

# **Towards a Theory-Neutral Evaluation Protocol for Consciousness Theories: The Consciousness Theory Evaluation Protocol (CTEP) v1.0**

Max Wharton

*Independent Researcher*

June 2026

## **Abstract**

The scientific study of consciousness is characterised by theoretical proliferation without agreed evaluative standards. Robert Kuhn's catalogue of consciousness theories identifies more than 340 distinct proposals, spanning incompatible ontological commitments and assessed, where they are assessed at all, by criteria that vary from theorist to theorist. The result is a field in which the same phenomena are claimed as confirming evidence by mutually incompatible frameworks, and in which theoretical adjudication lags badly behind theoretical production. This paper proposes the Consciousness Theory Evaluation Protocol (CTEP) as a theory-neutral instrument designed to address this gap.

CTEP organises evaluation around seven questions that any theory of consciousness must answer regardless of its ontological commitments: what consciousness is, how it relates to physical processes, what it does causally, what systems have it, why it has the character it does, what explains its presence in this universe, and how it could be empirically detected or falsified. These questions are assessed against four criteria: coherence, sufficiency, discriminability, and parsimony, where parsimony is evaluated at two levels distinguishing initial ontological cost

from downstream explanatory yield. The seven questions are organised into a three-tier dependency hierarchy reflecting the logical structure of the explanatory demands facing any theory. Tier 1 contains the axiomatic questions whose answers determine what answers are available downstream. Tier 2 tests how well those axioms map onto observable reality. Tier 3 assesses empirical tractability. The instrument produces a structured profile across twenty-eight scored cells, each requiring written justification, organised into a qualitative synthesis and research frontier assessment. The instrument produces a characteristic profile, not a verdict.

The protocol is applied as a proof of concept to four major theories: Global Neuronal Workspace Theory, Integrated Information Theory, Orchestrated Objective Reduction, and the Infinite Continuum framework. Each theory was evaluated independently by two large language model evaluators using identical source texts and instrument specifications. The resulting profiles are clearly differentiated, confirming that the instrument discriminates between theories rather than producing uniform outputs. A systematic evaluator difference was identified across all four theories, with the two evaluators diverging consistently on sufficiency assessments for causal function and phenomenal structure questions. This divergence is interpreted as a prior-correlated contested finding reflecting a genuine fault line in the field rather than evaluator error, and it motivates the multi-evaluator panel architecture described in the paper.

The paper proposes a full multi-evaluator implementation architecture in which a minimum panel of three independent evaluators with declared theoretical backgrounds produces profiles that are compared by a synthesis layer designed to distinguish stable findings, contested findings, and instrument artefacts. The mapping problem, the question of why a proposed mechanism produces specific qualitative character rather than merely correlating with it, is identified as the field's most persistent shared liability, appearing as an unresolved sufficiency gap in every theory

evaluated. CTEP v1.0 is offered as a starting point for community development. Critique, application, and refinement are explicitly invited.

## **1. Introduction**

The scientific study of consciousness is marked by a paradox. It is one of the most active areas of contemporary inquiry, yet it lacks the evaluative infrastructure that other mature scientific fields take for granted. Robert Kuhn's ongoing catalogue of theories of consciousness has identified more than 340 distinct theoretical proposals, a figure that almost certainly underestimates the true count given the volume of informal theoretical activity occurring outside peer-reviewed publication. These theories span incompatible ontological commitments: physicalist and non-physicalist, reductive and non-reductive, generator and filter, functionalist and structuralist, quantum and classical. They are assessed, when they are assessed at all, by criteria that vary from theorist to theorist and from paper to paper. The field has no agreed standard for what a theory of consciousness must explain, no agreed method for determining whether a proposed explanation is sufficient, and no agreed procedure for comparing theories across ontological divides. The result is a field in which theoretical proliferation outpaces theoretical adjudication, and in which the same phenomena are claimed as confirming evidence by mutually incompatible frameworks.

Recent efforts have begun to address this gap, and several bodies of work constitute the immediate prior art for the protocol proposed here. Hanson and Walker (2021) deploy the unfolding argument, derived from the Krohn-Rhodes theorem in automata theory, to produce a logical rather than empirical falsification of Integrated Information Theory and any causal structure theory of consciousness. Their central result is that any system IIT claims is conscious can be mathematically unfolded into a functionally identical system IIT classifies as not

conscious, a contradiction that refutes the theory on internal grounds. This is a significant contribution to the falsification literature, but its scope is deliberately narrow: it addresses causal structure theories only and does not apply to quantum substrate theories, filter and access models, panpsychist accounts, or hybrid frameworks. Ma and Wang (2026) propose three evaluation axes, physical feasibility, philosophical sufficiency, and empirical testability, as a neutral starting point for assessing theories of consciousness, and their framework is the closest antecedent to the instrument proposed here. Their coverage is similarly bounded: the framework is developed specifically for quantum theories and has not been extended to the full theoretical landscape. Herzog, Schurger, and Doerig (2026) provide the most recent comprehensive survey of competing theories, with contributions from originators of leading frameworks, offering a taxonomy of the field that complements Kuhn's (2024) catalogue. Signorelli, Szczotka, and Prentner (2021) propose a systematic classification of explanatory profiles for models of consciousness, identifying three analytical dimensions along which theories diverge and applying them to fourteen theories with provisional scoring. This is the most direct structural antecedent to CTEP: it moves from classification toward comparison and notes that theories often differ not merely in their conclusions but in their underlying conception of what an explanation of consciousness requires. Del Pin, Skora, Sandberg, Overgaard, and Wierzbichon (2021) demonstrate empirically that fewer than seven per cent of publications on IIT and GNW mention both theories, quantifying the siloing problem the field faces and advocating for systematic adversarial comparison as a corrective. Doerig, Schurger, and Herzog (2021) examine the hard criteria that empirical theories of consciousness must satisfy to count as scientific proposals. These are complementary efforts, and the present paper is intended to sit alongside rather than supersede them.

What none of these efforts provides is a unified evaluation instrument applicable across all theory categories: causal structure theories, quantum substrate theories, filter and access models, panpsychist accounts, and hybrid frameworks. Hanson and Walker handle logical falsification for one category. Ma and Wang handle three-axis evaluation for a second. Kuhn and Herzog et al. map the landscape without providing a scoring methodology. Signorelli et al. classify theories by explanatory type but do not evaluate whether a theory's answers to its chosen questions are adequate, only what kind of answers it offers. Del Pin et al. advocate for comparison but do not specify the questions theories must answer or the criteria answers must satisfy. The gap is a deployable, theory-neutral protocol that specifies what questions any theory must answer, what criteria those answers must satisfy, how the relative weight of different questions should be understood, and how evaluator bias should be identified and managed. The Consciousness Theory Evaluation Protocol (CTEP) proposed here is designed to fill that gap. It does not presuppose that consciousness is generated, filtered, emergent, fundamental, illusory, or relational. It imposes equivalent burdens on all theoretical positions by requiring each theory to answer seven questions organised into a three-tier dependency hierarchy, assessed against four evaluative criteria. It produces a structured profile revealing where a theory's explanatory power is concentrated and where its liabilities lie. Not a binary verdict, but a characteristic shape that can be compared across theories and tracked across theoretical versions. The protocol includes a multi-evaluator architecture designed to surface and manage evaluator bias, and proposes a synthesis layer capable of distinguishing stable findings from contested ones. CTEP v1.0 is a starting point. It is offered as a concrete proposal for the kind of evaluative infrastructure the field needs, and community critique, application, and refinement are explicitly invited.

## **2. Design Principles**

CTEP is built on four principles, each of which reflects a considered response to the specific problems that have undermined previous evaluation attempts.

The first principle is theory neutrality. The instrument imposes equivalent burdens on all theoretical positions. A theory that holds consciousness to be ontologically primary faces the same structured questions as a theory that holds it to be an emergent product of physical complexity. A panpsychist theory, a functionalist theory, an illusionist theory, and a relational theory are all required to answer the same seven questions against the same four criteria. The instrument does not assume that any family of theories has already solved any of the questions it poses. Where prior evaluation instruments have been built around a specific theoretical commitment, such as the hard problem of consciousness, integrated information, or global broadcast, CTEP treats those as positions theories may take rather than as presuppositions the instrument endorses.

The claim to theory neutrality requires a direct engagement with an objection raised by Beni (2026), who argues that theory-neutral methodology in consciousness science is incoherent. Drawing on accounts of theory-ladenness in the philosophy of science, Beni demonstrates that observation, experimental design, and data interpretation in consciousness research are structured by the theoretical commitments of the researchers, and that any attempt to purge theoretical bias produces naïve empiricism rather than genuine neutrality. This is a serious objection and it applies to CTEP. The instrument's response is not to deny the point but to distinguish between two levels at which theory-neutrality can be claimed. CTEP does not claim that evaluators are theory-neutral. It claims that the instrument's seven questions impose equivalent burdens on all theoretical positions. Whether consciousness is generated, filtered, emergent, or fundamental, the same questions must be answered against the same criteria. Evaluator priors will shape how

those questions are read and how answers are weighted, but the instrument itself does not presuppose any answer. The multi-evaluator protocol described in section 5 is a direct response to Beni's point: rather than eliminating evaluator theory-ladenness, CTEP makes it visible, tracks it systematically through evaluator declarations, and uses the pattern of divergence across evaluators to distinguish contested findings from stable ones. The goal is not purity but transparency.

The second principle is structured sufficiency. Every question in the instrument requires a substantive answer. Reframing, dissolution, and deferral are scored as sufficiency failures unless the dissolution argument is itself sufficient. A theory that answers a different question than the one asked, acknowledges a question without engaging it, or defers it to another discipline without providing a specifically phenomenological account, fails sufficiency for that question. This principle is grounded in a distinction drawn by Schurger and Graziano (2022) between a scientific law and a scientific theory. A law describes the conditions under which a phenomenon occurs; a theory explains why those conditions produce the phenomenon. On this account, most current accounts of consciousness are proposed laws rather than theories: they describe what accompanies consciousness without explaining why. The sufficiency criterion in CTEP is designed to surface this distinction by requiring answers that engage the explanatory question rather than redescribing the phenomenon. The related concept of the fallacy of misplaced objectivity, named by Signorelli and colleagues (2026), identifies the tendency of consciousness models to explain objective correlates while leaving the subjective structure of experience unaddressed. CTEP's sufficiency criterion applies pressure at precisely this point: it distinguishes between answering the question asked and answering a related but structurally different question.

The instrument is designed to make that substitution visible rather than allowing it to pass unnoticed.

The third principle is profile over verdict. The instrument produces a characteristic shape, not a score. The profile reveals where a theory's explanatory power is concentrated and where its liabilities lie, and it can be compared across theories and tracked as theories develop over time. This principle reflects the genuine state of the field. Consciousness is not a solved problem. An instrument that produced clean pass or fail verdicts would be telling us more about the instrument than about the theories. The profile approach acknowledges that different theories have made genuine progress on different questions, and that identifying those patterns is more useful than ranking theories on a single scale.

The fourth principle is versioning discipline. Every application of CTEP specifies the theory being evaluated, the specific version or source text used, the date of evaluation, the evaluator's declared background and theoretical commitments, and the version of the instrument applied. Theories develop. A CTEP profile is a snapshot of a specific version of a theory at a specific date, not a permanent judgement. This principle also applies to the instrument itself. CTEP v1.0 will be refined as it is applied to a broader range of theories and as evaluator experience accumulates. Version numbers matter on both sides of the evaluation.

### **3. The Instrument**

CTEP organises its evaluation around four criteria applied to seven questions arranged in a three-tier dependency hierarchy. This section specifies each component in full.

#### **3.1 The Four Criteria**



The following criteria apply to every question in every tier. They are defined once and applied consistently throughout.

Criterion A is Coherence. The theory's answer to a given question must be internally consistent and consistent with its answers to all other questions. A theory fails coherence if its answer to one question contradicts or undermines its answer to another, or if the answer contains a formal contradiction. Coherence is assessed within the theory on its own terms. The evaluator does not import external commitments when assessing this criterion.

Criterion B is Sufficiency. The theory's answer must actually address the question rather than reframe, defer, dissolve, or change the subject. A theory that answers a different question than the one asked, or that acknowledges the question without providing a substantive response, fails sufficiency. Partial credit is awarded where the theory provides a genuine but incomplete answer that advances the question without fully resolving it. Dissolution of a question, that is, arguing that the question is malformed, counts as a sufficiency response only if the dissolution argument is itself sufficient. Claiming a question is meaningless without argument fails sufficiency.

Criterion C is Discriminability. The theory's answer must generate at least one consequence, empirical, conceptual, or predictive, that distinguishes it from competitor theories. A theory that produces answers indistinguishable from all its competitors on a given question fails discriminability for that question, even if those answers are coherent and sufficient. Discriminability is assessed at two levels. Empirical discriminability means the answer generates testable predictions that competitors do not make. Conceptual discriminability means the answer generates distinct verdicts on thought experiments, edge cases, or boundary conditions that competitors handle differently. Both levels should be reported where applicable.

Criterion D is Parsimony. Parsimony is assessed at two levels and both must be reported in every cell where it is applied. Initial parsimony measures the ontological cost of the theory's commitments at that question: how many new entities, properties, laws, or primitives the theory requires beyond what is already accepted in background science and philosophy. Downstream parsimony measures the explanatory yield per unit of ontological commitment: how many phenomena, problems, or anomalies the theory's commitments explain or dissolve. A theory with high initial cost but exceptional downstream yield may be more parsimonious overall than a theory with low initial cost and narrow explanatory reach. Both levels must be reported. The evaluator must not collapse them into a single score or treat initial parsimony alone as the measure of a theory's ontological economy.

To illustrate: the Infinite Continuum framework carries a high initial parsimony cost by positing an ontologically real infinite-dimensional experiential phase space. Its downstream yield is correspondingly high, dissolving the hard problem, the cosmological admissibility problem, and the evolutionary convergence problem within a single ontological commitment. Whether that trade-off is favourable depends on the evaluator's prior commitments about what consciousness theories are required to explain. An evaluator who weights cosmological and phenomenological sufficiency heavily may find the trade-off acceptable. An evaluator who weights ontological economy as a near-absolute constraint will not. Both verdicts are legitimate applications of the criterion. The dual-level structure makes the disagreement visible and locatable rather than obscuring it behind a single parsimony score.

### **3.2 The Seven Questions and Three-Tier Hierarchy**

The seven questions are organised into three tiers. The tiers reflect the logical dependency between questions, not an arbitrary weighting scheme.

Tier 1 contains the axiomatic questions: those whose answers determine what answers are available at Tier 2 and Tier 3. A theory that fails sufficiency on a Tier 1 question has not engaged with the central demands of the field. Failures at Tier 1 are not rescued by strong performance at Tier 2 or Tier 3. The evaluator must flag Tier 1 sufficiency failures explicitly in the synthesis.

Tier 2 contains the implication questions: those that test how well the theory's Tier 1 axioms map onto observable reality. Failures at Tier 2 indicate that the theory's core commitments generate implausible or unsupported implications. Tier 2 failures are serious but not necessarily fatal. They may indicate that the theory requires refinement rather than rejection.

Tier 3 contains the epistemic question: whether the theory can be brought into contact with evidence. A theory that fails Tier 3 is not necessarily false. It may be correct but currently untestable. The evaluator must distinguish between theories that are untestable in principle and theories that are untestable with current technology. Permanent Tier 3 failure renders a theory scientifically inert regardless of its Tier 1 and Tier 2 performance.

Tier 1, Question 1: What is consciousness? (The Ontological Question). The theory must provide a substantive account of what consciousness is: its nature, its ontological status, and its relationship to other fundamental categories. A purely functional definition is not a complete answer. Saying consciousness is whatever plays the functional role of consciousness defers the question rather than answering it. The coherence cell asks whether the ontological account is consistent with the theory's accounts of distribution, interface, and phenomenal structure. The sufficiency cell asks whether the account addresses what consciousness is, or instead addresses what consciousness does, what correlates with it, or how it is measured. The discriminability cell asks whether the ontological account generates distinct consequences for how the theory handles

downstream questions. The parsimony cell requires reporting of initial ontological cost and downstream explanatory yield separately.

Tier 1, Question 3: How does consciousness relate to physical processes? (The Interface Question). The theory must provide a principled account of the relationship between consciousness and physical processes, whether that relationship is identity, causation, correlation, constitution, filtration, or something else. The account must specify the direction and nature of the relationship, not merely assert that a relationship exists. The coherence cell asks whether the interface account is consistent with the ontological and distribution accounts. The sufficiency cell asks whether the account specifies the mechanism or structure of the relationship, or merely labels it. The discriminability cell asks whether the interface account generates predictions about which physical systems are conscious and how consciousness varies with physical changes. The parsimony cell requires reporting at both levels.

Tier 1, Question 7: What does consciousness do? (The Causal Function Question). The theory must take an explicit position on the causal efficacy of consciousness. Does consciousness exert genuine downward causation on physical processes, or is it epiphenomenal? If causally efficacious, what specific work does it do that could not be done by physical processes alone? If epiphenomenal, the theory must explain why consciousness exists at all and why it has the character it does without reference to its causal role. The coherence cell asks whether the causal function account is consistent with the interface and ontological accounts. The sufficiency cell asks whether the account specifies what consciousness does that physical processes alone could not do. The discriminability cell asks whether the account generates distinct predictions and requires the theory to take a principled position on whether philosophical zombies are conceivable, possible, or incoherent. The parsimony cell requires reporting at both levels.

Tier 2, Question 2: What has consciousness? (The Distribution Question). The theory must specify which systems are conscious, which are not, and why. The account must follow from the theory's Tier 1 commitments rather than being stipulated independently. The theory must handle boundary cases including simple organisms, fetuses, patients in vegetative states, artificial systems, split-brain patients, and simple physical systems. The coherence cell asks whether the distribution account follows consistently from the ontological and interface accounts. The sufficiency cell asks whether the account provides a principled criterion applicable to novel cases. The discriminability cell asks whether the distribution account generates predictions that differ from competitor theories for at least some systems. The parsimony cell requires reporting at both levels.

Tier 2, Question 4: Why does consciousness have the character it does? (The Phenomenal Structure Question). The theory must explain why conscious experience has specific qualitative character and why it appears unified. This question has two mandatory components. Component A covers qualitative character: why does the proposed mechanism produce this specific phenomenology rather than some other phenomenology or none at all? Component B is the binding sub-question: how does the theory explain the apparent unity of conscious experience? The evaluator must explicitly address whether the theory's binding account is mechanical, definitional, or eliminativist. The coherence cell asks whether the phenomenal structure account is consistent with the interface and distribution accounts. The sufficiency cell requires both components to be addressed explicitly. A theory that ignores binding fails sufficiency on that component. The discriminability cell asks whether the account generates distinct predictions about phenomenal variation and binding breakdown. The parsimony cell requires reporting at both levels.

Tier 2, Question 5: What explains consciousness's presence in this universe at this time? (The Cosmological Question). The theory must account for the relationship between consciousness and cosmological and evolutionary history. The universe began approximately 13.8 billion years ago in a state that most theories would not characterise as richly conscious. Biological complexity emerged only recently and locally. The theory must explain consciousness's relationship to that history without contradiction. The coherence cell asks whether the cosmological account is consistent with the ontological and distribution accounts. The sufficiency cell asks whether the account provides a principled explanation of consciousness's relationship to cosmological history, or merely defers to physics or evolutionary biology without engaging with specifically phenomenological dimensions. The discriminability cell asks whether the account generates distinct predictions about when and where consciousness first appeared. The parsimony cell requires reporting at both levels.

Tier 3, Question 6: How could we tell? (The Empirical Question). The theory must specify at least one prediction that is empirically testable, either currently or in principle, and that discriminates between this theory and at least one significant competitor. This question has three mandatory components. Component A covers positive evidence: what observations or experiments would confirm the theory, specified precisely enough to distinguish confirmation from consistency. Component B covers falsification conditions: what observations or experiments would falsify the theory. A theory that cannot specify its own falsification conditions fails this component. Component C covers counter-evidence handling: how the theory accounts for existing apparent counter-evidence including evidence of unconscious processing, split-brain findings, and vegetative state cases. The coherence cell asks whether the empirical predictions are consistent with the theory's Tier 1 and Tier 2 commitments. The sufficiency cell

asks whether the predictions test the theory's distinctive claims rather than claims shared by all theories. The discriminability cell asks whether the predictions differ from those of at least one significant competitor. The parsimony cell requires reporting of experimental infrastructure cost and downstream empirical yield.

### **3.3 The Scoring Framework and Synthesis**

Each of the seven questions is assessed against each of the four criteria, producing twenty-eight evaluation cells. Each cell receives one of three scores. Pass means the theory provides a substantive, adequate response to this criterion for this question. Partial means the theory provides a genuine but incomplete response; the evaluator must specify what the response achieves and what it leaves unresolved. Fail means the theory does not provide an adequate response; the evaluator must specify whether the failure is one of omission, evasion, or contradiction. Every cell score must be accompanied by a written justification. Scores without justification are not valid applications of the instrument.

After completing all twenty-eight cells, the evaluator produces a structured synthesis addressing four questions. First, Tier 1 integrity: do the three Tier 1 questions cohere with each other? Second, characteristic profile: where is the theory's explanatory power concentrated and where are its liabilities? Third, comparative standing: how does the theory's profile compare to its closest competitors? Fourth, research frontier: what would it take to move the theory's Partial scores to Pass scores?

### **3.4 Versioning and Attribution**

Every application of CTEP must specify the theory being evaluated and the specific version or source text used, the date of evaluation, the evaluator's declared theoretical commitments and disciplinary background, and the version of the CTEP instrument applied. This requirement

applies equally to human and automated evaluators. Declared priors are not a disqualification criterion. They are a transparency mechanism that makes evaluator background visible and interpretable when profiles are compared across evaluators.

#### **4. Proof of Concept**

This section applies CTEP v1.0 to four theories: Global Neuronal Workspace Theory (GNW), Integrated Information Theory (IIT), Orchestrated Objective Reduction (Orch-OR), and the Infinite Continuum framework (IC). The evaluations were conducted by two large language model evaluators, Gemini and Grok, applying the finalised instrument specification independently to the same source texts. The use of LLM evaluators was a deliberate methodological choice, discussed further in section 5. Each evaluator declared their background and prior exposure before producing a profile. All twenty-eight cells were completed for each theory.

The profiles below are structured summaries drawing on both evaluator outputs. Where evaluators converge, findings are reported as stable. Where they diverge, the divergence is noted and its significance assessed. Full evaluator outputs are available as supplementary material.

One declaration is required at the outset. The IC framework is the author's own research programme. Both evaluators had prior exposure to IC materials before producing their profiles. This case therefore directly illustrates the declared prior requirement specified in section 3.4, and the IC profile should be read with that context in mind. The protocol's transparency mechanism is not a guarantee of neutrality. It is a record of the conditions under which a profile was produced.

##### **4.1 Global Neuronal Workspace Theory**



Source text: Dehaene and Changeux (2011) and Mashour et al. (2020), covering the neuronal implementation of Baars' Global Workspace Theory.

GNW's characteristic profile is concentrated in Tier 3 and in Q7 Causal Function. It is the most empirically tractable theory in the dataset. The interface account specifies precise biomarkers: a late post-300ms surge of electrical activity, non-linear frontoparietal ignition, and sustained prefrontal activation. These predictions have been tested directly in adversarial collaboration with IIT through the Cogitate Consortium. The causal function account is the strongest in the dataset on evolutionary grounds. GNW specifies what consciousness does that unconscious processing cannot: it breaks modularity, enabling flexible cross-domain integration, working memory maintenance, and novel behavioural responses. Both evaluators scored Q7 consistently.

The primary liability is in Tier 1 Q1 Ontological Sufficiency, which both evaluators scored Partial. GNW defines consciousness as global information availability. That is a functional definition of access consciousness. It does not address phenomenal consciousness at the foundational level. Both evaluators identified this as a structural deferral rather than a substantive ontological account. The same gap reappears in Q4 Phenomenal Structure Component A, where both evaluators noted that GNW correlates workspace entry with qualitative feel without explaining why broadcasting information should produce phenomenal experience rather than sophisticated unconscious processing.

The evaluators diverged on Q7 Causal Function Sufficiency. Gemini scored it Pass, accepting GNW's account of what global broadcasting accomplishes as sufficient. Grok scored it Partial, pushing back on whether the account explains why those functions require phenomenal experience rather than unconscious processing. This divergence appeared across multiple theories and is discussed as a systematic evaluator difference in section 5.

GNW's profile reveals a theory optimised for neuroscientific tractability and functional explanation at the cost of ontological depth. It is the structural inverse of IIT.

## **4.2 Integrated Information Theory**

Source text: IIT 3.0 and 4.0 as formalised by Tononi and colleagues, including Albantakis et al. (2023).

IIT's characteristic profile is concentrated in Tier 1 and Q4 Phenomenal Structure. Both evaluators scored Tier 1 consistently high across all three questions. The ontological account is substantive and non-deferring: consciousness is identical to a system's intrinsic capacity to exert cause-effect power upon itself. The interface account specifies a precise mathematical calculus. The phenomenal structure account, particularly on the binding sub-question, is the strongest in the dataset. Both evaluators recognised that IIT's integration postulate provides a mechanical and definitional solution to binding that no other theory in the proof of concept matches.

The primary liability is in Q6 Empirical Sufficiency, which both evaluators scored Partial. Calculating exact integrated information for any biologically complex system is computationally intractable using current methods. The theory is therefore forced to rely on proxy measures such as the Perturbational Complexity Index. Both evaluators noted that this insulates the core mathematical postulates from direct falsification, leaving a gap between the theory's precision and its testability.

The evaluators diverged on Q5 Cosmological Sufficiency and Q7 Causal Function Sufficiency. Gemini scored both as Pass. Grok scored both as Partial, identifying that IIT's cosmological account defers the question of why physical laws should favour high integrated information to

physics rather than providing a specifically phenomenological explanation. Both divergences follow the same pattern identified in the GNW run.

IIT's profile reveals a theory whose explanatory power is concentrated precisely where GNW is weakest, and whose liabilities are concentrated precisely where GNW is strongest. Together they form the clearest structural contrast in the dataset.

#### **4.3 Orchestrated Objective Reduction**

Source text: Hameroff and Penrose (2014) Physics of Life Reviews paper, with evaluator notes acknowledging developments through 2025 including quantum biology findings on microtubule superradiance and anesthetic effects.

Orch-OR's characteristic profile is concentrated in Q3 Interface, Q7 Causal Function, and Q5 Cosmological. The interface account specifies a precise biophysical mechanism: quantum superpositions in tubulin protein states, orchestrated by classical synaptic inputs, collapse at a threshold determined by the Diosi-Penrose criterion. The causal function account is the most explicit in the dataset on what consciousness does that physical processes alone cannot. Non-computable processing via objective reduction enables genuine mathematical understanding and intentional choice that deterministic classical computation cannot replicate. Both evaluators scored Q7 consistently as Pass across all criteria, the only theory in the dataset to achieve this.

The cosmological account is also among the strongest in the dataset. Both evaluators accepted the two-phase historical model: un-orchestrated abiotic objective reduction events stabilised physical constants in the early universe, and biological evolution later developed microtubule networks to orchestrate those events into high-fidelity access. This account dissolves the cosmological emergence problem without retrocausal commitment.

The most significant divergence between evaluators occurred on Q4 Phenomenal Structure Component A. Gemini scored it Partial, identifying a residual mapping gap between specific spacetime geometry configurations and specific qualitative character. Grok scored it Pass, accepting the quantum geometry account as sufficient. Gemini's reading is more consistent with the instrument specification. The theory gestures at Platonic values embedded at the Planck scale but does not provide a principled account of why a specific geometric configuration produces the specific feel of pain rather than some other qualitative state. That is a mapping gap, and Component A sufficiency requires more than a structural gesture toward its resolution.

The evaluators also diverged on Q2 Distribution Sufficiency and Q6 Empirical Sufficiency. Grok scored both as Partial, noting vague thresholds for edge cases in distribution and contested empirical status. Gemini scored both as Pass. Given that the empirical programme remains contested and the distribution criterion does not fully specify thresholds for simple organisms and fetuses, Grok's Partial verdicts are the more cautious and defensible readings.

Orch-OR's profile reveals a theory that occupies hybrid territory between physicalism and panprotopsychism, stronger on causal function and cosmological integration than any other theory in the dataset, but carrying the highest initial parsimony cost and facing ongoing empirical contestation.

#### **4.4 The Infinite Continuum Framework**

Source text: Wharton (2026d), formal paper v2, revised June 2026.

As noted above, this profile was produced by evaluators with prior exposure to IC materials. The declared prior requirement is directly illustrated by this case. A further caveat is necessary: the author of this paper is a proponent of the IC framework. Even with declared priors and

transparent evaluation methodology, author-as-proponent introduces a risk of leniency in how the profile is characterised and contextualised. The IC profile should be read with that limitation in mind. It is included in the proof of concept because it represents a genuinely distinct theory type, a filter model, that does not appear in the other three theories evaluated, and because honest self-application of the instrument is methodologically preferable to exemption. It is not included as an endorsement of the framework.

IC's characteristic profile is concentrated in Q1 Ontological, Q5 Cosmological, and Q6 Empirical. The ontological account executes a principled inversion: rather than explaining how physical complexity generates experience, it treats experience as the primary ontological ground and explains how physical constraint architectures govern access to a pre-existing experiential phase space. Both evaluators recognised this as substantively addressing the ontological question rather than deferring it. The cosmological account is stable across both evaluators, dissolving the cosmological emergence problem by treating the phase space as prior to the physical universe and explaining biological filters as late-arising access architectures rather than generators. The empirical programme is unusually aggressive for a theory with fundamentalist metaphysical commitments, deploying explicit falsification conditions and a multi-substrate hyperscanning protocol designed to adjudicate between competing positions within the framework.

The primary liability is in Q4 Phenomenal Structure Component A. Both evaluators scored it as a failure of sufficiency, though with different granularity. Grok scored Component A explicitly as Fail. Gemini scored it as a Partial within the broader Q4 Partial. The paper itself acknowledges the mapping problem as an open theoretical vulnerability. Grok's explicit Fail is the more disciplined verdict and the more consistent application of the instrument specification.

Q7 Causal Function Sufficiency was scored Partial by both evaluators on consistent grounds. The governor dynamics describe how the filter regulates access to the phase space but do not fully specify what consciousness itself accomplishes causally beyond providing the ground that the filter accesses. The parsimony assessment produced a notable divergence: Gemini scored initial cost as Extremely High, Grok as High. Reifying an infinite-dimensional experiential phase space as ontologically real territory represents the largest metaphysical commitment in the dataset, and Gemini's Extremely High is the more defensible verdict.

IC's profile reveals a theory whose strength is concentrated in dissolving the foundational problems that generator theories struggle with, at the cost of a high initial ontological commitment and two specific open liabilities: the mapping problem and the causal function question.

#### 4.5 Comparative Analysis

The four profiles are presented in summary form in Table 1. Scores are reported as Coherence/Sufficiency/Discriminability. Where evaluators diverged, the more conservative verdict is reported. Full evaluator profiles are available as supplementary material.

Question	GNW	IIT	Orch-OR	IC
Q1 Ontological	Partial/Pass/Pass	Pass/Pass/Pass	Pass/Pass/Pass	Pass/Pass/Pass
Q3 Interface	Pass/Pass/Pass	Pass/Pass/Pass	Pass/Pass/Pass	Pass/Pass/Pass
Q7 Causal Function	Pass/Partial/Pass	Pass/Partial/Pass	Pass/Pass/Pass	Pass/Partial/Pass
Q2 Distribution	Pass/Partial/Pass	Pass/Pass/Pass	Pass/Partial/Pass	Pass/Pass/Pass
Q4 Phenomenal	Pass/Partial/Pass	Pass/Partial/Pass	Pass/Partial/Pass	Pass/Partial/Pass
Q5 Cosmological	Pass/Partial/Pass	Pass/Partial/Pass	Pass/Pass/Pass	Pass/Pass/Pass

Q6 Empirical	Pass/Pass/Pass	Pass/Partial/Pass	Pass/Partial/Pass	Pass/Pass/Pass
--------------	----------------	-------------------	-------------------	----------------

*Table 1: CTEP v1.0 Summary Profiles (Coherence/Sufficiency/Discriminability)*

Two structural patterns emerge from the comparative data.

The first is the generator-filter axis. GNW and IIT are both generator theories but with inverted strength profiles. GNW is strong at Tier 3 and weak at Tier 1 ontological sufficiency. IIT is strong at Tier 1 and weak at Tier 3 empirical sufficiency. They share the same liability on Q4 Component A: neither provides a principled account of why the proposed mechanism produces specific qualitative character rather than merely correlating with it. Orch-OR and IC both attempt to ground consciousness in something more fundamental than functional organisation or information structure, and both achieve stronger cosmological accounts as a result. Both carry higher initial parsimony costs in return.

The second pattern is the mapping problem. Every theory in the dataset carries an unresolved liability on the relationship between its proposed mechanism and specific phenomenal character. GNW does not explain why broadcasting produces redness. IIT does not explain why a specific causal geometry produces pain rather than some other quality. Orch-OR does not explain why a specific spacetime configuration produces a specific feel. IC acknowledges the mapping problem explicitly as an open theoretical vulnerability. The instrument surfaces this pattern consistently across all four theories and across both evaluators. It is the field's most persistent shared liability and the research frontier where progress would be most consequential.

The comparative data also demonstrate the multi-evaluator point directly. Across all four theories, the two evaluators produced divergent scores on Q7 Causal Function Sufficiency and

on Q4 Component A Sufficiency in a consistent direction. One evaluator accepted structural or definitional accounts as sufficient. The other demanded explicit explanatory closure. That systematic difference reflects a genuine fault line in how sufficiency is weighted, a fault line that divides evaluators and by implication divides theorists more broadly. The instrument did not resolve that disagreement. It made it visible, located it precisely, and distinguished it from cells where evaluators converged regardless of background. Section 5 discusses what this means for the multi-evaluator protocol and for the synthesis layer.

#### **4.6 Systematic Evaluator Differences**

The two-evaluator structure of the proof of concept makes it possible to quantify the pattern of divergence across the dataset. Of the twenty-eight cells evaluated per theory, the two evaluators diverged on between three and five cells per theory, producing an overall divergence rate of approximately fifteen per cent across the full dataset of one hundred and twelve cell comparisons. Critically, the divergences were not randomly distributed. They clustered on two specific questions: Q7 Causal Function Sufficiency, where divergence occurred on three of the four theories evaluated, and Q4 Component A Sufficiency on phenomenal character, where divergence occurred on two of the four theories. No divergence occurred on Q3 Interface, Q2 Distribution Coherence, or Q6 Empirical Coherence across any theory. This pattern is not consistent with random evaluator noise. It is consistent with a systematic difference in how the two evaluators weight sufficiency on questions where the theory provides a structural or definitional account rather than a fully explicit explanatory account. Gemini consistently accepted structural accounts as sufficient. Grok consistently required explicit explanatory closure. This difference maps onto a genuine fault line in the philosophy of consciousness, between those who accept that identifying the right physical or mathematical structure is



sufficient explanation and those who require that the explanation show why the structure produces experience. Within the CTEP framework this is a contested finding: the divergence correlates with evaluator background and reflects genuine theoretical disagreement rather than instrument error. A panel including human evaluators with declared positions on this fault line would be expected to produce the same pattern of divergence. That is precisely what the synthesis layer is designed to detect and report.

## **5. The Multi-Evaluator Protocol**

### **5.1 The Bias Problem**

The instrument specified in section 3 reduces evaluator subjectivity by requiring written justification for every cell score and by specifying pass, partial, and fail conditions in advance. It does not eliminate subjectivity. Every evaluator, whether human or automated, brings prior commitments to the instrument: disciplinary training, theoretical allegiances, and background knowledge that shapes how questions are read and how answers are weighted.

This is not a flaw unique to CTEP. It is a structural feature of any evaluation that requires interpretive judgement, and consciousness theory evaluation requires a great deal of it. A philosopher trained in the analytic tradition will weight the sufficiency criterion on the ontological question differently than a neuroscientist whose primary concern is empirical tractability. A researcher sympathetic to panpsychist frameworks will read the distribution question differently than a committed functionalist. These differences are not errors. They reflect genuine disagreement about what matters in a theory of consciousness, and that disagreement is part of what the field is trying to resolve.

The bias problem takes a specific form when large language models are used as evaluators. LLMs are trained on large corpora that include philosophical and scientific literature on

consciousness, preprints, online discussions, and informal theoretical work. Their assessments reflect the distributional properties of that training data in ways that are not fully characterisable. Two LLMs trained on different corpora or with different fine-tuning will produce systematically different profiles for the same theory applied to the same instrument. In the proof of concept described in section 4, two LLM evaluators applied the CTEP to Global Neuronal Workspace Theory using the same instrument specification and the same source text. Their profiles diverged on specific cells, most notably on the causal function sufficiency cell and on phenomenal structure sufficiency, in ways that appear to reflect differences in how each model weighted structural versus explicitly explanatory accounts of sufficiency. Section 4.6 quantifies this pattern across all four theories. Those divergences are informative. They reveal that those cells sit in genuinely contested territory where evaluator background predicts the score. But they also demonstrate that a single evaluator, human or automated, cannot be treated as producing a definitive profile.

## **5.2 The Panel Architecture**

The full implementation of CTEP uses a minimum panel of three independent evaluators. The panel architecture is designed to produce three things that a single evaluator cannot: stable findings that hold across evaluator backgrounds, contested findings that reveal genuine theoretical disagreement, and artefact findings that reveal weaknesses in the instrument specification itself.

Each evaluator applies the instrument independently to the same version of the same theory using the same source text. There is no cross-contamination between evaluators during the evaluation process. Before beginning, each evaluator declares their theoretical commitments and

disciplinary background. The declaration is not a disqualification criterion. It is a transparency mechanism that makes priors visible and interpretable when profiles are compared.

A panel that combines human and LLM evaluators is stronger than a panel of humans alone. LLM prior distributions differ structurally from human prior distributions, and disagreements between human and LLM evaluators carry different information than disagreements between two human evaluators with different disciplinary backgrounds. A cell that receives consistent scores across human evaluators but divergent scores from LLM evaluators points to a different kind of problem than a cell that receives divergent scores from human evaluators with different theoretical commitments. The panel architecture is designed to make those distinctions visible.

### **5.3 The Synthesis Layer**

After all panel members have produced independent profiles, a synthesis layer compares them and produces a structured output identifying three categories of finding.

Stable findings are cells where all evaluators agree regardless of declared background. A stable Pass or stable Fail on a given cell is the strongest finding the protocol can produce. It indicates that the theory's performance on that criterion for that question is not sensitive to evaluator background, and can therefore be reported with the highest confidence the protocol affords.

Contested findings are cells where evaluator scores diverge and the divergence correlates with declared priors. A cell that a functionalist evaluator scores as Pass and a panpsychist evaluator scores as Fail on the sufficiency criterion is a contested finding. It does not indicate that one evaluator is wrong. It indicates that the question the cell is asking sits at a genuine fault line in the field, and that the theory's performance on that cell cannot be reported without acknowledging that the verdict depends on contested background commitments. Contested

findings are among the most valuable outputs the protocol produces, because they identify precisely where the field disagrees rather than obscuring that disagreement behind an averaged score.

Artefact findings are cells where evaluator scores diverge but the divergence does not correlate with declared priors. A cell that produces inconsistent scores across evaluators with similar backgrounds points to a problem with the instrument specification rather than genuine theoretical disagreement. Artefact findings are the primary mechanism by which CTEP improves across versions. When a cell consistently produces artefact divergence, the specification for that cell requires refinement.

The synthesis layer can be implemented by a human meta-evaluator, by an algorithmic aggregation model, or by a combination of both. The paper does not prescribe a specific implementation. It proposes the architecture as an open infrastructure problem and invites community contribution to its development. The synthesis layer is the component most likely to benefit from dedicated engineering effort, and it is the component that institutions with relevant technical infrastructure, including projects focused on logical falsification machinery for consciousness theories, are best positioned to build.

#### **5.4 Evaluator Declaration Requirements**

Every evaluator applying CTEP must declare the following before producing a profile: theoretical commitments, including the evaluator's position, if any, on the major fault lines in consciousness research; disciplinary background, including the evaluator's primary training and research focus; and prior exposure, including whether the evaluator has previously evaluated this theory, engaged with the theory's proponents, or has any professional or collaborative relationship with those proponents.

For LLM evaluators, declaration takes a different form. The model identifier, version, and any relevant fine-tuning or system prompt conditions should be recorded. Where possible, the evaluator should note the training data cutoff date, since evaluations of recently developed theories may be affected by limited training data coverage. Declarations are recorded alongside the profile and are available to the synthesis layer. They are not used to weight or discount individual evaluator profiles prior to synthesis. Weighting decisions, if any, are made at the synthesis stage with full transparency about the rationale.

## **6. Limitations and Future Development**

### **6.1 Current Limitations**

Three limitations of CTEP v1.0 should be stated clearly.

The instrument requires evaluator judgement at every cell. The specification reduces subjectivity by defining pass, partial, and fail conditions in advance and by requiring written justification for every score. It does not eliminate subjectivity. The multi-evaluator protocol described in section 5 mitigates this through panel structure and systematic comparison of independent profiles. It does not resolve it. Any evaluation of consciousness theories that requires interpretive judgement will be subject to evaluator influence. The honest position is that CTEP makes that influence visible and manageable rather than claiming to have removed it.

The instrument is currently calibrated against theories that make explicit ontological commitments. Theories that are deliberately non-committal, including some enactivist, relational, and deflationary approaches, may resist clean cell-by-cell scoring. A theory that withholds commitment on the ontological question as a principled methodological stance will score differently on the sufficiency criterion than a theory that simply fails to address the question, but the current specification does not fully distinguish between those cases. Future

versions of the instrument should develop explicit guidance for evaluating theories whose non-commitment is itself a theoretical position.

To illustrate the challenge provisionally: a strong illusionist position, such as that defended by Frankish (2016), holds that phenomenal consciousness as ordinarily conceived does not exist, and that what requires explanation is not experience but the tendency to believe in experience. Applied to CTEP, this position would score Q1 Ontological Sufficiency as a dissolution argument: it denies the explanandum rather than explaining it. Under the instrument's sufficiency criterion, this counts as a pass only if the dissolution argument is itself sufficient, which requires the theory to provide a substantive account of why the belief in phenomenal consciousness arises and why it is systematically mistaken. A theory that merely asserts the denial without that account fails sufficiency. That verdict is available within the current specification without modification, but the instrument's cell descriptions do not currently flag the dissolution pathway explicitly. Future versions should include explicit guidance on how to evaluate dissolution responses to ensure they are assessed as arguments rather than as evasions. Strong dissolution strategies employed by enactivist and relational frameworks raise the same issue: an optional Dissolution Pathway scoring note, appended to the sufficiency criterion specification, would ensure consistent handling across theory types.

The proof of concept in section 4 applies the instrument to four well-documented theories that have been extensively discussed in the literature. The instrument has not yet been systematically applied to the long tail of Kuhn's catalogue, which includes many theories that are obscure, informally stated, or documented only in brief summaries. Performance on that long tail is unknown. It is possible that the instrument specification requires modification to handle theories for which source texts are thin or background knowledge is limited.

## **6.2 Future Development Priorities**

Four development priorities follow from the current limitations and from the experience of building and stress-testing the instrument.

The first priority is community refinement of cell specifications. The twenty-eight cell specifications in section 3 were developed through iterative stress-testing against four theories. That is a narrow base. As the instrument is applied to a broader range of theories, specification weaknesses will become apparent. A version-controlled, publicly accessible specification document would allow the community to propose and debate refinements in a structured way.

The second priority is building the multi-evaluator synthesis layer. The architecture described in section 5.3 requires engineering effort that goes beyond what a single paper can deliver. The synthesis layer needs to handle structured comparison of independently produced profiles, classify findings as stable, contested, or artefact, and produce outputs that are interpretable without specialist knowledge. This is a tractable software engineering problem and a meaningful contribution to the field's evaluative infrastructure. Existing projects focused on logical falsification machinery for consciousness theories represent natural institutional homes for this work.

The third priority is integration with existing theory catalogues. Kuhn's catalogue of 340-plus theories, if made available in a structured format with sufficient theoretical content per entry, would provide an immediate application base for CTEP. A browser interface allowing users to select a theory from the catalogue, specify a source text, and produce a structured profile against the instrument would make the protocol accessible to researchers without specialist knowledge of every theory in the catalogue.

The fourth priority is longitudinal tracking. The versioning discipline built into CTEP means that the same theory can be evaluated at multiple points in its development, using the same instrument version and comparable evaluator panels, producing a record of how the theory's profile changes over time. Longitudinal tracking would allow the field to assess not only where a theory currently stands but whether it is making progress on its specific liabilities.

## **7. Conclusion**

The proliferation of consciousness theories documented by Kuhn's catalogue presents the field with a problem that is partly philosophical and partly organisational. The philosophical problem is real: consciousness is genuinely difficult, and theoretical diversity reflects genuine disagreement about what kind of thing consciousness is and what would count as explaining it. The organisational problem is distinct and more tractable. Even where genuine disagreement exists, the field currently lacks the shared evaluative infrastructure to make that disagreement productive. Theories are assessed by different criteria, on different questions, with different standards of sufficiency, and the result is a field in which theoretical proliferation and theoretical adjudication are badly mismatched.

The Consciousness Theory Evaluation Protocol proposed here is an attempt to address the organisational problem without prejudging the philosophical one. By specifying seven questions that any theory must answer, four criteria against which answers are assessed, and a three-tier dependency hierarchy that reflects the logical structure of the explanatory demands facing any theory of consciousness, CTEP provides a common framework within which theories with incompatible ontological commitments can be compared on equal terms. It does not tell us which theory is correct. It tells us where each theory's explanatory power is concentrated, where its liabilities lie, and where targeted theoretical or empirical work would make the most difference.



The proof of concept in section 4 applied the instrument to four major theories across eight independent evaluator profiles. Several findings warrant emphasis. The four theories produced clearly differentiated profiles rather than clustering around similar scores, which confirms that the instrument discriminates between theories rather than producing uniform verdicts. The mapping problem, the question of why a proposed mechanism produces specific qualitative character rather than merely correlating with it, emerged as a persistent shared liability across the four theories evaluated, appearing as an unresolved sufficiency gap in every profile produced. That finding is not an artefact of the instrument design. It is consistent with the broader observation in the literature that the relationship between proposed mechanisms and specific phenomenal character remains underspecified across competing frameworks, and it identifies a concrete research frontier where progress would simultaneously strengthen multiple theories.

The multi-evaluator runs produced a further finding that is independent of any individual theory. The two LLM evaluators diverged systematically on Q7 Causal Function Sufficiency and Q4 Component A Sufficiency across all four theories, in a consistent direction. One evaluator accepted structural or definitional accounts as sufficient. The other demanded explicit explanatory closure. That systematic difference reflects a genuine fault line in how sufficiency is weighted, a fault line that divides evaluators and by implication divides theorists more broadly. The instrument did not resolve that disagreement. It made it visible, located it precisely, and distinguished it from cells where evaluators converged regardless of background. That is what a well-specified evaluation instrument should do.

Three things follow from this work. First, CTEP v1.0 is a starting point. The instrument will require refinement as it is applied to a broader range of theories, particularly theories that are deliberately non-committal or that resist clean cell-by-cell scoring. The versioning discipline

built into the protocol is intended to support that refinement transparently. Second, the multi-evaluator synthesis layer described in section 5 is the most consequential development priority. The systematic evaluator differences identified in the proof of concept demonstrate both why the synthesis layer is needed and what it should be designed to detect. Building that infrastructure is a tractable engineering problem and a meaningful contribution to the field. Third, the mapping problem is the field's most tractable shared research frontier. Every theory in the dataset carries it as an unresolved liability. Progress on that question would move multiple theories from Partial to Pass on the same cell simultaneously, and identifying shared liabilities of that kind is one of the things a common evaluation framework is for.

CTEP v1.0 is offered as a concrete proposal. The field needs agreed evaluative infrastructure. This is one way to build it. Planned next applications include Higher-Order Thought theory, analytic panpsychism as represented by Goff (2019), and strong illusionism as represented by Frankish (2016), which together would extend the proof of concept across the generator-filter axis into relational and eliminativist territory. Community critique, application to theories not covered here, and development of the synthesis layer are all explicitly invited.

## **References**

- Albantakis, L., Barbosa, L., Findlay, G., Signorelli, C. M., Marshall, W., Tononi, G., et al. (2023). Integrated information theory (IIT) 4.0: Formulating the properties of phenomenal existence in physical terms. *PLOS Computational Biology*, 19(10), e1011465.
- Baars, B. J. (1988). *A cognitive theory of consciousness*. Cambridge University Press.
- Beni, M. D. (2026). Against (theory-neutral) method (in consciousness science). *Neuroscience of Consciousness*, 2026(1), niag003. <https://doi.org/10.1093/nc/niag003>

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.

Consortium, C., Ferrante, O., Gorska-Klimowska, U., et al. (2025). Adversarial testing of global neuronal workspace and integrated information theories of consciousness. *Nature*, 642, 133-142. <https://doi.org/10.1038/s41586-025-08888-1>

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

Dehaene, S., Changeux, J.-P., and Sergent, C. (2021). Experimental and theoretical approaches to conscious processing: Two decades later. *Neuron*, 109(12), 1949-1952.

Del Pin, S. H., Skora, Z., Sandberg, K., Overgaard, M., and Wierzbach, M. (2021). Comparing theories of consciousness: why it matters and how to do it. *Neuroscience of Consciousness*, 2021(2), niab019. <https://doi.org/10.1093/nc/niab019>

Doerig, A., Schurger, A., and Herzog, M. H. (2021). Hard criteria for empirical theories of consciousness. *Cognitive Neuroscience*, 12(2), 41-62.

Hanson, J. R., and Walker, S. I. (2021). Formalizing falsification for theories of consciousness across computational hierarchies. *Neuroscience of Consciousness*, 2021(2), niab014. <https://doi.org/10.1093/nc/niab014>

Herzog, M. H., Schurger, A., and Doerig, A. (eds.) (2026). *Scientific theories of consciousness: The grand tour*. Cambridge University Press. ISBN 9781009244688.

Frohlich, H. (1968). Long-range coherence and energy storage in biological systems. *International Journal of Quantum Chemistry*, 2(5), 641-649.

Hameroff, S., and Penrose, R. (1996). Orchestrated reduction of quantum coherence in brain microtubules: A model for consciousness. *Mathematics and Computers in Simulation*, 40(3-4), 453-480.

Hameroff, S., and Penrose, R. (2014). Consciousness in the universe: A review of the Orch OR theory. *Physics of Life Reviews*, 11(1), 39-78.

Hoel, E. (2021). The overfitted brain: Dreams evolved to assist generalization. *Patterns*, 2(5), 100244.

Kuhn, R. L. (2024). Landscape of consciousness: Toward a taxonomy of explanations and implications. *Progress in Biophysics and Molecular Biology*, 190, 28-169.

Ma, W., and Wang, Z. (2026). Quantum theories of consciousness: A critical review of feasibility, philosophical sufficiency, and empirical testability. *Frontiers in Psychology*, 17, 1730965.

Mashour, G. A., Roelfsema, P., Changeux, J.-P., and Dehaene, S. (2020). Conscious processing and the global neuronal workspace hypothesis. *Neuron*, 105(5), 776-798.

Melloni, L., Mudrik, L., Pitts, M., et al. (2023). An adversarial collaboration protocol for testing contrasting predictions of global neuronal workspace and integrated information theory. *PLoS One*, 18, e0268577. <https://doi.org/10.1371/journal.pone.0268577>

Nagel, T. (1974). What is it like to be a bat? *Philosophical Review*, 83(4), 435-450.

Penrose, R. (1989). *The emperor's new mind: Concerning computers, minds, and the laws of physics*. Oxford University Press.

Penrose, R. (1994). *Shadows of the mind: A search for the missing science of consciousness*. Oxford University Press.

Signorelli, C. M., Szczotka, J., and Prentner, R. (2021). Explanatory profiles of models of consciousness: Towards a systematic classification. *Neuroscience of Consciousness*, 2021(2), niab021.

Schurger, A., and Graziano, M. (2022). Consciousness explained or described? *Neuroscience of Consciousness*, 2022(1), niac001. <https://doi.org/10.1093/nc/niac001>

Signorelli, C. M., Cea, I., and Prentner, R. (2026). We need to explain subjective experience, but its explanation may not be mechanistic. *Neuroscience of Consciousness*, 2026(1), niaf034. <https://doi.org/10.1093/nc/niaf034>

Tononi, G. (2004). An information integration theory of consciousness. *BMC Neuroscience*, 5(1), 42.

Tononi, G. (2008). Consciousness as integrated information: A provisional manifesto. *Biological Bulletin*, 215(3), 216-242.

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.

Wharton, M. (2026a). *The Infinite Continuum*. Zenodo. <https://doi.org/10.5281/zenodo.19220913>

Wharton, M. (2026b). Mind before matter: A pressure paper against generator theories of consciousness (v3). Zenodo. <https://doi.org/10.5281/zenodo.20355468>

Wharton, M. (2026c). After the Lateness Problem (v3). Zenodo. <https://doi.org/10.5281/zenodo.20355585>

Wharton, M. (2026d). The Infinite Continuum: Consciousness as primary and the filter architecture of mind (formal paper v2). Zenodo. <https://doi.org/10.5281/zenodo.20498022>

Wharton, M., and Pender, M. (2026). The Locus of Consciousness (v4 draft). Unpublished manuscript.