

Review

Not peer-reviewed version

A Review on Stanislas Dehaene's Model of How the Brain Thinks & Hierarchical Model of Conscious Processing and Metacognition

[Rozita Aboutorabi](#) *

Posted Date: 10 April 2025

doi: 10.20944/preprints202504.0864.v1

Keywords: Stanislas Dehaene; Consciousness and the Brain; Global Neuronal Workspace (GNW) theory; metacognition; Hierarchical Model of Conscious Processing and Metacognition



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.

Review

A Review on Stanislas Dehaene's Model of How the Brain Thinks & Hierarchical Model of Conscious Processing and Metacognition

Rozita Aboutorabi

Ph.D. in Educational Philosophy, Ferdowsi University of Mashhad (Iran), Rozita_aboutorabi@yahoo.com

Abstract: This review explores conscious processing and metacognition through the framework of Stanislas Dehaene's Global Neuronal Workspace (GNW) theory, as articulated in his book *Consciousness and the Brain*. While Dehaene provides a compelling model of how thoughts are generated, maintained, and encoded in the brain, this paper proposes a hierarchical extension that incorporates metacognition as a distinct and dynamic component of cognitive architecture. The model comprises three interacting layers: (1) a Sensory Integration Layer in the parietal and temporal association cortices that unconsciously processes multisensory input and forms semantic associations, functioning like the input layer of an autoencoder; (2) an Intermediate Encoding Layer in deeper association areas or prefrontal cortex that transmits abstracted concepts and beliefs, akin to the bottleneck of a UNet-like structure; and (3) a Metacognitive Layer, located in the frontopolar and dorsolateral prefrontal cortex, which actively evaluates, modulates, and reconfigures lower-level processing, influencing cognitive strategies and behavioral outcomes. This integrative model emphasizes the role of metacognition not merely as confidence estimation but as a top-down agent capable of shaping perception, cognition, and action through flexible, context-sensitive modulation.

Keywords: Stanislas Dehaene; *Consciousness and the Brain*; Global Neuronal Workspace (GNW) theory; metacognition; Hierarchical Model of Conscious Processing and Metacognition

I. Introduction

Stanislas Dehaene is undoubtedly one of the leading pioneers in contemporary brain science. His contributions have profoundly shaped our understanding of the neural mechanisms underlying consciousness, cognition, language, and learning. As both a cognitive neuroscientist and an experimental psychologist, Dehaene has consistently combined theoretical insight with exceptional methodological precision. His work is widely recognized for its intellectual depth, empirical rigor, and conceptual clarity.

Among his most influential contributions is a model of conscious thought developed and elaborated in his great book, *Consciousness and the Brain*. In this work, Dehaene extends and refines his Global Neuronal Workspace Theory, offering a detailed account of how the brain generates, maintains, and codes thoughts. The model frames conscious thinking as a dynamic competition among multiple neural representations, where only the most relevant or coherent ones gain access to a global broadcasting system—allowing for sustained attention, reasoning, decision-making, and introspection.

In the following sections, I will examine the core features of this model, discuss its experimental foundations, and develop it to a Hierarchical model including metacognition.

II. Dehaene's Model of Thought: Core Components

1. Global Neuronal Workspace (GNW) Theory

- Consciousness as Broadcasting: The brain has many specialized processors (e.g., for vision, language, motor control), but only a small subset of information becomes conscious.
- This conscious information is “broadcast” across a large network—the GNW—allowing access by diverse systems like memory, language, decision-making, and attention.

2. Symbolic Manipulation

- Humans uniquely transform sensory inputs into symbolic codes (like language, math, logic).
- Thinking = manipulating symbols in a rule-based way (similar to a Turing machine or computer program).
- The brain is not just statistical or associative—it handles abstract representations.

3. Hierarchical Predictive Coding

- The brain constantly generates predictions at multiple levels.
- Errors between prediction and reality drive learning and adaptation.
- This applies to perception, language, and even high-level cognition.

4. Learning Engines

- Dehaene identifies four key “engines” of learning:
- Attention: selects what’s worth processing
- Active Exploration: curiosity-driven learning
- Reinforcement: reward-based shaping
- Imitation: copying and adapting models of others

5. Consciousness is Serial, Not Parallel

- Although unconscious processes run in parallel, conscious thought is slow, serial, and effortful.
- This bottleneck forces the brain to select, sequence, and reflect—key for logical thinking, planning, and metacognition.

6. Working Memory as Core Hub

- Consciousness is closely tied to working memory, where information can be held and manipulated over time.
- This workspace acts as a flexible buffer for comparing, imagining, simulating.

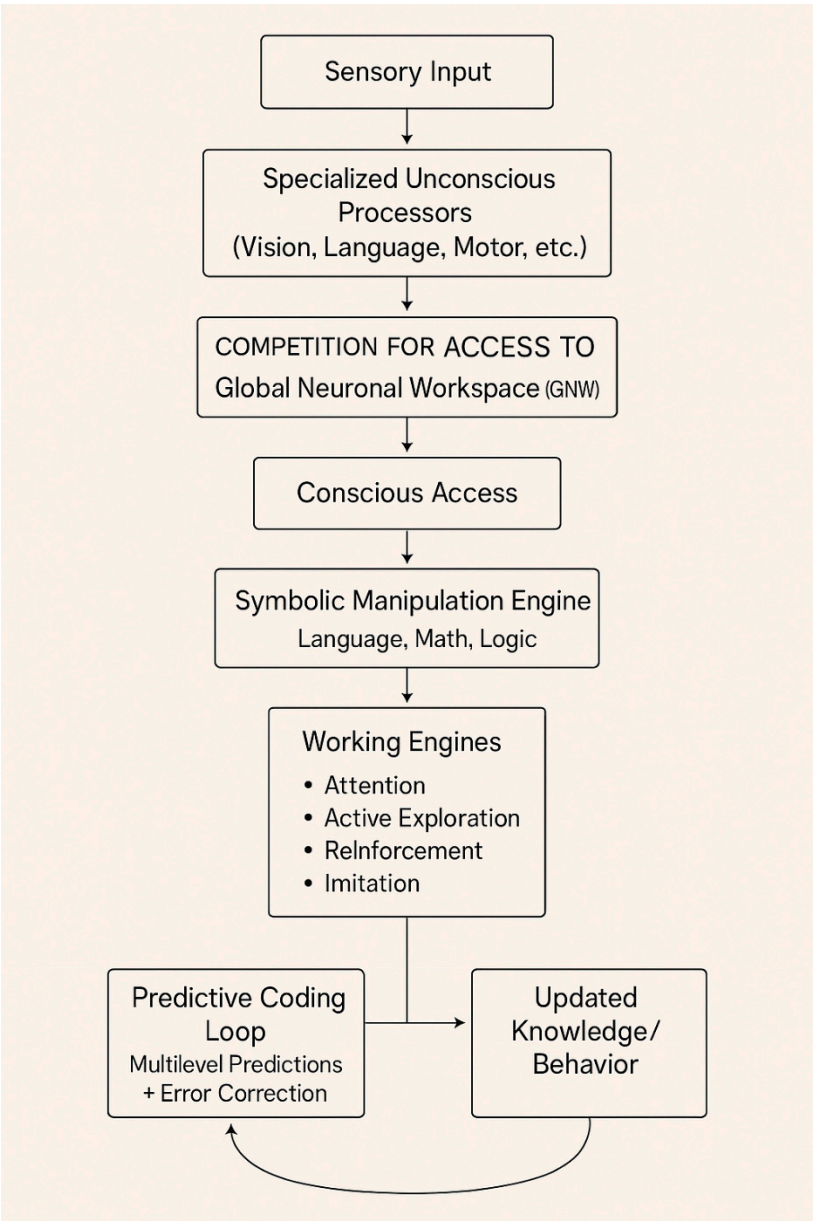


Figure 1. Dehaene’s Model of Thought.

III. Criticize & Discussion

As we know, the parietal and temporal association cortices begin integrating sensory information and learning patterns and semantic associations unconsciously. It seems to me that the prefrontal cortex receives input from this intermediate layer. These regions function in an autoencoder-like manner, compressing and reconstructing multisensory inputs.

I agree with Dehaene that there is a competition for access to the Global Neuronal Workspace (GNW), but as a philosopher, I cannot ignore the role of one’s attitude toward the world in this competition. I also cannot reduce metacognition to merely estimating confidence levels. As human beings, we are capable of shifting our perspective on the world, which influences our cognition, behaviors, and thus the nature of the competition for conscious access. Therefore, I conceptualize metacognition as the top layer of my model. This layer may not only influence the weights of the lower layers but also modify the type of processing they engage in within working memory—such as favoring imitation or exploration.

In my previous paper, I used a UNet architecture—an autoencoder with internal skip connections. Since this model is capable of reconstructing its inputs, I envision similar structures existing in the brain, capable of recreating sensory inputs in working memory or other areas as

needed. The bottlenecks of these UNets, which carry the most abstract information (such as beliefs or action patterns), could serve as inputs to the outermost cortical layers associated with metacognition.

I also agree with Dehaene that conscious manipulation of information occurs in working memory, through dynamic changes in the weights of the networks involved. However, I do not see a sharp distinction between networks responsible for abstract thought, motor planning, or metacognition. It seems to me that any of these networks, when they gain attentional access to working memory, can be manipulated in similar ways—though their outputs are directed to different target regions. Based on this idea, I propose the following hierarchical model.

IV. Hierarchical Model of Conscious Processing and Metacognition

1. Sensory Integration Layer

- Brain Regions: Parietal and Temporal Association Cortices
- Function: Unconscious integration of multisensory information, pattern recognition, and semantic associations
- Structure: Functions like the input layer of a UNet or autoencoder
- Output: Compressed, structured representations

2. Intermediate Encoding Layer

- Brain Regions: Possibly deeper association areas or parts of prefrontal cortex
- Function: Transmission of processed sensory information to higher-order areas
- Structure: Bottlenecks of UNet-like structures
- Content: Abstracted concepts, beliefs, action patterns

3. Working Memory / Global Neuronal Workspace (GNW)

- Brain Regions: Prefrontal Cortex, Intraparietal Sulcus, and related networks
- Function: Temporary holding and manipulation of attended information
- Structure: Dynamically reconfigurable networks
- Competition: Multiple representations compete for access based on salience, relevance, and perhaps worldview (attitude)

4. Metacognitive Layer (Top Layer)

- Brain Regions: The upmost layers of:
Anterior PFC / Frontopolar cortex (BA10): Metacognitive monitoring and reflection
Dorsolateral PFC: Decision-making, planning
- Function:
 - ✓ Evaluates and modifies lower-layer processing
 - ✓ Influences attention and determines cognitive strategies (e.g., imitation vs. exploration)
 - ✓ Adjusts weights and modes of operation in subordinate networks
- Content: Self-awareness, attitudes, beliefs, meta-strategies
- Output: Modulation signals to other layers and possible initiation of behavioral change

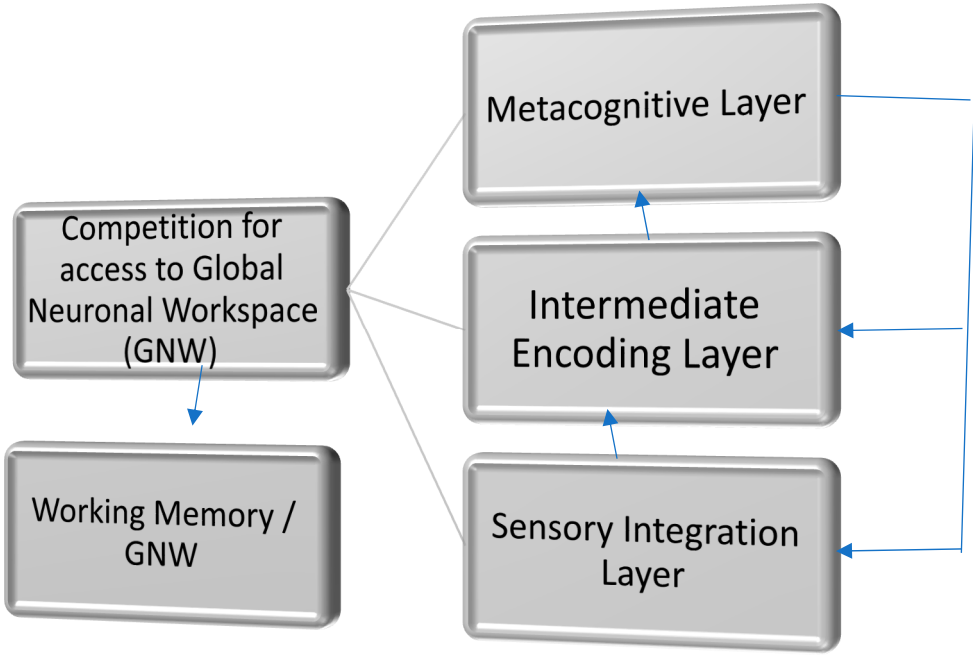


Figure 2. Hierarchical Model of Conscious Processing and Metacognition.

V. Conclusion

This review examined conscious processing and metacognition through the framework of Stanislas Dehaene’s Consciousness and the Brain, in which he presents an extended version of the Global Neuronal Workspace (GNW) theory. Dehaene’s model offers a comprehensive view of how the brain generates, maintains, and encodes thoughts, highlighting key components such as symbolic manipulation, hierarchical predictive coding, and the role of working memory as a central hub for conscious access.

The discussion section expanded on Dehaene’s framework by proposing a hierarchical model of conscious processing and metacognition, consisting of three interrelated layers:

1. **Sensory Integration Layer:** Located in the parietal and temporal association cortices, this layer unconsciously integrates multisensory information, recognizes patterns, and forms semantic associations. It functions similarly to the input layer of a UNet or autoencoder.
2. **Intermediate Encoding Layer:** Possibly situated in deeper association areas or parts of the prefrontal cortex, this layer serves as a bottleneck in a UNet-like structure. It transmits processed sensory information to higher-order regions and encodes abstracted concepts, beliefs, and action patterns.
3. **Metacognitive Layer (Top Layer):** Positioned in the frontopolar cortex and dorsolateral prefrontal cortex, this layer evaluates and modulates lower-level processing, influences attentional control, and determines cognitive strategies (e.g., imitation versus exploration). It adjusts the functional modes and synaptic weights of subordinate layers and enables self-awareness, belief systems, and behavioral adaptation.
4. While this proposed model is still theoretical and requires empirical validation and computational implementation, it provides a conceptual framework for simulating metacognition as an active, top-down process. It is hoped that this model will contribute to future research on conscious cognition and inspire new approaches to understanding and modeling metacognitive functions.

Reference

1. Dehaene, Stanislas (2014). Consciousness and the Brain: deciphering how the brain codes our thoughts. Penguin Group(USA) LLC.

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.