Intrinsic Quality Theory: A Geometric Theory of Phenomenal Experience

Diwank Singh Tomer

Draft v1.6 – February 2026

Abstract

We propose Intrinsic Quality Theory (IQT), a framework identifying phenomenal quality with the intrinsic physical state of bounded spacetime regions. Unlike functionalist theories, IQT treats quality as an identity, not an emergent property of information processing. We formalize this using Algebraic Quantum Field Theory (AQFT) on causal sets, defining a region's quality as the pair of local algebra and restricted state, up to physically induced isomorphism. We provide a constraint-based argument that this pair is the unique mathematical object satisfying the requirements any candidate for "intrinsic nature" must meet, and state an explicit Intrinsicness Principle committing IQT to a precise criterion of sameness of quality.

A central commitment is the perspectival relativity of quality: intrinsic content is relative to the locus (scale, position, extent) of the region, with no privileged "view from nowhere." We develop an effective-theory framework that bridges fundamental and neural scales, and show how the geometry of effective causal diamonds generates temporal phenomenology —including the asymmetry between panoramic spatial experience and fleeting temporal experience—without additional postulates.

This perspectival relativity yields a non-ad-hoc mechanism for the composition of experience—grounded in correlation structure, not a special binding force—and dissolves the subject-selection debate by separating universal quality from narrative selfhood. The theory's logical structure separates a core postulate (the identity of quality with the algebra-state pair), bridge hypotheses (conditions for self-threads and narrative operators), and operationalizations (neural metrics approximating bridge quantities). We introduce a narrative operator connecting the formalism to phenomenological report, implemented as a control-theoretic sufficient statistic over self-thread trajectories, and

present a worked toy pipeline from state trajectories through coarse-graining to belief states and report variables.

We specify three pre-registerable experimental protocols—titrated-propofol anesthesia, within-hemisphere overlapping parcellations with perturbational and intrinsic-synergy arms, and psychedelic temporal phenomenology supported by a mechanistic toy model—with defined metrics, predicted response profiles, and explicit failure conditions at each theoretical layer. The primary axis of divergence from Integrated Information Theory is IQT's democracy of diamonds—multiple overlapping regions bearing quality simultaneously—versus IIT's exclusion postulate. Our overlap protocol combines observational, intrinsic-synergy, and perturbational (single-pulse electrical stimulation) evidence to distinguish genuinely semi-autonomous overlapping loci from modular task responses within a single irreducible complex.

IQT says phenomenal quality isn't something brains produce—it's what physical states are from the inside. But the kind of consciousness we care about (coherent experience, selfhood, and report) is not everywhere: it requires specific persistent self-threads and finite-bandwidth narrators—that can be measured, disrupted, and tested. The "self" is the time-extended thread that both persists and gains control over report channels; the feeling of time passing is what finite-bandwidth updating feels like from the inside.

Plain-English Summary

IQT proposes that consciousness isn't something the brain produces—it's what physical states are, from the inside. Every bounded region of spacetime has an intrinsic character (its "quality"), defined by the mathematical pair of its local algebra and state.

The theory uses established physics (AQFT on causal sets) rather than inventing new math. Quality is relative to scale: a neuron's quality differs from a brain region's, and both are real. This "perspectival relativity" dissolves classic puzzles about who is conscious and how experiences combine.

Three levels: Quality (universal, even rocks have it), Experience (requires temporal persistence, like a mouse), and Report (requires a narrative self, like a human). The paper proposes three concrete experiments to test IQT against rival theories like IIT and GNW.

The Core Theses

1. Consciousness is not produced — it's what physics already is, from the inside.

Every bounded region of spacetime has a complete physical description: all the measurements you could make inside that region, and all the values those measurements would give you. IQT calls this the "algebra-state pair." The theory's central bet is that this mathematical object — which physics already uses — is phenomenal quality. Not "gives rise to," not "correlates with," not "produces." Is. The redness of red and the algebra-state pair of the relevant brain region are two descriptions of the same thing, the way "the morning star" and "the evening star" are two descriptions of Venus.

This isn't arbitrary. IQT shows that if you ask "what properties must the intrinsic nature of a physical region have?" — completeness, nesting consistency, covariance, self-containedness — exactly one mathematical object in all of physics satisfies all four. The algebra-state pair. There is no runner-up. The identification with consciousness is still a bet, but it's a bet on the only horse in the race.

2. Quality is everywhere, but experience and selfhood are not.

A rock has quality. A thermostat has quality. This sounds absurd until you see the three-level distinction that makes it precise:

- Level 0 — Quality: Every bounded spacetime region has it. The rock's quality is extremely simple — a featureless, low-complexity state. It doesn't "experience" anything. Quality at this level is like temperature: everything has one, but that doesn't make everything hot.

• Level 1 — Experience: Requires a self-thread — a quality-stream that persists over time and is causally semi-autonomous (what happens inside it is partly determined by its own past, not just outside forces). A mouse has this. A thermostat doesn't, because its states don't maintain complex, self-sustaining temporal structure.

• Level 2 — Report: Requires a narrative operator — a self-thread that builds a compressed self-model and controls a readout channel (speech, behavior). This is the "I" that can say "I see red." Humans have this. Mice probably don't.

The key move: the Hard Problem ("why does anything feel like something?") is dissolved at Level 0 by the identity thesis. The interesting scientific questions live at Levels 1 and 2, and those are empirical questions about what kinds of physical systems sustain self-threads and narrative operators.

3. Quality is relative to scale — and no scale is privileged.

Your visual cortex has a quality. Your whole brain has a different, richer quality. A single neuron in your visual cortex has its own, simpler quality. All three are equally real. None is "the real one."

This is perspectival relativity: there is no God's-eye view that picks out the "correct" level at which consciousness happens. A neuron's quality doesn't include information about cross-regional binding (it can't "see" what the motor cortex is doing). The whole-brain quality includes all of that but at coarser resolution. They're different perspectives on the same physical system, like how a city looks different from street level versus satellite view — both views are real, neither is the truth.

The radical implication: overlapping regions can all bear quality simultaneously. The region covering your left visual cortex plus your auditory cortex has its own quality. So does the region covering just your left visual cortex. So does the region covering your auditory cortex plus your prefrontal cortex. They overlap, and they're all real. IQT calls this the democracy of diamonds — every causally bounded region gets a vote.

4. The shape of a spacetime region determines the character of its quality.

A region that's spatially wide but temporally shallow (like the human brain's effective "diamond" — ~15 cm across, ~100ms–1s deep) has an algebra rich in spatial correlators but poor in temporal ones. By the identity thesis, that is a quality with panoramic spatial experience and fleeting temporal experience. That's exactly what human consciousness feels like: a wide visual field, a detailed soundscape, a distributed body sense — but the present feels instantaneous and the past is gone.

This isn't a coincidence IQT explains after the fact. It's a prediction: change the shape of the diamond, and you change the character of experience. Psilocybin widens the temporal depth (time dilates). Ketamine narrows it (time fragments). DMT flattens the timescale hierarchy (time dissolves). These are testable.

5. Standard physics connects this fundamental theory to the lab.

The identity thesis lives at the level of relativistic quantum field theory, operating at scales far below anything a brain scanner can see. Brains are warm, wet, and classical for practical purposes. Every consciousness theory that invokes fundamental physics faces this gap. IQT bridges it the same way particle physics does: through effective field theory.

The idea is straightforward. A particle physicist doesn't need to solve string theory to predict what a collider will measure. They coarse-grain — zoom out from the fundamental description to a simpler one that keeps only what matters at the relevant energy scale. The Standard Model is an effective theory of whatever the fundamental physics turns out to be, and it works spectacularly at collider energies.

IQT does the same thing for brains. The fundamental algebra (quantum fields on causal sets) gets coarse-grained down to an effective neural-scale algebra. The electrode spacing of an EEG determines spatial resolution. Amplifier bandwidth determines frequency content. Sampling rate determines temporal granularity. These instrumental facts fix the effective algebra — not the theorist's preferences.

The crucial claim: this effective algebra is still a *local algebra* in the formal sense. It inherits the structural properties (isotony, covariance, well-defined restriction) that make the identity thesis, perspectival relativity, and the composition story work. So the entire interpretive framework carries over at brain scale. The metrics in the experimental protocols — persistence, coherence, readout dominance — aren't ad hoc proxies. They're computable functionals of a state on a genuine local algebra, just at coarser resolution.

What prevents circularity is that the coarse-graining is constrained by physics, not fitted to results. The embedding must respect locality (an observable in one region can't secretly map to a different region), preserve algebraic structure (products and adjoints are faithfully represented), and be determined by instrumentation (what the electrodes physically access, not what would make predictions come out right). The pre-registration protocols lock down parcellation, frequency bands, and analysis pipelines before data collection, specifically to block post-hoc adjustment.

Without this bridge, IQT is philosophy. With it, the three experimental protocols are principled tests of a theory formulated in the language of local algebras — not loose analogies between formalism and measurement.

Also, see , at the end of Section 5.

How to Read This Paper

This paper serves multiple audiences. If you’re short on time or approaching from a specific discipline, Appendix A provides tailored reading guides. Here are three recommended paths:

Philosophers: Start with §1 (core thesis and constraint argument), then §4 (subject selection and phenomenological puzzles), then §6 (connections to Russellian monism, IIT, and process philosophy). The formal apparatus of §2–3 can be treated as a black box on first reading.

Neuroscientists: Begin with §5 (empirical protocols and metrics), then read §2.6 (effective-theory bridge from formalism to neural data), then §4 (levels of consciousness and split-brain predictions). The mathematical foundations of §2.0–2.5 and §3 can be consulted as needed.

Physicists and mathematicians: Read §1–3 sequentially for the formal architecture, then §5 for empirical grounding, then §7 for open mathematical problems (presheaf cohomology, boundary data, causal-set QFT).

Skeptic's Map: Where to Disagree

is built in layers. You can reject a later layer while keeping earlier ones. This map locates the precise point of disagreement for different critical positions.

If you reject QI (the identity of intrinsic nature with phenomenal quality): You can still take the constraint argument (§1.1) as a novel characterization of "intrinsic nature" in AQFT. The formalism of §2-3 stands as a contribution to the metaphysics of local physics, independent of consciousness. You are a property dualist; §1.4 addresses what this costs you.

If you accept QI but reject democracy (you want an exclusion principle): You accept that quality is the algebra-state pair but insist only one region per brain bears consciousness. The single change: add an exclusion postulate selecting a maximal complex. You are now doing IIT with a different ontology. Protocol 2 (§5.2) is designed to adjudicate between your position and ours.

If you accept QI and democracy but reject N (the narrative operator as bridge to report): You accept that multiple overlapping diamonds bear quality but doubt that the narrative-operator framework correctly links quality to reportable experience. The empirical predictions of §5 are what fail for you, not the ontology. The bridge hypotheses (§4.4–4.5) are where your disagreement lives, and they are the most revisable component of the theory.

If you accept everything but doubt the effective-theory bridge (§2.6): You think the jump from AQFT to neural-scale metrics is too large. Fair. §7 flags this as an open problem. The empirical protocols test the effective-level predictions; the fundamental-to-effective bridge is a separate research program.

1. The Core Thesis: Identity, Not Supervenience

The “Hard Problem” of consciousness (Chalmers, 1995) arises from the apparent gap between extrinsic descriptions of structure (physics) and intrinsic apprehension of content (experience). IQT closes this gap by asserting an identity thesis:

Intrinsic Quality

is the physical state of a bounded spacetime region.

This is not a claim of correlation or supervenience. It is the stronger claim that phenomenal quality and physical state are the same mathematical object under two descriptions. The term “Intrinsic Quality

” is chosen deliberately: substantial to indicate that it is something rather than nothing, and quality to indicate an intrinsic character determined by physical content.

1.1 Why Identity? A Constraint-Based Argument

The identity thesis is not a definitional stipulation. It follows from a constraint analysis: we ask what properties any candidate for “the intrinsic nature of a physical region” must possess, and show that the local algebra-state pair is the unique object satisfying all of them.

Constraint 1: Completeness. The intrinsic nature of a region must fix everything that is physically true of that region—all expectation values, all dispositional properties, all correlations internal to the region. The restriction satisfies this by construction: it determines the expectation value of every observable in .

Constraint 2: Consistency under nesting (isotony). The intrinsic nature of a subregion must be derivable from the intrinsic nature of any containing region by a well-defined operation. The restriction operation satisfies this: if , then .

Constraint 3: Covariance. The intrinsic nature must transform appropriately under the symmetries of the theory. States on local algebras inherit the covariance of the net.

Constraint 4: Perspectival completeness. The intrinsic nature must be complete relative to the region, without requiring reference to anything outside. Only the restriction to the local algebra satisfies this, since it discards all information about observables localized elsewhere.

Uniqueness claim. Any mathematical object satisfying Constraints 1-4 on a net of algebras is equivalent to the pair of local algebra and restricted state. The point is that these constraints—each independently motivated by what we mean by "intrinsic nature"—converge on a single, already well-understood mathematical object.

This convergence establishes a conditional: if a physical region has an intrinsic nature, then it is

(equivalent to) the local algebra-state pair. What it does not establish is that intrinsic nature has anything to do with phenomenal quality. That identification is IQT's core postulate—the Quality Identity—and it carries the theory's explanatory weight. We isolate it here to make the logical structure transparent:

Quality Identity (QI). The intrinsic nature of a bounded spacetime region —the local algebra-state pair uniquely identified by Constraints 1-4—is phenomenal quality.

QI is not derivable from the constraints. It is a metaphysical postulate, one additional step beyond the constraint argument. But it is not arbitrary. Its justification is abductive: QI is the simplest postulate that, when conjoined with the constraint convergence, generates a unified account of composition (§3), subject selection (§4.4), temporal phenomenology (§2.6), and concrete empirical predictions (§5)—four explanatory burdens that rival positions either leave unaddressed or handle with separate, unconnected mechanisms.

Without QI, the constraint argument delivers an elegant characterization of “intrinsic nature” with no connection to the mind. With QI, the same characterization becomes a theory of consciousness that dissolves the Hard Problem, predicts the structure of phenomenological dissolution under anesthesia, and distinguishes itself empirically from IIT and GNW.

The reader should evaluate QI by its explanatory yield, not by whether it “follows” from the constraints. It does not follow. What follows is that no other candidate is available. QI bets that this is not a coincidence.

Constraint dialectic: rival proposals. Two alternative constraint sets merit explicit rebuttal.

The causal-powers alternative (IIT’s move). IIT identifies consciousness not with the state but with the cause-effect structure—the set of causal constraints a mechanism in a state exerts over its past and future (Oizumi, Albantakis, & Tononi, 2014; Albantakis et al., 2023). This amounts to replacing Constraint 1 with a completeness condition over counterfactual interventions. We reject this on parsimony grounds: the cause-effect structure is derivable from the state on the algebra together with a specified dynamical evolution (e.g., an automorphism group, Hamiltonian, or channel family). If the state already fixes all observables and the dynamics is specified, then causal-influence measures are calculable consequences, not additional ontological facts.

The modal/counterfactual objection. One might argue that intrinsic nature must include not just the actual state but the laws governing its evolution. We partially concede this: the algebra does encode the “space of possibilities” for D, and IQT uses it (quality is defined as a state on an algebra). What IQT denies is that the global dynamical laws or global state are part of D’s intrinsic nature. The local algebra and the local state jointly constitute the intrinsic nature; everything beyond that is extrinsic-to-D.

The Constraint Argument in Plain English

The constraint argument asks: if a region of spacetime has an “intrinsic nature,” what properties must that nature have? Four requirements emerge, each independently motivated by what we mean by “from the inside.”

Completeness: the intrinsic nature must capture everything physically true about the region—every measurement outcome, every correlation. Nesting consistency (isotony): the nature of a smaller region should be derivable from the nature of any containing region, just by restricting your attention.

Covariance: it must transform properly under physical symmetries—rotating or translating the region shouldn’t change its intrinsic character, only its description. Perspectival completeness: it must be self-contained, depending on nothing outside the region.

The algebra-state pair—the mathematical object that assigns expectation values to every observable localized in the region—is the unique object in standard physics satisfying all four. The argument does not prove that intrinsic nature is consciousness. That identification is the separate Quality Identity postulate. What the argument proves is that if you’re looking for intrinsic nature, there is exactly one candidate.

1.2 The Definition of Quality

Definition (Intrinsic Quality) . The quality of a region D is:

under physically induced -isomorphism. That is, two regions and have the same quality if and only if there exists a -isomorphism induced by a symmetry of the net (a spacetime symmetry, causal-set automorphism, or field relabeling that maps to) such that .

Why physically induced isomorphisms only. In continuum AQFT, local von Neumann algebras for bounded regions are typically hyperfinite Type III factors (Haag, 1996). By Connes’ classification, all hyperfinite Type III factors are abstractly isomorphic. If arbitrary -isomorphisms were allowed in the equivalence relation, “same quality” would reduce to “same state up to an arbitrary relabeling of observables”—an overly permissive criterion that could

trivialize phenomenological distinctions between physically different regions.

Restricting to net-induced isomorphisms ensures that “same quality” means “same algebra-state pair up to physically meaningful relabeling”—the kind of identification warranted by an actual symmetry of the physics, not an accidental algebraic coincidence. A visual cortex diamond and an auditory cortex diamond may have abstractly isomorphic algebras (both Type III), but there is no net automorphism mapping one to the other, so their qualities are compared state-by-state on their respective (distinct, physically situated) algebras.

This restriction preserves the intended spirit of the equivalence: quality is insensitive to mere coordinate relabeling but sensitive to genuine physical differences.

What the Definition Actually Says

Two regions have “the same quality” if there is a physical symmetry—a genuine transformation of spacetime—that maps one to the other while preserving the state. The restriction to physically induced isomorphisms is critical.

Without it, the mathematics produces a surprising result: in continuum quantum field theory, almost all local algebras for bounded regions are abstractly isomorphic (they are all “hyperfinite Type III factors”). If any abstract relabeling counted as “same quality,” then a visual cortex region and an auditory cortex region could count as identical—clearly wrong. Restricting to physical symmetries ensures that “same quality” means “genuinely related by a symmetry of nature,” not “accidentally similar in abstract structure.” Quality is insensitive to coordinate relabeling but sensitive to real physical differences.

1.3 The Intrinsicness Principle

Intrinsicness Principle (IQT). Two physical situations are intrinsically identical for a region D if and only if they agree on —the pair up to net-induced * isomorphism. All other differences—including differences in global purification, embedding, and the physics of causally disconnected regions—are extrinsic to D and do not affect D's quality.

This is a substantive metaphysical commitment. It entails that two regions with identical local algebra-state pairs have identical quality even if one is entangled with its environment in a fundamentally different way. The global entanglement structure is real, but it is extrinsic to D—it belongs to the quality of a larger diamond encompassing both D and its entanglement partner.

A critic may object: "Two identical reduced states can arise from radically different global purifications. Shouldn't the intrinsic nature reflect that difference?" IQT's answer is no. The difference between purifications is a fact about the complement of D, not about D itself. If intrinsic nature were sensitive to the complement, it would violate Constraint 4. The Intrinsicness Principle is the formal expression of IQT's core conviction: quality is from the inside, and "the inside" means the local algebra and nothing else.

Worked example: Alice sees her child, Alice sees a stranger. Suppose Alice looks at her daughter, then later looks at a stranger of similar build. A critic objects: "Surely the visual cortex diamond has the same quality in both cases—same retinal geometry, same V1 response—yet the experiences obviously differ. Doesn't this show that quality must depend on context beyond the local state?"

The objection fails at its first premise. The excitations are physically distinct starting at the cornea. Alice's daughter and the stranger present different facial geometries, different skin textures, different micro-expressions. These differences propagate: distinct photoreceptor activation patterns, distinct retinal ganglion cell spike trains, distinct V1 orientation maps, distinct V4 shape and color responses. By the time signals reach even early visual cortex, the algebra-state pair Q(Dvisual) already differs between the two cases. The local quality of the visual cortex diamond is not the same. No appeal to global context is needed.

But suppose we stipulate, for the sake of argument, a case where the visual cortex diamond genuinely has an identical algebra-state pair in two situations—perhaps Alice views two photographs that are pixel-identical but one depicts her daughter (unbeknownst to her visual cortex at that instant). The felt difference between “seeing my child” and “seeing a stranger” in such a case does not reside in the visual cortex diamond. It resides in a larger diamond—one whose algebra includes the correlators binding visual activation to hippocampal memory traces, amygdalar emotional valence, and prefrontal contextual framing. The quality of this larger diamond differs in the two cases because the cross-boundary correlators differ: the visual representation is entangled with different memory and emotional states. The Intrinsicness Principle is not violated. The “feeling of seeing my child” is the quality of the brain-scale diamond, not the visual-cortex-scale diamond. Perspectival relativity tells us both are real; the error is in expecting the smaller diamond’s quality to carry information that belongs, by construction, to the larger one.

1.4 Epistemic Status of the Identity

The logical structure of IQT is:

Constraint argument (§1.1): The local algebra-state pair is the unique candidate for the intrinsic nature of a bounded spacetime region.

Quality Identity (QI): This intrinsic nature is phenomenal quality.

Bridge hypotheses (§4.4–4.5): Self-threads and narrative operators specify which qualities correspond to “experience” and “report.”

Operationalizations (§5.0.1): P_j, K_jl, R, Syn approximate bridge-hypothesis quantities in neural data.

Step 1 is a mathematical result (given the axioms of AQFT). Step 2 is a metaphysical postulate. Steps 3–4 are empirical hypotheses. The protocols of §5 primarily test steps 3–4; step 2 is constrained indirectly by whether the bridge can be made predictive. A critic who accepts step 1 but rejects step 2 is coherent—they are a property dualist, holding that intrinsic nature and phenomenal quality are distinct. But this coherence is purchased at a cost:

What QI buys that property dualism does not. (i) Composition for free: if quality is the algebra-state pair, then composition is the extension problem (§3)—a well-posed mathematical question with known structure. Property dualism requires a separate composition principle for phenomenal properties, with no guidance from physics. (ii) Subject selection dissolved: if quality is universal and self-threads are functional, the boundary between conscious and non-conscious is empirical, not metaphysical (§4.4). Property dualism must explain why phenomenal properties attach to some intrinsic natures and not others. (iii) Temporal phenomenology: the geometry of causal diamonds generates the structure of temporal experience (§2.6). Property dualism must posit a separate mechanism for time-consciousness. (iv) Parsimony: QI introduces no new ontological category. It identifies two apparently distinct categories (physical intrinsic nature, phenomenal quality) as one. Property dualism introduces an additional category and must then explain its relationship to the physical.

The alternative—accepting that the local algebra-state pair is the unique candidate for intrinsic nature but denying that intrinsic nature has anything to do with phenomenal quality—is logically coherent but explanatorily idle. QI is justified not as a derivation but as the most productive interpretation of the constraint convergence.

1.5 Terminological Discipline: Quality, Experience, Report

Quality (Level 0): The intrinsic physical algebra-state pair of any bounded spacetime region. Universal. Every causal diamond has quality. A rock has quality; it does not “experience” anything in the sense relevant to the Hard Problem debate.

Experience (Level 1): The quality of a self-thread—a temporally persistent, causally semi-autonomous quality-stream (formalized in Section 4.4). A mouse has experience. Experience requires temporal coherence and causal autonomy, but not language, reflection, or self-modeling.

Report (Level 2): The subset of experience controlled by a narrative operator—a self-thread with dominant readout access to report channels, capable of recursive self-modeling. Humans have reportable experience. Report generates the everyday sense of "I."

The "so rocks are conscious?" objection conflates Levels 0 and 2. Rocks have quality (Level 0). They do not have self-threads (Level 1) or narrative selves (Level 2). The three levels are ontologically distinct, with each higher level requiring additional structural conditions that most physical systems do not satisfy.

A note on "phenomenal." Throughout this paper, "phenomenal quality" means the intrinsic character of a region—"what it is"—not "what it is like for a subject." The latter requires a subject, which requires Level 1+ structure (temporal persistence, causal autonomy). Level 0 quality is phenomenal in the minimal Nagelian sense: there is something that the region is, intrinsically, whether or not any subject apprehends it. We expect pushback from philosophers who reserve "phenomenal" exclusively for subjective experience. Our usage is deliberate: it signals that quality is not merely structural or dispositional but has intrinsic character—the very character that, at Level 1+, constitutes the "what it is like" of subjective experience. The terminological choice reflects a theoretical commitment, not a conflation.

1.6 Rejection of Emergentism

Quality does not "arise" from complex computation. A Planck-scale spacetime region has quality—the algebra-state pair of that minimal region. We term this a *qualon*, analogous to a phonon: not a new fundamental entity, but a description of what the underlying physical structure is at that scale. The qualon is not an atom of experience; it is an atom of quality.

IQT: Three Levels of Phenomenal Structure Figure 1: IQT Three-Level Hierarchy – Quality > Experience > Report *So rocks are conscious?* conflates Level 0 with Level 2

1.7 The Intrinsic/Extrinsic Flip

Physics describes systems extrinsically—through interaction relations, measurement outcomes, and observational correlates. Phenomenology is the intrinsic existence of those same systems. The "Hard Problem" is an artifact of treating the extrinsic description as exhaustive—of mistaking partial structural descriptions for the complete specification.

Distinction from selfhood. "What it is like to be" (Nagel, 1974) implies subjectivity and self-modeling. distinguishes this from "what it is"—the bare qualitative character of a region. An apple has intrinsic Quality ; it does not have a self-model.

Interpreting Section 1

2. Formalism: Quality as Local State

2.0 Status of the Formalism

Standard AQFT (established). The Haag-Kastler axioms define a net of -algebras on regions of Minkowski spacetime satisfying isotony, microcausality, and covariance. State restrictions are well-defined.

AQFT on causal sets (conjectural but active). The extension to discrete spacetime is an active research program (Sorkin, 2003; Johnston, 2008; Surya, 2019; Dowker & Glaser, 2013). A fully rigorous construction of the net satisfying all Haag-Kastler analogues is not yet established.

Specific assumptions. 1. Net existence: satisfying isotony and microcausality.

2. Split property (approximate): sufficiently separated regions approximately tensor-factorize.

3. Faithful state: is faithful and normal on the quasi-local algebra.

Independence from causal sets. The philosophical framework does not require causal sets specifically. Any theory with local algebras and state restrictions supports the same interpretive structure.

2.1 The Geometric Substrate

Let be a locally finite partially ordered set (a causal set). A region is a causal diamond:

The category of regions. Regions are causal diamonds and their finite unions under causal completion. When , we mean the causal completion: the smallest diamond containing both.

2.2 The Algebraic Definition

To each diamond D, associate a local algebra . Quality is the pair up to net-induced *-isomorphism (Section 1.2).

Relation to reduced states. In finite-dimensional QM with tensor-product Hilbert space, corresponds to the reduced density matrix . In AQFT, the algebraic restriction is the more general operation. For the neuroscience applications, the reduced-state picture is adequate.

2.3 Key Properties

Isotony. , and

Microcausality. Spacelike-separated algebras commute. Commutativity does not imply statistical independence: can entangle commuting algebras.

Covariance. Symmetric physical situations have identical quality under the mapping.

2.4 A Finite-Dimensional Toy Model

Setup.

Three qubits A, B, C. Local algebras: , etc.

State 1 (GHZ): . State 2 (mixture): .

Isotony: is derivable from by further restriction.

Same local qualities, different global quality: Both states yield . Yet differs: tripartite coherence is present in the GHZ state and absent in the mixture.

Intrinsicness Principle applied: Qubit A has the same quality in both situations (same algebra, same state). The global entanglement difference is extrinsic to A.

2.5 Perspectival Relativity of Quality

The most distinctive commitment of is the perspectival relativity of quality: intrinsic content is relative to the region whose algebra defines the restriction. There is no “view from nowhere.”

Consider nested diamonds: neuron-scale cortical-column-scale whole-brain . The qualities are consistent (by isotony) but not identical. contains structure ontologically absent from —the correlations between neuron-scale sub-diamonds. This is not coarse-graining: the larger algebra supports observables that simply do not belong to any sub-algebra.

Box 2.A — Perspectival relativity: analogy and disanalogy

Analogy (special relativity). No privileged reference frame; length is relative to frame. no “true quality”—only qualities-relative-to-diamonds.

Disanalogy (no Lorentz invariants). In SR, frame-independent invariants exist. In the “invariant” is not a shared quality but the consistency condition: local qualities form a coherent presheaf on the poset of diamonds. Coherence, not content, is absolute.

No global subject claim. The global state is a mathematical bookkeeping device, not a quality anyone experiences. No bounded region contains all of spacetime, so there is no “cosmic perspective” from which the global state is a quality. The universe as a whole does not have a point of view—only bounded regions do. This rules out a single cosmic consciousness without needing an exclusion postulate.

Finite multiplicity. Democracy does not mean an uncountable infinity of overlapping quality-bearing regions. Because spacetime is discrete (a causal set with finitely many elements in any bounded region), the number of sub-diamonds is combinatorially large but finite. At the neural scale, the number of meaningfully distinct effective diamonds is further constrained by the parcellation of neural tissue into functional units. Many simultaneous quality-bearing regions exist—but the multiplicity is quantized, not continuous.

Scale-dependent phenomenology. As diamond scale increases, qualitatively novel dimensions of content become available—larger algebras support cross-regional correlators absent from sub-algebras. Perceptual binding transitions may correspond to thresholds where specific correlator families become sufficiently strong to dominate the larger diamond’s quality. We do not claim mathematical discontinuity—states vary continuously—but the space of phenomenological distinctions expands at larger scales.

Substrate dependence. Two systems—biological and silicon—could have identical qualities at one scale while diverging at another. The substrate question becomes scale-dependent: same at which level of restriction?

2.6 The Effective-Theory Framework: From Fundamental to Neural Scales

The formalism of §2.0–2.5 defines quality in terms of -algebras and state restrictions on causal sets. The empirical program of §5 measures EEG voltages, phase-locking indices, and transfer entropies. This section bridges the two levels.

The effective-theory bridge. The fundamental local algebra for a brain-scale diamond contains degrees of freedom at every scale from the Planck length upward—far more than any measurement can access. Standard effective-field-theory reasoning (Wilson, 1975; Polchinski, 1984) motivates identifying a tractable subalgebra of effective observables. Let be the effective neural algebra—a finite-dimensional -algebra whose generators correspond to mesoscopic neural quantities: local field potentials, population firing rates, and their spatiotemporal correlators. The bridge between fundamental and effective scales is an embedding:

1:

an injective *-homomorphism (subalgebra inclusion) identifying each effective observable with its fundamental counterpart—the specific combination of sub-neural degrees of freedom that the mesoscopic quantity physically aggregates. This is a genuine algebraic embedding: is multiplicative (), preserves the* -operation, and is injective, so is realized as a subalgebra of . The effective state is defined by restriction:

That is, the expectation value of any effective observable equals the expectation value of its fundamental image under . Coarse-graining in the Schrödinger picture is handled by the conditional expectation : , the unique completely positive, identity-preserving, normal projection onto the subalgebra (in the finite-dimensional case, is a CPTP map on density matrices). The two pictures are equivalent; we work primarily with because the Heisenberg-picture formulation makes the identity-theoretic content explicit: effective observables are fundamental observables, not approximations to them.

Effective quality. The effective quality of a diamond D at neural scale is:

By the coarse-graining monotonicity property (P3, §3.3), is a lower bound on the structure present in : coarse-graining can only lose correlational information, never create it. The effective quality is therefore a conservative approximation of the full quality—sufficient for empirical contact, even though it discards sub-neural detail.

Effective causal diamonds. At the neural scale, the relevant "regions" are not individual causal diamonds on the fundamental causal set but effective diamonds—bounded spacetime regions whose spatial extent corresponds to a neural population (cortical column, Brodmann area, or distributed network) and whose temporal extent corresponds to an integration window (tens to hundreds of milliseconds). The effective algebra for such a region is generated by the coarse-grained observables localized within that spatiotemporal volume.

The key structural fact is that is still a local algebra in the AQFT sense—it satisfies isotony (nesting of effective diamonds inherits from nesting of fundamental ones) and approximate microcausality (effective observables localized in spacelike-separated neural populations approximately commute at the timescales relevant to consciousness). The effective theory therefore supports the same interpretive framework as the fundamental theory: quality is defined as a state on a local algebra, and all the properties of §2.3—including perspectival relativity—carry over.

Constraints on the embedding. A natural objection: if is chosen to fit the data, the theory is circular—it predicts whatever the theorist arranges it to predict. The objection has force against an unconstrained . But is not unconstrained. Three requirements fix its structure:

Physical locality. The image of must respect the net structure: an effective observable localized in region D must map to a fundamental observable localized in D. The embedding cannot identify an effective observable in one region with a fundamental observable in a disjoint region. This is not a convention—it is inherited from isotony.

Algebraic homomorphism. must be an injective, unital *-homomorphism —i.e., a genuine subalgebra embedding. This is a stronger requirement than mere complete positivity: it ensures that preserves products () and adjoints (), so the effective algebra inherits its algebraic structure faithfully from the fundamental algebra. The coarse-graining map (conditional expectation onto) is then automatically completely positive and identity-preserving; in the finite-dimensional picture, it is CPTP on density matrices. Arbitrary linear maps are not admissible.

Instrumental determination. The generators of —the effective observables—are fixed by the physical characteristics of the measurement apparatus: electrode spacing determines spatial resolution, amplifier bandwidth determines the frequency content, and sampling rate determines temporal granularity. The theorist chooses the parcellation scheme (which electrodes form a “module”), but the observables within each parcel are determined by what the instrument physically accesses, not by what would make the theory’s predictions come out right.

The situation is analogous to collider physics. A particle physicist’s detector geometry—calorimeter segmentation, tracker resolution, trigger thresholds—determines the effective field theory used to interpret the data. Nobody objects that this makes QCD circular. The detector constrains the effective description; the effective description generates predictions; the predictions are tested against the data the detector collects. The same logic applies here: the EEG/iEEG apparatus constrains , the theory generates predictions for functionals of (the metrics P, K, R of §5.0.1), and the predictions are tested against the recorded signals. (The analogy is structural—the same logic of effective description—not a claim that enjoys the same evidential standing as QCD. The fundamental theory here is conjectural; the effective-level predictions are what the experiments test.)

What would make the framework circular is post-hoc adjustment of ι to accommodate disconfirming results—changing the parcellation, bandwidth, or temporal window after seeing that the predicted metric trajectory fails. The pre-registration protocols of §5 are designed to block precisely this move: parcellation, frequency bands, integration windows, and statistical thresholds are fixed before data collection.

Perspectival Relativity: Nested Causal Diamonds

Figure 2: Nested Causal Diamonds with Perspectival Relativity (section 2.5)

Why the quantum formalism when neural scales are classical? A reasonable skeptic will ask: if the effective neural algebra is finite-dimensional and, for practical purposes, classical, why invoke AQFT at all? Why not simply use classical information theory? The answer is structural, not empirical. The quantum algebraic framework provides the scaffolding that ensures the effective theory inherits the correct compositional properties from the fundamental theory—isotony, covariance, the well-defined restriction operation that grounds perspectival relativity. Classical information theory can compute the metrics of §5, but it cannot explain why those metrics have the interpretive significance claims for them: that significance flows from the effective algebra being a genuine local algebra in the AQFT sense, not merely a convenient statistical model. The empirical predictions of §5 do not depend on quantum coherence at neural scales. The formalism is not a claim about quantum brains; it is a guarantee that the effective-theory bridge respects the same structural principles as the fundamental theory it coarse-grains.

Temporal phenomenology from diamond geometry. The geometry of an effective causal diamond—its spatial extent and temporal depth—determines which correlators belong to its algebra. A diamond with temporal depth supports observables involving temporal correlations up to , but not beyond. The quality of such a diamond therefore includes temporal structure up to scale and is silent about longer-timescale structure.

This generates temporal phenomenology without additional postulates. The "specious present"—the experienced duration of "now"—corresponds to the temporal depth of the effective diamond whose quality constitutes the self-thread at that moment. Wider diamonds (longer) yield an expanded present; narrower diamonds yield a compressed, fragmented present. The asymmetry between panoramic spatial experience and fleeting temporal experience follows from the geometry: the brain's effective diamond is spatially wide (encompassing distributed cortical populations) but temporally shallow (integration windows of hundreds of milliseconds, not hours).

Crucially, compounds that alter effective integration timescales—by modulating recurrent cortical coupling, synaptic time constants, or network gain—change the geometry of the effective diamond and thereby change temporal phenomenology. This is the mechanism underlying Protocol 3 (§5.3): psilocybin widens the effective diamond's temporal depth (time dilation), ketamine narrows it (time fragmentation), and DMT flattens the timescale hierarchy (time dissolution).

The shape-phenomenology mapping. The temporal phenomenology discussed above—how compounds shift the experienced "now"—is a special case of a more general claim: the shape of an effective diamond determines the character of the quality it bears. A self-thread (§4.4) is a directed family of overlapping effective diamonds . At each update, the narrative operator (§4.5) compresses the current diamond's quality into an internal model . Because the temporal integration window is finite (ms–1 s for human narrators), captures only a thin temporal slice of the self-thread's worldtube. The next update, , captures a different slice—one that partially overlaps the previous but has shed its earliest content and acquired new content at the leading edge. This sequential shedding-and-acquiring is what temporal passage is, phenomenologically. It is not that reality moves; it is that the narrator's window slides. The "flow" of time is the content generated by a finite-bandwidth system tracking its own trajectory through a four-dimensional quality landscape.

Directionality versus passage. A clarification is needed about the relationship between the narrator's directionality and the phenomenology of passage, because these are distinct.

The narrative operator NNN does not first exist and then get "pointed forward" by thermodynamic constraints. Narration is computation—compression, prediction, sufficient-statistic formation over a self-thread—and at the scales where narrators operate this is a physically sustained process, not an abstract mapping. But it is a mistake to infer from this that *all* computation is constitutively irreversible. In principle, microdynamics can be time-reversal invariant, and logically reversible computation can be made arbitrarily low-dissipation in the quasi-static limit. What forces an arrow at macroscopic scales is not “computation in general,” but *bounded, persistent computation*: a narrator is a finite-bandwidth system that must continually update a bounded internal model while remaining coupled to action and report.

That boundedness matters. A narrator cannot retain an ever-growing full record of its own worldtube. It must overwrite, discard, compress, and commit to a lossy sufficient statistic. Those are logically irreversible steps, and Landauer ties them to entropy export. Sustained narration therefore presupposes an entropy sink and a nonequilibrium gradient—conditions under which reliable records, error correction, and stable model-updating are physically available. In a universe with a low-entropy past and higher-entropy future (ours), these conditions obtain in the thermodynamic-forward direction. A “backward narrator” is not a logical contradiction; it would require the opposite boundary condition (a low-entropy future) together with exquisitely tuned correlations that make record-formation and control stable in the reverse direction. For the class of systems that instantiate narration as a robust macroscopic process, the thermodynamic arrow is constitutive: it is part of what makes narration physically realizable, not merely a preference layered on top.

The felt character of passage, however, is not produced by this directionality. Even fixing the arrow of narration, passage requires finite bandwidth. An idealized negligible-loss narrator—one that could hold the entire forward trajectory of its self-thread “at once,” without continually overwriting and re-encoding—would still be an arrowed physical process, but it would not generate a moving-present texture. The passage-feeling comes from finite bandwidth: because the temporal integration window τ\tauτ is finite (hundreds of milliseconds to ~\sim~1s for human narrators), NNN captures only a thin temporal slice of the self-thread’s worldtube. The next update captures a different slice—partially overlapping the previous, but having shed content at the trailing edge and acquired content at the leading edge. This sequential shedding-and-acquiring is what temporal passage is, phenomenologically. It is not that reality moves; it is that the narrator's bounded update does.

Two features of temporal experience are therefore perspectival in different ways. The direction of felt time is constituted by the physical preconditions for narration (bounded, persistent computation in nonequilibrium conditions). The texture of felt passage—the sense that the present is moving—is generated by finite narrative bandwidth (lossy update over a finite τ\tauτ). Neither feature is an additional dynamical ingredient in the diamond itself. (See §7 for the further stratification involving time-reversal symmetry and CP-violation at the fundamental level.)

Figure 5: Three Layers of Temporal Directionality

The human self-thread's effective diamond is spatially wide relative to its temporal depth. Its spatial extent is roughly the diameter of the brain (cm), determined by the reach of cortico-cortical connectivity. Its temporal extent is roughly (ms–1 s), determined by the narrative operator's integration window. This geometric asymmetry has a direct algebraic consequence. The effective algebra of a spatially wide, temporally thin diamond is rich in spatial correlators—cross-regional coherence patterns, long-range binding, the full repertoire of simultaneous multimodal integration—but comparatively poor in temporal correlators. The algebra contains few observables relating the state at the diamond's early boundary to the state at its late boundary, because the diamond does not extend far enough in time for complex temporal structure to develop within it.

By QI, this algebraic asymmetry is the phenomenology of "panoramic space, fleeting time." Human experience is spatially rich—we perceive a wide, simultaneous visual field, a three-dimensional auditory scene, a distributed body schema—because the effective diamond's algebra supports extensive spatial correlators. Human experience is temporally thin—the present feels instantaneous, the past is gone, the future has not arrived—because the algebra supports only limited temporal correlators. The felt asymmetry between space and time in experience is not a mystery requiring separate explanation; it is a direct consequence of the diamond's shape under the identity thesis.

Different organisms have different effective causal speeds and different integration windows,

Figure 3: The Effective-Theory Bridge – from fundamental AQFT to empirical neural metrics (section 2.6)

yielding differently shaped effective diamonds and therefore qualitatively different temporal phenomenology—not merely different "amounts" of consciousness but different kinds of temporal experience. A hypothetical system that was temporally extended but spatially thin would have the reverse asymmetry: rich temporal structure, impoverished spatial structure. It would not experience "time flowing" in our sense at all. The shape-phenomenology mapping thus generates a space of possible phenomenologies indexed by diamond geometry, within which human experience occupies a specific region.

Box 2.B — The shape-phenomenology mapping: scope and limits

What the mapping claims. The algebraic content of the effective diamond's quality—specifically, which families of correlators the algebra supports and how the state populates them—determines the character of spatial and temporal phenomenology. No additional "binding force," "temporal flow mechanism," or "stream of consciousness" is needed beyond the identity thesis applied to the diamond's algebra-state pair. Panoramic spatial experience, fleeting temporal experience, and the sense of passage all follow from the shape of the diamond and the finite bandwidth of the narrative operator.

What the mapping does not claim. The mapping does not claim that time is unreal, that passage is illusory, or that causal order is phenomenologically inert. The partial order on the causal set is structurally present in the formalism and does real work: it determines which diamonds exist, which nesting relations hold, and which correlators the algebra contains. Time is structurally special (via causal order) but not additionally dynamic (no moving present, no ontological becoming). The felt dynamism of passage is perspectival—generated by the narrator's finite-bandwidth tracking, not by reality's furniture. The narrator's directionality is itself constituted by the thermodynamic arrow: computation at macroscopic scales is irreversible, so narration runs forward not by choice but by constitution. The passage-feeling is a further consequence of finite bandwidth within that forward-running process (§2.6, "Directionality versus passage").

The circularity pressure point. A critic asks: "You've shown that a spatially wide diamond has rich spatial correlators. But why does that feel like panoramic space rather than buzzing confusion?" This is the Hard Problem reasserting itself within IQT. The answer: rich spatial correlators don't cause panoramic phenomenology—they are panoramic phenomenology, by the identity thesis. This is where the framework's bet is most naked. §2.7 addresses it head-on.

2.7 The Hardest Objection

We flag the circularity pressure point above as IQT's strongest dialectical vulnerability, and address it directly rather than in passing.

The objection, stated at full strength: "The identity thesis is doing all the work. You show that the algebra-state pair has rich structure, then declare that this structure is phenomenology. But the question was never whether the structure exists—it was why structure feels like anything at all. Your answer ('it just is, by identity') is either a tautology or a refusal to engage."

Our response has three parts.

First, yes, at some point QI is a metaphysical posit. We do not apologize for this. Every theory of consciousness contains a metaphysical posit at its core—IIT posits that integrated information is consciousness; GNW posits that global broadcast is conscious access; higher-order theories posit that meta-representation is awareness. The question is not whether a theory contains a posit but whether the posit earns its keep. QI earns its keep through the bridge structure: it generates composition for free (§3), dissolves subject selection (§4.4), predicts temporal phenomenology from diamond geometry (§2.6), and yields three pre-registerable empirical protocols (§5) with defined failure conditions. No other identification of phenomenal quality with a mathematical object currently delivers this explanatory yield.

Second, the objection presupposes that there should be an explanation of why structure feels like something—an explanatory demand that, if taken seriously, generates an infinite regress. Any candidate explanation ("structure feels like something because of X") invites the further question "why does X make it feel like something?" The identity thesis terminates this regress by denying the premise: there is no gap between structure and feeling to be bridged. The algebra-state pair does not cause phenomenology and phenomenology does not emerge from the algebra-state pair. They are the same thing described in two vocabularies—one extrinsic (physics), one intrinsic (phenomenology).

Third, the posit is not empty because it is falsifiable at the bridge level. If QI is correct and the bridge hypotheses of §4.4–4.5 are approximately right, then the empirical predictions of §5 follow. If those predictions fail, QI is not directly refuted (the bridge might be wrong), but its explanatory yield—the very thing that justifies it—is undermined. A QI that cannot generate successful bridge predictions is a QI that has stopped earning its keep. The theory is designed so that the metaphysical posit has empirical consequences, even though the consequences are mediated by revisable bridge hypotheses.

We state this openly because we believe that honest engagement with the hardest objection builds more trust than burying it in a box. Critics who accept the formal structure but resist QI are

Figure 4: Shape-Phenomenology Mapping – diamond geometry determines experience character (section 2.7)

coherent; we ask only that they specify what alternative identification of phenomenal quality they propose, and whether it delivers comparable explanatory yield.

Connection to Protocol 3. The shape-phenomenology mapping generates the specific predictions of §5.3. Compounds that widen the effective diamond's temporal depth (psilocybin: enhanced recurrent coupling larger) shift the algebra toward richer temporal correlators, expanding the experienced "now." Compounds that narrow it (ketamine: disrupted recurrent coupling smaller) impoverish temporal correlators, fragmenting temporal experience. The peak position ($\S 5.0.1$) is the empirical proxy for diamond shape, and its compound-specific shifts are the testable consequence of the mapping.

The bridge to empirical metrics. The EEG/iEEG signals measured in §5 are observables in _eff(D). The metrics defined in §5.0.1 are computable functionals of the effective state _eff:

measures the temporal persistence and dimensionality of the effective state within module j—approximating the self-thread persistence of the effective diamond.

measures the phase-coherence between modules—approximating the cross-boundary correlator strength in the effective quality of a diamond spanning both modules.

measures directed information flow—approximating the readout dominance of the narrative self-thread.

This bridge is held together by the same logic that connects any fundamental theory to its experimental tests: through a chain of effective descriptions, each justified by scale separation. The bridge is not a derivation from AQFT axioms to electrode voltages—no theory of consciousness currently achieves this. What it provides is a principled account of why neural-scale operationalizations are legitimate tests of a theory formulated in the language of local algebras: because the effective algebra is a local algebra, at the appropriate scale.

What the effective-theory framework does not provide. It does not derive the specific embedding from first principles—the constraints above fix the form of , not its detailed content. It does not specify which neural observables enter the effective algebra—this is determined by instrumentation and the parcellation choices of §5.0.1. And it does not close the gap between effective and fundamental quality: the relationship between Q_eff(D) and Q(D)—how much phenomenological structure is lost in coarse-graining—is an open problem (§7, “The coarse-graining bridge”). The effective-theory framework is a bridge, not a derivation. But it is a principled bridge, grounded in standard physics methodology and constrained by physical locality, quantum consistency, and instrumental reality—not an ad hoc identification of formalism with measurement.

The Effective-Theory Bridge in Plain Terms

The move from fundamental algebras to neural-scale observables is not a weakness or a hack—it is standard physics methodology. Every successful physical theory connects to experiment through effective descriptions: particle physics uses effective field theories, condensed matter uses renormalization group methods, and fluid dynamics emerges from molecular interactions. The embedding plays exactly this role for

Three constraints discipline the bridge. Physical locality: the effective algebra respects the causal structure of the underlying spacetime—an observable localized in one brain region cannot be secretly mapped to a different region. Quantum consistency: the coarse-graining operation can only lose information, never create correlations that were not present at the fundamental level. Instrumental reality: the generators of the effective algebra correspond to quantities that neuroscientific instruments actually measure—electrode voltages, firing rates, phase relationships.

The crucial payoff: because the effective algebra is still a local algebra (it satisfies the same structural axioms as the fundamental one), the full interpretive framework carries over. Perspectival relativity, composition via correlation structure, unity via mutual information—all of these apply at neural scale, not by analogy, but by mathematical inheritance. The effective theory is not a metaphor for the fundamental theory; it is a principled coarse-graining of it.

Interpreting Section 2

This section gives the theory its mathematical bones. A causal diamond is the set of all spacetime points causally between events and . To each diamond, assigns a local algebra (the space of possible observations) and a state (what's actually happening). The quality of the diamond is this algebra-state pair.

The toy model (three qubits) makes it concrete: two states can look identical locally but differ globally when you include correlations. Quality lives at every scale, and each scale reveals structure invisible to smaller scales.

Perspectival relativity (Box 2.A) is the key insight: there's no "true" quality, only quality relative to a diamond. Like reference frames in relativity, but without frame-independent invariants—only coherence across scales.

§2.6 (the effective-theory framework) bridges the gap between abstract AQFT and neural

measurements. The effective algebra is a coarse-grained version of the fundamental one, retaining mesoscopic neural observables. Crucially, this effective algebra is still a local algebra—so the full interpretive framework (including perspectival relativity) applies at neural scale.

Temporal phenomenology falls out of diamond geometry: the temporal depth of an effective diamond determines the "specious present." This is what Protocol 3 tests.

3. The Composition Problem: Holism via Correlation Structure

3.1 The Mechanism

Let be diamonds with causal completion . The restrictions and do not determine . The missing information is the pattern of mixed correlators—

for . The toy model (Section 2.4) illustrates: GHZ and mixture agree on all marginals but differ on the three-body state.

3.2 The Composition Theorem (Sketch)

Extension set:

Ext is generally not a singleton. Distinct extensions correspond to distinct correlation structures. A key open problem is characterizing the minimal composition datum such that .

The Composition Problem in Plain English

If you know the quality of two separate regions, do you know the quality of the region containing both? No. The missing ingredient is correlation structure—how the parts are entangled or correlated with each other.

The toy model from §2.4 illustrates: two quantum states (GHZ and mixture) look identical at every single-qubit level but differ globally because of different three-body correlations. Many possible global states are compatible with the same local information—knowing the parts underdetermines the whole. IQT's answer to the composition problem: the "glue" binding parts into wholes is correlation structure, not a special binding force or an additional postulate.

3.3 The Unity Functional

The claim "correlators are the glue" is only a problem statement without a measure of how much a diamond's quality exceeds the sum of its parts. We define such a measure.

Definition. For a diamond D and a bipartition with D the causal completion of , the correlation excess is the quantum mutual information:

where S is the von Neumann entropy. The unity of D is the minimum over all bipartitions:

Scope. The definition uses von Neumann entropy, which is well-defined for the effective finite-dimensional algebra -eff(D) and for the finite-dimensional toy models of §2.4. At the fundamental level, where local algebras are Type III (§1.2), the von Neumann entropy of restrictions is not directly available (there is no trace, and naive entropy diverges). A fundamental-level unity functional would require modular-theoretic tools—e.g., Araki relative entropy S(ω_AB | ω_A ⊗ ω_B), which is well-defined under the split property and recovers quantum mutual information in the finite-dimensional limit. Constructing such a functional rigorously is an open mathematical problem (§3.4). All empirical predictions of §5 use only the effective-scale U, where the definition is unproblematic.

Properties.

(P1) Vanishes for product states. If for some bipartition, then U(D; so . A diamond whose parts are uncorrelated across every bipartition has zero unity.

(P2) Increases with cross-boundary dependence. is the total (classical + quantum) correlation between and in the state . Adding entanglement or classical correlations across the boundary strictly increases U.

(P3) Coarse-graining monotonicity. Under restriction to the effective subalgebra (§2.6), the

mutual information is non-increasing: . This follows from the data-processing inequality applied to the conditional expectation . For a fixed bipartition , when acts locally on each partition factor, coarse-graining can only lose correlational structure, never create it. The effective-scale unity is a lower bound on the fundamental unity. (When the parcellation itself changes across scales, the comparison requires matching bipartitions via ; the claim is that the effective-scale mutual information for the image bipartition is bounded by the fundamental-scale mutual information for the original bipartition.)

(P4) Compatible with democracy. is defined for every diamond. Multiple overlapping diamonds can simultaneously have high U. There is no exclusion step: we do not select the diamond with maximal U and declare all others non-existent. Every diamond has both a quality and a unity . The unity quantifies how much that diamond's quality exceeds the quality of its parts; it does not determine whether the diamond "counts."

Distinction from . is structurally similar to IIT's both measure irreducibility of a system over its minimum information partition. The differences are: (i) U is defined on the state, not on the cause-effect structure (transition probability matrix). U measures how much the actual state correlates across parts; measures how much the causal repertoire is irreducible. (ii) U does not trigger exclusion. In IIT, the system with maximal is the only conscious one; in every diamond has and . (iii) The relationship between U and is an open mathematical question. For product states both vanish; for maximally entangled states both are high. Whether they can diverge substantially is worth investigating.

Toy model values. For the three-qubit system of §2.4, consider the bipartition A|BC. GHZ state (pure on ABC): , , , so bits. The same value obtains for any single-qubit split; hence bits. Mixture : , , , so bit; bit. Product state : .

The GHZ and mixture have identical local qualities (§2.4) but different unity—quantum coherence in the GHZ contributes additional cross-boundary correlation beyond the classical mixture. U tracks total (classical + quantum) correlation and is sensitive to the pattern of that correlation across bipartitions. This is exactly right: unity tells you how much a diamond's quality exceeds its parts' qualities, not what the quality is. Two states can share all single-site qualities yet differ in compositional structure; U quantifies that difference.

Neural operationalization. At the effective neural scale, von Neumann entropy becomes Shannon entropy on coarse-grained variables, and U becomes the minimum mutual information over parcellations. The cross-module coherence metric (§5.0.1) serves as an operationally tractable proxy. The dwPLI-based captures one family of cross-boundary dependencies (phase coupling); a complete MI estimate would additionally require amplitude and cross-frequency terms. We use dwPLI because it is robust to volume conduction and computationally tractable, and because it provides a conservative lower bound on total correlation—consistent with the coarse-graining monotonicity of P3. The connection: high K across all module pairs high U for the brain-scale diamond the brain-scale quality is highly unified. Anesthesia reducing K (Protocol 1) corresponds to reducing U, fragmenting the brain-scale diamond's unity while leaving local qualities intact.

The Unity Functional in Plain English

measures how much a region's quality exceeds the sum of its parts—essentially, how tightly bound together the region is. It works by finding the weakest link: the way of splitting the region in two that produces the least mutual information. If even the weakest-link split still shows strong correlations, the region is highly unified.

U is zero for regions whose parts are completely independent, and increases as cross-boundary correlations strengthen. It is analogous to IIT's but without the exclusion postulate: every region gets a U score, and multiple overlapping regions can all be highly unified simultaneously. At neural scales, U becomes ordinary mutual information computed on coarse-grained brain signals, with the cross-module coherence metric K from §5.0.1 as its empirical proxy.

3.4 What Remains Open

quantifies unity but does not solve the full composition problem. The full problem is: given and , characterize the extension set and the composition datum that selects a specific extension. This is a well-posed mathematical problem with known partial results in finite dimensions (quantum marginal problem; Klyachko, 2004; Schilling et al., 2013) and open questions in the algebraic setting.

What U provides is a diagnostic: it tells you how much compositional structure a diamond has, and it does so with a measure that (a) vanishes when composition is trivial, (b) is compatible with democracy, (c) connects to empirical metrics, and (d) does not require solving the full extension problem. It is not a replacement for the full composition theory. But it is substantially more than “correlators are the glue.”

3.5 Phenomenological Translation

The quality of a brain-scale self-thread is the quality of a large diamond whose “new” content consists of cross-boundary correlators binding local activations. The unity U of this diamond quantifies how much this binding contributes beyond the parts. Altered states transform global quality without changing local quality by altering correlation structure—and correspondingly altering U—while leaving local activations intact.

Interpreting Section 3

The composition problem: how do parts combine into a unified whole? IQT’s answer: through correlations. The quality of a composite region includes cross-boundary correlators that are absent from any sub-region’s quality. These correlators are the “glue.”

The unity functional measures how much a diamond’s quality exceeds the sum of its parts. It’s similar to IIT’s , but without exclusion—every diamond gets a U score, and multiple diamonds can be highly unified simultaneously.

Neural translation: high K_jl (phase coherence between brain modules) means high unity of the brain-scale diamond. Anesthesia reduces K, fragmenting unity while leaving local qualities intact.

4. Resolving Classic Problems

4.1 Infinite Regress

eliminates the observer-of-the-observer regress. Every observer is a bounded region with quality determined by its own algebra-state pair. The regress terminates at the qualon.

4.2 Mary's Room

Before leaving the room, Mary knows everything expressible as expectation values for observables A in the algebras of other people's visual cortex diamonds—she has the complete extrinsic description. What she lacks is not information but instantiation: no diamond in her own brain has ever had an algebra-state pair in the equivalence class . When she sees red, a new causal diamond D is created in her visual cortex whose quality . She gains a new quality—a new element in her own presheaf of local states—not a new proposition about the world.

The distinction from standard identity-theory responses is formal: Mary's gain is the instantiation of a specific equivalence class of algebra-state pairs in her cortex, not merely "a new way of knowing the same fact." The equivalence class Q_red is a mathematically precise object; its instantiation in Mary's brain is a physical event with a definite location, duration, and causal structure.

Anticipated counter: "new mode of representation, not new instantiation." A critic may argue that Mary gains a new representational format for facts she already possessed—indexical, demonstrative, or phenomenal knowledge of the same physical facts. IQT's response: the distinction between "new representation of old fact" and "new instantiation of quality" is precisely what IQT dissolves. The representational format is the quality. A visual-cortex algebra-state pair in class Q_red is not a representation of redness—it is redness, at the scale of that diamond. The critic's move tacitly assumes that quality is something over and above the physical state that a representation can "point to" more or less directly. Under QI, there is no such separation. Mary's new representational capacity and her new quality are the same event described in two vocabularies.

4.3 Substrate Dependence

As an identity thesis, IQT implies distinct substrates typically instantiate distinct qualities, even under similar functional organization. This is a metaphysical corollary; empirical leverage focuses on within-substrate manipulations (Section 5).

4.4 The Subject Selection Problem: Dissolution, Not Solution

The subject selection problem conflates which regions have quality (all) and which regions constitute a self (a functional question).

Eliminativism objection. is not eliminativism. It affirms that quality exists universally and

experience (Level 1) exists wherever self-threads persist. What denies is that the unity reported by the narrative self is a metaphysical fact rather than a content of the narrative self-model. The experienced unity is real as phenomenology—an artifact as metaphysics.

Three-level analysis:

Level 0: Quality (universal). Every causal diamond D possesses quality . No exclusion principle. By perspectival relativity, each diamond's quality is equally real.

The democracy of diamonds is not a multiplication of subjects. It is the elimination of the need to select one. The alternative—drawing a principled boundary around the "real" conscious region—requires an exclusion principle. Exclusion principles require justification: why this region and not that one? IIT answers with maximal irreducibility; 's response is that the question is malformed. All diamonds have quality (Level 0). The interesting phenomenological question is which diamonds sustain self-threads (Level 1) and which self-threads control report (Level 2)—and those are functional questions with empirical answers, not metaphysical ones requiring a priori exclusion criteria.

Level 1: Self-threads (persistence + autonomy). A self-thread is a directed family whose qualities maintain coherence across successive temporal slices:

sustained across timescales much longer than the individual diamond's duration. A self-thread also exhibits causal autonomy: conditioned on its boundary, the interior's future is approximately independent of the exterior (the formal analogue of a Markov blanket).

Level 2: Narrative self (readout dominance + recursion). The self-thread with maximal causal control over report channels, plus recursive self-modeling. The sense of being a single subject is the content of this thread's self-model, not an independent metaphysical fact.

4.5 The Narrative Operator: From Quality to Report

Definition. N maps evolving restrictions along a self-thread to an internal generative model:

where contains self-indexed variables for prediction and control.

Implementation class: control-theoretic sufficient statistic. N extracts the minimal representation sufficient to: (i) predict sensory inflow, (ii) select persistence-maintaining actions, (iii) distinguish self from environment.

In active inference terms (Friston, 2010; Friston et al., 2017), N is the recognition model of a free-energy-minimizing agent whose generative model includes a self-variable.

Formal properties: Dimensionality reduction (raw quality trajectory). Temporal integration over ("specious present"). Report coupling ; Level 2 requires $r(t) > r\_threshold$. Recursive self-reference: contains a representation of itself.

Status of N as a theoretical schema. The definition above specifies what N must do (predict, select, distinguish) and what formal properties it must have (dimensionality reduction, temporal integration, report coupling, recursion). It does not specify how it does so in the human brain. The worked pipeline of §4.5.1 demonstrates that N has definite mathematical structure, but it is a toy: six qubits are not a cortex. The specific implementation of N—the particular recurrent architecture, the precise coarse-graining map, the dynamics of the belief-state update—is a problem for computational neuroscience, not for the foundational theory. constrains N (it must operate on quality trajectories and produce a control-theoretic sufficient statistic); it does not derive N from first principles.

This is analogous to how thermodynamics constrains heat engines without specifying any particular engine design. A specific computational model of N, instantiated in a biophysically realistic neural-field simulation, is an open problem (§7) and a priority for future work.

Status assessment: N is the weakest link. We state this plainly. The core postulate (QI) is a clean metaphysical commitment. The formalism (§2–3) inherits established mathematics. The empirical protocols (§5) are concrete and pre-registerable. But N—the narrative operator that bridges quality to report—is currently a schema, not a model. The toy pipeline below demonstrates that N has definite mathematical structure, but six qubits are not a cortex. What a biophysically realistic N would require: a recurrent neural-field model (Wilson-Cowan or neural-mass-model class) with realistic cortico-cortical connectivity, where the self-thread is identified as a slow manifold in the field dynamics and the narrative operator is implemented as a low-rank filter on trajectories along that manifold, with report coupling measured as Granger-causal influence from the filter's output to motor-effector populations. This implementation is computationally tractable with current methods (neural-mass models with nodes, fitted to individual subjects' connectivity data) and would generate quantitative and predictions for Protocol 1. We flag this as the single highest-priority theoretical development for IQT

How the Narrative Operator Actually Works

The narrative operator deserves more than a label, because it's doing specific mathematical work — not just waving at "the self."

A self-thread (Level 1) is a sequence of overlapping effective diamonds whose quality persists over time. The narrative operator N (Level 2) is what turns that quality-stream into something that can report on itself. It works in five steps, each a standard operation in computational neuroscience:

Physical state. At each moment, the self-thread's effective diamond has a full physical state — all the neural activity across the relevant brain regions.

Coarse-graining. Extract the interface variables: sensory inputs, motor outputs, and the cross-boundary correlations bridging them. This is just dimensionality reduction — going from the full state to the variables the system can actually use.

Bayesian filtering. Compress the history of interface variables into a *belief state* — the minimal summary sufficient to predict what happens next. This is variational inference: a low-dimensional representation that tracks the high-dimensional state.

Self-model. Extract slowly varying features of the belief-state trajectory. This is where "I" lives — a compact representation of the system's own dynamics over time. It's recursive: the self-model influences action, which changes sensory input, which updates the belief state, which updates the self-model.

Report coupling. The self-model drives motor output. Readout dominance — how much the self-model controls the report channel — is what distinguishes a narrative self (Level 2) from a mere self-thread (Level 1).

Each step is a standard mathematical operation (partial trace, variational inference, recurrent compression, linear readout). The paper includes a worked toy version with six qubits standing in for sensory, internal, and motor populations — not a brain, but a proof that the pipeline has definite mathematical structure rather than being a verbal sketch.

The narrative operator is constrained by IQT's framework — the self-model is a compressed representation of quality, not something separate from it. But its *implementation* draws on active inference and predictive processing. This is deliberate: IQT treats those frameworks as correct descriptions of how brains build self-models, while denying that the self-model is what consciousness *is*. The self-model is how Level 2 report works. Consciousness (quality) is already there at Level 0.

4.5.1 Worked Toy Pipeline: From to Report

This pipeline demonstrates that N has definite mathematical structure. It uses density matrices for clarity; the same structure can be instantiated with classical stochastic states or coarse-grained neural field states. No claim of macroscopic quantum coherence in cortex is intended.

Step 1: Physical state trajectory. Six qubits: sensory , internal , motor . At each timestep , quality .

Step 2: Coarse-graining to interface variables. partial traces to sensory, motor, and cross-boundary joint states. These are themselves qualities of sub-diamonds.

Step 3: Bayesian filtering to belief state. |) - a variational inference update computing a sufficient statistic for predicting future sensory input. Dimensionality: for .

Step 4: Self-model. — a recurrent map extracting slow-varying features of the belief-state trajectory. Recursive: influences action sensory input

Step 5: Report coupling. . Readout dominance: .

Pipeline summary: . Each arrow: partial trace, variational inference, recurrent compression, linear readout.

Neural operationalization. In Section 5: multivariate neural state; channel subsets; dimensionality-reduced neural trajectories; prefrontal representations; directed information from prefrontal to motor cortex.

The Toy Pipeline in Plain English

This pipeline demonstrates how raw physical states get compressed into a self-model that can report. Five steps: (1) Start with the full physical state of a small system (six qubits standing in for sensory, internal, and motor neural populations). (2) Coarse-grain by extracting the interface variables

—sensory inputs, motor outputs, and the correlations bridging them. (3) Compress the history into a belief state via Bayesian filtering—the minimal summary sufficient to predict what happens next. (4) Extract slowly changing features of the belief trajectory into a self-model. This is where "I" lives—a compact representation of the system's own dynamics over time. (5) The self-model drives motor outputs (report).

The pipeline is a toy—six qubits are not a brain—but it demonstrates that the narrative operator has concrete mathematical structure: a chain of dimensionality reductions from physical state to reportable self-model. Each arrow (partial trace, variational inference, recurrent compression, linear readout) is a standard operation in computational neuroscience.

Box 4.A — Split-brain: sanity check, not exclusion test

Post-callosotomy, hemispheres are disjoint (non-overlapping) regions. Both IQT and IIT can consistently accommodate two disjoint conscious systems. The split-brain case confirms that semi-autonomous self-threads exist but does not test IIT's exclusion postulate, which concerns overlapping candidates. Protocols targeting overlap directly are in Section 5.

IQT interprets callosotomy as decoupling two self-threads that were previously correlated. The left hemisphere's continued report of unity is evidence that the narrator's self-model has not been updated—not evidence of actual unity.

4.5.2 Consequences

The drifting self. In flow states, for the narrative thread drops below threshold while a sensorimotor self-thread dominates without generating narrative report.

Anesthesia. General anesthesia disrupts temporal persistence of self-threads, not local qualities. Without a persistent thread, there is no self—not because quality disappears, but because the temporal structure required for selfhood dissolves.

The democracy of diamonds. No fact of the matter about “how many experiencers” are present. The framework predicts a spectrum of selfhood, not a binary threshold.

Interpreting Section 4

This section shows IQT’s resolving power across classic consciousness puzzles. The infinite regress terminates because every region already has quality—no external observer needed. Mary gains a new quality (a new algebra-state class instantiated in her cortex), not new information about the world.

The subject selection problem is dissolved, not solved: all diamonds have quality (Level 0), self-threads are persistent quality-streams (Level 1), and the narrative self is the thread controlling report (Level 2). The sense of being a single unified subject is content of the Level 2 self-model.

The narrative operator is a theoretical schema constrained by but implemented by computational neuroscience—like thermodynamics constraining heat engines without specifying

Figure 5: Toy Pipeline – from full physical state to report via six-qubit model (section 4.5.1) designs.

5. Empirical Program: Concrete Protocols and Falsification Conditions

5.0 Discriminative Strategy

What the protocols test and what they do not.

has a layered logical structure (§1.4). The empirical program engages different layers with different directness:

Core Postulate (CP). is phenomenal quality.

Bridge Hypotheses (BH). Self-threads (Level 1) are quality-streams satisfying persistence and causal semi-autonomy conditions. Narrative operators (Level 2) are self-threads with dominant readout access. Democracy of diamonds: multiple overlapping self-threads can coexist.

Operationalizations (OP). P_j approximates self-thread persistence; K_jl approximates cross-boundary correlator strength; R approximates readout dominance; Syn approximates irreducibility.

The protocols primarily test BH and OP: do the predicted patterns of modular persistence, independence, and synergy appear in neural data under the conditions BH specifies? CP—the identity of quality with the algebra-state pair—is constrained indirectly: if BH generates successful predictions, this supports (but does not prove) the framework from which BH is derived; if BH fails systematically, the framework loses its empirical motivation. A protocol failure falsifies the conjunction . It does not, by itself, falsify CP alone—the identity thesis could survive with revised bridge hypotheses, just as general relativity survives revised models of stellar interiors. Conversely, CP is not unfalsifiable merely because the protocols do not test it in isolation. No fundamental postulate in any science is tested in isolation; it is tested through the predictive success of the theoretical structure it anchors.

We state this transparently so the reader can evaluate the evidence at the appropriate level.

Integrated Information Theory (IIT) (Tononi, 2004, 2008; Oizumi et al., 2014; Albantakis et al., 2023) identifies consciousness with integrated information and imposes an exclusion postulate: among overlapping candidate subsystems, only the maximally irreducible one exists as a conscious entity.

Global Neuronal Workspace Theory (GNW) (Baars, 1988; Dehaene & Changeux, 2011; Mashour et al., 2020) identifies consciousness with global broadcasting: a content becomes conscious when recurrent amplification "ignites" and the representation becomes globally accessible.

’s primary discriminative axis (against IIT): overlapping substrates. IIT’s exclusion postulate forbids multiple overlapping complexes. ’s democracy of diamonds permits them. This divergence concerns overlapping regions sharing physical substrate—not disjoint regions (where both theories allow multiple conscious systems).

’s secondary axis (against GNW): experience without broadcast. predicts Level 1 experience without global ignition. We acknowledge this is difficult to operationalize and focus our protocols on the more tractable IIT contrast.

What the protocols can and cannot adjudicate. ’s metrics operationalize self-thread persistence and causal autonomy. IIT’s exclusion is defined in terms of irreducible cause-effect power . These are different quantities. We do not claim our metrics directly measure . We claim they provide evidence that is differentially expected under versus under exclusion-type views, supplemented by a synergy-based irreducibility surrogate (§5.2) that more directly engages the irreducibility question. We also specify what would be needed to test exclusion in its own terms (computing for overlapping candidates), which is currently computationally intractable for systems of this size but may become feasible.

The operationalization gap. The formalism of §2 defines quality using C*-algebras and state restrictions. The empirical metrics of §5.0.1 use dwPLI, participation ratios, and transfer entropy computed on electrophysiological signals. The connection between these two levels is mediated by the effective-theory framework of §2.6: EEG/iEEG signals are observables in the effective neural algebra , and the metrics P_j, K_jl, and R are computable functionals of the restricted state . This bridge is held together by the same logic that connects any fundamental theory to its experimental tests—through a chain of effective descriptions, each justified by scale separation and coarse-graining. The bridge is not a derivation from AQFT axioms to electrode voltages; no theory of consciousness currently achieves this. What the effective-theory framework provides is a principled account of why neural-scale operationalizations are legitimate tests of a theory formulated in the language of local algebras: because the effective algebra is a local algebra, at the appropriate scale. The empirical program tests the effective-level predictions of The fundamental-to-effective bridge (§7, "The coarse-graining bridge") is a separate, complementary research program.

5.0.1 Metric Definitions

Thread-persistence metric . For cortical module , at time-window scale :

where temporal mutual information between successive windows of width w: and effective dimensionality via participation ratio of the covariance eigenspectrum:

The product rewards structured, high-dimensional stability—predictable AND complex. This avoids the confound where anesthetic-induced slow oscillations increase raw predictability (high TMI) while collapsing the signal into low-dimensional stereotyped patterns (low d_eff).

We compute across a range of w (100ms, 250ms, 500ms, 1s, 2s) and use the area-under-curve AUC_P_j = ∫ P_j(w) dw as the persistence summary. This multi-scale approach avoids dependence on a single arbitrary window size and captures persistence structure across timescales relevant to different conscious phenomena.

Cross-module coherence metric . For modules (j, l):

where dwPLI is the debiased weighted phase-lag index (Vinck et al., 2011), robust to volume conduction. Averaged over frequency bands (theta, alpha, beta, gamma).

Readout dominance metric R. Directed information from prefrontal to motor cortex:

We cross-validate directed influence with multiple estimators (transfer entropy, spectral Granger causality, state-space models) and include controls for common-driver confounds (conditional mutual information conditioning on shared input). Reported results will include concordance across estimators.

What the Metrics Measure, in Brain Terms

Three metrics form the empirical backbone of the protocols. P (thread persistence) measures whether a brain region maintains structured, high-dimensional activity over time—not just any activity, but complex activity that predicts its own future. It is the product of temporal predictability (TMI) and signal richness (d_eff). This catches the case where anesthesia creates highly predictable but low-dimensional slow waves (high TMI, low d_eff low P).

K (cross-module coherence) measures phase synchronization between distant brain regions, approximating the strength of cross-boundary correlators in the effective diamond. R (readout dominance) measures directed causal flow from prefrontal to motor cortex—how much the narrative self-thread controls the report channel.

Together: P tells you which modules are "alive" as self-threads, K tells you how bound together they are (the unity of the brain-scale diamond), and R tells you whether the narrative self is in charge of report.

5.1 Protocol 1: Titrated Propofol Anesthesia (Multi-Component Dissolution)

Rationale. This protocol tests whether phenomenological dissolution under anesthesia proceeds as multi-component fragmentation—distinct cortical modules losing self-thread persistence at different depths—or as single-locus degradation.

We are precise about what this tests. 's democracy of diamonds predicts that the brain sustains multiple semi-autonomous quality-bearing subsystems, and that anesthesia degrades these at different rates because different correlator families have different pharmacological vulnerability. This predicts multi-component fragmentation as a natural consequence of the framework. A single-locus degradation pattern—where all modules lose persistence in lockstep—is more naturally consistent with views positing a single conscious system (whether IIT's single maximally irreducible complex or GNW's single global workspace). However, we do not claim that IIT or GNW logically require lockstep degradation; a single system can degrade non-uniformly. The protocol provides evidence that is differentially expected under versus under single-locus views, not a decisive refutation of any specific competitor.

Subjects. , healthy volunteers, informed consent.

Apparatus. High-density EEG (256-channel) or MEG. Target-controlled infusion (TCI) of propofol. Structured phenomenological interview.

Procedure. Propofol in 10 steps (c_0 = baseline to c_9 = LOC). At each c_k: 5-minute steady state, 3-minute recording, multi-modal stimulus battery, brief awakening with structured interview probing five dimensions: spatial orientation (Y_spatial), body awareness (Y_body), temporal flow (Y_temporal), narrative continuity (Y_narrative), emotional valence (Y_valence), each 0–10.

Pre-registered analysis plan. Prior to data collection, register:

(i) exact electrode-to-module assignment for J = 5 modules (visual, auditory, somatosensory, prefrontal/executive, default-mode),

(ii) P_j computation parameters (w range, AUC method),

(iii) primary hypothesis test (within-subject correlation of AUC_P_j trajectories across modules),

(iv) significance thresholds for success/failure conditions.

IQT prediction (multi-component fragmentation):

The discriminative signature is not merely that modules differ in their degradation rate—heterogeneous pharmacological sensitivity could produce that under any framework. The signature is conditional independence under perturbation at intermediate depths: that at some propofol concentration c_k, there exist modules A and B such that A retains high P while B collapses, and perturbations to A alter A's dynamics without propagating to B beyond a pre-registered effect-size threshold (partial $\eta^2 < 0.05$), after controlling for arousal (pupil diameter) and spectral confounds (total power, spectral slope). This conditional independence of semi-autonomous persistence is what distinguishes democracy from mere heterogeneity within a single system.

Specific predictions:

AUC_{P_j} metrics decline at different rates and thresholds across modules. Prefrontal declines earliest; primary sensory modules decline last.

Phenomenology components Y_j track the P_j decline of corresponding modules.

At intermediate depths, the system exhibits partial fragmentation: some modules retain high P_j while others have collapsed, with reduced K between high-P and low-P modules.

Modality-specific perturbations (increasing auditory stimulus complexity) affect $P_{auditory}$ without propagating to P_{visual} at intermediate depths. This is the critical conditional-independence test: pre-register the perturbation effect-size threshold and the cross-module propagation criterion before data collection.

Metric validity and confounds.

(i) Oscillatory confound: Propofol induces frontal alpha/slow-delta that increases raw TMI. The $P_j = TMI \times d_{eff}$ product controls for this: d_{eff} collapses under propofol even if TMI increases. We report TMI and d_{eff} separately to verify.

(ii) Pharmacokinetic confound: Non-uniform receptor distribution could produce heterogeneous P_j decline for pharmacological reasons. Control: replicate with sevoflurane. IQT predicts fragmentation persists across agents (though module ordering may change).

(iii) Non-monotonic coherence: We target K_{j} specifically on long-range inter-module dwPLI, not local within-module coherence, to isolate cross-boundary correlators.

(iv) Interview noise: Complemented by no-report measures (ERPs, pupillometry, skin conductance).

Evidence for IQT over single-locus views: At intermediate depths, ICA reveals ≥2 separable high-P components with low mutual K, responding independently to modality-specific perturbations. A single-component model fitting equally well supports single-locus views.

Evidence against IQT P_j decline in parallel across all modules (within-subject Pearson $r > 0.9$ across module pairs) in ≥80% of subjects. No partially-fragmented intermediate state observed.

5.2 Protocol 2: Within-Hemisphere Overlapping Parcellations (Direct Overlap Test)

Rationale. This protocol directly targets the overlap divergence between IQT and IIT's exclusion postulate. We define two regions that share physical substrate and test whether both simultaneously sustain independent high-persistence dynamics—which IQT permits and exclusion forbids.

Critically, we include a synergy-based irreducibility surrogate to address the "union complex" escape route: an IIT proponent could grant that A and B show independent task responses while maintaining that A∪B is the single conscious complex (satisfying exclusion). The synergy metric distinguishes "two semi-autonomous loci" from "one irreducible complex with modular task responses."

The overlap construction. Left posterior cortex monitored by intracranial EEG (iEEG, high-density grid, minimum 64 electrodes over posterior cortex). Define two overlapping parcellations:

Region A: electrodes covering visual association cortex + lateral parietal cortex (V4/IT + LIP). Minimum 15 electrodes; minimum 5 in the overlap zone (LIP).

Region B: electrodes covering lateral parietal cortex + posterior prefrontal cortex (LIP + FEF/dlPFC). Minimum 15 electrodes; same 5+ overlap electrodes.

Regions share the lateral parietal component. Electrode-to-region assignment is pre-registered based on anatomical landmarks before task data collection.

Subjects. Epilepsy patients undergoing iEEG for surgical planning ($n \geq 8$). Pre-registered inclusion: adequate posterior coverage; exclusion: seizure focus within target regions.

Procedure. Two concurrent cognitive tasks:

Task α (engages Region A): Visual object categorization—fine-grained discrimination (V4/IT) with spatial attention (LIP).

Task β (engages Region B): Visuospatial working memory—spatial location maintenance (LIP) with saccade planning (FEF/dlPFC).

Four conditions:

C1 (Single-task α). Only Task α. Measure P_A, P_B, K_{AB}, Syn_{AB}.

C2 (Single-task β). Only Task β. Measure same.

C3 (Dual-task, both hard). Both tasks simultaneously at high difficulty.

C4 (Dual-task, both easy). Both tasks simultaneously at low difficulty. (Attentional-demand control: if independence holds in C4 but breaks in C3, the effect is attentional, not exclusion-related.)

Task difficulty parametrically varied across trials (easy/medium/hard) for perturbation analysis.

Metrics.

Overlap contamination control. P_A and P_B share electrodes in the LIP overlap zone, which can generate spurious independence or dependence artifacts. We address this with three complementary strategies: (i) Exclusive-electrode P: compute P_A^{excl} and P_B^{excl} using only non-overlapping electrodes. These are the primary metrics. (ii) Inclusive-electrode P: compute P_A^{incl} and P_B^{incl} using all electrodes including the overlap. (iii) Overlap-only P: compute $P_{overlap}$ using only the shared LIP electrodes. [107] predicts $P_{overlap}$ contributes structured information to both exclusive sets independently; if $P_{overlap}$ merely drives both P_A^{incl} and P_B^{incl}, this will be visible as P_A^{excl} and P_B^{excl} being low while P^{incl} values are high. All three are reported; the primary analysis uses exclusive-electrode metrics.

K_{AB} = dwPLI between the non-overlapping electrode subsets of A and B only.

Perturbation independence index Π: $|\mathrm{Corr}(\Delta \mathrm{Difficulty}_\alpha, \Delta P_B^{excl})| + |\mathrm{Corr}(\Delta \mathrm{Difficulty}_\beta, \Delta P_A^{excl})|$, where ΔDifficulty is the trial-by-trial difficulty change and ΔP is the corresponding persistence change. Low Π means perturbations don't propagate across regions.

Synergy measures: task-relative and intrinsic. We compute two distinct synergy metrics. The distinction matters: an IIT proponent can reasonably argue that task-relative synergy is about behavioral relevance, not about the irreducibility of experience. We therefore supplement task-relative synergy with an intrinsic measure that engages the irreducibility question more directly.

Task-relative synergy Syn^{task}_{AB}. Using partial information decomposition (Williams & Beer, 2010; Mediano et al., 2021), decompose the mutual information between neural activity and task performance:

where Y_{task} is the joint task-performance variable and X_A, X_B are the neural signals from exclusive electrode subsets. This measures whether AB carries joint behavioral information beyond the parts. It is a necessary but not sufficient indicator of irreducibility: low Syn^{task} is consistent with semi-autonomous loci but does not rule out irreducibility in a non-task-relevant sense.

Intrinsic synergy via tripartite O-information. The overlap zone is not just a methodological nuisance—it is the theoretically interesting structure, the shared substrate that [107] claims participates in two quality-bearing regions simultaneously. We exploit this by defining three non-overlapping neural variables:

X_{AO} = signals from Region A's exclusive electrodes; X_O = signals from the overlap zone (shared LIP electrodes); X_{BO} = signals from Region B's exclusive electrodes.

O-information (Rosas et al., 2019) over three or more variables is:

where $TC = \Sigma H(X_i) - H(X_{AO}, X_O, X_{BO})$ is the total correlation and $DTC = H(X_{AO}, X_O, X_{BO}) - \Sigma H(X_i \mid X_{i})$ is the dual total correlation. (Note: for two variables, $TC = DTC = I(X_1; X_2)$, so $\Omega = 0$ identically. The tripartite split is essential.)

Negative Ω indicates synergy-dominated dynamics; positive Ω indicates redundancy-dominated

Figure 6: Protocol 2 – Overlapping parcellations and tripartite O-information structure (section 5.2)

dynamics.

This tripartite structure directly tests whether the overlap zone acts as a binding node (creating synergistic dependence between the exclusive zones, consistent with a single irreducible union) or as a shared resource (contributing independently to each exclusive zone's processing, consistent with democracy). We additionally compute the two bipartite O-information values and —these are identically zero by construction and serve as sanity checks.

Estimation. We specify the estimator class to avoid ambiguity: Gaussian copula estimator (Ince et al., 2017), which handles high-dimensional continuous data without discretization. Each electrode subset is first dimensionality-reduced to its top k principal components (k chosen by participation ratio, matching d_{eff}), yielding variables of manageable dimension (typically per subset). O-information is computed on these reduced representations. We report results for multiple k values to verify robustness. Confidence intervals via block bootstrap (block length , to respect temporal autocorrelation).

Complementary: predictive synergy about the system's own future. As a second intrinsic measure, we compute synergy in the predictive sense:

how much synergistic information do and jointly carry about —the overlap zone’s own future state? Using PID:

Syn^pred = Syn(X_ {A ~ O} (t), X_ {B ~ O} (t) X_ {O (t +))

This is IIT-adjacent: it asks whether the exclusive zones jointly and irreducibly constrain the overlap zone’s future beyond what either constrains alone. High Syn^pred means the exclusive zones are causally entangled through the overlap—consistent with a single integrated complex. Low Syn^pred means each exclusive zone’s influence on the overlap is independent of the other—consistent with democracy. Target variable with (within the gamma-band integration timescale).

Perturbational protocol (the decisive causal test). Observational metrics—however sophisticated—are always debatable. Exclusion is fundamentally about causal power: whether a system’s causal structure is irreducible. The perturbational arm is therefore the central test of this protocol, not a supplement to the synergy analysis. We exploit the fact that iEEG electrodes can deliver focal electrical stimulation.

Single-pulse electrical stimulation (SPES): In a subset of trials during C3, deliver a brief (1ms, 1-5mA) single-pulse stimulus to one electrode in Region A’s exclusive zone, and separately to one in Region B’s exclusive zone. Measure the evoked response propagation using cortico-cortical evoked potentials (CCEPs).

prediction (causal containment): SPES in A’s exclusive zone evokes responses primarily within A (including the overlap) with minimal propagation to B’s exclusive zone, and vice versa. Perturbation containment: the evoked response in the non-stimulated region’s exclusive electrodes is $< 20\%$ of the evoked response in the stimulated region’s exclusive electrodes (pre-registered amplitude ratio).

Exclusion prediction (union complex): SPES in either zone evokes responses throughout AuB, because A and B are causally integrated components of a single irreducible complex.

Exclusion prediction (switching): SPES propagation pattern alternates depending on which complex currently “owns” the substrate at that moment.

This perturbational arm transforms the protocol from purely observational to causal, directly engaging IIT’s emphasis on causal power. The SPES containment result is the single most informative outcome of Protocol 2: if perturbations in A do not propagate to B’s exclusive zone during simultaneous high-P dual-task performance, this is strong evidence for causally semi-autonomous overlapping loci—the core prediction of democracy.

Controls for SPES confounds. (i) Volume conduction and stimulation spread: compare SPES-evoked responses during C3 with those during rest (no task). If propagation patterns differ between task and rest, the task-state containment is not an artifact of stimulation geometry. (ii) Non-specific arousal effects: SPES can produce widespread transient arousal. Control by analyzing early (<50ms) versus late (>100ms) evoked components separately; causal containment should be evident in early components, which reflect direct connectivity rather than arousal. (iii) Current spread to deep structures: document stimulation parameters and model current spread; exclude trials where estimated current density at the non-target region exceeds 10% of the target-region density.

Complementary observational and synergy metrics. The SPES results are interpreted alongside the following observational metrics, which provide convergent evidence:

Winner-take-all switching analysis. If exclusion operates via temporal switching between competing complexes, this should be detectable as anti-correlated fluctuations in excl and excl at fine temporal resolution. We test for this explicitly: Compute excl(t) and

P_B^excl(t) in sliding 500ms windows during C3. Test for anti-correlation: Pearson r(P_A^excl(t), P_B^excl(t)) < -0.3 would indicate switching. Compare against a null model of dual-task resource competition: if anti-correlation disappears when controlling for overall arousal/workload (pupil diameter, global power), the effect is attentional, not exclusion-related. If anti-correlation persists after controlling for arousal, it supports switching dynamics. Distinguish switching from gradual resource sharing using Hidden Markov Model (HMM) analysis: fit a 2-state HMM to [P_A^excl(t), P_B^excl(t)] and test whether the data are better described by discrete state transitions (switching) or continuous co-variation (resource sharing). BIC comparison between HMM and a smooth bivariate model determines which account the data favor.

IQT prediction (coexisting overlapping quality-bearing regions):

In C3 (dual-task, hard), both P_A^excl and P_B^excl remain within 80% of their single-task baselines.

$\Pi < 0.3$ in C3: perturbations to one task don’t propagate to the other region’s exclusive electrodes.

Tripartite , X_O, X_{B}O) is near zero or positive (redundancy-dominated or independent): the overlap zone does not create synergistic binding between the exclusive zones. Syn^pred is low: each exclusive zone’s influence on the overlap zone’s future is independent of the other.

SPES containment: perturbations in A’s exclusive zone do not propagate to B’s exclusive zone, and vice versa.

No switching: P_A^excl(t) and P_B^excl(t) are not anti-correlated; HMM is not preferred over smooth co-variation model.

Performance on both tasks maintained above chance in C3 (with some dual-task cost).

Exclusion-compatible predictions:

Union complex: is strongly negative (synergy-dominated triad—the overlap zone creates irreducible binding), Syn^pred is high, Syn^task_AB is high, SPES propagates throughout AB, high . P_A and P_B may both remain high but are coupled components of a single complex.

Switching: P_A^excl(t) and P_B^excl(t) anti-correlate (r < -0.3), HMM preferred over smooth model, fluctuates between negative and near-zero as the dominant complex alternates. SPES propagation pattern depends on current switching state.

Third complex: Both P_A and P_B show lower signatures of integrated processing compared to a detectable third region. (Harder to test; honest acknowledgment.)

Evidence for IQT

SPES contained (the primary result), plus both P^excl high, low, near zero or positive, Syn^pred low, no switching. This converging pattern—causal containment confirmed by perturbation, supported by observational independence and low synergy—is not naturally accommodated by exclusion.

Evidence against IQT

SPES propagates uniformly across AB (the primary disconfirmation), or P_A^excl and P_B^excl anti-correlate in C3 (switching), $\Pi > 0.7$, strongly negative (synergy-dominated triad), Syn^pred high. Any of these would support an exclusion-type dynamic. SPES propagation is the most decisive: if a focal perturbation in A reliably evokes responses in B's exclusive zone during dual-task performance, the two regions are causally integrated in a way that democracy does not predict.

Caveats

Overlap contamination: addressed by exclusive-electrode analysis; reported alongside inclusive metrics for transparency. (ii) Neither nor Syn^pred is . captures statistical higher-order dependence; Syn^pred captures predictive causal synergy. Both are surrogates for the full cause-effect structure that IIT's irreducibility requires. Computing for systems of this size (channels) is currently intractable. We state the gap honestly and supplement with the perturbational arm (SPES), which engages causal structure directly. (iii) SPES can produce non-specific arousal effects: control by comparing SPES-evoked responses during C3 with those during rest (no task). (iv) Dual-task performance costs from shared attentional resources: C4 (both easy) controls for this. (v) iEEG coverage determined by clinical need: minimum coverage criteria pre-registered. (vi) O-information estimation via Gaussian copula requires approximate multivariate normality; we verify with Mardia's test and report copula-transformed residual diagnostics. Power analysis should target trials per condition for observational metrics, SPES trials per stimulation site.

Protocol 2: The Logic of the Overlap Test

This is the decisive experiment for democracy versus exclusion. The setup: define two brain regions that physically overlap—they share electrodes in lateral parietal cortex. Give each region its own cognitive task (visual categorization for one, spatial working memory for the other). The question: can both regions sustain independent high-quality processing simultaneously, despite sharing substrate?

Three converging lines of evidence: (a) Observational—do both regions maintain high P independently, and do perturbations to one task fail to propagate to the other? (b) Synergy analysis—

does the overlap zone create irreducible binding between the regions (supporting a single complex, as exclusion predicts) or act as a shared resource (supporting democracy)? (c) Perturbational—when you electrically stimulate one region, does the evoked response stay contained or spread throughout both?

The single-pulse stimulation (SPES) result is the most decisive test: if a focal electrical pulse in region A does not propagate to region B's exclusive electrodes during simultaneous high-performance processing, the regions are causally semi-autonomous—exactly what democracy predicts and what exclusion does not naturally accommodate.

5.3 Protocol 3: Psychedelic Temporal Phenomenology (Diamond Geometry Test)

Rationale. IQT predicts that the shape of effective causal diamonds determines temporal phenomenology (§2.6). Compounds that alter temporal experience should produce measurable shifts in the multi-scale persistence profile —a direct link between diamond geometry and phenomenological report that no competing theory predicts in this form.

Mechanistic toy model: why peak tracks integration window. The prediction that w_peak tracks temporal phenomenology requires more than a verbal argument. We demonstrate the mechanism in a minimal dynamical system.

Consider a network of coupled oscillators with state , governed by: , where is the intrinsic timescale of unit i, is the coupling matrix, is a sigmoid nonlinearity, and is noise. The coupling matrix has a block structure with two modules, each internally connected with strength within and cross-connected with strength between.

The critical parameter is the effective integration timescale , which depends on and : stronger within-module coupling and longer intrinsic timescales produce larger . We compute for this system across window sizes .

Result 1 (widening): Increasing (by increasing or) shifts w_peak rightward. The system maintains coherent, high-dimensional structure across longer temporal windows, expanding the effective "now." This corresponds to time dilation: the "specious present" expands.

Result 2 (narrowing): Decreasing (by decreasing while preserving) shifts w_peak leftward. Temporal correlations decay faster; structure is concentrated at short timescales. This corresponds to time fragmentation.

Result 3 (flattening): Setting coupling to be scale-free (1/f-like coupling spectrum, achieved by with) removes any preferred timescale. flattens—no peak. This corresponds to time dissolution.

Result 4 (control): Uniformly reducing signal amplitude (analogous to sedation/reduced arousal) decreases at all w without shifting w_peak. This is the key dissociation: arousal changes scale vertically; integration-window changes shift it horizontally. The midazolam active control in the empirical protocol tests this dissociation.

The toy model establishes that peak position is mechanistically determined by the effective temporal integration window of the dynamical system, not by banal confounds like spectral power or noise level. The psychedelic predictions follow if psilocybin widens (plausible: 5-HT2A agonism increases cortical integration timescales; Muthukumaraswamy et al., 2013, who report broadband cortical desynchronization—interpreted here as an expansion of the effective temporal window, not a simple loss of temporal structure, since desynchronization of narrow-band oscillations can coexist with enhanced broadband temporal integration across longer windows), ketamine narrows (plausible: NMDA antagonism disrupts recurrent temporal integration; Muthukumaraswamy et al., 2015), and DMT produces scale-free dynamics (plausible: massive 5-HT2A agonism disrupts the hierarchical timescale organization; Timmermann et al., 2019). Preller et al. (2019) provide additional mechanistic support: LSD-induced altered states involve enhanced thalamocortical driving, particularly from thalamus to posterior cortex, consistent with a pharmacological reshaping of the effective diamond’s geometry via altered thalamocortical gain rather than a simple global excitability change. These pharmacological mappings are independently testable predictions, not ad hoc assumptions.

Protocol 3’s Toy Model: Why the Persistence Peak Tracks Temporal Experience

The toy model—50 coupled oscillators—demonstrates that the peak of the persistence curve is mechanistically set by the system’s effective temporal integration timescale, not by confounds like overall signal strength. Four results establish this.

Widening the integration timescale shifts the peak rightward (longer temporal coherence, expanded "now"). Narrowing it shifts the peak leftward (faster decay, fragmented time). Making coupling scale-free removes the peak entirely (no preferred timescale, dissolved time). Crucially, simply reducing amplitude—like sedation—scales vertically without shifting the peak. This dissociation between arousal and integration-window changes is what the midazolam control condition tests in the actual

experiment: if a non-psychedelic sedative scales P without shifting its peak while psychedelics shift it, the effect is specific to temporal integration, not a banal arousal confound.

Subjects. n ≥ 25 per compound, healthy volunteers, standard psychedelic research screening (no personal or family history of psychosis), informed consent. Within-subject crossover design with placebo control and minimum 2-week washout.

Apparatus. High-density EEG (256-channel). Standardized phenomenological interview including temporal experience subscales (adapted from 5D-ASC and the Temporal Experience of Pleasure Scale). Continuous self-report of subjective time passage via an analog dial.

Compounds and predicted temporal profiles:

Psilocybin (0.2 mg/kg oral). Reported phenomenology: time dilation—the present feels expanded, moments stretch. IQT prediction: P_j(w) peak shifts toward larger w. The system maintains coherent, high-dimensional structure across longer temporal windows, expanding the effective "now." The temporal integration window of the self-thread widens. AUC_P_j increases at large w, decreases at small w relative to placebo.

Sub-anesthetic ketamine (0.5 mg/kg IV over 40 min). Reported phenomenology: time fragmentation—experience feels choppy, discontinuous, moments don't cohere. IQT prediction: P_j(w) peak shifts toward smaller w. Temporal coherence breaks at longer windows while short-window structure is preserved or enhanced. AUC_P_j increases at small w, decreases at large w relative to placebo.

DMT (IV infusion, extended state). Reported phenomenology: time dissolution—temporal structure collapses entirely, “eternity” or “timelessness.” IQT prediction: P_j(w) curve flattens—no preferred timescale of coherence. AUC_P_j shows no clear peak across the w range.

Pre-registered analysis plan. (i) For each subject and compound, compute P_j(w) for five modules (as in Protocol 1) across w = 100ms, 250ms, 500ms, 1s, 2s, 4s. (ii) Primary metric: w_peak = the window size at which P_j(w) is maximized, averaged across modules. (iii) Primary hypothesis: w_peak(psilocybin) > w_peak(placebo) > w_peak(ketamine). (iv) Secondary hypothesis: P_j(w) kurtosis under DMT is lower than under all other conditions (flatter curve). (v) Correlation: within-subject, within-trial correlation between w_peak and continuous self-reported time passage rate. (vi) Secondary confound-control model: regress w_peak on compound condition after controlling for total spectral power, spectral slope (1/f exponent), and movement-artifact residual variance. IQT predicts compound effects on w_peak survive this regression.

Discriminative logic. IIT’s exclusion postulate and Φ do not predict specific relationships between temporal phenomenology and the multi-scale persistence profile—Φ is a single scalar, not a function of temporal window size. GNW predicts global ignition or not, without generating multiscale temporal structure predictions. IQT uniquely predicts that the shape of the persistence-vs-window-size curve tracks the character of temporal experience because both are determined by the geometry of the effective causal diamond.

Evidence for IQT w_peak shifts in predicted directions across compounds, and w_peak correlates with self-reported temporal experience (r > 0.4, within-subject).

Evidence against IQT w_peak does not differ across compounds, or shifts in the wrong direction, or shows no correlation with temporal phenomenology.

Confounds and controls. (i) Spectral power / noise / entropy confound: the danger that w_peak tracks banal EEG changes rather than temporal integration. The toy model (Result 4) demonstrates that arousal-driven amplitude changes shift P(w) vertically (scaling), not horizontally (w_peak shift). We test this dissociation empirically: if midazolam (non-psychedelic sedative) scales P(w) without shifting w_peak, while psychedelics shift w_peak, the confound is ruled out. (ii) Pharmacological confound: different compounds affect different neurotransmitter systems. The toy model specifies the mechanistic parameter (τ_eff) each compound should modulate, generating independently testable pharmacological predictions. (iii) Arousal confound: drowsiness shifts P_j(w). Control: monitor arousal via pupillometry and exclude drowsy epochs; the midazolam condition directly tests whether arousal changes alone produce w_peak shifts. (iv) Movement artifacts: psychedelic states can increase movement. Control: ICA-based artifact rejection with manual review.

5.4 Supplementary: Split-Brain (Sanity Check)

The split-brain paradigm tests whether disjoint cortical regions sustain independent self-threads. This is a sanity check for the self-thread concept, not a test of exclusion. IIT can accommodate two disjoint complexes.

Procedure. Callosotomy patients perform independent visual-motor tasks with each hemisphere, plus Wada-test patients as supplement.

IQT prediction: Both hemispheres maintain high P_h with independent perturbation responses. If this fails (one hemisphere consistently dominates), it challenges IQT's self-thread framework

Protocol 3: Predicted Persistence Profiles Under Psychedelic Compounds Horizontal shift = integration window change; Vertical shift = arousal change Figure 7: Protocol 3 – Predicted P(w) persistence profiles under psychedelic compounds (section 5.3)

broadly.

5.5 Additional Predictions (Qualitative)

Flow states. R drops while P for sensorimotor modules remains high—narrative self attenuates while experience persists.

Animal consciousness. A spectrum of self-threads without language. Different correlator architectures → different kinds of quality, not just different "amounts."

Interpreting Section 5

The empirical program tests IQT's bridge hypotheses, not the core postulate directly. Three pre-registerable protocols target different aspects of the theory: Protocol 1 (anesthesia) tests multi-component fragmentation of self-threads; Protocol 2 (overlap) directly tests the exclusion divergence with IIT using synergy, perturbation, and switching analyses; Protocol 3 (psychedelics) tests whether diamond geometry determines temporal phenomenology.

Each protocol specifies what counts as evidence for and against IQT and what confounds are controlled for. The discriminative logic is transparent: these are not magic-bullet experiments but carefully designed tests that are differentially expected under IQT versus competing frameworks. Each protocol is differentially diagnostic, not a decisive single experiment—the power lies in the pattern of results across all three.

The operationalization gap—from C*-algebras to electrode voltages—is bridged by the effective-theory framework of §2.6, not by derivation from axioms.

5.6 What Partial Failure Would Mean

A theory that can only succeed or fail wholesale is not a serious empirical program. IQT is designed so that different patterns of partial failure have different diagnostic implications, pointing to specific revisions rather than wholesale abandonment.

Scenario A: Protocol 1 succeeds (fragmentation) but Protocol 2 fails (SPES propagates uniformly). This would mean that modules do lose persistence at different depths (consistent with democracy at an effective scale), but the perturbational test shows greater causal integration between overlapping regions than democracy predicts. The diagnosis: democracy may hold at a coarser spatial scale than the electrode grid can resolve, or the bridge hypotheses identifying P_j with persistence are partially correct but the identification of observational modules with causal diamonds needs revision. The revision: adjust the effective-theory bridge (§2.6) to account for stronger inter-module causal coupling, perhaps by replacing hard diamond boundaries with soft boundaries weighted by causal influence. QI is untouched; the bridge hypotheses are what need work.

Scenario B: Protocol 2 succeeds (causal containment) but Protocol 3 fails (w_peak does not shift). Democracy of overlapping loci is confirmed, but diamond geometry does not determine temporal phenomenology as predicted. The diagnosis: QI and democracy are empirically supported, but the specific mapping from τ_eff to phenomenal time (the "diamond geometry determines temporal experience" bridge) is wrong. The revision: temporal phenomenology may depend on additional parameters beyond τ_eff—perhaps on the spectral structure of the state within the diamond, not just the diamond's geometry. This motivates a richer effective-theory bridge without threatening the core ontology.

Scenario C: Protocol 3 succeeds (w_peak shifts) but Protocol 1 fails (parallel decline, no fragmentation). Diamond geometry determines temporal experience, but anesthesia degrades all modules in lockstep rather than fragmenting them. The diagnosis: the self-thread construct is either wrong or the anesthetic mechanism does not target individual modules but acts globally (e.g., via thalamic shutdown). The revision: self-threads may be less modular than hypothesized, or propofol's mechanism of action targets a shared bottleneck (e.g., thalamocortical loops) that affects all modules simultaneously. This would weaken the multi-component fragmentation story but not the overlap divergence or the temporal phenomenology predictions.

Scenario D: All protocols fail. QI's explanatory yield is undermined. The identity thesis is not directly refuted (it is metaphysical), but a QI that generates no successful bridge predictions has stopped earning its keep. At this point, the constraint argument of §1.1 remains a contribution to the metaphysics of local physics, but the consciousness-specific claims would need fundamental rethinking.

The point is that IQT's three protocols are not three independent coin-flips. Each tests a different bridge hypothesis, and the pattern of successes and failures is diagnostic of where the theory needs revision. This is the hallmark of a progressive research program in the Lakatosian sense: the core (QI) is protected by a belt of revisable bridge hypotheses, and empirical results tell you which belt element to revise.

The Ramifications, with IIT and GNW contrasts:

A brief orientation on the two main rivals:

Integrated Information Theory (IIT) identifies consciousness with integrated information (Φ) and imposes an *exclusion postulate*: among all overlapping candidate systems, only the one with maximal Φ is conscious. One brain, one conscious complex, period.

Global Neuronal Workspace Theory (GNW) identifies consciousness with global broadcasting: a content becomes conscious when it "ignites" a recurrent prefrontal-parietal network and becomes available to all cognitive processes. Consciousness is about access, not about what regions are made of.

Ramification 1: Anesthesia fragments your self-threads — it doesn't "turn off" consciousness.

IQT predicts that propofol doesn't flip a consciousness switch. It disrupts the temporal persistence of self-threads, and it does so module by module at different doses. Your visual cortex's self-thread might break down at a different propofol concentration than your auditory cortex's. The prediction is a staggered, multi-component fragmentation — not a single threshold.

- IIT predicts a single transition: the integrated information Φ of the whole-brain complex drops below some threshold, and consciousness disappears in one event. There's one complex, and it either exists or it doesn't.

• GNW predicts that consciousness disappears when global broadcasting fails — when the prefrontal-parietal “workspace” can no longer ignite. This is more of a gateway mechanism: the workspace is either on or off, so the transition should be relatively sharp and correlated with frontal activity collapse.

• IQT predicts graded, region-specific dissolution. Some modules lose persistence before others. The “lights” don’t go out all at once — they go out room by room. This is testable by tracking persistence metrics independently across brain regions during titrated propofol infusion.

Ramification 2: Multiple overlapping conscious regions coexist in your brain right now.

This is the biggest empirical wedge between IQT and IIT. Right now, the region of your brain handling vision + audition has its own quality. So does vision + motor. These overlap (they share the visual cortex), and both are real simultaneously.

• IIT says this is impossible. The exclusion postulate insists that among all overlapping candidate systems, only the one with maximal Φ exists as a conscious entity. The others are “excluded.” Your brain has one conscious complex, period.

• GNW doesn’t directly address this — it’s about access, not substrate boundaries. It’s agnostic on whether overlapping regions can independently bear consciousness.

• IQT says overlapping regions are all conscious, and you can test this. If you electrically zap a small area in Region A, does the response propagate to Region B’s exclusive territory, or does it stay contained? If it stays contained — even while both regions are independently performing complex tasks — they’re causally semi-autonomous. That’s democracy, and exclusion can’t easily accommodate it.

Ramification 3: Time doesn't flow — your narrator's window slides.

The "passage of time" isn't a feature of reality. It's a feature of being a finite-bandwidth narrator tracking your own trajectory through a four-dimensional quality landscape. Your narrative operator can only hold ~100ms–1s of content at once. Each update sheds trailing content and acquires leading content. That sequential shedding-and-acquiring is the feeling of passage.

An infinite-bandwidth narrator would still run forward (computation is irreversible — there's no "backward narration" at macroscopic scales), but it would hold its entire trajectory at once. It wouldn't feel time passing. It would just… have the whole story simultaneously.

• IIT doesn't have a mechanism for temporal phenomenology. It characterizes the cause-effect structure of a system at a moment but doesn't predict how the shape of temporal integration should feel.

• GNW explains temporal experience through the dynamics of workspace access — contents enter and leave the global broadcast — but this is a functional description of what happens, not a structural account of why it feels like passage.

• IQT generates temporal phenomenology directly from geometry: the temporal depth of your effective diamond determines the duration of your "now," and the finite bandwidth of your narrative operator generates the passage-feeling. This predicts specific, measurable shifts in temporal experience under psychedelics.

Ramification 4: The composition problem dissolves — no "binding glue" is needed.

How do separate brain regions produce unified experience? This is the "combination problem" that haunts every theory of consciousness. IQT's answer: the quality of the whole is not the sum of the qualities of the parts. The "missing ingredient" is correlation structure — the pattern of entanglement and statistical dependence between regions. This isn't a mysterious binding force. It's a mathematical fact about how states on algebras work: knowing the state of Region A and the state of Region B doesn't tell you the state of the combined region A+B. The correlations between them carry additional quality that belongs to the whole but not to either part.

• IIT handles composition through its Φ measure and exclusion — the maximally irreducible system is the conscious one. But it faces the combination problem in reverse: how do micro-level experiences combine into macro-level experience? IIT's mathematical framework addresses this but at enormous computational cost, and the exclusion postulate means most candidate compositions are declared non-existent.

• GNW sidesteps the combination problem by defining consciousness as global access rather than as a compositional property. It doesn't need to explain how parts combine because it identifies consciousness with a functional role (being broadcast), not with a structural property of regions.

• IQT dissolves it: composition is just the extension problem from quantum information theory — a well-posed mathematical question with known structure. The "glue" is correlations. No new physics, no binding force, no special mechanism.

Ramification 5: There's no fact of the matter about "how many conscious beings" are in a brain.

IQT says this question is ill-posed. There are many overlapping regions, each with quality. Some sustain self-threads (Level 1). Some of those self-threads run narrative operators (Level 2). The boundaries between "subjects" are empirical and fuzzy, not metaphysical and sharp.

• IIT gives a definite answer: exactly one conscious complex per brain (the one with maximal Φ). This is clean but possibly wrong — and it creates awkward edge cases (what if two candidate complexes have nearly identical Φ?).

• GNW gives a different definite answer: consciousness is the content currently in the global workspace. There’s one workspace, so there’s one stream of consciousness at a time (though its content changes rapidly).

• IQT says: your left hemisphere and right hemisphere each sustain their own self-threads. They’re normally tightly correlated, producing a sense of unity (the narrative operator builds a unified self-model). Split-brain surgery decouples them, and the left hemisphere’s narrator — still running its old self-model — continues reporting unity it no longer has. The continued report of unity is the evidence against actual unity, not for it.

Ramification 6: The theory is designed to be wrong in specific, informative ways.

This might be the most important ramification for evaluating IQT as science. The theory is built in layers: core postulate → bridge hypotheses → operationalizations → experimental protocols. Each layer can fail independently, and the pattern of failure is diagnostic.

If the anesthesia protocol works but the overlap protocol fails, the identity thesis survives but the democracy-of-diamonds claim needs revision. If the psychedelic protocol works but anesthesia fails, diamond geometry determines temporal experience but self-thread fragmentation isn’t how anesthesia operates. If everything fails, the identity thesis hasn’t earned its keep.

• IIT is notoriously difficult to test because computing Φ for biologically realistic systems is computationally intractable. The theory makes predictions in principle but can’t currently cash them out at brain scale.

• GNW is testable (and has been tested extensively through the “ignition” paradigm), but its predictions are about access and reportability — it’s harder to distinguish from “GNW describes the functional architecture of report” versus “GNW explains consciousness.”

• IQT pre-registers three protocols with defined metrics, predicted response profiles, and explicit failure conditions at each theoretical layer.

It tells you in advance what partial failure means and which part of the theory it indicts.

6. Connections to Existing Work

Integrated Information Theory (Tononi, 2004, 2008; Oizumi et al., 2014; Tononi et al., 2016; Albantakis et al., 2023). Central divergence: exclusion postulate. IIT asserts that among overlapping candidates, only the maximally irreducible one exists. ’s democracy of diamonds denies exclusion. Additionally, IIT identifies consciousness with cause-effect structure; identifies quality with the algebra-state pair. Protocol 2 is designed to test the overlap divergence directly via synergy-based irreducibility surrogates, while being honest about the limitations of proxy metrics. For a more direct test, computing for overlapping candidates would be needed—currently computationally intractable but a priority for future work.

Causal Set Theory (Bombelli et al., 1987; Sorkin, 2003; Dowker & Glaser, 2013). The discrete spacetime substrate of

Relational Quantum Mechanics (Rovelli, 1996). extends RQM: the relational description is intrinsic quality, not merely all we can say.

Quantum Reference Frames (Giacomini, Castro-Ruiz, & Brukner, 2019). Perspectival relativity is structurally analogous to QRF’s frame-relative descriptions.

Russellian Monism (Russell, 1927; Chalmers, 2015). formalizes Russellian Monism with the Intrinsicness Principle. Perspectival relativity adds a dimension Russell did not anticipate.

Contemporary panpsychism and panqualityism (Goff, 2019; Seager, 2020). is most precisely classified as a form of Russellian panqualityism: quality is universal, but experience (Level 1) and phenomenal consciousness in the folk sense (Level 2) are not. Goff’s constitutive Russellian panpsychism faces the combination problem; addresses combination via correlation structure (§3). Seager’s natural attractor theory shares ’s emphasis on intrinsic nature but lacks ’s formal apparatus.

Process Philosophy (Whitehead, 1929). provides the formal precision process philosophy aspired to.

GNW and Higher-Order Theories (Baars, 1988; Dehaene & Changeux, 2011; Mashour et al., 2020; Rosenthal, 2005). In ’s framing, these describe Level 2 phenomena—correct about functional architecture, silent on Level 0 and Level 1.

Active Inference (Friston, 2010; Friston et al., 2017; Clark, 2013; Hohwy, 2013). The narrative operator draws on active inference for its implementation class. Section 4.5.1 makes the connection explicit.

Mathematical frameworks for consciousness theories (Kleiner & Tull, 2021). Kleiner and Tull develop a category-theoretic framework for formalizing and comparing consciousness theories. ’s presheaf structure (§7) fits naturally into this framework, with the quality functor as a concrete instance of their general scheme. This connection could facilitate formal comparison between and IIT within a shared mathematical language.

Interpreting Section 6: Where Fits (and Doesn’t)

is a form of Russellian panqualityism—but one with actual physics under the hood and a correlation-based solution to the composition problem. Classic Russellian monism posits intrinsic natures without specifying their mathematical structure; identifies them as algebra-state pairs and inherits the full machinery of AQFT.

The central divergence from IIT: where IIT’s exclusion postulate selects a single maximally irreducible substrate, ’s democracy of diamonds allows every causally bounded region to bear quality simultaneously. This is not a bug but a feature—it dissolves the subject selection problem by making subject boundaries empirical rather than metaphysical.

GNW and higher-order theories describe Level 2 (the narrative self, reportable awareness) correctly but are silent on Levels 0 and 1. treats them as accurate accounts of readout and report, not of quality itself. Active inference supplies the natural implementation class for the narrative operator N of §4.5—predictive processing is how brains build the self-model, not what consciousness is.

Perspectival relativity is what distinguishes from its philosophical ancestors. Neither classic Russellian monism nor process philosophy had the formal tools to express the idea that quality is relative to the locus of observation without a privileged "view from nowhere." AQFT provides those tools.

7. Open Problems and Call for Collaboration

The boundary data problem. What is the minimal composition datum ? Status: well-posed; partial finite-dimensional results exist; open in AQFT.

Sheaf-theoretic formulation. Can be formalized as a presheaf on the poset of causal diamonds? Status: natural; connects to boundary data problem.

Quantitative N. Generate numerical P_j(c_k) and R(c_k) predictions from a biophysically realistic neural-field model with propofol pharmacodynamics. Status: computationally tractable; requires parameter calibration.

computation for overlapping candidates. Direct test of exclusion requires computing IIT's for overlapping candidate subsystems. Currently intractable for systems of empirically relevant size. Developing efficient approximations is a priority shared with the IIT community.

Protocol execution. Priority: Protocol 1 (anesthesia), given available subject pool. Protocol 2 requires iEEG patients. Protocol 3 (psychedelic temporal phenomenology) requires psychedelic research licensing. Collaboration sought with clinical anesthesiology, epilepsy surgery, and psychedelic research groups.

CP-violation, time reversal, and quality-equivalence. IQT defines the quality of a region DDD as the algebra-state pair Q(D)=(A(D),ω|A(D))Q(D) = (\mathcal{A}(D), \omega|_{\mathcal{A}(D)})Q(D)=(A(D),ω|A(D)) up to physically induced identifications (§1.2). As defined, this is orientation-free: a region is a bounded set of events; it does not come with an intrinsic “forward reading” versus “backward reading.” Asking whether “forward” and “backward” differ therefore cannot mean “two different qualities for the same (A(D),ω|A(D))(\mathcal{A}(D), \omega|_{\mathcal{A}(D)})(A(D),ω|A(D)).” It can only mean: are the *actual physical situation* and its *time-reversed counterpart* identified by a genuine symmetry of the theory, or not?

To make that comparison precise, one must represent time reversal as a physically induced transformation of the net. In standard QFT, time reversal is implemented antiunitarily; at the algebraic level it acts as an anti-linear structure-preserving map on observables (an anti-*automorphism / anti*-isomorphism, depending on conventions). The natural refinement of §1.2 is therefore: “same quality” is invariance of the algebra-state pair under the physically induced symmetry group of the theory, allowing anti-linear symmetries where the physics requires them. If TTT were an exact symmetry in this sense, then the time-reversed image of a situation would lie in the same quality-equivalence class: the time-reversed state on the time-reversed image of the region would be physically identified with the original.

In our universe, however, the Standard Model exhibits CP-violation (e.g., in neutral kaon and B-meson systems). Under the usual assumptions of the CPT theorem, CP-violation implies TTT-violation: there is no pure time-reversal symmetry that maps the actual dynamics (and its states) to their TTT-images while holding the rest fixed. Put in IQT terms: there is generally no physically induced TTT-map that identifies (A(D),ω|A(D))(\mathcal{A}(D), \omega|_{\mathcal{A}(D)})(A(D),ω|A(D)) with its time-reversed counterpart as the *same quality* in the sense of §1.2. The asymmetry is therefore real at Level 0: a physical situation and its time-reversed counterfactual need not share a quality-equivalence class.

This point is conceptually clarifying but phenomenologically inert at effective neural scales. Whatever microphysical TTT-asymmetries exist, CP-violating corrections to the effective neural algebra are vanishingly small compared to the thermodynamic asymmetries that dominate macroscopic computation. The felt directionality of time traces not to weak-sector TTT-violation but to the thermodynamic arrow that constitutes record formation, error correction, and stable narration (§2.6). The layered picture is:

(i) **Microphysical asymmetry (Level 0):** time reversal is not a symmetry-equivalence in our world; a situation and its TTT-image are, in general, distinct in quality-equivalence class.

(ii) **Thermodynamic arrow (precondition for Level 2):** nonequilibrium boundary conditions make bounded, persistent computation—and therefore narration—physically realizable and robust in one temporal direction.

(iii) **Finite narrative bandwidth (proximate phenomenology):** the phenomenal character of passage is generated by lossy update over a finite integration window τ\tauτ, producing sequential shedding-and-acquiring along a self-thread.

These layers are separable in thought experiments. In a hypothetical universe with exact TTT-symmetry but the same thermodynamic boundary conditions, a situation and its TTT-image would be symmetry-equivalent at Level 0, yet narrators would still run in the thermodynamically selected direction and passage would still be felt (the texture would remain bandwidth-generated). In a universe with microphysical TTT-violation but global thermal equilibrium, the Level 0 asymmetry would remain while the macroscopic conditions for narration would fail to arise. The stratification therefore clarifies what is doing which explanatory job: microphysics can break symmetry-equivalence, thermodynamics makes narrators possible, and finite bandwidth makes passage feel like passage.

Figure 6: Thought-Experiment Universes — How the Layers Come Apart

The spectrum of qualia (speculative). Can specific algebraic structures be mapped to specific phenomenological structures? Status: speculative; long-term target.

This framework is sufficiently constrained to be wrong, sufficiently formal to be testable, and sufficiently novel to warrant investigation. Section 5 proposes concrete ways to find out.

Interpreting Section 7: What Comes Next

is mature enough for concrete experiments but leaves open problems across three fronts. Mathematical: the boundary data problem (how initial/final conditions constrain diamond quality), the presheaf cohomology conjecture (whether quality is locally determined), and rigorous construction of AQFT on causal sets. Computational: realistic approximations of the unity functional for biologically plausible system sizes, and tractable surrogates for comparisons with IIT. Empirical: execution of the three pre-registered protocols with appropriate sample sizes and analysis pipelines.

The strongest near-term opportunities are Protocol 1 (propofol-induced fragmentation of self-threads, which has the most straightforward clinical logistics and the clearest specific predictions) and neural-field simulations of the effective-diamond story (testing whether simulated coarse-graining reproduces the predicted multi-scale persistence profiles).

This is not a finished edifice—it is a scaffold designed to be tested, revised, or replaced. The theory is sufficiently constrained to be wrong, and sufficiently formal that being wrong would be informative. The open problems listed here are invitations, not admissions of defeat: each represents a well-posed research program where progress would illuminate not just but the science of consciousness more broadly.

Appendix A: Audience-Specific Reading Guides

A.1 For Mathematical Physicists

Constraint-based uniqueness argument (§1.1); quality as algebra-state pair under net-induced isomorphism with Type III discussion (§1.2); Intrinsicness Principle with worked example (§1.3); perspectival relativity as presheaf structure (§2.5); composition as extension/gluing problem (§3.2) with unity functional U(D) defined via quantum mutual information over minimum bipartitions (§3.3); connection to quantum marginal problem (§3.4). Status of formalism: §2.0.

A.2 For Neuroscience and Consciousness Science

Unity functional as concrete composition measure distinct from (§3.3). Narrative operator as control-theoretic sufficient statistic with worked pipeline (§4.5.1). Three pre-registerable protocols with CP/BH/OP layering (§5.0): Protocol 1 (anesthesia), Protocol 2 (overlap with tripartite O-information, predictive synergy, and single-pulse electrical stimulation arms), Protocol 3 (psychedelic temporal phenomenology with mechanistic toy model). Multi-scale persistence metric robust to anesthetic oscillatory confounds. Discriminative predictions against IIT (overlap) and GNW (experience without broadcast).

A.3 For Philosophers of Mind

Identity theory with Intrinsicness Principle and worked example (§1.3); honest epistemic status (§1.4); constraint dialectic (§1.1); dissolution of subject selection via three-level analysis (§4.4) with eliminativism response and deflationary democracy framing. Democracy of diamonds replaces binary subject/non-subject distinction. Positioning relative to Goff, Seager, and Russellian panqualityism (§6).

References

Albantakis, L., Barbosa, L., Findlay, G., et al. (2023). Integrated information theory (IIT) 4.0. PLoS Computational Biology, 19(10), e1011465.

Baars, B. J. (1988). A Cognitive Theory of Consciousness. Cambridge University Press.

Bombelli, L., Lee, J., Meyer, D., & Sorkin, R. (1987). Space-time as a causal set. Physical Review Letters, 59(5), 521-524.

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

Chalmers, D. J. (2015). Panpsychism and panprotopsychism. In T. Alter & Y. Nagasawa (Eds.), Consciousness in the Physical World. Oxford University Press.

Clark, A. (2013). Whatever next? Behavioral and Brain Sciences, 36(3), 181-204.

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

Dowker, F., & Glaser, L. (2013). Causal set d’Alembertians for various dimensions. Classical and Quantum Gravity, 30(19), 195016.

Friston, K. (2010). The free-energy principle: a unified brain theory? Nature Reviews Neuroscience, 11(2), 127-138.

Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P., & Pezzulo, G. (2017). Active inference: a process theory. Neural Computation, 29(1), 1-49.

Giacomini, F., Castro-Ruiz, E., & Brukner, C. (2019). Quantum mechanics and the covariance of physical laws in quantum reference frames. Nature Communications, 10, 494.

Goff, P. (2019). Galileo's Error: Foundations for a New Science of Consciousness. Pantheon Books.

Haag, R. (1996). Local Quantum Physics (2nd ed.). Springer.

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

Ince, R. A. A., Giordano, B. L., Kayser, C., Rousselet, G. A., Gross, J., & Schyns, P. G. (2017). A statistical framework for neuroimaging data analysis based on mutual information estimated via a Gaussian copula. Human Brain Mapping, 38(3), 1541-1573.

Johnston, S. (2008). Particle propagators on discrete spacetime. Classical and Quantum Gravity, 25(20), 202001.

Kleiner, J., & Tull, S. (2021). The mathematical structure of integrated information theory. Frontiers in Applied Mathematics and Statistics, 6, 602973.

Klyachko, A. A. (2004). Quantum marginal problem and representations of the symmetric group. arXiv:quant-ph/0409113.

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

Maudlin, T. (1989). Computation and consciousness. Journal of Philosophy, 86(8), 407-432.

Mediano, P. A. M., Rosas, F. E., Luppi, A. I., et al. (2021). Greater than the parts: A review of the information decomposition approach to causal emergence. Philosophical Transactions of the Royal Society A, 380, 20210246.

Muthukumaraswamy, S. D., Carhart-Harris, R. L., Moran, R. J., et al. (2013). Broadband cortical desynchronization underlies the human psychedelic state. Journal of Neuroscience, 33(38), 15171-15183.

Muthukumaraswamy, S. D., Shaw, A. D., Jackson, L. E., et al. (2015). Evidence that subanesthetic doses of ketamine cause sustained disruptions of NMDA and AMPA-mediated frontoparietal connectivity in humans. Journal of Neuroscience, 35(33), 11694-11706.

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

Oizumi, M., Albantakis, L., & Tononi, G. (2014). From the phenomenology to the mechanisms of consciousness: IIT 3.0. PLoS Computational Biology, 10(5), e1003588.

Polchinski, J. (1984). Renormalization and effective Lagrangians. Nuclear Physics B, 231(2), 269-295.

Preller, K. H., Razi, A., Zeidman, P., et al. (2019). Effective connectivity changes in LSD-induced altered states of consciousness in humans. Proceedings of the National Academy of Sciences, 116(7), 2743-2748.

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

Rosas, F. E., Mediano, P. A. M., Gastpar, M., & Jensen, H. J. (2019). Quantifying high-order interdependencies via multivariate extensions of the mutual information. Physical Review A, 100(3), 032310.

Rovelli, C. (1996). Relational quantum mechanics. International Journal of Theoretical Physics, 35(8), 1637-1678.

Russell, B. (1927). The Analysis of Matter. Kegan Paul.

Schilling, C., Gross, D., & Christandl, M. (2013). Pinning of fermionic occupation numbers. Physical Review Letters, 110(4), 040404.

Seager, W. (2020). The Routledge Handbook of Panpsychism (editor). Routledge.

Sorkin, R. D. (2003). Causal sets: Discrete gravity. In Lectures on Quantum Gravity. Springer.

Surya, S. (2019). The causal set approach to quantum gravity. Living Reviews in Relativity, 22, 5.

Timmermann, C., Roseman, L., Schartner, M., et al. (2019). Neural correlates of the DMT experience assessed with multivariate EEG. Scientific Reports, 9, 16324.

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

Tononi, G. (2008). Consciousness as integrated information. Biological Bulletin, 215(3), 216–242.

Tononi, G., Boly, M., Massimini, M., & Koch, C. (2016). Integrated information theory. Nature Reviews Neuroscience, 17(7), 450–461.

Vinck, M., Oostenveld, R., van Wingerden, M., Battaglia, F., & Pennartz, C. M. A. (2011). An improved index of phase-synchronization for electrophysiological data. NeuroImage, 55(4), 1548–1565.

Whitehead, A. N. (1929). Process and Reality. Macmillan.

Williams, P. L., & Beer, R. D. (2010). Nonnegative decomposition of multivariate information. arXiv:1004.2515.

Wilson, K. G. (1975). The renormalization group: Critical phenomena and the Kondo problem. Reviews of Modern Physics, 47(4), 773–840.