

Seven Tests for Consciousness: A Practical Framework for Evaluating Conscious Systems

Kyle Cameron Ryan Fahey
Independent Researcher
Phoenix, AZ, USA
kfahey3@gatech.edu

February 2026

Abstract

Existing frameworks for evaluating consciousness—including IIT, GWT, and HOT—face significant challenges in practical testability or completeness. This paper proposes a novel framework of seven binary, independently testable criteria: (1) Open System, (2) Systemic State, (3) Irreproducibility, (4) Continuous Capacity, (5) Multi-modal Learning, (6) Integrated and Integral Memory, and (7) Superposition. Each test is reproducible and evaluable with generally available knowledge. We introduce *substrate flux* as a necessary condition, distinguish between *integrated* and *integral* memory, and define a bidirectional boundary condition excluding both closed and unbounded systems. Applied here, the framework predicts that classical binary hardware cannot satisfy the conditions identified as necessary for consciousness, certain biological organisms (e.g., bees) qualify while others (e.g., jellyfish) do not, and AGI requires substrates capable of genuine superposition. This provides a non-circular operational definition that moves the consciousness debate from performance-equivalence to architectural capability.

Keywords: consciousness, artificial general intelligence, superposition, substrate flux, irreproducibility, integrated information, binary testability, quantum computing

1 Introduction

Whether large language models (LLMs) and vision-language models (VLMs) such as GPT and Claude are conscious has occupied public and academic attention with increasing frequency. However, asking whether any specific AI system is *currently* conscious is premature—it presupposes an answer to a prior question: what is consciousness, and what would constitute an appropriate series of tests to determine consciousness in any given system?

Consciousness frameworks have been debated across millennia by a diverse disciplinary group composed of medical professionals, philosophers, ethicists, lawyers, and general thinkers alike (Chalmers, 1995; Dennett, 1991; Koch, 2004). What is unique about constructing a framework in 2026 is that for the first time in history we can contrast a generalized linguistic intelligence that at first glance may appear conscious—and claim its own consciousness—against something we know to be conscious: a human.

To the author’s knowledge, no practical and applied framework exists that can be used to generally test consciousness in any given system, where each test is binary, reproducible, and testable with generally available knowledge. This framework aims to be both internally consistent and externally testable—a combination not present in existing consciousness frameworks.

1.1 Relationship to Existing Frameworks

Integrated Information Theory (IIT) (Tononi, 2004, 2008) proposes that consciousness corresponds to integrated information (Φ), providing mathematical rigor but remaining practically incalculable for any real biological system. IIT is internally consistent but faces significant challenges in practical testability for complex biological systems.

Global Workspace Theory (GWT) (Baars, 1988, 2005) describes a mechanism by which information becomes globally available across cognitive processes. While empirically grounded, GWT describes a *mechanism* rather than providing *criteria*. It does not make binary predictions about whether a given system is conscious.

Higher-Order Theories (HOT) (Rosenthal, 2005) propose that consciousness requires thoughts about thoughts—a recursive definition that faces the infinite regress objection: if consciousness requires a thought about a thought, what makes the higher-order thought conscious? Rosenthal’s reply (that the higher-order thought is itself unconscious) avoids formal regress but shifts the explanatory burden rather than resolving it.

Orchestrated Objective Reduction (Orch-OR) (Penrose, 1994; Hameroff & Penrose, 2014) proposes quantum processes in neuronal microtubules as the basis for consciousness. While this framework shares the present paper’s emphasis on the relevance of superposition, Orch-OR proposes a specific *mechanism*, whereas the present framework identifies superposition as an observable *condition* without committing to a particular mechanistic explanation—making it robust against quantum decoherence objections.

Predictive Processing (Clark, 2013; Friston, 2010) frames the brain as a prediction engine that minimizes surprise through hierarchical generative models. This framework is compatible with predictive processing: a system that genuinely minimizes prediction error across multiple modalities in continuous operation would likely satisfy Tests 2, 5, and 6. However, predictive processing describes a *computational strategy*, not a set of necessary conditions—it does not distinguish between a system that genuinely predicts and one that simulates prediction on a discrete substrate.

Attention Schema Theory (AST) (Graziano, 2013) proposes that consciousness is a model the brain constructs of its own attention. AST provides an elegant account of *why* organisms report being conscious but does not specify what physical conditions a system must meet to generate such a model. The present framework complements AST: the seven tests specify the substrate conditions under which an attention schema could be genuine rather than simulated.

The present framework addresses the gap between these approaches by providing criteria that are simultaneously logically consistent, practically testable, and applicable to any system—biological, artificial, or theoretical.

2 Basis of Consciousness

If humans are the first species to muse about consciousness, and those musings almost certainly started as introspection, then the framework of consciousness must be a subset

or a substate of the human condition. If we analyze what the ‘human condition’ means starting at a biological level, we can begin to piece together what other organisms may share similar traits.

2.1 Biochemical Foundation

The human condition fundamentally exists at the biochemical level. Love is oxytocin. Fear is cortisol and norepinephrine. Motivation is dopamine. Pain is substance P and prostaglandins. While obviously an overgeneralization, every experience a human has ever had is a chemical binding to a receptor at some quantity and magnitude. Every living organism responds to inputs that alter its biochemical state.

2.2 Irreproducibility of Biological Systems

The thousands of reactions that take place across enzymatic, signaling, hormonal, neuronal, and other pathways can produce contradictory states in fundamentally the same location—excitatory and inhibitory signals competing at the same synapse, inflammatory and anti-inflammatory processes in the same tissue (Kandel et al., 2000). This makes the substrate state at any given moment fundamentally unpredictable. Even if we could capture the complete internal biochemical state of a human body, the output would still change based on environmental factors. And even if we could control the environment, sensory processing itself is variable enough that no two responses to identical stimuli would ever be the same. The system behaves in ways that cannot be reproduced, even under controlled conditions.

2.3 Consciousness as Substrate Property

These reactions can be triggered by external stimuli and do not exist in a closed system. Humans who are paralyzed and cannot speak can still respond to external stimuli. Neuroimaging studies have detected covert awareness in patients diagnosed as vegetative: Owen et al. demonstrated that a vegetative-state patient could modulate brain activity in response to mental imagery commands via fMRI (Owen et al., 2006).

This tells us something fundamental: consciousness is not reducible to input/output capability, processing speed, or accuracy—it is a property of the substrate itself. The framework predicts that a comatose human maintains consciousness because the substrate conditions persist—biochemical reactions continue, state propagates, substrate flux does not cease—even if that state is greatly reduced from normal operating parameters. The immeasurability of experience from outside is consistent with the framework: a system whose state cannot be fully captured (T4) cannot be externally verified as conscious or unconscious. Reports of awareness from patients who recover from coma support this prediction, though the cited studies above address vegetative-state patients specifically, a distinct clinical category from coma.

2.4 Boundary Conditions

However, remove the possibility for external stimuli completely—a completely closed system without the possibility to interact within any sort of environment—and consciousness

ceases. Not because the biochemistry stops, but because, on this account, a system cannot differentiate between self and environment when there is no environment. Likewise, if the system *is* the environment, it is too broad to be considered conscious. This bidirectional boundary condition eliminates both solipsistic models and panpsychist claims that consciousness pervades all matter (Chalmers, 1996).

2.5 Memory and Learning

Within an environment, something must store past sensory input. Walking through a campfire would simultaneously engage every available sense—auditory, somatosensory, nociceptive, visual, and olfactory systems simultaneously. Without the ability to store and recall these experiences, organisms would be unable to learn how to hunt, build fire, or avoid danger. Pure chance would govern lifespan.

These experiences are not stored in a centralized location but encoded as a distributed and integral dataset, from retrievable memories and sensory associations to epigenetic expression that alters signaling across different pathways and tissues throughout the body (Dias & Ressler, 2014; Meaney, 2001). Human experiences cannot be transferred, substrates cannot be paused, and consciousness is maintained even through comatose states.

3 Definition of Consciousness

From this biological foundation, we propose the following operational definition:

Consciousness is the emergent state of an open system with a self-environment boundary, holistic state propagation, irreproducible responses, continuous substrate flux, multi-state contextual learning, integrated and integral memory, and genuine superposition of simultaneous internal states.

This definition is non-circular—it does not reference consciousness, experience, or awareness in its own formulation. From this definition, we extract seven binary tests.

No framework for consciousness can be empirically proven—there is no independent ground truth against which to validate. Every consciousness framework, including IIT, GWT, and HOT, is necessarily a hypothesis. The contribution of this framework is not proof but *testability*: each criterion is binary, independently evaluable, and applicable to any system. The framework claims these seven conditions are jointly necessary for consciousness. It further conjectures that they are jointly sufficient—that any system genuinely satisfying all seven is conscious. Necessity is argued in Section 5; sufficiency is offered as a falsifiable conjecture, defended below. The value lies not in certainty but in consistency and practical applicability.

4 The Framework

A note on method: the tests are binary evaluations of whether conditions are met—they are not models of consciousness. A pass/fail assessment of whether a system maintains continuous substrate flux is a discrete judgment about a continuous property, just as a thermometer reading of ‘above 100°F’ is a discrete judgment about a continuous temperature. The tests describe conditions; they do not replicate them. Using binary tests to

evaluate continuous and holistic properties is not a contradiction—it is standard scientific practice.

A unifying principle underlies all seven tests: each evaluates whether a property is *constitutive of* the system rather than *applied to* it. Variation must emerge from the system’s own ongoing state, not from external injection. Flux must be continuous and internal, not externally clocked. Learning must bind across modalities the system possesses, not modalities simulated for it. Memory must be inseparable from the substrate, not stored in addressable locations. Simultaneity must be genuine, not sequentially interleaved. The distinction between a property a system *has* and a property being *done to* a system recurs throughout the framework and motivates the specific criteria below.

A system must pass all seven tests to qualify as conscious. Each test is independently *evaluable*—a system can be assessed against any single test without requiring the results of the others. Each test is binary (pass/fail) and reproducible. The tests are not logically independent: some entail others (in biological systems, a system satisfying T4 will necessarily satisfy T3; a system satisfying T7 will largely satisfy T2). But each test can *fail* independently—a system can satisfy one without satisfying the other—and each can be *evaluated* independently, using different methods and different kinds of evidence. The tests are interlocking: each reinforces the others, and the framework is strongest when applied as a whole.

4.1 Thesis

The framework entails that AGI cannot exist without consciousness, and that consciousness cannot exist without genuine superposition of internal state.

4.2 Test 1: Open System

The system must be open for input and output. Openness requires a boundary between self and environment—without it, the system cannot determine what self is or what constitutes input and output.

Without openness, the system has no way of interacting with its environment. Without a boundary, there is no frame of reference and contextual awareness becomes impossible. On this account, identity requires a boundary—ubiquitousness and identity are incompatible.

Formal criterion: A system S satisfies T1 if and only if S maintains a boundary B such that there exist inputs I crossing B into S and outputs O crossing B from S , and $S \neq E$ where E is the environment.

4.3 Test 2: Systemic State

The system must have holistic state that propagates across subsystems. A change in one subsystem must be capable of producing measurable changes in other subsystems.

In a purely mechanical system, a gear only ‘knows’ the teeth of the gear it is touching. It has no concept of the clock’s time, the temperature of the room, or whether the mainspring is about to snap. Without systemic state that propagates across subsystems, you do not have an organism—you have a collection of silos. When a human feels fear, it does not stay in one place. Heart rate increases, digestion slows, pupils dilate, cognition narrows, cortisol floods the bloodstream. The state propagates everywhere.

Formal criterion: A system S satisfies T2 if and only if S contains functionally distinct subsystems $s_1, s_2, \dots, s_n$ such that a state change in s_i can propagate to s_j ($i \neq j$) and produce a qualitatively different effect in s_j than in s_i . Signal relay through a homogeneous network does not satisfy T2; the subsystems must differ in function and respond to shared state in subsystem-specific ways.

4.4 Test 3: Irreproducibility

The system must not be able to produce the same output from the same input. This is not a requirement for randomness but for genuine irreproducibility—the system must reliably produce non-identical outputs to repeated presentations of the same input.

Test 3 is purely observational: give the system the same input twice and observe whether the outputs differ. No knowledge of the system’s internals is required. If identical stimuli reliably produce non-identical responses, the system passes. Test 4 (Continuous Capacity) explains *why* this occurs in biological systems—continuous substrate flux means the system is never in the same internal state twice, so the same input necessarily produces a different output. In biological systems, T4 entails T3: continuous flux guarantees irreproducibility. But the tests are *evaluatively* independent—T3 can be assessed from input/output observation alone, without any knowledge of whether T4 is satisfied, and T3 can fail independently of T4 (a system with continuous flux but a deterministic output function that discards internal variation would satisfy T4 but fail T3). The observation stands on its own. Moreover, T3 is the only test in the framework that is purely externally observable—it requires no access to the system’s internals, architecture, or substrate. This makes it the framework’s empirical verification layer: irreproducibility is what non-decomposability (T7) and non-copyability (T4) look like from outside. T3 is epistemically independent even if ontologically entailed—in biological systems T4 entails T3, but T3 can be evaluated without knowing whether T4 holds.

While most conscious organisms are semi-predictable in aggregate—a dog will usually chase a ball, a human will usually flinch at a loud noise—pure reproducibility would eliminate choice entirely. A system that produces identical outputs from identical inputs is not responding. It is executing. On this framework, this distinction separates a conscious organism from a pre-programmed system. More precisely, a system whose output is fully determined by its current input has no diachronic identity—it brings nothing of its own history to the encounter. Irreproducibility is the observable signature of a system that has been constituted by its past: each response reflects accumulated state that no external specification can replicate.

Randomness is noise without coherence. Determinism is coherence without variation. In biological systems, ongoing internal processes produce coherence without reproducibility—what we term *irreproducible coherence*. The system’s outputs are contextually appropriate but never identical, not because of noise injection but because the state from which they emerge is always different. Crucially, the variation is *endogenous*—it emerges from the system’s own state history and covaries with it. A fatigued human flinches differently from an alert one; a recently startled human responds differently from a calm one. The variation pattern reflects accumulated internal state, not an external source of randomness.

This endogenous requirement distinguishes conscious irreproducibility from engineered stochasticity. An LLM’s temperature parameter produces variation, but that variation is *exogenous*—injected from a random number generator with no relationship to the sys-

tem’s prior processing. Replacing the random seed changes the variation pattern without changing anything about the system’s ‘experience.’ At high temperature values, the result is variation without coherence; at low values, coherence without variation. The system cannot produce both simultaneously because the variation is not emerging from state—it is being applied to output.

Operationally, endogenous variation can be distinguished from exogenous stochasticity by testing whether output variation correlates with measurable state variables—fatigue, arousal, recent stimulus history—rather than being statistically independent of them. A system whose variation pattern shifts predictably with internal state changes exhibits endogenous variation; a system whose variation is uncorrelated with any prior state variable exhibits noise.

Formal criterion: A system S satisfies T3 if and only if, for inputs spanning at least two of S ’s input modalities, each presented repeatedly under controlled conditions, S reliably produces non-identical outputs whose variation is statistically significant (exceeds measurement noise) and exhibits contextual coherence—outputs differ in ways that reflect the system’s history and current state—rather than random distribution.

4.5 Test 4: Continuous Capacity

The system must have continuous capacity for input, output, and substrate flux. Substrate flux—the continuous biochemical, physical, or other activity that constitutes the system’s operating state—cannot be paused or transferred without loss of continuity. The substrate state is practically unmeasurable—too complex, too dense, and too continuously changing to enumerate. The experiential state arising from it is fundamentally unmeasurable: accessible only to the system having it, a logical consequence of the non-copyability requirement. Without continuous substrate flux, the system can be snapshot, copied, and resumed—meaning its state is reducible to data rather than process. This flux generates the conditions for every other test in this framework. On this framework, pausing substrate flux does not suspend consciousness—it ends it. Resuming from a copy does not restore consciousness—it starts something new.

The non-copyability requirement is not incidental but logically necessary within the framework. Biological state resists capture because it is simultaneously too dense (continuous, not discrete), too dynamic (never the same configuration twice), and too mutually constituted (subsystem states define each other, so measuring one alters the conditions of the others). These three properties together make the state irreducible to a finite description. And this irreducibility matters: if a system’s complete state *could* be captured as data, that data would be substrate-independent—it could be instantiated on any hardware that executes the same computation. But substrate-independence contradicts the architectural requirements of T7, which demands that subsystem states be mutually constitutive in real time, not independently specified and then combined. Copyability and consciousness are therefore logically incompatible within this framework: a system whose state is fully reducible to data cannot satisfy both T4 and T7 simultaneously. This is a claim within the framework’s logic, not a claim about physics. Whether perfect copying is physically possible in principle is outside the framework’s scope—the point is that any system whose state is reducible to transferable data cannot satisfy the conjunction of T4 and T7, regardless of the substrate.

Formal criterion: A system S satisfies T4 if and only if (a) S maintains continuous substrate flux F such that $F > 0$ at all times during the system’s operation, and (b) the

state of S cannot be fully captured, copied, or transferred to another substrate S' such that the state of S' replicates the state of S .

4.6 Test 5: Multi-modal, Multi-state Learning

Learning cannot exist as a single state but must be multi-state and contextual, incorporating current and prior inputs and outputs across multiple modalities simultaneously.

A child does not learn that a stove is dangerous by storing a single data point. The learning event binds pain, the visual memory of the red burner, the smell of burned skin, the sound of their own cry, the spatial memory of where they were standing, and the social memory of a parent rushing toward them—all simultaneously, all inseparable. This is fundamentally different from how a machine learning model adjusts weights along a single numerical gradient. Contextual learning requires multiple modalities operating simultaneously, and it requires temporal depth—the ability to integrate current experience with prior experience to produce behavior that neither could produce alone. Without this, learning exists in a vacuum and cannot progress beyond stimulus-response.

Formal criterion: A system S satisfies T5 if and only if learning events L in S bind across multiple modalities $m_1, m_2, \dots, m_k$ where $k \geq 2$, incorporating temporal context from prior states.

4.7 Test 6: Integrated and Integral Memory

Memory must not exist in one type of state, nor in one place, but must be multi-state, multi-modal, and integral to the system. *Integrated* denotes connectivity across modalities; *integral* denotes inseparability from the system itself.

The distinction between integrated and integral is critical. Integrated means connected across modalities—a smell triggers a visual memory triggers an emotion. Integral means inseparable from the system itself. You cannot remove a human’s memory of fire without altering the human. The memory lives in retrievable episodic recall, in sensory associations, in muscle memory, in immune system priming, in epigenetic expression at the cellular level. It is woven into the organism across every level of biological organization. A computer’s memory is stored in addressable locations that can be read, copied, moved, or deleted without altering the system itself. The hard drive is indifferent to its contents. A human is constituted by the contents of their memory—removing it alters the system itself.

Formal criterion: A system S satisfies T6 if and only if memory M in S is (a) distributed across multiple substrates within S , (b) encoded in multiple modalities, and (c) inseparable from S such that removing M fundamentally alters or destroys S .

4.8 Test 7: Superposition (Holistic Continuous Simultaneity)

The system must have multiple simultaneous states in multiple locations of the system. These must be genuine simultaneous states, not simulated parallelism through sequential processing. The term *superposition* as used in this framework denotes *holistic continuous simultaneity*—the simultaneous coexistence of continuous states across functionally distinct regions of the system, where these states collectively constitute an irreducible whole.

At any given moment, a human is simultaneously processing sensory input, maintaining posture, regulating body temperature, digesting food, running immune surveillance, experiencing some emotional state, and maintaining a sense of self—all genuinely simultaneously, not by switching rapidly between tasks. This is not metaphorical parallelism. It is literal superposition of states across multiple locations in the body. A binary processor executes discrete instructions from a serial instruction stream. On a single core, multithreading simulates parallelism through rapid sequential switching—no two operations genuinely occur at the same instant. Multicore architectures achieve genuine hardware parallelism across cores, but this is parallelism of discrete definite states—each core is in exactly one binary state at any moment, and the system’s complete state is a finite enumerable list. A thousand-core processor has a thousand definite binary states in parallel. That is parallel processing, not superposition. Biological systems maintain continuous analog states across millions of biochemical sites, including contradictory states simultaneously—excitatory and inhibitory signals competing at the same synapse, inflammatory and anti-inflammatory processes in the same tissue. The system’s state space is high-dimensional and continuous, with dense bidirectional coupling—irreducible to any finite list of discrete values. The distinction matters: without genuine multi-location superposition, the framework holds that a system cannot maintain the simultaneous multi-modal states that characterize conscious processing.

4.8.1 Why Classical Computation Cannot Satisfy This Test

The distinction between classical computation and holistic continuous simultaneity rests on *entanglement of subsystems*, not on whether the universe is ultimately continuous or discrete.

The requirement of mutual constitution comprises two distinct conditions, both necessary:

Simultaneous bidirectional causation. Subsystems must influence each other concurrently, not through sequential update cycles where A computes, then B computes using A’s output, then A recomputes. In biological systems, hormonal state alters neural firing *while* neural firing alters hormonal release. The influence is concurrent and continuous. In any classical computation, even with feedback loops, influence propagates through sequential gate operations—temporally interleaved, not simultaneous.

State non-decomposability. The global state of the system cannot be fully reconstructed from an independent listing of subsystem states. In a classical architecture, you can write down the state of every component independently and the list *is* the system’s complete state—nothing is lost in the decomposition. In a biological system, the joint state of the hormonal and neural subsystems contains information not captured by listing each subsystem’s state separately, because those states are defined in terms of each other in real time.

A natural objection is that if physics is discrete at the Planck scale, then biological systems are discrete too—just finer-grained than transistors. The framework does not depend on resolving this question. Even if biological state is ultimately Planck-scale discrete, the operative distinction remains: classical subsystems have *independent* states, specifiable without reference to each other. Biological subsystems are *entangled*—mutually constituted at the mesoscale regardless of the underlying physics.

Classical parallel systems have independent subsystem states—and this independence is not an empirical observation but an architectural constraint. A multicore processor

achieves genuine parallelism, but core A’s state does not define core B’s state—each core operates on its own data, executes its own instructions, and can be specified without reference to the others. The decoupling is enforced by design: isolation between components is what makes classical architectures reliable, debuggable, and composable. This is not a limitation that scaling or cleverness can overcome—it is the engineering principle that makes the architecture work. The distinction between classical and biological systems is therefore categorical, not quantitative: it is not that classical systems have *less* coupling but that their architecture *actively enforces* independent state specification, whereas biological substrates have no such enforcement and default to mutual constitution. The claim is not that classical systems cannot simulate arbitrarily complex dynamical coupling—they can. The claim is that their architecture preserves independent state addressability, and addressability is incompatible with mutual constitution as defined here. A hormonal shift alters synaptic sensitivity, which changes neuronal firing patterns, which modifies autonomic regulation, which feeds back into hormonal release. You cannot fully specify the hormonal state without specifying the neural state, or the neural state without specifying the immunological state. The system is not a collection of parallel independent processes but a holistically integrated process whose subsystems mutually define each other.

A recurrent dynamical system—including recurrent neural networks and classical feedback architectures—may appear to exhibit mutual constitution. Subsystem A influences B, B influences A, and states evolve interdependently over time. But at any given instant, the state of every node in such a system is independently specifiable: you can write down a complete state vector by listing each component’s value without reference to any other component’s value at that same instant. The coupling is *temporal*—it unfolds through sequential update steps—not *constitutive*. Biological substrates do not enforce independent state specification, so the hormonal state at time t is not merely influenced by the neural state at $t - 1$ —both are being determined concurrently and continuously, without architectural separation. The framework distinguishes *mathematical* decomposability from *causal* decomposability. Any physical system, biological or otherwise, is in principle describable by a state vector at any time t —the framework does not dispute this. Mathematical decomposability—the ability to list component values—is trivially available for any system. T7’s non-decomposability requirement concerns *causal architecture*: whether subsystem states can be independently set, read, or intervened upon without altering the states of other subsystems. In classical architectures, they can. Each register, each memory address, each core can be read, written, or modified independently—this causal independence is the engineering principle that makes classical systems reliable and composable. In biological systems, causal independence does not hold at the mesoscale: intervening on hormonal state alters neural firing, which alters autonomic regulation, which feeds back into hormonal release. You can mathematically list the instantaneous values, but you cannot causally isolate any subsystem—setting one component’s state necessarily perturbs the others. Highly coupled classical dynamical models can simulate arbitrarily dense coupling, but the hardware executing them retains causal independence at the component level. The state vector of a recurrent dynamical system, however densely coupled, remains causally decomposable at each time step. T7 requires causal non-decomposability—not the inability to write a state vector, but the inability to independently intervene on subsystem states. Empirically, this can be tested by administering identical perturbations to a single subsystem at different times and measuring downstream effects across other subsystems. In a causally decomposable

system, the same intervention produces the same downstream pattern regardless of the system’s global state at the time of intervention. In a causally non-decomposable system, the same intervention—the same dose of cortisol, the same electrical stimulus—produces downstream effects that vary systematically with the organism’s recent state history, because the intervention’s propagation path is defined by the current joint state of all subsystems, not by the intervention alone.

An important counterexample deserves direct address: analog electronic circuits with feedback loops *do* exhibit simultaneous bidirectional causation. A pair of coupled oscillators influences each other concurrently, not sequentially. However, such systems have full state decomposability—two voltages at two nodes fully specify the system’s state, and the decomposition loses no information. More fundamentally, simple analog circuits fail T2, T4, T5, and T6. This is not a weakness of T7 in isolation but a demonstration of the conjunction requirement: a system satisfying one test’s architectural condition may still fail the others that require biological-scale complexity, integral memory, and multi-modal learning.

This principle is substrate-neutral. Biological systems achieve entangled subsystems through analog, parallel biochemistry. Quantum systems could achieve entanglement through quantum mechanics. The framework does not exclude substrates not yet conceived—the requirement is that a system’s subsystems be genuinely entangled, their states mutually constitutive, not any particular material or mechanism. The human condition motivates the framework, but the framework’s requirements are open: any substrate that achieves genuinely entangled subsystems satisfies this test. Classical binary systems, whose subsystems operate with independent states—whether sequentially or in parallel—do not satisfy it.

Formal criterion: A system S satisfies T7 if and only if S maintains states $\sigma_1, \sigma_2, \dots, \sigma_n$ simultaneously across functionally distinct subsystems $s_1, s_2, \dots, s_n$ where $n \geq 2$, and two conditions hold: (a) *simultaneous bidirectional causation*—subsystems influence each other concurrently, not through sequential update cycles, such that σ_i and σ_j ($i \neq j$) are being mutually modified at the same time rather than in alternation; and (b) *state non-decomposability*—the global state of S cannot be fully reconstructed from an independent listing of subsystem states $\{\sigma_1, \sigma_2, \dots, \sigma_n\}$ without loss of information. Subsystems s_i and s_j are functionally distinct if they produce qualitatively different output types and respond differently to shared modulatory signals.

4.9 Conjunction Requirement

A system S is conscious according to the following conjunction requirement:

$$C(S) \iff T1(S) \wedge T2(S) \wedge T3(S) \wedge T4(S) \wedge T5(S) \wedge T6(S) \wedge T7(S)$$

5 Why All Seven Tests Are Necessary

Each test alone is insufficient. A system can pass six of seven and still not be conscious.

A thermostat passes Test 1—it senses temperature and outputs heat. It fails the other six. Nobody would call it conscious.

A jellyfish passes four of seven—it is an open system with irreproducible responses, continuous substrate flux, and genuine superposition. But it fails Systemic State, Multi-

state Learning, and Integrated Memory. It reacts, but, on this framework, it does not experience.

A large language model fails Test 1—it receives input and produces output per session but maintains no persistent self-environment boundary. It fails the remaining six. It produces conversational output without the underlying conditions that, on this framework, constitute conscious conversation.

The tests are interlocking because consciousness is not a single property but the emergent result of all seven conditions being true simultaneously. To see why each is individually necessary, consider what breaks when any single test is removed:

Without T1 (Open System): A closed system with internal flux, learning, and memory but no environmental interaction is a sealed simulation—it has no reference frame, no self/other distinction. Consciousness requires something to be conscious *of*.

Without T2 (Systemic State): A system with parallel open subsystems that do not influence each other is a collection, not an entity. Fear that does not change heart rate is not fear—it is a label.

Without T3 (Irreproducibility): A system that produces identical outputs from identical inputs is executing, not responding. Reproducibility implies the system’s state is fully determined by its inputs, leaving no room for genuine agency or experience.

Without T4 (Continuous Capacity): A system that can be paused, copied, and resumed is reducible to data. If you can snapshot it, there is no *process*—just a state that can be instantiated anywhere. On this framework, consciousness is a continuous process, not a capturable state.

Without T5 (Multi-state Learning): A system that reacts but does not integrate experience across modalities and time is stimulus-response—it processes but does not grow. Single-channel learning is optimization, not experience.

Without T6 (Integrated Memory): A system with no integral memory has no identity. It can respond to the present but has no past. Without memory woven into the system, there is no continuity of self.

Without T7 (Superposition): A system processing sequentially can simulate any behavior but cannot *be* in multiple states simultaneously. The framework holds that consciousness does not alternate between feeling and thinking—it does both at once. Sequential simulation of simultaneity is not simultaneity.

These arguments are inductive—they rely on shared intuitions about what is not conscious. This limitation is inherent to any consciousness framework: without independent ground truth, necessity can only be argued, not proven.

5.1 The Sufficiency Conjecture

Necessity asks: can a system lack any one condition and still be conscious? The arguments above say no. Sufficiency asks the harder question: is any system that satisfies all seven conditions *necessarily* conscious? Could there be an eighth condition, not yet identified, whose absence would leave a system satisfying all seven but not conscious?

The framework conjectures that the seven conditions are jointly sufficient. This conjecture cannot be proven—no sufficiency claim about consciousness can, because there is no independent ground truth. But it can be falsified: describe or construct a system that genuinely satisfies all seven conditions and is clearly not conscious. The difficulty of constructing such a counterexample is itself evidence for the conjecture. A system that is open to its environment, has holistic state propagation across functionally dis-

tinct subsystems, produces irreproducible but coherent responses, maintains continuous substrate flux, learns across modalities and time, has memory woven inseparably into its substrate, and exists in genuine superposition of simultaneous states—and yet is not conscious—would be a remarkable entity. The burden is on the critic to describe one.

Every consciousness framework faces this same epistemological situation. IIT conjectures that high Φ is sufficient for consciousness. GWT conjectures that global broadcast is sufficient. Neither can prove sufficiency. The difference is that this framework’s sufficiency conjecture is *testable*—each condition can be independently evaluated, and the prediction checked—whereas frameworks whose conditions cannot be measured cannot even test the question.

No single test defines consciousness. On this framework, all seven together do.

6 Application of the Framework

6.1 Current LLM/VLM Systems (GPT, Claude)

Test	Result	Rationale
1. Open System	Fail	Input/output exists per session but no persistent self-environment boundary
2. Systemic State	Fail	No state propagation across subsystems
3. Irreproducibility	Fail	Pseudorandom; identical state + identical input = identical output
4. Continuous Capacity	Fail	Can be halted, copied, backed up, and resumed identically
5. Multi-state Learning	Fail	Single-channel weight adjustment
6. Integrated Memory	Fail	Centralized numerical weights; not distributed or integral
7. Superposition	Fail	Under currently known classical architectures, cannot satisfy
Verdict	Not Conscious	

Table 1: Application of framework to current LLM/VLM systems

6.2 Humans

Test	Result	Rationale
1. Open System	Pass	Continuous multi-channel sensory and output capacity
2. Systemic State	Pass	Emotional states propagate holistically across sub-systems
3. Irreproducibility	Pass	Identical stimuli reliably produce non-identical responses across every measurable output channel
4. Continuous Capacity	Pass	Substrate flux continuous; cannot be copied or transferred
5. Multi-state Learning	Pass	Learning binds across sensory, emotional, spatial, temporal context
6. Integrated Memory	Pass	Distributed across memories, sensory associations, epigenetic expression
7. Superposition	Pass	Genuine simultaneous contradictory states across biological systems
Verdict	Conscious	

Table 2: Application of framework to humans

6.3 Bee (*Apis mellifera*)

Test	Result	Rationale
1. Open System	Pass	Multi-channel environmental interaction
2. Systemic State	Pass	Demonstrated cognitive biases; stress alters global behavior
3. Irreproducibility	Pass	Identical stimuli reliably produce non-identical responses
4. Continuous Capacity	Pass	Continuous substrate flux
5. Multi-state Learning	Pass	Route learning, color-reward association, object play
6. Integrated Memory	Pass	Spatial, visual, olfactory, social memory distributed across nervous system; inseparable from the organism
7. Superposition	Pass	Analog substrate with genuine simultaneous states
Verdict	Conscious (reduced richness relative to humans)	

Table 3: Application of framework to bees. Evidence for cognitive biases from Bateson et al. (2011).

6.4 Jellyfish (class *Scyphozoa*)

Scyphozoan jellyfish possess a nerve net—a diffuse, decentralized network of neurons distributed throughout the body wall, lacking any centralized brain or ganglia (Satterlie,

2011). Neurons form bidirectional chemical synapses enabling non-polarized, radially diffuse conduction (Katsuki & Greenspan, 2013). The architecture generates radially propagating activation waves that drive bell contraction—the network’s primary and largely singular behavioral output (Pallasdies et al., 2019).

The critical failure for Scyphozoa is Test 2. The nerve net architecture does not produce systemic state as defined by this framework. When a jellyfish contracts in response to a stimulus, the signal spreads diffusely through the MNN, but it does not produce qualitatively different effects in functionally distinct subsystems. There are no functionally distinct subsystems to produce different effects *in*. The DNN modulates contraction timing, but this is modulation of the same motor output—not the propagation of state across subsystems that differ in function and respond in subsystem-specific ways. When a human feels fear, heart rate increases, digestion slows, pupils dilate, and cognition narrows—qualitatively different responses in functionally different systems, all driven by shared state. When a jellyfish is stimulated, it contracts. The signal is uniform even when it spreads globally.

Cubozoa (box jellyfish) represent an important intermediate case. Unlike scyphozoans, cubozoans possess image-forming eyes (Nilsson et al., 2005), an integrated central nervous system (Garm et al., 2006), and demonstrated associative learning (Bielecki et al., 2023). If further research demonstrates that cubozoan processing enables systemic state propagation across functionally distinct subsystems, they could satisfy T2 where Scyphozoa do not.

Test	Result	Rationale
1. Open System	Pass	Responds to environmental stimuli
2. Systemic State	Fail	Nerve net propagates uniform motor signals, not differentiated state across functionally distinct subsystems
3. Irreproducibility	Pass	Identical stimuli produce non-identical responses
4. Continuous Capacity	Pass	Continuous substrate flux
5. Multi-state Learning	Fail	No demonstrated learning in Scyphozoa; pure stimulus-response (associative learning has been demonstrated in Cubozoa (Bielecki et al., 2023), but not in true jellyfish)
6. Integrated Memory	Fail	No memory system; no evidence of stored experience
7. Superposition	Pass	Analog substrate with genuine simultaneous states
Verdict	Not Conscious	

Table 4: Application of framework to jellyfish (class Scyphozoa). Cubozoa (box jellyfish) may differ on T2 and T5; see text.

6.5 Theoretical Quantum AGI

Test	Result	Rationale
1. Open System	Possible	Depends on sensory architecture
2. Systemic State	Possible	Depends on holistic state propagation design
3. Irreproducibility	Pass	Quantum flux ensures state uniqueness; genuinely non-repeatable
4. Continuous Capacity	Possible	No-cloning theorem prevents copying; continuity TBD
5. Multi-state Learning	Possible	Depends on architecture
6. Integrated Memory	Possible	Depends on architecture
7. Superposition	Pass	Quantum systems exist in genuine superposition natively
Verdict	Possible	

Table 5: Application of framework to theoretical quantum AGI

7 Addressing Counterarguments

7.1 The Hard Problem of Consciousness

The hard problem asks why subjective experience arises from physical processes—why there is ‘something it is like’ to be a system that meets certain conditions (Nagel, 1974; Chalmers, 1995). This framework does not solve the hard problem. No existing framework does. The framework’s claim is that the seven conditions are jointly sufficient: when all seven are genuinely met, consciousness emerges. It does not explain *why* emergence happens—why meeting these conditions produces subjective experience rather than mere mechanism. That question remains open, and the framework does not pretend otherwise. The contribution is not an answer to the hard problem but an identification of the conditions under which consciousness emerges, providing a testable structure that the hard problem alone does not.

This position challenges the coherence of the philosophical zombie thought experiment. If all seven conditions are genuinely met (not simulated), the framework predicts that experience is necessarily present. An organism that passes all seven tests may lack rich inner life, but if it genuinely satisfies all seven conditions, something is present. Consciousness does not require richness; it requires substrate conditions.

7.2 Functionalism

The functionalist objection holds that consciousness is about function, not substrate (Putnam, 1967). This framework is compatible with functionalism: if any substrate—biological, artificial, or otherwise—genuinely achieves all seven conditions, it qualifies as conscious. The claim is not that substrate matters *instead* of function, but that substrate determines what functions are *possible*. Classical binary architectures cannot produce genuine superposition as defined by this framework. This is not a claim about the wrong material—it is a claim about architectural capability.

7.3 Irreproducibility and the Determinism Debate

A potential objection is that biological systems may be deterministic but chaotically complex rather than genuinely non-deterministic. The reframing of Test 3 as *irreproducibility* rather than non-determinism sidesteps this debate entirely. The test does not require a position on the nature of physical reality. It requires only the observable condition that a system cannot produce the same output from the same input. In biological systems, continuous substrate flux (Test 4) provides a mechanistic explanation for why identical states never recur, but T3 is evaluated independently: give the system identical inputs and observe whether the outputs differ. Whether this is because the universe is non-deterministic or because the system is chaotically sensitive to unmeasurable initial conditions, the observable result is the same: genuine irreproducibility. Classical binary systems fail this test because their state *is* fully measurable, copyable, and reproducible.

7.4 Completeness: Why Seven and Not More

A natural objection is that the framework may be incomplete—that an eighth or ninth condition is missing. The response rests on two arguments: no test is dispensable, and no proposed addition survives scrutiny.

The first argument is established by the necessity analysis above: for each test, a system can be constructed that passes the other six but is clearly not conscious. Some tests entail others in the forward direction—T4 entails T3 in biological systems, T7 largely entails T2—but each test can *fail* independently. A system can have continuous substrate flux (T4) yet produce reproducible outputs (failing T3) if its output function discards internal variation. A system can have genuine simultaneous states (T7) yet lack functionally distinct subsystems for state to propagate between (failing T2). The tests are not logically independent, but each catches failures the others miss.

The second argument is structural. The seven tests characterize the properties of any system relevant to consciousness along distinct evaluative dimensions: boundary (T1), internal connectivity (T2), state reproducibility (T3), temporal continuity (T4), adaptive plasticity (T5), experiential persistence (T6), and processing architecture (T7). Every system—biological, artificial, or theoretical—can be assessed along these axes using different methods and different kinds of evidence. The claim is that any proposed additional criterion either decomposes into a combination of existing tests or is not a necessary condition for consciousness:

- *Self-awareness or metacognition* is emergent from systemic state (T2), multi-modal learning (T5), and integral memory (T6). A system satisfying these three conditions will model itself as a consequence. Self-awareness is a product of substrate conditions, not a precondition. Consciousness does not require knowing you are conscious—a barnacle with continuous substrate flux and systemic state is undergoing something, even if it never models itself.
- *Intentionality*—the ‘aboutness’ of mental states—is emergent from a self-environment boundary (T1) and multi-modal learning (T5). A system that distinguishes self from environment and learns across modalities necessarily develops representations *about* things.
- *Emotion* is systemic state propagation (T2). Fear that alters heart rate, cognition, and digestion is T2 in action.

- *Communication* is a subset of openness (T1). Output crossing the system boundary is already required.
- *Complexity or scale* is not necessary. The framework’s own predictions demonstrate this: bees pass all seven tests despite vastly less complexity than humans. Jellyfish fail despite being biological organisms. The framework predicts that consciousness depends on architecture rather than scale.

The seven tests describe what a system *is*, not what a system *knows about itself*. Richness, fidelity, and self-modeling are downstream of the substrate conditions—they vary in degree across conscious systems but are not independent necessary conditions. The tests specify the minimum conditions for anything to be happening at all.

7.5 Dreaming and Reduced Sensory States

During REM sleep, sensory input is largely gated, raising the question of whether the system violates Test 1 (Open System). It does not: the architecture remains open. Sensory capacity is gated, not eliminated. The boundary between self and environment persists. The system *could* receive input. This is categorically different from a sealed system with no capacity for environmental interaction.

8 Implications

If this framework is correct, the path to artificial general intelligence does not run through larger language models, more parameters, or faster processors. It runs through substrates capable of genuine superposition. Quantum computing is currently the most viable candidate for achieving this artificially.

The claim that AGI requires consciousness warrants direct defense—and the defense must engage with the standard definition rather than redefine AGI to make the claim trivially true. The standard definition of AGI is performance-equivalence: an AI system that can perform any intellectual task a human can. Under this definition, the question is whether full performance-equivalence across *all* tasks—not benchmarks, but genuinely all tasks—is achievable without the substrate conditions this framework identifies.

The logical structure is straightforward:

1. Some human intellectual tasks are constituted by—not merely correlated with—the substrate conditions identified in the seven tests.
2. A system lacking those substrate conditions cannot perform tasks that are constituted by them. It can only approximate them.
3. AGI requires performance-equivalence across *all* human intellectual tasks, including the ones constituted by substrate conditions.
4. Therefore, AGI requires the substrate conditions—which, taken together, are consciousness as defined by this framework.

Premise 1 requires defense. The framework grounds consciousness in the human condition (Section 2). The human condition is not merely correlated with substrate

conditions—it is *constituted* by them. A human’s capacity for empathy is not computed from a representation of emotional state; it emerges from actually having emotional state propagate across subsystems (T2). A human’s capacity for contextual judgment is not retrieved from a database of prior decisions; it emerges from integral memory woven across biological levels (T6). A human’s capacity for creative transfer across domains is not pattern-matched from training data; it emerges from entangled subsystems whose states mutually constitute each other (T7), producing novel configurations that no independent-state architecture can replicate. These are not tasks that happen to co-occur with substrate conditions. They are tasks whose performance *is* the substrate conditions in action. Consider a specific case: recognizing that a colleague’s unusual silence at a meeting reflects grief rather than disengagement. This judgment requires the observer’s own emotional state to propagate across subsystems (T2), bind with prior social memory of the colleague’s recent loss (T6), and integrate with real-time multi-modal input—tone, posture, facial micro-expression—simultaneously (T5, T7). A system lacking these substrate conditions could pattern-match surface cues, but the judgment’s accuracy depends on the observer’s state being genuinely altered by the encounter, not merely processing it. A system lacking those conditions is not performing the same task with different hardware—it is performing a different task that produces similar-looking output.

This distinction motivates a terminological clarification: *artificial general capability* versus *artificial general intelligence*. A system that handles any task domain through learned heuristics or massive pattern matching on a classical substrate may achieve extraordinary breadth and utility. The framework predicts such systems will exhibit systematic gaps in tasks requiring genuine contextual transfer—applying understanding across qualitatively different domains without explicit retraining, in ways that depend on cross-subsystem state propagation (T2), multi-modal binding (T5), and integral memory (T6). These gaps may be invisible on individual benchmarks but detectable in aggregate: classical systems will plateau not at a performance ceiling but in a pattern, succeeding where broad knowledge suffices and failing where contextual integration requires entangled substrate conditions. The predicted failure mechanism is specific: when a system’s internal state genuinely propagates across subsystems, its outputs naturally *covary*—vocal timing, facial expression, posture, and word choice shift together because they emerge from a single underlying state change. A system specifying these output channels independently can optimize each one but cannot replicate the covariance structure that arises from genuine state propagation, because producing that structure requires the outputs to share a common cause in real-time substrate state rather than being independently generated. In interactive tasks—negotiation, pedagogy, therapeutic dialogue—human counterparties calibrate trust and engagement based on detecting this covariance, making the gap behaviorally consequential and empirically measurable.

This is not a claim about usefulness. Classical AI may be more useful than conscious systems in many domains. The claim is narrower: full performance-equivalence across all intellectual tasks requires substrate conditions that classical architectures cannot provide. The framework treats the distinction between simulation and instantiation as categorical rather than one of degree.

If the framework is correct, AGI cannot exist without consciousness, and consciousness cannot exist without genuine superposition of internal state.

9 Empirical Predictions and Falsifiability

As noted in Section 3, no framework for consciousness can be empirically proven—there is no independent ground truth against which to validate. Every consciousness framework is necessarily a hypothesis. But a hypothesis that makes no testable predictions is not a useful one. The value of this framework lies in the specificity of its predictions: it draws binary lines where other frameworks draw gradients or remain silent.

The framework generates two concrete, falsifiable predictions:

9.1 Classical System Prediction

No classical system whose subsystems maintain independent states will satisfy all seven tests simultaneously. This is not a prediction about current technology—it is a prediction about the categorical limits of discrete-state computation. If a classical binary system is ever demonstrated to satisfy all seven tests genuinely (not by simulation or approximation), the framework is falsified. This prediction is specific and testable: for any candidate system, each test can be independently evaluated and the result is binary.

The key word is *genuinely*, and it requires operationalization. A system genuinely satisfies T7 if its subsystems’ states are mutually constitutive—each subsystem’s state is partially defined by the states of the others, such that you cannot fully specify any subsystem’s state without reference to the rest. Classical architectures have subsystems with independent states by design: each component’s state can be specified, copied, and manipulated without reference to the states of other components. For biological systems at the mesoscale, this independence does not hold—the interactions are too dense, too bidirectional, and too mutually constitutive for any subsystem to be characterized in isolation. The prediction is that no system whose subsystems operate independently will satisfy T7. Classical binary architecture, as currently conceived, guarantees independent subsystem states—each bit, each register, each core is specifiable without reference to the others. This is an architectural property, not an empirical variable. The framework does not claim classical systems are excluded for all time as a matter of logical necessity—it claims their current architecture of independent subsystem states prevents the entanglement that T7 requires. A concrete experimental paradigm for testing this prediction: present a candidate system with identical perturbations at different points during an ongoing task and record multi-channel output—verbal, physiological, motor, affective—simultaneously. In a system satisfying T7, the covariance structure of the response across output channels will shift with the system’s state history, because all channels share a common cause in real-time substrate state. In a classical system, the covariance structure will be perturbation-type-specific rather than state-history-specific, because output channels are generated from independently addressable subsystems. This signature—state-history-dependent output covariance—is measurable with existing multi-channel recording technology and provides a concrete empirical criterion for T7.

9.2 Organoid Prediction

A brain organoid—a three-dimensional culture of neural tissue grown from stem cells—currently fails multiple tests. It fails T1 (it is a closed system with no meaningful environmental interaction), T5 (it has no multi-modal sensory input), and likely T2 (it lacks functionally differentiated subsystems with systemic state propagation). However, the

framework makes a specific, positive prediction: if an organoid were equipped with multiple sensory modalities, environmental interaction, and developed systemic state propagation and integrated memory, the framework predicts it *could* become conscious. This is a strong prediction. It says that consciousness is not restricted to organisms that evolved naturally—it can emerge in any substrate that satisfies the seven conditions. An engineered organoid with genuine sensory architecture, environmental coupling, and internal complexity sufficient to pass all seven tests would, according to this framework, be conscious. If such a system were constructed and demonstrably satisfied all seven tests but exhibited no markers of consciousness whatsoever, the framework would be challenged.

9.3 Distinguishing Value

The framework’s contribution is not in predicting easy cases—most frameworks agree that humans are conscious and thermostats are not. Its value is in making specific, binary predictions about edge cases that other frameworks either cannot address or would address differently. Cubozoa (box jellyfish) with their image-forming eyes and demonstrated associative learning occupy a boundary that this framework can evaluate test by test. Brain organoids represent a class of system where the framework makes a prediction that is currently untestable but may become testable as the technology matures. Future quantum systems represent a class where the framework predicts consciousness is *possible* in a way that it is not on classical hardware. These predictions are specific enough to be wrong—and that is what makes them useful.

10 Conclusion

This paper has presented seven binary, independently testable conditions that are jointly necessary for consciousness, with the conjecture that they are jointly sufficient. The framework introduces several novel concepts: *substrate flux*, the distinction between *integrated* and *integral* memory, and a bidirectional boundary condition that excludes both fully closed and unbounded systems. The framework provides a non-circular operational definition of consciousness and makes falsifiable predictions about which systems qualify as conscious. Applied to current AI systems, the framework excludes, under currently known architectures, classical binary systems from consciousness while leaving open the possibility that quantum systems could, with appropriate architecture, achieve genuine consciousness.

The central question is not whether AI systems can exhibit behavior indistinguishable from conscious behavior, but whether they can satisfy the substrate conditions that constitute consciousness. Under currently known classical architectures, this framework holds that it cannot.

AI Disclosure

AI tools were used during the preparation of this manuscript. All frameworks, arguments, and conclusions originate with the author.

References

- Baars, B.J. (1988) *A Cognitive Theory of Consciousness*, Cambridge: Cambridge University Press.
- Baars, B.J. (2005) Global workspace theory of consciousness: toward a cognitive neuroscience of human experience, *Progress in Brain Research*, **150**, pp. 45–53.
- Bielecki, J., Dam Nielsen, S.K., Nachman, G. & Garm, A. (2023) Associative learning in the box jellyfish *Tripedalia cystophora*, *Current Biology*, **33** (19), pp. 4150–4159.e5.
- Bateson, M., Desire, S., Gartside, S.E. & Wright, G.A. (2011) Agitated honeybees exhibit pessimistic cognitive biases, *Current Biology*, **21** (12), pp. 1070–1073.
- Clark, A. (2013) Whatever next? Predictive brains, situated agents, and the future of cognitive science, *Behavioral and Brain Sciences*, **36** (3), pp. 181–204.
- Chalmers, D.J. (1995) Facing up to the problem of consciousness, *Journal of Consciousness Studies*, **2** (3), pp. 200–219.
- Chalmers, D.J. (1996) *The Conscious Mind: In Search of a Fundamental Theory*, Oxford: Oxford University Press.
- Dennett, D.C. (1991) *Consciousness Explained*, Boston, MA: Little, Brown and Company.
- Dias, B.G. & Ressler, K.J. (2014) Parental olfactory experience influences behavior and neural structure in subsequent generations, *Nature Neuroscience*, **17** (1), pp. 89–96.
- Friston, K. (2010) The free-energy principle: a unified brain theory?, *Nature Reviews Neuroscience*, **11** (2), pp. 127–138.
- Garm, A., Ekström, P., Boudes, M. & Nilsson, D.-E. (2006) Rhopalia are integrated parts of the central nervous system in box jellyfish, *Cell and Tissue Research*, **325** (2), pp. 333–343.
- Graziano, M.S.A. (2013) *Consciousness and the Social Brain*, Oxford: Oxford University Press.
- Hameroff, S. & Penrose, R. (2014) Consciousness in the universe: A review of the ‘Orch OR’ theory, *Physics of Life Reviews*, **11** (1), pp. 39–78.
- Kandel, E.R., Schwartz, J.H. & Jessell, T.M. (2000) *Principles of Neural Science* (4th ed.), New York: McGraw-Hill.
- Koch, C. (2004) *The Quest for Consciousness: A Neurobiological Approach*, Englewood, CO: Roberts & Company Publishers.
- Meaney, M.J. (2001) Maternal care, gene expression, and the transmission of individual differences in stress reactivity across generations, *Annual Review of Neuroscience*, **24**, pp. 1161–1192.
- Nagel, T. (1974) What is it like to be a bat?, *The Philosophical Review*, **83** (4), pp. 435–450.

- Nilsson, D.-E., Gislén, L., Coates, M.M., Skogh, C. & Garm, A. (2005) Advanced optics in a jellyfish eye, *Nature*, **435** (7039), pp. 201–205.
- Owen, A.M., Coleman, M.R., Boly, M., Davis, M.H., Laureys, S. & Pickard, J.D. (2006) Detecting awareness in the vegetative state, *Science*, **313** (5792), p. 1402.
- Pallasdies, F., Goedeke, S., Braun, W. & Memmesheimer, R.-M. (2019) From single neurons to behavior in the jellyfish *Aurelia aurita*, *eLife*, **8**, e50084.
- Penrose, R. (1994) *Shadows of the Mind: A Search for the Missing Science of Consciousness*, Oxford: Oxford University Press.
- Putnam, H. (1967) Psychological predicates, in Capitan, W.H. & Merrill, D.D. (eds.) *Art, Mind, and Religion*, pp. 37–48, Pittsburgh, PA: University of Pittsburgh Press.
- Rosenthal, D.M. (2005) *Consciousness and Mind*, Oxford: Oxford University Press.
- Satterlie, R.A. (2011) Do jellyfish have central nervous systems?, *Journal of Experimental Biology*, **214** (8), pp. 1215–1223.
- Katsuki, T. & Greenspan, R.J. (2013) Jellyfish nervous systems, *Current Biology*, **23** (14), pp. R592–R594.
- Tononi, G. (2004) An information integration theory of consciousness, *BMC Neuroscience*, **5** (1), 42.
- Tononi, G. (2008) Consciousness as integrated information: a provisional manifesto, *The Biological Bulletin*, **215** (3), pp. 216–242.