Essay

Explaining Consciousness: Two Leading Neurological Models

D. John Doyle *

Posted Date: 10 March 2026

doi: 10.20944/preprints202603.0738.v1

Keywords: Integrated Information Theory (IIT); Global Workspace Theory (GWT); Human and Artificial Consciousness

 doi:10.20944/preprints202603.0738.v1

Essay

Explaining Consciousness: Two Leading Neurological Models
Running Title: Explaining Consciousness

D. John Doyle

Case Western Reserve University; djdoyle@hotmail.com

Abstract

The question of how consciousness arises from physical systems remains one of the most profound challenges in neuroscience and philosophy. This essay examines two leading models that attempt to explain the emergence of consciousness from both biological and synthetic neural networks: Integrated Information Theory (IIT) and Global Workspace Theory (GWT). Each offers a distinct approach—one grounded in intrinsic informational structure, the other in functional accessibility and cognitive architecture. By comparing their principles, empirical support, and criticisms, this essay aims to clarify how these models contribute to our understanding of consciousness and its potential replication in artificial systems. Recent adversarial testing reveals that both theories face substantial empirical challenges, suggesting the field may need to resolve fundamental conceptual questions before definitive adjudication between theories becomes possible.

Keywords: Integrated Information Theory (IIT); Global Workspace Theory (GWT); Human and Artificial Consciousness

---

1. Introduction

As artificial intelligence advances and neuroscience probes deeper into the workings of the brain, the question of how consciousness arises from physical substrates grows more urgent. Can machines ever be conscious? What makes the human brain capable of subjective experience? Two of the most influential theories attempting to answer these questions are Integrated Information Theory (IIT) and Global Workspace Theory (GWT). These theories differ fundamentally in their approach: IIT proposes that consciousness is an intrinsic property of certain information configurations, while GWT views consciousness as an emergent property of information broadcasting across cognitive modules.[1–6] Understanding their strengths, limitations, and empirical support is essential for advancing consciousness science and for evaluating the possibility of machine consciousness.

2. Integrated Information Theory (IIT)

Developed by Giulio Tononi and supported by Christof Koch, IIT proposes that consciousness corresponds to the degree of integrated information within a system.[1,2,5] This is quantified by a value called Φ (phi), which measures how much a system is both functionally differentiated and causally integrated. According to IIT, consciousness is not about behavior or outputs—it is about the internal structure of information. Regions of the brain, such as the posterior cortex, are believed to have high Φ and are thus linked to conscious states.[5,7]

From a synthetic perspective, if a machine or neural network achieved high Φ, it could—according to IIT—possess a form of consciousness. This position makes IIT one of the few theories that allows for machine consciousness independent of human-like behavior.

Major Criticisms and Challenges

However, IIT faces substantial theoretical and empirical challenges. A comprehensive critique by Merker et al. argues that IIT's Φ measure actually quantifies "the causal efficacy with which

 doi:10.20944/preprints202603.0738.v1

differentiated networks accomplish global information transfer" rather than consciousness itself.[1] This leads to what critics call the "panpsychism problem": IIT's framework would attribute consciousness to systems such as power grids, gene-regulation networks, and electronic circuit boards—entities that intuitively lack subjective experience.[1] This represents what critics characterize as a "functional misattribution" at the heart of IIT's identity claim between integrated information and consciousness.

Herzog et al. present the "Unfolding Argument," which suggests IIT leads to a form of "dissociative epiphenomenalism" where consciousness has no causal power—ironically undermining IIT's own axioms based on first-person experience.[2] If consciousness is merely a property of certain causal structures, it becomes unclear how it could influence behavior or cognition.

The computational intractability of calculating Φ for real neural systems represents another severe limitation. Even practical approximations have proven insufficient to discriminate certain states of consciousness without additional parameters.[3] Kim et al. found that estimating Φ from high-density EEG during different states of consciousness in humans remains extremely challenging, limiting IIT's empirical testability.[3]

Most critically, recent adversarial testing published in Nature found that IIT's predictions about sustained synchronization within the posterior cortex were not supported by intracranial recordings during conscious perception.[4] The lack of sustained network connectivity that IIT predicts should specify consciousness represents a significant empirical challenge to the theory's core claims.

3. Global Workspace Theory (GWT)

Originally proposed by Bernard Baars and later developed by Stanislas Dehaene and others, GWT views consciousness as a function of information broadcasting.[3,4,6] In this model, various cognitive modules (e.g., vision, memory, language) process information independently until a particular input is selected for global broadcast via a central 'workspace.' This broadcast makes the information accessible across the entire cognitive system, enabling conscious awareness.

Neuroscientific support for GWT includes brain imaging studies showing widespread activation, especially in the prefrontal cortex, during conscious perception.[6–10] GWT is attractive to AI researchers because it closely resembles certain architectures in machine learning, particularly attention mechanisms and transformer models.

Empirical Challenges and Theoretical Limitations

However, GWT faces its own substantial empirical challenges. The 2025 adversarial collaboration testing both IIT and GWT simultaneously found a "general lack of ignition at stimulus offset and limited representation of certain conscious dimensions in the prefrontal cortex".[4] This directly contradicts GWT's core prediction about widespread prefrontal activation during conscious perception. The absence of sustained "global ignition" in prefrontal regions challenges the theory's fundamental mechanism.

GWT is also criticized for explaining access consciousness—what we can report—without fully addressing phenomenal consciousness, or the subjective quality of experience (qualia).[11–13] The distinction between phenomenal consciousness (P-consciousness, or "what it is like" to have an experience) and access consciousness (A-consciousness, or the ability to report that experience) has been influential in consciousness studies.[11–14] Recent work suggests these may represent two necessary conditions for consciousness rather than two distinct types, but GWT primarily addresses the access component.[13]

Furthermore, the relationship between attention and consciousness in GWT is more complex than the theory initially suggested. GWT has been criticized for conflating these processes, and recent evidence suggests certain types of phenomenal consciousness may occur without the attentional mechanisms GWT requires.[5,6,15] Pitts et al. argue that attention and consciousness can be dissociated under certain experimental conditions, challenging GWT's assumption that attentional selection is necessary for conscious access.[5]

 doi:10.20944/preprints202603.0738.v1

4. Phenomenal Versus Access Consciousness: A Critical Distinction

Understanding the distinction between phenomenal and access consciousness is essential for evaluating both IIT and GWT. Phenomenal consciousness refers to the subjective, qualitative aspects of experience—the "what it is like" to see red, feel pain, or hear music.[11–14] Access consciousness refers to the cognitive availability of information for verbal report, reasoning, and behavioral control.[11,12]

Recent intracranial recordings reveal that consciousness-related neural activity shows a bimodal temporal distribution, with early activity (potentially corresponding to phenomenal consciousness) and late activity (potentially corresponding to access consciousness) that are largely anatomically separated—except in the lateral prefrontal cortex, which may link the two processes.[16] Early awareness-related activity appears in posterior sensory regions with short latencies, while late activity appears in prefrontal regions with longer latencies.[16]

This finding challenges both theories' predictions. For IIT, which emphasizes the posterior "hot zone" as the primary substrate of consciousness, the question arises: does high Φ in posterior cortex correspond only to phenomenal consciousness, or to consciousness more broadly? For GWT, which emphasizes prefrontal broadcasting, the early posterior activity suggests that some aspects of conscious experience may precede global workspace ignition.[16]

Some researchers argue that phenomenal and access consciousness should not be understood as two different types of consciousness, but as two necessary conditions for consciousness.[13] This conceptual shift may help reconcile theories that emphasize sensory processing (like IIT's focus on posterior cortex) with those that emphasize cognitive access (like GWT's focus on prefrontal broadcasting).

5. Comparative Analysis: Neuroanatomical and Functional Perspectives

IIT and GWT differ significantly in focus and methodology. IIT is structural and metaphysical, arguing that consciousness is an intrinsic property of certain information configurations. GWT, in contrast, is functional and cognitive, describing consciousness as an emergent property of information sharing and coordination across cognitive domains.

Neuroanatomical Evidence

The neuroanatomical evidence presents a complex picture. Koch et al. argue that the neural correlates of consciousness are primarily localized to a posterior cortical "hot zone" that includes sensory areas, rather than to a fronto-parietal network involved in task monitoring and reporting.[5,7] This supports IIT's emphasis on posterior cortex. However, computational modeling by Ihalainen et al. found that the most consistent out-of-sample predictions of the state of consciousness come from frontoparietal connections, suggesting that while the posterior hot zone is important for explaining the contrast between conscious awareness and unconsciousness, frontoparietal connectivity may be crucial for maintaining consciousness.[17]

Studies of patients with disorders of consciousness reveal that both posterior regions (particularly the posterior cingulate cortex) and frontoparietal connectivity are critical for consciousness.[18–20] The posterior cingulate cortex appears to function as a crucial hub, with its cross-hemispheric connectivity predicting levels of consciousness in traumatic brain injury patients.[18] Impaired consciousness is linked to changes in effective connectivity of the posterior cingulate cortex within the default mode network, with reduced self-inhibition and increased oscillations in this region correlating with decreased consciousness.[20]

Functional Mechanisms

While IIT may allow for consciousness in highly integrated machines with no external behavior, GWT emphasizes modular interaction and global accessibility, making it more applicable to computational modeling. However, neither theory fully resolves what philosopher David Chalmers termed "the hard problem of consciousness"—explaining why physical processes give rise to subjective experience at all.[21–24]

 doi:10.20944/preprints202603.0738.v1

IIT attempts to address the hard problem by proposing that integrated information is identical to consciousness, but critics argue this merely restates the problem rather than solving it.[1,2] GWT largely sidesteps the hard problem by focusing on the functional mechanisms of access consciousness, leaving the question of why these mechanisms produce subjective experience unanswered.[22,23]

Empirical Testing and Current Challenges

The 2025 adversarial collaboration represents a landmark attempt to empirically distinguish between IIT and GWT using intracranial recordings during visual awareness tasks.[4] The results were sobering for both theories: IIT's predictions about sustained posterior synchronization were not confirmed, and GWT's predictions about prefrontal global ignition were only partially supported. Both theories face substantial challenges in accounting for the full spatiotemporal dynamics of consciousness-related neural activity.

A comprehensive review by Mudrik et al. found that "at this stage, there is more controversy than agreement between the theories, pertaining to the most basic questions of what consciousness is, how to identify conscious states, and what is required from any theory of consciousness".[23] This suggests the field may need to resolve fundamental conceptual foundations before empirical testing can definitively adjudicate between theories.

6. Alternative Theoretical Perspectives

While IIT and GWT represent two of the most prominent theories, the landscape of consciousness science includes other influential approaches. Higher-Order Theories (HOT) propose that consciousness requires higher-order representations of first-order mental states—that is, we are conscious of a mental state when we have a thought about that state.[22] Recurrent Processing Theory (RPT) emphasizes recurrent connections between cortical areas rather than global broadcasting, suggesting that consciousness arises from sustained recurrent processing within sensory cortex.[22] Predictive Processing frameworks view consciousness as hierarchical inference and prediction error minimization, where conscious experience reflects the brain's best guess about the causes of sensory input.[22,25]

Each of these theories offers different insights into the mechanisms of consciousness, and future progress may require integrating elements from multiple theoretical frameworks rather than selecting a single winner.[22,23,26]

7. Conclusions

Integrated Information Theory and Global Workspace Theory offer two of the most sophisticated models for explaining consciousness in both biological and synthetic contexts. IIT emphasizes informational integration and proposes that consciousness is an intrinsic property of certain causal structures, with the posterior cortex serving as a primary substrate. GWT focuses on functional broadcasting and proposes that consciousness arises from global accessibility of information across cognitive modules, with prefrontal cortex playing a key role.

Both theories have generated substantial empirical research and have implications for the development of conscious AI and for deepening our understanding of the human mind. However, recent adversarial testing reveals that both face substantial empirical challenges. IIT's predictions about sustained synchronization within posterior cortex were not confirmed, and GWT's predictions about prefrontal global ignition were only partially supported.[4]

The distinction between phenomenal and access consciousness remains central to evaluating these theories. Recent intracranial recordings suggest a bimodal temporal distribution of consciousness-related activity, with early posterior activity potentially corresponding to phenomenal consciousness and late prefrontal activity potentially corresponding to access consciousness.[16] The lateral prefrontal cortex may serve as a critical link between these two processes.

 doi:10.20944/preprints202603.0738.v1

Neither theory currently provides a complete account of consciousness, and fundamental disagreements persist about what consciousness is, how to identify conscious states, and what is required from any theory of consciousness.[23] Both theories largely sidestep the hard problem of consciousness—explaining why physical processes give rise to subjective experience—though they do so in different ways.[21–24]

Future progress in consciousness science will likely require not only refined empirical testing but also resolution of fundamental conceptual questions. The field may benefit from integrating insights from multiple theoretical frameworks, including higher-order theories, recurrent processing theory, and predictive processing approaches.[22,26] Together, these efforts represent a growing scientific attempt to approach one of the oldest philosophical mysteries with empirical rigor, even as the ultimate nature of consciousness remains elusive.

References

1. Merker B, Williford K, Rudrauf D. The integrated information theory of consciousness: a case of mistaken identity. **Behav Brain Sci.** 2021;45:e41. doi:10.1017/S0140525X21000881.
2. Herzog MH, Schurger A, Doerig A. First-person experience cannot rescue causal structure theories from the unfolding argument. **Conscious Cogn.** 2022;98:103261. doi:10.1016/j.concog.2021.103261.
3. Kim H, Hudetz AG, Lee J, Mashour GA, Lee U. Estimating the integrated information measure Phi from high-density electroencephalography during states of consciousness in humans. **Front Hum Neurosci.** 2018;12:42. doi:10.3389/fnhum.2018.00042.
4. Ferrante O, Gorska-Klimowska U, Henin S, et al. Adversarial testing of global neuronal workspace and integrated information theories of consciousness. **Nature.** 2025;642(8066):133-142. doi:10.1038/s41586-025-08888-1.
5. Pitts MA, Lutsyshyna LA, Hillyard SA. The relationship between attention and consciousness: an expanded taxonomy and implications for "no-report" paradigms. **Philos Trans R Soc Lond B Biol Sci.** 2018;373(1755):20170348. doi:10.1098/rstb.2017.0348.
6. Mashour GA, Roelfsema P, Changeux JP, Dehaene S. Conscious processing and the global neuronal workspace hypothesis. **Neuron.** 2020;105(5):776-798. doi:10.1016/j.neuron.2020.01.026.
7. Koch C, Massimini M, Boly M, Tononi G. Neural correlates of consciousness: progress and problems. **Nat Rev Neurosci.** 2016;17(5):307-321. doi:10.1038/nrn.2016.22.
8. Pal D, Dean JG, Liu T, et al. Differential role of prefrontal and parietal cortices in controlling level of consciousness. **Curr Biol.** 2018;28(13):2145-2152.e5. doi:10.1016/j.cub.2018.05.025.
9. León-Domínguez U, León-Carrión J. Prefrontal neural dynamics in consciousness. **Neuropsychologia.** 2019;131:25-41. doi:10.1016/j.neuropsychologia.2019.05.018.
10. Panagiotaropoulos TI. An integrative view of the role of prefrontal cortex in consciousness. **Neuron.** 2024;112(10):1626-1641. doi:10.1016/j.neuron.2024.04.028.
11. Naccache L. Why and how access consciousness can account for phenomenal consciousness. **Philos Trans R Soc Lond B Biol Sci.** 2018;373(1755):20170357. doi:10.1098/rstb.2017.0357.
12. Overgaard M. Phenomenal consciousness and cognitive access. **Philos Trans R Soc Lond B Biol Sci.** 2018;373(1755):20170353. doi:10.1098/rstb.2017.0353.
13. Mudrik L, Faivre N, Pitts M, Schurger A. On a confusion about there being two types of consciousness. **Trends Cogn Sci.** 2025. doi:10.1016/j.tics.2025.11.012.
14. Amir YZ, Assaf Y, Yovel Y, Mudrik L. Experiencing without knowing? Empirical evidence for phenomenal consciousness without access. **Cognition.** 2023;238:105529. doi:10.1016/j.cognition.2023.105529.
15. Simione L, Di Pace E, Chiarella SG, Raffone A. Visual attention modulates phenomenal consciousness: evidence from a change detection study. **Front Psychol.** 2019;10:2150. doi:10.3389/fpsyg.2019.02150.
16. Fang Z, Dang Y, Li X, et al. Intracranial neural representation of phenomenal and access consciousness in the human brain. **Neuroimage.** 2024;297:120699. doi:10.1016/j.neuroimage.2024.120699.
17. Ihalainen R, Gosseries O, de Steen FV, et al. How hot is the hot zone? Computational modelling clarifies the role of parietal and frontoparietal connectivity during anaesthetic-induced loss of consciousness. **Neuroimage.** 2021;231:117841. doi:10.1016/j.neuroimage.2021.117841.

 doi:10.20944/preprints202603.0738.v1

18. Zhang H, Dai R, Qin P, et al. Posterior cingulate cross-hemispheric functional connectivity predicts the level of consciousness in traumatic brain injury. **Sci Rep.** 2017;7(1):387. doi:10.1038/s41598-017-00392-5.
19. Di Perri C, Bahri MA, Amico E, et al. Neural correlates of consciousness in patients who have emerged from a minimally conscious state: a cross-sectional multimodal imaging study. **Lancet Neurol.** 2016;15(8):830-842. doi:10.1016/S1474-4422(16)00111-3.
20. Crone JS, Schurz M, Höller Y, et al. Impaired consciousness is linked to changes in effective connectivity of the posterior cingulate cortex within the default mode network. **Neuroimage.** 2015;110:101-109. doi:10.1016/j.neuroimage.2015.01.037.
21. Tucker DM, Luu P, Johnson M. Neurophysiological mechanisms of implicit and explicit memory in the process of consciousness. **J Neurophysiol.** 2022;128(4):872-891. doi:10.1152/jn.00328.2022.
22. Seth AK, Bayne T. Theories of consciousness. **Nat Rev Neurosci.** 2022;23(7):439-452. doi:10.1038/s41583-022-00587-4.
23. Mudrik L, Boly M, Dehaene S, et al. Unpacking the complexities of consciousness: theories and reflections. **Neurosci Biobehav Rev.** 2025;170:106053. doi:10.1016/j.neubiorev.2025.106053.
24. Grossberg S. Towards solving the hard problem of consciousness: the varieties of brain resonances and the conscious experiences that they support. **Neural Netw.** 2017;87:38-95. doi:10.1016/j.neunet.2016.11.003.
25. Pennartz CMA. What is neurorepresentationalism? From neural activity and predictive processing to multi-level representations and consciousness. **Behav Brain Res.** 2022;432:113969. doi:10.1016/j.bbr.2022.113969.
26. Winters JJ. The temporally-integrated causality landscape: reconciling neuroscientific theories with the phenomenology of consciousness. **Front Hum Neurosci.** 2021;15:768459. doi:10.3389/fnhum.2021.768459.