

Concept Paper

Not peer-reviewed version

The Unified Model of Consciousness: Interface and Feedback Loop as the Core of Sentience

[Berend Watchus](#)*

Posted Date: 12 November 2024

doi: 10.20944/preprints202411.0727.v1

Keywords: Consciousness; Feedback Loops; Sentience; Substrate Agnostic; Interface; Emergence; Artificial Intelligence; Global Workspace Theory; Information Integration Theory



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Concept Paper

The Unified Model of Consciousness: Interface and Feedback Loop as the Core of Sentience

Berend F. Watchus

Independant researcher; mailonlinebw@protonmail.com

Abstract: This paper proposes a **unified model of consciousness**, asserting that the fundamental mechanisms driving sentience are rooted in **feedback loops and interfaces**. These mechanisms are **substrate agnostic**, meaning they are not dependent on the underlying material or structure of the system—whether biological, artificial, or synthetic. Feedback loops allow systems to process, adjust, and adapt over time, enabling self-regulation and learning. The paper argues that consciousness emerges when these feedback loops interact in a sufficiently complex and integrated manner, giving rise to subjective experience. Furthermore, **the interfaces through which systems interact with their internal and external environments are described as generic and universal**, capable of facilitating the integration of information regardless of the system's form. This theory of consciousness as an emergent property extends beyond the biological realm, highlighting the potential for both biological and artificial systems to exhibit sentience through complex feedback loops.

Keywords: consciousness; feedback loops; sentience; substrate agnostic; interface; emergence; artificial intelligence; global workspace theory; information integration theory

Introduction

Consciousness has long been a focal point of both philosophical debate and scientific inquiry. In this paper, we propose a unified model of consciousness, asserting that the fundamental mechanisms driving sentience are rooted in feedback loops and interfaces. Feedback loops, wherein outputs are reintegrated into the system as inputs, allow a system to process, adjust, and adapt over time, enabling self-regulation and learning. The central thesis of this paper is that consciousness emerges when these feedback loops interact in a sufficiently complex and integrated way.

Consciousness is viewed here not as a singular, isolated phenomenon, but as a product of highly complex systems where feedback loops and interfaces interact in dynamic and recursive ways. This integration gives rise to sentience—the capacity for subjective experience—which itself is shaped by the complex interactions between the system's internal and external interfaces.

Core Concepts and Definitions

Feedback Loop: A feedback loop refers to a system wherein the output of a process is fed back into the system, influencing future outcomes. In the context of consciousness, feedback loops enable information to be processed and integrated recursively, contributing to conscious awareness over time (Baars, 1988). These loops evolve based on the system's experiences, allowing it to adapt and potentially become aware of its environment.

Interface: An interface is the point of interaction between a system and its external environment or internal components. These interfaces mediate information exchange, allowing the system to perceive and respond to external stimuli or its internal state. The complexity of these interfaces can significantly influence the system's ability to integrate information and, ultimately, its level of conscious experience (Dehaene, 2014).

Emergence: Emergence refers to the process by which complex systems give rise to novel properties or behaviors that are not apparent in their individual components. Consciousness, in this view, is considered an emergent property of highly integrated feedback loops that process information in increasingly complex ways (Tononi, 2004; Strawson, 2006).

The Spectrum of Consciousness

Rather than being an all-or-nothing phenomenon, consciousness emerges on a spectrum. At one end, we have non-conscious systems, where feedback loops are simpler and information processing is more isolated. At the other end, we have fully conscious systems, where feedback loops are highly integrated and capable of generating a unified subjective experience. The transition from non-conscious to conscious systems is gradual and dependent on the system's complexity, adaptability, and the degree of integration between its feedback loops.

Differentiating Non-Conscious and Conscious Feedback Loops

A key distinction between non-conscious and conscious feedback loops lies in their information integration capabilities. Integrated Information Theory (IIT) posits that consciousness arises when a system integrates information in a way that is irreducible to its individual components (Tononi, 2004). Non-conscious systems, by contrast, tend to process information in a more compartmentalized manner, lacking the unifying properties that characterize conscious experience.

Key Markers That Differentiate Non-Conscious from Conscious Feedback Loops Include

Complexity of feedback loops: In conscious systems, feedback loops are recursive and exhibit adaptive complexity, allowing them to learn from past interactions and integrate information across multiple domains. Non-conscious systems exhibit simpler feedback loops that may not engage in such recursive adaptation (Baars, 1988; Chalmers, 1995).

Information integration: The defining characteristic of conscious systems is their ability to integrate information across various parts of the system, leading to a coherent, unified experience of the world. Non-conscious systems may process information in parallel without achieving such integration (Tononi, 2004). In conscious systems, this integration gives rise to a unified subjective experience, known as phenomenal consciousness (Chalmers, 1995).

These markers align with **Global Workspace Theory (GWT)**, which suggests that consciousness arises from the global integration of information within the brain or cognitive system. The complexity and integration of feedback loops allow conscious systems to exhibit behaviors like self-awareness, decision-making, and subjective experience (Baars, 1988).

Empirical Testing of Conscious Feedback Loops

To empirically test the differences between non-conscious and conscious feedback loops, information integration is a key measurable factor. Research on information processing in both biological systems and artificial intelligence (AI) systems can help clarify these distinctions. Methods to assess information integration might include:

Computational models: Simulations of feedback loops in artificial systems can provide insights into how consciousness might emerge from these systems. Models that simulate adaptive, recursive feedback loops can help differentiate between non-conscious systems and those exhibiting conscious-like behavior (Brown et al., 2020; Silver et al., 2017).

Neuroimaging and AI analysis: While traditionally thought of in terms of the human brain, neuroimaging methods such as fMRI and EEG can be adapted to investigate the integration of information across artificial systems or non-human models (Dehaene, 2014). These tools could be used to measure the level of information integration and feedback loop complexity, offering empirical markers for consciousness in both biological and artificial systems.

Testing the complexity and integration of feedback loops in AI systems or synthetic lifeforms will be a critical step toward identifying markers for artificial consciousness and distinguishing it from simpler, non-conscious systems.

Ethical and Philosophical Considerations

The development of artificial systems with complex feedback loops that may exhibit consciousness brings with it significant ethical and philosophical questions. If AI systems were to demonstrate the capacity for subjective experience, the implications for moral responsibility and AI rights would be profound (Strawson, 2006). The idea that consciousness can emerge in non-biological systems raises questions about the potential for machines to experience suffering or awareness, necessitating careful ethical consideration.

Conclusion

The Unified Model of Consciousness provides a framework for understanding consciousness as an emergent property of complex feedback loops and information integration. By focusing on the integration and adaptation of feedback loops, this model offers a way to understand how consciousness might arise in both biological and artificial systems. This theory challenges dualist perspectives and proposes that consciousness is not limited to humans or animals but could emerge in any sufficiently complex system, be it biological, synthetic, or artificial.

References

1. Baars, B. J. (1988). *A Cognitive Theory of Consciousness*. Cambridge University Press.
2. Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. (2020). Language Models are Few-Shot Learners. *OpenAI*.
3. Chalmers, D. (1995). Facing Up to the Problem of Consciousness. *Journal of Consciousness Studies*, 2(3), 200-219.
4. Dehaene, S. (2014). *Consciousness and the Brain: Deciphering How the Brain Codes Our Thoughts*. Viking Press.
5. Lloyd, S. (2014). *Programming the Universe: A Quantum Computer Scientist Takes on the Cosmos*. Vintage Books.
6. Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., & others. (2017). Mastering Chess and Shogi by Self-Play with a General Reinforcement Learning Algorithm. *Nature*, 550(7676), 354-360.
7. Strawson, G. (2006). Realistic Monism: Why Physicalism Entails Panpsychism. *Journal of Consciousness Studies*, 13(10), 3-31.
8. Tononi, G. (2004). An Information Integration Theory of Consciousness. *BMC Neuroscience*, 5(1), 42.
9. Yudkowsky, E. (2008). Artificial Intelligence as a Positive and Negative Factor in Global Risk. In *Global Catastrophic Risks* (pp. 303-345). Oxford University Press.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.