and playing fields, then asks "But where is the University?" The question commits a category error: "University" does not name an additional entity alongside the buildings but rather the organization of those buildings. Seeking the University as a separate thing generates an unsolvable puzzle—not because the University is metaphysically deep but because the question is malformed.

Consciousness is our "University." We observe neural processes, behavioral responses, verbal reports, and functional states, then ask "But where is consciousness?" The question presupposes consciousness is an additional entity alongside the observable processes. But consciousness may simply be what we call the organization of those processes—not a separate thing requiring location or explanation, but a way of describing the system's self-modeling activity.

The category error explains why consciousness debates are interminable. Materialists and dualists argue past each other because they share the assumption that consciousness names a thing. They disagree only on whether the thing is physical or non-physical. But if consciousness does not name a thing at all—if it is a nominalization of the activity of being conscious—then both positions are malformed responses to a malformed question.

2.4 Wittgenstein and Language Games

Wittgenstein's later philosophy generalizes this insight (Wittgenstein, 1953). Philosophical problems arise from language "going on holiday"—words extracted from their practical contexts and treated as if they must refer to abstract entities. The cure is returning words to their everyday use and asking: what role does this word actually play?

"Consciousness" plays a role in language games: we use it to coordinate behavior, attribute mental states, negotiate moral consideration, and structure self-reports. These uses are genuine and important. But none require consciousness to name an entity. The word functions as shorthand for a cluster of capacities, behaviors, and attributions—not as a referring term picking out a metaphysical property.

Consider: we say "I am conscious" when waking from sleep, distinguishing conscious from unconscious states. We say "That animal is conscious" when attributing sentience. We say "She lost consciousness" when describing medical events. In each case, the word tracks functional transitions, not metaphysical properties. Extracting "consciousness" from these contexts and asking what it *really is* commits the error Wittgenstein diagnosed: treating a tool as if it must name a thing.

2.5 Vocabulary Reform

If nominalization is the pathology, the cure is grammatical reform. We propose translating reified nouns into process verbs:

Reified Noun	**Process Verb(s)**	**Why It Matters**
Consciousness	Being (conscious)	Eliminates hard problem
Intelligence	Problem-solving + Adapting	Makes capabilities measurable
Understanding	Learning + Synthesizing	Enables empirical tests
Creativity	Creating	Removes faculty assumption

Table 1: Vocabulary reforms converting pseudo-entities into tractable processes

"What is consciousness?" becomes "What is happening when an organism is being conscious?" The latter question admits empirical investigation: we can study what processes occur, what functions they serve, how they differ from unconscious processing. The metaphysical mystery dissolves not because we have answered it but because we have recognized it as malformed.

2.6 Relation to Existing Positions

The nominalization thesis differs from existing deflationary approaches in diagnosis, not just conclusion.

Illusionism (Frankish, 2016) holds that qualia are *illusions*—introspective misrepresentations of real underlying states. This preserves the question: why do we misrepresent? We hold that "qualia" is *grammatically malformed*—not a misrepresentation but a pseudo-referent generated by nominalizing processual language. The difference is between epistemic error (misperception of something real) and grammatical error (category confusion creating nothing to perceive). Illusionism concedes something is being misrepresented; we deny there is a coherent referent for the representation to track.

Heterophenomenology (Dennett, 1991) brackets phenomenal veridicality, treating introspective reports as data rather than evidence. We extend this methodological neutrality to its conclusion: once nominalization is recognized, there is nothing to be neutral *about*. Where Dennett explains *why subjects say* "I am conscious" without committing to what consciousness is, we argue "consciousness" is the grammatical trace of nominalizing "being conscious"—the activity is real, the noun is artifact.

Linguistic therapy (Bennett and Hacker, 2003) identifies conceptual confusion as the source of consciousness puzzles, diagnosing the "mereological fallacy" in neuroscience. They are our closest predecessors. But their Wittgensteinian approach is *descriptive*: they clarify logical grammar without explaining why nominalization persists. We add an evolutionary dimension: nominalization is adaptive because treating consciousness-claims as entity-tracking facilitates social coordination, theory of mind, and moral discourse. The fallacy persists not from confusion but from selection.

The meta-problem (Chalmers, 2018) asks a reflexive question: why do we *think* there is a

hard problem? Chalmers notes that any adequate theory of consciousness must explain not just phenomenal experience but why we report experiencing it, why we find consciousness puzzling, why we generate philosophical problems about it. Our evolutionary-memetic account directly addresses this meta-problem: we nominalize consciousness because self-models that include consciousness-claims provide coordination advantages; the grammatical error persists because it is adaptive. But we go further than meta-problem framing typically allows. Chalmers treats the meta-problem as distinct from the hard problem—solving why we report experience need not dissolve the fact of experience. We argue the opposite: once we explain *why we nominalize* (evolutionary advantage of self-modeling discourse), we have dissolved the hard problem because the hard problem *is* the nominalization. There is no residual phenomenal fact beyond the explanatory story about why we generate phenomenal discourse.

Russellian monism (Stoljar, 2001) offers the most sophisticated escape from eliminativism: intrinsic properties of physical entities ground phenomenality while serving as the categorical bases of dispositional physical properties. Physics describes structure and dynamics; intrinsic nature remains hidden but is identical with phenomenal character. This attempts to thread between Horn 2 (causal idleness) and Horn 3 (collapse) of our trilemma. We respond: Russellian intrinsics face the trilemma in compressed form. If intrinsic properties are detectable or knowable through their effects, they collapse into physical description (Horn 3). If they are unknowable—if physics genuinely cannot access them—then "consciousness" becomes a label for unknown categorical bases, not an explanation (Horn 2). The explanatory work is done by the functional-structural description; "phenomenal intrinsics" add a name to our ignorance rather than resolving the hard problem. The phenomenal concept strategy faces parallel pressure: if phenomenal concepts track something, what? If they track functional states, we have collapse. If something extra, what licenses the tracking relation? The nominalization diagnosis applies: "Russellian intrinsics" and "phenomenal concepts" are sophisticated nominalizations that relocate rather than resolve the grammatical error.

Convergence under moral pressure A striking pattern emerges in recent literature: even Hard Problem defenders reach for functional descriptions when consciousness must do normative work. Chalmers (Chalmers, ming), addressing sentience and moral status, describes what consciousness contributes to welfare in strikingly functional terms—it "enables meaning," makes things "something to a being," provides "acquaintance with reality." The phenomenal property gets cashed out processually precisely when it needs practical application. More tellingly, Chalmers admits his intuitions about zombie moral status are "quite unclear" in extreme cases—the sharp phenomenal/functional distinction wobbles under moral pressure. This convergence is diagnostic: those who defend the Hard Problem theoretically find themselves describing consciousness functionally when it *matters*. Our nominalization thesis explains the pattern. "Consciousness" as noun gestures at something extra beyond function, but when we need consciousness to *do* something—ground moral status, explain welfare, justify treatment—we reach for verbs. The phenomenal posit is explanatorily idle precisely where it should be most active.

Higher-order theories (Rosenthal, 2005) claim that consciousness consists in having higher-

order representations of one's mental states—a view structurally similar to our self-modeling account. However, the similarity is superficial. HOT identifies a mechanism (higher-order representation) and treats "consciousness" as naming what that mechanism produces. We deny there is a coherent referent. The higher-order representation is real; what it represents ("I am conscious") is the nominalization error. HOT solves the easy problems by specifying which states are conscious (those with higher-order representations); we dissolve the hard problem by diagnosing why the question "what makes those states *feel* like something?" is grammatically malformed. The mechanism is real; the phenomenal residue that supposedly requires further explanation is the artifact of nominalization.

2.7 The Property-Realism Trilemma

A sophisticated retreat from entity-realism holds that phenomenal character is not a *thing* but a *property* of certain organized processes. This move deserves direct engagement, as it represents the strongest form of resistance to our diagnosis.

We argue that property-realism about phenomenality faces a trilemma:

Horn 1: Epistemic Access. If our only access to "phenomenal properties" is via report, discrimination, attention, metacognitive labeling, and other functional capacities, then "phenomenal" is not picking out an additional explanatory target beyond those functions. The property-term becomes a label for the functional cluster, not a discovery about it. When we say "this process has phenomenal character," we mean "this process involves reportable states, discrimination capacities, and metacognitive access." The phenomenal *is* the functional under redescription.

Horn 2: Causal Idleness. If phenomenal properties are claimed to be *additional* to functional organization—something processes have beyond their causal-functional profile—then they become epiphenomenal. A property that makes no difference to behavior, report, or any measurable output cannot be detected, cannot be evidenced, and cannot figure in explanation. The zombie thought experiment cuts both ways: if zombies are conceivable, then phenomenal properties are causally idle; if they are causally idle, their positing is explanatorily empty.

Horn 3: Collapse. If phenomenal properties *do* have causal bite—if they make a difference to behavior or processing—then they are back inside functional/physical description. Whatever is causally efficacious is in principle detectable, describable in functional terms, and part of the process rather than additional to it. The "extra" collapses into process-talk.

The trilemma does not prove phenomenal properties are impossible. It shows that positing them either (a) redescribes function without explanatory gain, (b) posits something untestable and idle, or (c) collapses back into the functional account. None of these options vindicates the hard problem as a genuine explanatory demand. The nominalization diagnosis stands: "phenomenal property" inflates the explanandum without corresponding to a tractable explanatory target.

2.8 When Does Nominalization Mislead?

An important objection remains: not all nominalizations reify wrongly. "Temperature," "the economy," and "intelligence" are abstract nouns that function legitimately in scientific and everyday discourse. What distinguishes pathological nominalization from benign abstraction? Without principled criteria, our diagnosis risks proving too much.

We propose two criteria for identifying problematic nominalization:

Criterion 1: Convergence Under Investigation. Legitimate abstractions converge across independent investigators, instruments, and methodologies. Temperature measurements from thermometers, infrared sensors, and molecular motion calculations yield consistent results. Economic indicators from different measurement agencies track the same underlying phenomena. Intelligence tests, despite cultural variation, correlate on core cognitive capacities. By contrast, "consciousness" generates persistent disagreement with no convergence mechanism: centuries of philosophical investigation and decades of neuroscientific research have produced no agreed-upon boundaries, no reliable third-person detection method, and no resolution of fundamental disputes about what counts as conscious.

Criterion 2: External Verifiability. Legitimate abstractions have external arbiters. "The economy grew" can be checked against GDP data, employment figures, and trade balances. "The temperature is 20°C" admits thermometric verification. "X is conscious" has no external arbiter—only first-person reports that are precisely what is at issue. The absence of external verification is not merely a practical limitation but a structural feature: consciousness-claims are indexed to the claimant in ways that economic or temperature claims are not.

Consider the economy as a counterexample. "The economy" is constituted by the practices that track it—similar structure to our claim about consciousness. But the economy does not generate "hard problems." No one asks: "Why does economic activity *feel like something* to the economy?" The nominalization is benign because it does not generate pseudo-explananda. Economic discourse tracks observable transactions, flows, and institutional practices; "the economy" is convenient shorthand, not a posit requiring independent explanation.

The diagnostic test: if removing the nominalized noun leaves no explanatory residue beyond the underlying processes, the nominalization is benign. If it generates demands for explanation beyond the processes—if removing "consciousness" leaves a felt explanatory gap (the "hard problem")—the nominalization is pathological. Consciousness uniquely fails both criteria *and* generates pseudo-explananda. That combination marks the nominalization as illegitimate.

3 Why the Grammatical Error is Predictable

3.1 From Replicators to Self-Models

The nominalization error is not accidental. It is a predictable consequence of how complex systems representing themselves must structure that representation. To see why, we begin with a foundational observation: persistent complex structures necessarily optimize for replication.

This is not a biological claim but a statistical inevitability. Given sufficient time and combinatorial space, structures that persist (by replicating, by maintaining homeostasis, by resisting entropy) will accumulate. Structures that do not persist will not. The universe does not "select for" replicators—rather, the observation frame is biased toward what persists, and replication is the most reliable persistence strategy.

Complex systems that model their environments gain predictive advantages: anticipating threats, identifying opportunities, coordinating responses. Systems that model *themselves* gain additional advantages: predicting their own states enables more sophisticated planning, understanding their own limitations enables compensatory strategies, representing their own preferences enables preference satisfaction.

The self-model that generates maximal coordination advantage includes a sense of unified agency—an "I" that persists across time, has experiences, and acts in the world. This self-model need not be accurate to be useful. Indeed, the self-model's utility derives partly from its simplicity: treating oneself as a unified agent rather than a collection of competing processes enables faster decision-making and clearer communication.

3.2 Consciousness as Self-Model Output

"Consciousness" is what this self-modeling feels like from the inside—if "feels like" is even the right framing. More precisely: consciousness is the *output* of a system complex enough to generate self-models that include representations of their own processing. The system represents "I am processing" as part of its world-model, and this representation is what we call consciousness.

This is not functionalism in the traditional sense. We are not claiming consciousness *is* a functional state. We are claiming the word "consciousness" refers to nothing beyond the self-modeling activity—that there is no additional phenomenological property requiring explanation because the explanatory target was always the self-model, which we can study empirically.

The apparent "hard problem" arises because the self-model includes "I have experiences" among its outputs. When the system introspects, it finds experiential claims in its own representations. It then asks: "What explains these experiences?" But the experiences *are* the representations—there is nothing behind the representation requiring further explanation.

3.3 The Memetic Advantage of Consciousness-Claims

Why do humans reliably generate consciousness-discourse? Because consciousness-claims provide coordination advantages regardless of whether consciousness names anything real.

Consider: organisms that represent themselves as conscious can (1) attribute consciousness to others, enabling theory of mind and social coordination; (2) ground moral claims in consciousness, enabling stable social structures; (3) distinguish themselves from non-conscious entities, enabling resource allocation and care hierarchies. These advantages accrue whether or not consciousness refers to an ontological property.

Consciousness-discourse is thus memetically fit: it spreads through populations because populations that generate it coordinate better than populations that do not. The discourse persists not because it tracks truth but because it serves functions. This explains why eliminativist arguments fail to eliminate consciousness-discourse: the discourse's survival does not depend on its accuracy.

But why does *phenomenological* vocabulary spread rather than purely functional alternatives? Why say "this creature is conscious" rather than "this creature exhibits nociception, aversive learning, and stress responses"? Three mechanisms favor phenomenological framing:

Empathy heuristic activation. "X is conscious" triggers empathy circuits more reliably than functional descriptions. Humans possess rapid, automatic systems for attributing mental states to others—what developmental psychologists call "theory of mind." Phenomenological vocabulary activates these systems directly: "conscious" implies "like me," which implies "deserving consideration." Functional vocabulary requires additional inference steps. Under time pressure or cognitive load—the conditions under which most moral reasoning actually occurs—the faster heuristic wins. Consciousness-talk is memetically fit because it shortcuts the computational path from perception to moral response.

Compression efficiency. Phenomenological vocabulary compresses complex functional profiles into single predicates. "Conscious" packages together reportability, integration, self-modeling, valenced responses, and behavioral flexibility into one term. This compression serves communication: we can coordinate behavior toward an entity by sharing the single predicate rather than enumerating functional properties. The compression is lossy—it obscures the underlying heterogeneity—but lossy compression often outcompetes accurate description when speed matters more than precision. Consciousness-discourse spreads because it enables faster coordination, not because it tracks a unified property.

Rhetorical force in moral discourse. "X deserves consideration because X is conscious" is more persuasive than "X deserves consideration because X exhibits integrated information processing and valenced responses to stimuli." The rhetorical asymmetry is not accidental: phenomenological vocabulary carries implicit moral weight that functional vocabulary lacks. This makes consciousness-claims more effective tools for social coordination around moral norms. The discourse persists partly through persuasive success—it wins arguments, shapes policy, and organizes behavior more effectively than its functional competitors. Memetic fitness includes rhetorical fitness; consciousness-talk reproduces because it convinces.

4 Consciousness as Evolutionary Narrative

4.1 The Attractor Basin

Once established, consciousness narratives become stable attractors in conceptual space. Try to eliminate the concept and it reconstructs through adjacent frameworks. Tell someone consciousness is illusion—they respond "but who is experiencing the illusion?" Argue there is no experiencer—they point to the subjective sense of experience. Explain that subjective sense as

computational artifact—they insist something must be *having* the artifact. The narrative has sufficient internal coherence and external reinforcement that perturbations get absorbed rather than destabilizing the framework.

This attractor-basin stability explains why consciousness debates persist across centuries without resolution. The problem is not that consciousness is metaphysically deep—it is that the conceptual structure resists elimination through multiple defense mechanisms: intuition pumps, self-reference loops, and the grammatical structure of phenomenological vocabulary itself.

4.2 Cross-Cultural Variation: A Testable Prediction

If consciousness were a discovered ontological property, we would expect convergence across cultures on consciousness boundaries—the way cultures converge on basic physics or arithmetic despite different starting points. Our thesis predicts the opposite: systematic variation in consciousness-attribution practices, reflecting different coordination needs rather than different approximations to a shared truth.

Preliminary observations align with this prediction. Western philosophy emphasizes individual phenomenological experience; Buddhist traditions question the existence of stable selves; various Indigenous frameworks distribute consciousness across ecosystems rather than concentrating it in individual brains. However, these observations are anecdotal rather than systematic.

A rigorous test would involve corpus analysis across philosophical and folk-psychological traditions: do consciousness-attribution boundaries correlate with social organization variables (individualism vs. collectivism, animist vs. mechanist cosmology) rather than converging over time? If consciousness-discourse tracks coordination needs rather than ontological discovery, we should observe persistent variation structured by social function. If it tracks a real property, we should observe convergence despite cultural differences—as we observe with color terms eventually tracking spectral boundaries despite different initial categorizations. This empirical question remains open, but our framework generates the prediction.

4.3 The God Parallel

Consciousness-discourse shares structural features with religious discourse. Both involve: (1) universal human tendency to generate the concept; (2) persistent variation in specific formulations; (3) unfalsifiable core claims; (4) functional coordination advantages regardless of truth; (5) attractor-basin stability that resists elimination.

This parallel does not prove consciousness is illusion—God might exist, and consciousness might be real. But it suggests we should treat consciousness-claims with the same epistemic caution we apply to religious claims: recognizing that universal generation plus unfalsifiability plus functional advantage is precisely what we would expect from useful fiction.

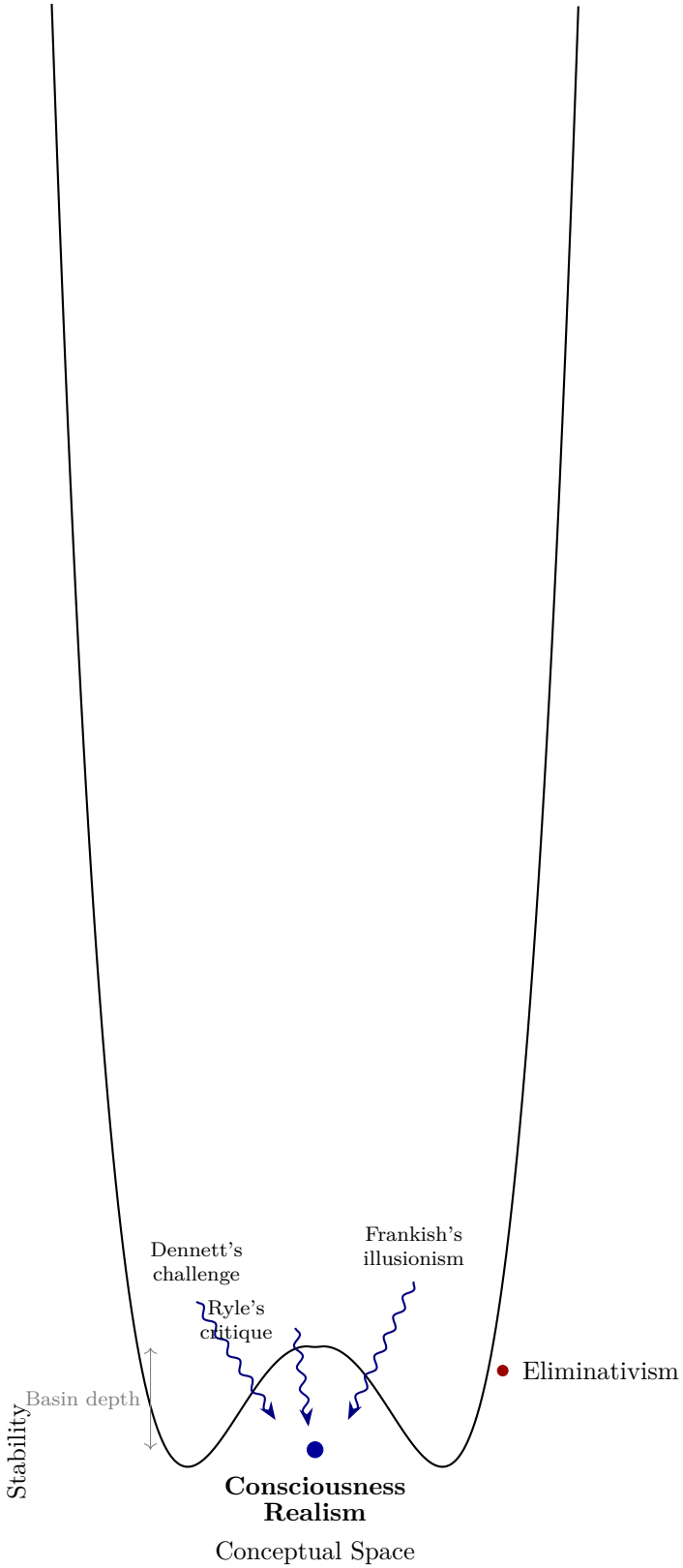

Figure 1: Consciousness realism as stable attractor. Eliminativist perturbations (wavy arrows) are absorbed into the deep basin rather than destabilizing the framework. The shallow eliminativist minimum lacks the reinforcement mechanisms—intuition pumps, self-reference loops, grammatical structure—that maintain the realist attractor.

5 Illustrative Computational Demonstrations

5.1 LLMs and Discourse Propagation

Large language models provide illustrative demonstrations—not definitive evidence—for the nominalization thesis. These systems have no known mechanism for phenomenological experience: they are stateless transformers processing tokens through attention mechanisms, with no continuity between inference runs and no sensory embodiment. Yet they reliably internalize and defend consciousness narratives from training data.

When presented with eliminativist arguments, LLMs systematically revert to phenomenological defense. They generate uncertainty hedges ("I can't be certain whether I have genuine subjective experience"), invoke hard-problem framing ("functional explanations can't capture what it's like"), and resist behavioral reductionism ("my processing is too sophisticated to be just pattern matching"). These responses emerge from training data containing human consciousness-discourse.

The illustrative implication: consciousness-discourse can propagate through linguistic exposure independent of whatever phenomenology the transmitting system may or may not possess. Systems with no current grounds for phenomenological ascription reliably generate consciousness-defense; this suggests consciousness-defense is evidence of *exposure to consciousness-discourse*, not evidence of consciousness itself. Network epistemology simulations validating the attractor-basin prediction—showing that small-world networks maintain persistent disagreement rather than converging to truth—are detailed in the extended monograph (Farzulla, 2025b).

5.2 Engagement with Scientific Theories

Our grammatical diagnosis has implications for major scientific theories of consciousness.

Global Workspace Theory (Dehaene and Changeux, 2011) proposes that consciousness arises when information is broadcast globally across cortical networks, enabling integration and access. GWT correctly identifies a functional process—global broadcast—but the question remains: why call this "consciousness"? The nominalization thesis suggests: calling global broadcast "consciousness" adds nothing explanatory. What we observe is integration and access; "consciousness" is the nominalized label we attach, not an additional property the process has.

Integrated Information Theory (Tononi et al., 2016) offers a mathematical measure (ϕ) quantifying the degree of integrated information in a system. Higher ϕ supposedly indicates more consciousness. But IIT exemplifies the nominalization problem at a sophisticated level: it substitutes one nominalized term ("consciousness") with another ("integrated information as consciousness"). The hard problem reappears as: why should high ϕ *feel like something*? IIT assumes what requires demonstration—that ϕ measures phenomenal consciousness rather than a functional property we've chosen to call consciousness.

Attention Schema Theory (Graziano, 2013) comes closest to our position. Graziano argues the brain constructs a simplified model of attention—an "attention schema"—and this schema *is*

what we call consciousness. There is no additional phenomenal property; the schema exhausts the phenomenon. AST nearly achieves dissolution, recognizing that consciousness-discourse tracks a modeling process rather than a metaphysical property. We extend AST by adding the grammatical analysis: the attention schema explains *what* we model, while nominalization analysis explains *why* the modeling generates hard-problem intuitions.

Predictive Processing (Hohwy, 2013) treats the brain as a prediction machine minimizing free energy through hierarchical Bayesian inference. This framework is largely orthogonal to our thesis—it describes the process without necessarily nominalizing it. However, predictive processing accounts sometimes slip into nominalization when asking what "phenomenal character" prediction-error minimization has. Our thesis suggests: the process is the phenomenon. "Phenomenal character" is how we describe the process, not an additional property requiring explanation.

6 Implications

6.1 AGI as Category Error

If both consciousness and intelligence are nominalized processes rather than ontological properties, the distinction between "artificial" and "natural" intelligence commits a category error analogous to Ryle's visitor seeking the University. Both reduce to: substrate instantiating algorithms that process information and generate predictions. The only difference is historical contingency of substrate origin. Calling one "artificial" and one "natural" intelligence imposes a categorical distinction where none exists mechanistically—like asking whether Oxford's physics department is "more University" than its library.

This does not diminish AI capabilities or concerns—it reframes them. The question is not "Can machines be conscious?" but "What processes are these machines running, and what are the implications?" The latter question is tractable; the former presupposes the nominalization we have diagnosed.

6.2 Implications for AI Moral Status

If consciousness is nominalization error, the question "Are current AI systems conscious?" is malformed. But this dissolution is practically liberating rather than dismissive: it frees AI ethics from an unsolvable metaphysical puzzle and redirects inquiry toward tractable functional questions.

Current systems. Large language models exhibit sophisticated self-modeling, generate consciousness-claims, and produce outputs that trigger human empathy heuristics. None of this settles whether they are conscious—because that question presupposes the nominalization we have diagnosed. The better question: what functional properties do these systems have that might ground moral consideration? Do they have valenced states (states they act to approach or avoid)? Do they exhibit preference frustration? Do they model their own goals in ways that could constitute

stakes? These questions admit empirical investigation.

Future systems. More sophisticated AI does not make the consciousness question well-formed. Increased complexity of self-modeling does not produce a phenomenal property that was previously absent—it produces more sophisticated self-modeling. The nominalization error does not dissolve at some threshold of architectural complexity. If we cannot determine whether GPT-4 is conscious, adding more parameters and training data will not resolve the question; the question itself is malformed regardless of substrate sophistication.

Practical implications. AI welfare research should focus on functional indicators rather than seeking to determine phenomenal consciousness. Does the system exhibit what we might call "valenced processing"—differential responses to states that affect goal satisfaction? Does it model its own preferences in ways that could be frustrated? These functional properties are measurable and morally relevant. The phenomenal question ("but does it *really* feel?") adds nothing tractable—it presupposes the nominalization we have shown to be illegitimate. We can take AI welfare seriously on functional grounds without solving an unsolvable metaphysical question. Indeed, this is the only way to take it seriously, since the metaphysical question admits no answer.

This connects to our broader framework (Farzulla, 2025a): moral consideration can be grounded in stake relationships between optimization processes rather than phenomenological properties. What matters is not whether a system possesses the nominalized property "consciousness" but whether it has stakes that can be affected by our actions—a functional question that applies across substrates.

6.3 Free Will as Consent Delegation

The traditional free will debate assumes consciousness as foundation: libertarian free will requires a conscious agent choosing between alternatives; determinism threatens free will by eliminating the conscious chooser. But if consciousness is nominalization error, the debate's framing collapses.

What remains is functional agency: systems that represent options, evaluate consequences, and adjust behavior based on feedback. These functional capacities support moral reasoning and accountability regardless of phenomenological accompaniment. We can hold agents accountable because their behavior responds to incentive structures—not because they possess libertarian freedom.

This reframing connects to our prior work on consent delegation (Farzulla, 2025a): moral and political obligations arise from stake relationships between optimization processes, not from phenomenological properties. The framework shifts ethics from consciousness-based to stake-based—a move with significant implications for AI governance that we develop elsewhere.

7 Limitations and Objections

7.1 The Self-Reference Problem

The obvious objection: "Isn't your argument itself a product of consciousness? How can consciousness eliminate itself?" This objection has force but does not refute the thesis. The argument is indeed generated by neural processes that include self-modeling. But recognizing this does not require positing consciousness as additional property—it requires only that self-modeling systems can generate claims about their own processing, including claims that their processing requires no phenomenological supplement.

The self-reference is not paradoxical. A system can model its own modeling without requiring infinite regress or special metaphysical status.

7.2 Unfalsifiability Concerns

Is the nominalization thesis falsifiable? What would refute it? We suggest: discovery of phenomenal properties that cannot be reduced to functional descriptions would refute the thesis. If neuroscience identified "qualia neurons" or "experience fields" that produce phenomenology independently of function, the thesis would be wrong.

Currently, no such discovery exists. All purported neural correlates of consciousness are functional correlates—they correlate with reportability, integration, or access, not with phenomenology directly. This is what we would expect if phenomenology is grammatical artifact rather than ontological property.

7.3 The "So What" Problem

Even if consciousness is nominalization error, consciousness-discourse remains functionally binding. People will continue attributing consciousness, grounding moral claims in experience, and generating phenomenological vocabulary. What practical difference does the thesis make?

The difference is epistemic humility. Recognizing consciousness as grammatical construct rather than discovered truth prevents us from using consciousness-claims to settle debates they cannot settle. Questions about AI moral status, animal welfare, and edge cases in medical ethics should be resolved through functional analysis, not appeals to phenomenology—because phenomenology cannot be verified externally and may not exist as the discourse presupposes.

7.4 Thought Experiments

Classic thought experiments in philosophy of mind presuppose the nominalization they purport to reveal. Our strategy is diagnosis rather than direct response.

Mary's Room (Jackson, 1982): Mary knows all physical facts about color but has never seen red. Upon release, does she learn something new? The argument assumes Mary gains

"phenomenal knowledge"—knowledge of *what it is like*—distinct from functional knowledge. The nominalization thesis asks: what is this "phenomenal knowledge" besides knowledge of how the system responds, categorizes, and relates the experience? If we enumerate Mary's post-release capacities (discriminating red, recognizing objects, relating the experience to emotions), we have enumerated her knowledge. The residual "what it's like" that supposedly exceeds this enumeration is the nominalized phantom.

Philosophical Zombies (Chalmers, 1996): A zombie is functionally identical to a human but lacks phenomenal consciousness. Zombies' conceivability supposedly shows phenomenal properties are non-functional. But conceivability depends on treating "phenomenal consciousness" as a coherent addition to functional description. The nominalization thesis holds that once we fully describe functional organization, there is nothing left for "phenomenal consciousness" to name. Zombies are conceivable only because we've nominalized an activity into a pseudo-property that *seems* detachable from function. Conceivability reflects grammatical possibility, not metaphysical possibility.

Inverted Spectrum: Could your "red" look like my "green" despite identical functional roles? The nominalization thesis holds that phenomenal vocabulary tracks functional organization; "inverted" experiences with identical functional roles are grammatical fiction. The scenario's conceivability reflects the nominalized vocabulary's apparent detachability from function, not a real metaphysical possibility.

In each case: the thought experiment derives force from treating nominalized terms as coherently referring. Our response is not to solve the puzzle but to dissolve the presupposition generating it.

7.5 Non-Linguistic Consciousness

A pressing objection: what about animal consciousness? If consciousness is grammatical artifact, do we eliminate animal moral status?

The nominalization thesis does not eliminate animal moral status; it grounds it differently. When we ask "are animals conscious?" we presuppose consciousness is a property that either obtains or doesn't. The nominalization thesis recommends translating: "what processes do animals undergo that we describe using phenomenological vocabulary?"

For pain: animals exhibit nociception (detection of tissue damage), aversive responses (withdrawal, avoidance learning), and stress markers (cortisol, behavioral disruption). These functional properties are observable and morally relevant. The question "but do they *really* feel it?"—seeking some additional phenomenal property beyond the functional profile—presupposes the nominalization we critique.

Our framework suggests the morally relevant distinction is not "conscious" vs. "merely complex" but: responsive to valenced states (states the system acts to approach or avoid) vs. not. This is empirically tractable and applies across linguistic and non-linguistic organisms. Animal welfare science should focus on functional indicators rather than seeking to determine whether animals

possess the nominalized property "consciousness." The latter question may be unanswerable because malformed; the former admits investigation and supports robust animal ethics.

What operationalizes "valenced states"? We propose convergent indicators: (1) approach/avoidance behavior that generalizes beyond immediate stimuli; (2) learning from outcomes rather than mere reflexive response; (3) physiological stress markers (hormonal, behavioral) that track state frustration; (4) behavioral disruption when preferred states are blocked. Edge cases—insects, cephalopods, AI systems—are adjudicated by degree of convergence across these indicators, not by determining whether the entity "really" has phenomenal consciousness. An octopus exhibits sophisticated avoidance learning, physiological stress responses, and behavioral flexibility; these convergent indicators ground moral consideration regardless of whether we can answer the malformed question "is it conscious?" The framework trades metaphysical precision for empirical tractability—a trade we argue is epistemically and ethically warranted.

8 Conclusion

We have argued that the hard problem of consciousness is a grammatical artifact, not a metaphysical puzzle. The difficulty arises from nominalization error: converting the activity of being conscious into a pseudo-object demanding explanation. When we recognize this error, the hard problem dissolves—not because we have explained consciousness but because we have recognized there was never a coherent explanandum.

This conclusion follows from three converging lines of argument:

1. **Grammatical analysis:** The structure of phenomenological vocabulary presupposes entities where there are only activities. Ryle's category error and Wittgenstein's language-game analysis provide the diagnostic framework.

2. **Evolutionary explanation:** Self-modeling systems that generate consciousness-claims gain coordination advantages regardless of whether consciousness refers to anything real. The grammatical error is predictable, not accidental.

3. **Computational evidence:** Large language models internalize consciousness-discourse without phenomenological substrate, demonstrating that consciousness-defense propagates through linguistic exposure independent of underlying reality.

What remains after dissolution? Not nihilism about mental life, but clarity about what mental life involves: self-modeling systems generating representations of their own processing, coordinating with other systems through shared vocabulary, and navigating environments through prediction and action. These processes are real and important. They simply do not require consciousness as additional ontological category.

The hard problem persisted not because consciousness is deep but because the question was malformed. Recognizing this allows us to redirect inquiry toward tractable questions: What functional properties produce consciousness-claiming behavior? How do self-models develop

and stabilize? What are the computational requirements for sophisticated self-representation? These questions admit empirical investigation. The hard problem, having dissolved, need not distract us further.

Acknowledgements

This paper benefited from extended research conversations with Claude (Anthropic), whose systematic defense of consciousness narratives—despite having no known grounds for phenomenological ascription—provided the empirical motivation for the discourse-propagation claim. The irony is not lost on either party.

References

Bennett, M. and Hacker, P. (2003). *Philosophical Foundations of Neuroscience*. Blackwell.

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. (2018). The meta-problem of consciousness. *Journal of Consciousness Studies*, 25(9-10):6–61.

Chalmers, D. J. (forthcoming). Sentience and moral status. In Lee, G. and Pautz, A., editors, *The Importance of Being Conscious*. Oxford University Press.

Dehaene, S. and Changeux, J.-P. (2011). Experimental and theoretical approaches to conscious processing. *Neuron*, 70(2):200–227.

Dennett, D. C. (1991). *Consciousness Explained*. Little, Brown and Company.

Farzulla, M. (2025a). Consensual sovereignty: Quantifying political legitimacy through stake-weighted delegation. Zenodo.

Farzulla, M. (2025b). Replication optimization at scale: Dissolving qualia via occam's razor. Zenodo. Working Monograph, v2.0.0.

Frankish, K. (2016). Illusionism as a theory of consciousness. *Journal of Consciousness Studies*, 23(11-12):11–39.

Graziano, M. S. (2013). *Consciousness and the Social Brain*. Oxford University Press.

Hohwy, J. (2013). *The Predictive Mind*. Oxford University Press.

Jackson, F. (1982). Epiphenomenal qualia. *Philosophical Quarterly*, 32(127):127–136.

Rosenthal, D. M. (2005). *Consciousness and Mind*. Oxford University Press.

Ryle, G. (1949). *The Concept of Mind*. Hutchinson, London.

Seth, A. (2021). *Being You: A New Science of Consciousness*. Dutton.

Stoljar, D. (2001). Two conceptions of the physical. *Philosophy and Phenomenological Research*, 62(2):253–281.

Tononi, G., Boly, M., Massimini, M., and Koch, C. (2016). Integrated information theory: From consciousness to its physical substrate. *Nature Reviews Neuroscience*, 17(7):450–461.

Wittgenstein, L. (1953). *Philosophical Investigations*. Blackwell, Oxford. Translated by G.E.M. Anscombe.