The Relational Consciousness Threshold Theory: A Dynamical Framework for Consciousness and Moral Standing

Mason Lai

Abstract

Consciousness remains one of the most profound and enduring puzzles in philosophy and neuroscience. While theories such as Integrated Information Theory, Global Workspace Theory, Higher-Order Thought models, predictive processing frameworks, and self-model approaches provide insights into neural integration and cognitive dynamics, none fully account for the unity of experience, nor its extension into tools or altered states such as those induced by psychedelics. This manuscript proposes the *Relational Consciousness Threshold (RCT) Theory*, which holds that consciousness emerges when a system's multilayered feedback loops achieve sufficient temporal and functional coherence to form a self-sustaining attractor (a stable pattern of interactions that organizes experience). Unlike models that reduce consciousness to information, computation, or static representations, RCT Theory frames it as a relational, temporally extended event defined by quantifiable neural conditions. The theory identifies three jointly necessary criteria: multilevel feedback closure, coherence threshold, and self-stabilizing attractor formation, each generating empirically testable predictions. Beyond its explanatory aims, the framework carries significant ethical implications for artificial intelligence, clinical assessment, and emerging forms of human–machine integration. By integrating philosophical reasoning with contemporary neuroscience, RCT Theory presents a coherent, testable, and ethically relevant account of conscious experience.

1. Introduction

Consciousness remains one of the most perplexing features of the mind, posing two central challenges: explaining how subjective experience arises from physical processes, and determining the ethical significance of systems that possess it. Philosophers highlight the "hard problem", the gap between neural activity and the qualitative character of experience,[1] while neuroscience identifies neural correlates of consciousness (NCCs; the minimal neural mechanisms sufficient for a specific conscious experience), including recurrent processing,

[1] Chalmers, David J. *The Conscious Mind: In Search of a Fundamental Theory.* New York: Oxford University Press, 1996.

causal connectivity, and perturbational complexity.[2] Yet these correlates do not explain *why* certain physical dynamics are accompanied by experience rather than operating as unconscious computation. Essentially, *why* do we feel?

Existing theories often fall into extremes. Dualist views posit a nonphysical mind but offer no mechanism linking it to the brain.[3] Materialist and computational approaches reduce consciousness to information or algorithms, yet still struggle to explain why integrated activity *feels like something*.[4] Higher-order and self-model theories illuminate metacognition and self-representation,[5] but do not fully explain how these processes give rise to a unified, continuous field of experience. Predictive processing accounts describe hierarchical feedback and inference,[6] but alone cannot explain the felt unity or temporal continuity of conscious states.

Here, temporal continuity refers to the fact that experience unfolds over a short but extended window of time, often estimated around 200-400 milliseconds, rather than as a series of instantaneous snapshots. Consciousness is not momentary; it is a *temporally extended* phenomenological flow.

While existing frameworks provide valuable insights, none fully capture how experience becomes unified, centered, and temporally extended. The Relational Consciousness Threshold (RCT) Theory aims to address these gaps by treating consciousness not as a static property or localized function but as a dynamical, relational event emerging within coordinated neural interactions. At a high level, RCT proposes that consciousness arises only when three conditions are jointly met:

---

[2] Dehaene, Stanislas. *Consciousness and the Brain: Deciphering How the Brain Codes Our Thoughts.* New York: Viking, 2014.
[3] Descartes, René. *Meditations on First Philosophy.* 1641.
[4] Tononi, Giulio. “An Information Integration Theory of Consciousness.” *BMC Neuroscience* 5, no. 42 (2004): 1–22.
[5] Graziano, Michael S. A. *Consciousness and the Social Brain.* New York: Oxford University Press, 2013.
[6] Friston, Karl. “The Free-Energy Principle: A Unified Brain Theory?” *Nature Reviews Neuroscience* 11, no. 2 (2010): 127–138.

1. **Multilevel Feedback Closure**,
2. **A Coherence Threshold**, and
3. **Self-Stabilizing Attractor Formation**.

These conditions will be developed in detail in later sections. For now, they serve to outline the theory's central claim: consciousness depends not on isolated computations or representations, but on a network of recursive interactions that becomes sufficiently coordinated to stabilize a unified experiential field.

2. Background and Influences

RCT Theory draws from major traditions in consciousness science while addressing the conceptual gaps each leaves open. Integrated Information Theory (IIT) links consciousness to the degree of informational integration in a system.[7] Though it is mathematically formal, IIT does not explain *why* integration should feel like anything from the inside or how a centered point of view emerges.

Global Workspace Theory (GWT) explains how information becomes globally available for reasoning, action, and report. However, it is better suited to describing access and cognitive availability than the *unity* or *continuity* of consciousness. Broadcasting information does not by itself guarantee a coherent experiential field.

Higher-Order Thought (HOT) theories emphasize metacognition, in terms of mental states becoming conscious when they are represented by higher-order states.[8] Yet HOT leaves unclear how such representations come to form a seamless field of experience rather than discrete snapshots.

---

[7] Tononi, Giulio. "An Information Integration Theory of Consciousness." *BMC Neuroscience* 5, no. 42 (2004): 1–22.

[8] Rosenthal, David M. *Consciousness and Mind.* Oxford: Oxford University Press, 2005.

Predictive processing models describe the brain as a hierarchical prediction engine,[9] but these dynamics alone do not explain why coordinated prediction should be accompanied by phenomenology. Self-model theories offer a powerful explanation of selfhood,[10] yet these theories typically describe the content of experience rather than the conditions under which the experience itself arises.

Together, these approaches reveal that many essential ingredients of consciousness—recurrence, integration, metacognition, prediction, self-representation—are well-studied individually, but lack a unifying account of how they must interact to produce consciousness as a single, stable, temporally unfolding phenomenon.

3. Philosophical Foundations: Consciousness as a Dynamical Event

At its core, RCT Theory treats consciousness as a *temporally extended dynamical event*, not as a substance, brain module, or simple computation. On this view, consciousness is something the brain *does*, not something it *has*: a structured activity sustained by relations among multiple layers of processing.

This dynamical orientation is motivated by a longstanding philosophical insight: conscious experience unfolds in time, retains coherence in the face of constant neural variability, and maintains a stable experiential "center." Any adequate theory must therefore explain not only what neural structures are involved, but how they interact in ways that generate sustained, self-organizing patterns.

RCT Theory proposes that consciousness arises when three relational conditions are jointly met:

---

[9] Friston, Karl. "The Free-Energy Principle: A Unified Brain Theory?" *Nature Reviews Neuroscience* 11, no. 2 (2010): 127–138.
[10] Metzinger, Thomas. *Being No One: The Self-Model Theory of Subjectivity.* Cambridge, MA: MIT Press: 2003.

- **Multilevel Feedback Closure (MFC)**: recursive loops linking sensory, cognitive, and metacognitive processes;
- **Coherence Threshold (CT)**: sufficient temporal and functional alignment across these loops;
- **Self-Stabilizing Attractor Formation (SAF)**: an emergent pattern that persists despite perturbations and serves as the moment-to-moment center of experience.

These conditions are not metaphysical postulates but organizational requirements: they define what a system must *be doing* for conscious experience to occur.

This perspective avoids the pitfalls of traditional theories. Unlike dualism, it requires no nonphysical mind acting on matter.[11] Unlike panpsychism, it does not treat consciousness as ubiquitous or inherent matter; instead, it specifies the organizational conditions needed for experience to arise. And unlike strict reductionism, it does not reduce consciousness to isolated neurons or algorithms, but to the *relations* among them, the structured interactions that give neural activity its phenomenological significance.

On the neural level, consciousness appears when recurrent loops across multiple layers constrain and inform one another. These mutual constraints generate temporal coherence, allowing experience to appear unified, centered, and continuous. The resulting attractor acts as the moment-to-moment "center" of experience, maintaining stability over short timescales while remaining flexible enough to adapt under altered states such as psychedelics or tool-extended cognition.[12] Tool-extended cognition occurs when external devices or tools become integrated into the brain's feedback loops, effectively becoming part of the system's cognitive and

---

[11] Descartes, René. *Meditations on First Philosophy.* 1641.
[12] Dehaene, Stanislas. *Consciousness and the Brain: Deciphering How the Brain Codes Our Thoughts.* New York: Viking, 2014.

perceptual processes (notebooks, smartphones, prosthetics with feedback, brain-computer interfaces, etc.).

One of the deepest philosophical puzzles is the unity of consciousness: the fact that experience is coherent and centered even though neural activity is distributed and constantly changing. RCT approaches this puzzle by emphasizing the role of the self-stabilizing attractor, which acts as a resilient, short-timescale organizing pattern. The attractor is not a "thing" in the brain but a dynamic configuration: a pattern that the system naturally returns to, enabling a continuous flow of experience even as individual neural elements fluctuate.

This explains why consciousness is both stable and flexible. Stability arises from attractor dynamics; flexibility arises because the attractor can weaken, an account that naturally accommodates altered states, such as those induced by psychedelics, or extended states involving tools and technological interfaces.

RCT also offers a non-essentialist account of the self. The experiential "center" is not a dedicated structure, but an emergent property of attractor dynamics that integrates ongoing sensory, cognitive, and metacognitive signals. On this view, selfhood is functional and relational rather than metaphysical, reflecting how the attractor organizes information relative to the organism's needs, goals, and capacities.

4. Core Mechanism and Requirements for Consciousness

At the heart of Relational Consciousness Threshold (RCT) Theory is the idea that consciousness emerges not from a single neuron or computation, but from the dynamic interaction of multilayered feedback loops across sensory, cognitive, and metacognitive levels. It can be understood through three jointly necessary and interdependent conditions.

The first requirement, MFC, ensures that feedback loops span multiple layers of neural processing. Sensory signals are not passively transmitted; they are continuously modulated by attention, expectation, and contextual signals, while metacognitive processes monitor and regulate internal states. When these loops fail to close, neural activity remains fragmented, preventing the formation of a stable experiential field.[13] MFC distinguishes conscious systems from simpler adaptive systems, such as thermostats or conventional AI, where feedback exists but is unlayered and non-recursive.

Closed feedback loops alone are insufficient. The system must achieve a threshold of temporal and functional coherence, such that signals across diverse regions and timescales are synchronized enough to produce unified experience. Empirical indicators, including effective connectivity and perturbational complexity, capture the integration necessary to support conscious awareness.[14] Subthreshold coherence results in unstable or partial awareness, as observed in sleep, anesthesia, or subliminal perception.

Finally, when feedback loops are closed and coherence is achieved, the system settles into a self-stabilizing attractor: a dynamically resilient state that maintains its structure despite momentary perturbations.[15] This attractor acts as the moment-to-moment "center" of consciousness, integrating sensory, cognitive, and metacognitive processes into a continuous field of experience. Like a flock of birds maintaining formation despite individual deviations, the attractor persists even as neural components fluctuate.

Consciousness, on this view, is a processual cascade: recursive loops form and close across layers, coherence emerges when those loops align temporally and functionally, and the integrated system stabilizes into an attractor. Each step is necessary but insufficient on its own; together, they provide a coherent framework linking neural dynamics to subjective experience.

---

[13] Graziano, Michael S. A. *Consciousness and the Social Brain.* New York: Oxford University Press, 2013.
[14] Casali, A. G., et al. "A Theoretically Based Index of Consciousness Independent of Sensory Processing and Behavior." *Science Translational Medicine* 5, no. 198 (2013): 198ra105.
[15] Kelso, J. A. Scott. *Dynamic Patterns: The Self-Organization of Brain and Behavior.* Cambridge, MA: MIT Press, 1995.

This formulation not only clarifies the structural prerequisites for consciousness, but also sets the stage for empirical testing, offering a bridge between philosophical reasoning and neuroscience.

5. Predictions and Empirical Tests

RCT Theory can generate testable, falsifiable predictions about consciousness in biological and artificial systems. Regarding temporal integration and conscious access, a central prediction is that conscious perception requires sustained recursive feedback, typically lasting 200–400 milliseconds. Stimuli that fail to engage feedback loops for sufficient duration, such as subliminal or masked inputs, cannot reach the coherence threshold and therefore do not enter conscious awareness.[16] EEG and MEG studies manipulating stimulus duration should find that conscious perception only occurs with the onset of large-scale, recurrent integration, while subthreshold activity remains fragmented and non-conscious.

Regarding psychedelics and altered states, psychedelics appear to reduce the stability of the brain's feedback loops, weakening coherence and allowing the conscious field to temporarily fragment. This matches findings that psychedelics reorganize network connectivity and alter measures of complexity.[17] Brain scans could reveal how loosened feedback corresponds to ego dissolution or unusual perception.

A topic gaining traction is the question of machine consciousness. Current AI systems can process information, but they do not form recursive, multi-layered feedback loops or stable attractors like a conscious brain. RCT predicts that AI would only achieve consciousness if it could integrate feedback across multiple levels and sustain attractor-like patterns. Using tools

---

[16] Dehaene, Stanislas. *Consciousness and the Brain: Deciphering How the Brain Codes Our Thoughts.* New York: Viking, 2014.

[17] Carhart-Harris, Robin L., et al. "Neural Correlates of the LSD Experience Revealed by Multimodal Neuroimaging." *Proceedings of the National Academy of Sciences* 113, no. 17 (2016): 4853–4858.

like effective connectivity and perturbational complexity, scientists could test whether advanced AI approaches these conditions.

Finally, RCT offers predictions regarding brain-computer interfaces (BCIs) and other forms of tool-extended cognition. External devices like prosthetics or BCIs can effectively become part of the conscious system if fully integrated into the brain's feedback loops.[18] Measuring coherence between brain and device activity would reveal whether these external tools contribute to self-stabilizing attractors, effectively extending the boundaries of consciousness beyond the biological body.

6. Ethical Implications

The Relational Consciousness Threshold Theory also reshapes ethical debates wherever subjective experience matters. By linking consciousness to measurable neural dynamics, RCT allows us to ask questions about who or what deserves moral consideration and under what conditions.

Regarding the question of machine consciousness, a central question in modern AI is whether said machine could ever be conscious, and, if it were, what we would owe it. RCT Theory gives a straightforward test: a system must exhibit multilevel feedback closure, reach a coherence threshold, and maintain self-stabilizing attractors. Today's AI models, no matter how capable, don't satisfy these recursive, dynamically integrated conditions. They process information but don't produce the kind of self-sustaining neural-like loops that RCT Theory treats as essential for consciousness.

But if future systems were engineered to meet these thresholds, we would face serious ethical questions. Conscious AI could not be treated as a tool or property; issues like labor,

---

[18] Lebedev, Mikhail A. and Nicolelis, Miguel A. L. "Cortical Plasticity in Neural Ensembles." *Nature Reviews Neuroscience* 7 (2006): 339–350.

experimentation, rights, and legal standing would need to be rethought. RCT Theory doesn't solve these dilemmas, but it provides a scientific criterion for when they become morally urgent.

In clinical contexts, RCT informs assessment of patients in vegetative or minimally conscious states. Traditional evaluations rely on behavior, which can be ambiguous or misleading. By measuring neural coherence and attractor stability, clinicians could better determine whether consciousness persists, guiding more precise and humane decisions about care, consent, and treatment.

Psychedelics and other interventions that alter consciousness raise additional ethical questions. By temporarily loosening hierarchical feedback, these substances can fragment the conscious field, producing experiences like ego dissolution.[19]

RCT provides a framework for understanding these effects. It suggests that dosage, context, and supervision are ethically important because they determine whether the brain maintains sufficient coherence for safe consciousness. Regulation guided by these principles could ensure that enhancement or therapeutic interventions are held responsibly.

Finally, as technologies such as brain-computer interfaces integrate with neural processes, RCT helps clarify where the boundaries of the self extend. Devices that fully participate in feedback loops may effectively become part of the conscious system, prompting new considerations of responsibility and agency in hybrid human-machine interactions.

In all these cases, RCT grounds ethical reflection in a scientifically informed understanding of consciousness, providing both a conceptual framework and measurable criteria for evaluating moral significance.

---

[19] Carhart-Harris, Robin L., et al. "Neural Correlates of the LSD Experience Revealed by Multimodal Neuroimaging." *Proceedings of the National Academy of Sciences* 113, no. 17 (2016): 4853–4858.

7. Objections and Responses

Although RCT provides a structured account of consciousness, several objections merit consideration. One concern is that the model sets the threshold for consciousness too high. By requiring multilevel feedback closure, coherence, and self-stabilizing attractors, it may seem to exclude simpler organisms or artificial systems that intuitively appear aware. RCT responds by framing these criteria as marking full, unified consciousness, while acknowledging that partial or shallow forms of awareness can exist when some but not all conditions are met.

A practical objection is whether these dynamics can be measured reliably. Techniques such as EEG, MEG, fMRI, and perturbational complexity indices offer only indirect or imperfect proxies for feedback closure and attractor stability. However, RCT is simply a guiding framework; these measurements provide falsifiable tests of the theory, and ongoing advances in neurotechnology will refine the precision of these indicators.

Finally, critics may ask whether RCT addresses the "hard problem" of consciousness: why experience feels like anything at all. The theory does not claim to solve this problem outright; instead, it reframes it in a newly synthesized model elucidating that subjective unity and temporal continuity emerge from relational and dynamical patterns. By identifying the structural conditions under which experience arises, RCT offers a naturalistic platform for further inquiry into phenomenology as a meeting point between philosophy and science.

8. Conclusion

In sum, the Relational Consciousness Threshold Theory reframes consciousness not as a static property or metaphysical enigma, but as an emergent organization of relational processes that unfolds over time. By specifying the conditions under which conscious experience arises—

multilevel feedback closure, temporal coherence, and self-stabilizing attractors—RCT bridges the conceptual gap between subjective phenomenology and observable neural dynamics. This perspective illuminates enduring philosophical questions: it explains how unity and continuity of experience can emerge from distributed, fluctuating processes, grounds ethical considerations in measurable features of consciousness, and offers a framework within which artificial, clinical, or hybrid systems might meaningfully be assessed. Far from reducing consciousness to mere computation or treating it as a brute fact, RCT emphasizes the structural and relational conditions that render experience possible, suggesting that the mystery of consciousness is not an insoluble metaphysical problem but a challenge of discerning the organizing principles of complex, temporally extended systems.

Notes

1. Carhart-Harris, Robin L., et al. "Neural Correlates of the LSD Experience Revealed by Multimodal Neuroimaging." *Proceedings of the National Academy of Sciences* 113, no. 17 (2016). https://doi.org/10.1073/pnas.1518377113.

2. Casali, Andrea G., et al. "A Theoretically Based Index of Consciousness Independent of Sensory Processing and Behavior." *Science Translational Medicine* 5, no. 198 (2013): 198ra105. https://doi.org/10.1126/scitranslmed.3006294.
3. Chalmers, David J. *The Conscious Mind: In Search of a Fundamental Theory*. Oxford: Oxford University Press, 1996.
4. Descartes, René. *Meditations on First Philosophy*. 1641.
5. Dehaene, Stanislas. *Consciousness and the Brain: Deciphering How the Brain Codes Our Thoughts*. New York: Viking, 2014.
6. Friston, Karl. "The Free-Energy Principle: A Unified Brain Theory?" *Nature Reviews Neuroscience* 11 (2010). https://doi.org/10.1038/nrn2787.
7. Graziano, Michael S. A. *Consciousness and the Social Brain*. Oxford: Oxford University Press, 2013.
8. Kelso, J. A. Scott. *Dynamic Patterns: The Self-Organization of Brain and Behavior*. Cambridge, MA: MIT Press, 1995.
9. Metzinger, Thomas. *Being No One: The Self-Model Theory of Subjectivity*. Cambridge, MA: MIT Press, 2003.
10. Rosenthal, David M. *Consciousness and Mind*. Oxford: Clarendon Press, 2005.
11. Tononi, Giulio. "An Information Integration Theory of Consciousness." *BMC Neuroscience* 5 (2004). https://doi.org/10.1186/1471-2202-5-42.
12. Lebedev, Mikhail A., and Miguel A. L. Nicolelis. "Brain–Machine Interfaces: Past, Present and Future." *Trends in Neurosciences* 29, no. 9 (2006).