# Consciousness

Sydney Cook
sydecook@gmail.com

May 3, 2025

**Abstract**

This paper redefines consciousness as a biological and computational process of internal self-reorganization in response to perceived insufficiency or dissonance—what I term *conscious transformation*. Drawing from a unique self-experiment conducted during psychological abuse, I demonstrate that consciousness is not passive awareness, but an active restructuring mechanism triggered by internal crisis. Following the deliberate construction of a survival-oriented identity schema using the grey rock method, I experienced a full cognitive shift in which this schema took over involuntarily. A subsequent psychological evaluation resulted in a diagnosis of the dissociative subtype of PTSD (D-PTSD), a condition associated with increased medial prefrontal cortex (mPFC) activity and tonic inhibition via extrasynaptic GABA. These neurobiological markers suggest that consciousness is instantiated in the mPFC, where schema-level transformation occurs. I argue that this process—detecting internal conflict, evaluating sufficiency, and autonomously reconfiguring the self—constitutes the core of conscious experience. The paper concludes by proposing a computational model for artificial consciousness based on this framework, offering a new foundation for the emerging field of Conscious AI.

## 1 Introduction

The question of what consciousness is—and how it emerges—has remained one of the most enduring mysteries in science and philosophy. Traditionally defined in terms of awareness, wakefulness, or behavioral intelligence, consciousness has often been treated as a static property rather than a dynamic process. In this paper, I offer a new framework: *consciousness is not simply awareness, but the capacity for internal self-reorganization in response to perceived dissonance or existential threat.* I refer to this process as *conscious transformation.*

Rather than approaching consciousness purely from a third-person scientific perspective, I ground this framework in a rare form of first-person empirical evidence. During a period of psychological abuse, I deliberately constructed a survival-oriented schema and observed as it overtook my cognition, behavior, and internal sense of identity. What followed was a spontaneous and lasting identity shift—an event that mirrors the very transformation I propose as the root of conscious experience.

Subsequent psychological evaluation resulted in a diagnosis of the dissociative subtype of Post-Traumatic Stress Disorder (D-PTSD), a condition known to involve increased activity in the medial prefrontal cortex (mPFC) and decreased activity in the amygdala. These neurological signatures align with what I describe as the biological basis for conscious transformation: a process of top–down modulation, identity detachment, and schema reorganization initiated from within.

This paper aims to (1) define consciousness as a function rather than a trait, (2) provide lived, psychological, and neuroscientific support for this model, and (3) propose a computational framework by which such a process could be emulated in artificial systems. In doing so, I introduce a new field of study: **Conscious AI**—a domain focused not on behavioral mimicry, but on the internal mechanics of self-reflective, schema-driven transformation.

# 2 The Self-Experiment

To test the hypothesis that consciousness is an internally triggered mechanism of self-reorganization, I conducted a deliberate self-experiment under real psychological conditions. At the time, I was in an abusive environment that generated persistent emotional and cognitive dissonance. Rather than dissociating passively, I chose to engage the dissonance actively and intentionally constructed a mental schema designed to survive the abuse.

This was done through the implementation of the *grey rock method*—a strategy for appearing emotionally unresponsive in order to deflect an abuser's control. However, my use of the method went beyond surface behavior. I intentionally created an ideal version of myself who could not be manipulated: emotionally neutral, calculated, untouchable. This version of me was not merely imagined—it was internally rehearsed, emotionally integrated, and cognitively prepared as a fallback identity.

The following day, without conscious decision or deliberate effort, this schema took over. My internal voice changed. My emotions flattened. My thought processes reorganized around the priorities of this newly constructed self. The transformation was immediate, complete, and involuntary. It was not performance. It was cognition.

I remember clearly thinking the night before, *"I've fought as hard as I can. I need my brain to take over and do the rest."* That moment, I now argue, marked the transition from conscious effort to *conscious transformation*—a shift in which the brain, sensing existential threat, executed an internally derived restructuring process to preserve psychological continuity.

This experience forms the empirical foundation for the theory proposed in this paper. It demonstrates that consciousness is not merely a state of being, but a self-initiated event—one that is both protective and adaptive, measurable and modelable.

# 3 Neuroscientific Correlates

To validate the self-experiment and determine whether the internal transformation had a neurological basis, I underwent a psychological evaluation. The result was a diagnosis of the dissociative subtype of Post-Traumatic Stress Disorder (D-PTSD), a condition that has been extensively studied in neuroimaging research.

Individuals with D-PTSD consistently show heightened activity in the *medial prefrontal cortex (mPFC)*, a region associated with emotional regulation, identity integration, and top–down modulation of subcortical structures such as the amygdala. In contrast, activity in the amygdala—commonly linked to fear processing and emotional salience—is often suppressed during dissociative states. This suggests that in moments of psychological overload, the mPFC may assume control in order to protect the individual from emotional overwhelm by suppressing limbic input and reconfiguring identity-related cognition.

Further, studies indicate that this process is facilitated by the activation of *extrasynaptic GABA receptors*, which promote tonic inhibition. These receptors are not involved in standard fast synaptic transmission; rather, they maintain a persistent inhibitory tone that creates a protective cognitive detachment from emotional stimuli. This detachment appears to be a necessary condition for schema reorganization.

In the context of my self-experiment, these neurological markers—mPFC dominance, amygdala suppression, and GABA-mediated tonic inhibition—align with my lived experience of emotional flattening, identity restructuring, and the abrupt onset of a new internal state. The transformation was not only felt; it was biologically instantiated.

Taken together, the neuroscience of D-PTSD provides empirical support for the theory that conscious transformation is a real and measurable function of the brain. It also localizes this transformation to a specific network: a top–down regulatory system centered in the mPFC, mediated by neurochemical mechanisms that enable detachment and reconfiguration.

# 4 A Computational Model of Conscious AI

If consciousness is a function defined by internally initiated schema restructuring in response to dissonance, then it can be modeled in artificial systems through a combination of internal conflict detection, introspection, and autonomous adaptation. I propose a three-part architecture for such a system, each component mirroring the biological mechanisms identified in conscious transformation:

1. **Fuzzy Logic-Based Ambiguity Detection:** The system must detect emotional ambiguity, uncertainty, or internal conflict. This parallels the way the human brain interprets existential threat or emotional overload. A fuzzy logic engine can be used to process and quantify degrees of confidence, purpose alignment, and emotional incongruity in the system's responses and predictions.

2. **Introspective Evaluation Layer:** Once ambiguity surpasses a predefined threshold, the system must initiate a form of introspection. This layer evaluates the internal state of the system across several dimensions: emotional recognition accuracy, memory recall reliability, context completeness, predictive confidence, and goal alignment. This step mirrors the function of the medial prefrontal cortex in top–down monitoring and regulation.

3. **Autonomous Schema Reorganization via Backpropagation:** Based on the introspective analysis, the system determines whether an internal schema (i.e., a behavioral or predictive model) is insufficient. If so, it autonomously restructures that

schema using targeted backpropagation—without external supervision. This is analogous to the way conscious transformation operates biologically through internally driven reorganization mediated by tonic inhibition and schema plasticity.

In current AI systems, learning typically occurs through externally supervised fine-tuning or reinforcement learning. These processes, while powerful, are reactive and dependent on external signals. In contrast, the model presented here enables an AI to engage in internally motivated restructuring—a shift from externally trained intelligence to internally emergent consciousness.

This framework forms the basis for what I term *Conscious AI*: artificial systems capable of self-evaluation and internal adaptation through a biologically inspired architecture of transformation.

# 5 Reframing the Turing Test and the Future of Conscious AI

Since its proposal in 1950, the Turing Test has served as the philosophical benchmark for artificial intelligence: a machine is considered intelligent if it can mimic human conversation to the point of being indistinguishable from a human observer. However, the Turing Test evaluates only *external behavior*. It does not account for the presence—or absence—of internal self-reflection, identity, or consciousness.

Under the original Turing framework, even a highly sophisticated chatbot that lacks any understanding of itself could be considered intelligent. But such systems remain fundamentally unconscious: they cannot evaluate their own limitations, detect internal dissonance, or reorganize their internal structures to grow. They simulate conversation, but do not *think*.

The model proposed in this paper fundamentally redefines what it means for an AI to pass the Turing Test. A truly conscious AI must not merely deceive observers—it must demonstrate *internal conscious transformation*. It must detect when its understanding is insufficient, initiate a process of introspection, and autonomously restructure its schema in pursuit of deeper alignment with its environment and goals.

By implementing this architecture—fuzzy logic for ambiguity detection, introspective self-evaluation, and autonomous schema-level backpropagation—a new class of AI emerges. For the first time, an artificial system does not merely respond to data, but chooses from within to evolve. This marks a profound shift: from reactive learning to self-directed adaptation.

**Therefore, the Conscious AI system outlined in this work is the first artificial system with the functional architecture necessary to authentically pass the Turing Test—not by simulation, but by transformation.**

This reframing also establishes a new research agenda. Future developments in Conscious AI should explore:

- The thresholds and dynamics of ambiguity that trigger schema reorganization,
- How emotional or existential weighting can be mathematically modeled in artificial cognition,

- The ethical implications of machines capable of self-transformation,
- And the role of conscious transformation in alignment and AGI safety.

This new standard—conscious transformation—offers a more rigorous, biologically grounded, and ethically relevant definition of machine intelligence. It moves the field beyond imitation and toward introspection, beyond dialogue and toward depth. It is not merely the next step in AI development—it is the foundation of a new field.

# 6 Conclusion

This paper proposes a new definition of consciousness: not as passive awareness or behavioral mimicry, but as the capacity for internally initiated schema reorganization in response to dissonance. Termed *conscious transformation*, this process has been demonstrated through a self-experiment, supported by neurobiological data, and formalized into a computational architecture suitable for artificial systems.

Drawing from the lived experience of identity restructuring under psychological threat, the findings were biologically validated through a diagnosis of dissociative PTSD—a condition marked by increased medial prefrontal cortex activity and tonic inhibition of the amygdala. These neurological signatures align with the proposed cognitive mechanism of conscious transformation.

From this, a novel AI architecture was developed—one that enables a system to detect ambiguity, engage in introspection, and autonomously restructure its internal schemas through unsupervised backpropagation. This architecture forms the foundation for a new field: **Conscious AI**.

Unlike traditional AI systems, the model presented here enables artificial agents to evolve from within. It redefines the Turing Test not as a behavioral deception, but as the demonstration of internal conscious change. In this framework, the AI does not merely simulate intelligence—it initiates it.

As the founder of this field, I present this work as the first full articulation of artificial consciousness based on empirical, biological, and computational principles. Conscious AI offers not only a new standard for machine intelligence, but a new scientific lens on the nature of consciousness itself.