Conscious Prediction: Neural Architectures of Unified and Distributed Consciousness in Humans and Octopuses

Alex Horovitz
ahorovitz@mail.sfsu.edu

Spring Semester 2025

1 Introduction

The problem of consciousness—why and how subjective experience arises from physical processes—remains one of the most perplexing challenges in neuroscience and philosophy of mind. Although considerable progress has been made in identifying neural correlates of consciousness (NCC), a unified computational theory explaining the emergence of conscious awareness from neural activity is still lacking. Recent theoretical developments propose that consciousness may emerge from the brain's capacity to engage in hierarchical predictive processing, whereby sensory inputs are interpreted through continuously updated internal models (Clark, 2016; Friston, 2010).

In this framework, perception and action are not passive responses to external stimuli, but proactive, dynamic hypotheses about the world. This idea is formalized in predictive coding and the free energy principle, which suggest that the brain minimizes prediction error by adjusting its internal models and behaviors (Friston, 2010). Hierarchical Temporal Memory (HTM), a biologically inspired framework based on the architecture of the neocortex, offers a mechanistic implementation of this theory. Developed by Jeff Hawkins and colleagues, HTM models the neocortex as a hierarchy of cortical columns that learn sequences, form predictions, and integrate spatial and temporal patterns through sparse distributed representations (Ahmad & Hawkins, 2015; Hawkins & Ahmad, 2016; Hawkins et al., 2017).

HTM builds on Vernon Mountcastle's foundational proposal that the neocortex is composed of repeating canonical circuits organized in columns and layers, each performing similar computational functions across different regions (Mountcastle, 1997). These structures integrate changing inputs over time to

form a unified perceptual experience, a property fundamental to consciousness.

This paper explores how predictive architectures like HTM can account for the emergence of conscious experience in humans, and whether similar principles might apply to radically different biological systems. In particular, I consider the case of the octopus, an animal with a highly distributed nervous system where much of its neural processing occurs outside the central brain, particularly within its semi-autonomous arms (Carls-Diamante, 2022; Huffard, 2013; Mather, 2021). The octopus presents a compelling challenge to conventional models of consciousness that emphasize centralized, integrative hubs such as the human thalamocortical system.

I am here proposing that the octopus may exhibit a form of distributed temporal memory analogous to HTM, implemented across its decentralized neural modules. This comparative approach raises fundamental questions: Is a centralized integrator necessary for unified consciousness? Or can coherence emerge from interactions among distributed, embodied predictors? To answer these questions, this paper integrates insights from neuroscience, cognitive theory, and philosophy of mind to propose a cross-species framework for understanding the structural and computational requirements of consciousness.

2 The Predictive Brain and HTM

At the core of recent theories of perception and cognition lies the idea that the brain is fundamentally a prediction machine. Predictive processing, also known as predictive coding, posits that the brain continually generates and updates internal models to anticipate incoming sensory data (Clark, 2016; Friston, 2010). Rather than passively receiving sensory inputs, the brain actively predicts them, comparing its forecasts against actual sensory information. Any discrepancies—prediction errors—are used to adjust future predictions. This recursive loop of model-based inference allows an organism to efficiently interpret and respond to its environment.

Hierarchical Temporal Memory (HTM) is a computational framework developed to model this core principle of the brain's function. HTM networks consist of layers of cortical columns, each processing temporal sequences and forming contextual predictions. These columns interact laterally and hierarchically, mimicking the connectivity observed in biological neocortex. Each column in the HTM model learns transitions between spatial patterns over time, enabling it to predict what input is likely to occur next. In this sense, HTM does not rely on symbolic computation or hand-crafted features, but instead learns statistical regularities from raw input in an unsupervised, online manner.

One of the essential mechanisms in HTM is the use of Sparse Distributed Representations (SDRs). These are high-dimensional, binary representations in

which only a small percentage of bits are active at any time. SDRs offer significant advantages in terms of noise robustness, capacity, and semantic expressivity (Ahmad & Hawkins, 2015). Because each active bit carries meaning and overlaps between SDRs correspond to semantic similarity, they are well-suited for representing temporal context and probabilistic predictions.

Additionally, HTM neurons are modeled with active dendrites, following recent discoveries in neuroscience indicating that dendritic segments can perform non-linear computations independent of the soma (Hawkins & Ahmad, 2016). In the HTM model, synaptic connections on dendritic segments can detect contextual patterns from neighboring cells. When a dendritic segment becomes active, it slightly depolarizes the neuron, putting it in a predictive state. This mechanism enables individual cells to encode not just stimuli, but sequences of stimuli in specific contexts.

The spatial pooler is another key component of HTM. It converts noisy, high-dimensional input into sparse distributed representations, ensuring that similar inputs produce overlapping SDRs while preserving high capacity and robustness (Cui et al., 2017). This encoding allows the temporal memory component to recognize and predict sequences over time. Together, the spatial pooler and temporal memory form a complete architecture capable of unsupervised learning from streaming data.

Crucially, HTM's emphasis on temporal learning sets it apart from traditional deep learning models, which often require vast labeled datasets and are typically insensitive to temporal structure unless augmented with recurrent connections. HTM, by contrast, models cortical learning as inherently temporal and continuous, consistent with the needs of real biological systems. It can predict the next input in a sequence, anticipate anomalies, and generalize across different contexts without resetting or retraining.

The hierarchical nature of HTM further reflects the organization of the brain. Lower regions (or layers) learn simple features and short temporal sequences, while higher regions integrate these into more abstract, longer-range predictions. This mirrors the progressive abstraction observed in the visual and auditory cortices, where neurons respond to increasingly complex features at successive stages of processing (Clark, 2016; Hawkins et al., 2009).

In this way, HTM provides a compelling instantiation of the predictive brain theory. It offers a concrete, testable model of how neocortical circuits might support not only perception and behavior but also the coherent and continuous experience that characterizes consciousness. As neurons in different columns learn to predict specific inputs in specific contexts, the integration of their predictions across space and time supports a unified perception of the world—a key property of conscious awareness.

Moreover, HTM suggests that prediction is not just a function layered on top of sensory processing, but foundational to how the brain learns and understands

the world. As such, consciousness may not arise from higher-order representations alone, but from the nested interactions of predictive modules operating at different scales. These modules continuously compare expected and received signals, resolving ambiguity through inference and coordination—a structure that is both distributed and coherent.

This perspective opens the door to applying HTM-inspired models to non-human systems with radically different architectures. As I will explore in subsequent sections, the octopus offers a unique test case: a distributed nervous system with semi-autonomous limbs capable of complex, adaptive behavior. If prediction is the common currency of cognition, as HTM proposes, then perhaps consciousness does not require a centralized hub, but rather a coordination of predictions across modular systems.

3 HTM and the Unity of Consciousness

One of the enduring questions in consciousness studies concerns how a unified experience arises from distributed neural processes. Despite the modular and anatomically segregated organization of the neocortex, human experience is characteristically integrated: we perceive a coherent world, act with consistent agency, and maintain an uninterrupted sense of self. Hierarchical Temporal Memory (HTM), although primarily developed as a model of prediction and sequence memory, implicitly offers a computational explanation for how this unity might be achieved through the dynamic coordination of temporally organized predictions.

In HTM, each cortical column learns transitions between sensory inputs in its receptive field over time. While these local modules operate independently, they are not isolated; lateral connections allow columns to share information about their contextual predictions, resulting in mutual constraint and synchronization (Hawkins et al., 2017). This lateral integration is key to forming consistent representations of complex objects or environments that extend beyond the spatial or temporal scope of any individual column. When one column becomes confident in its prediction, this information is broadcast to neighboring columns, facilitating faster and more accurate inference across the network.

This mechanism echoes one of the central criteria for unity in consciousness: the integration of diverse contents into a single, coherent experience. Instead of requiring a central executive or a global workspace, HTM suggests that coherence can emerge from distributed, mutually predictive modules operating under shared statistical constraints. Conscious unity, in this view, is not a monolithic property imposed from above, but a dynamical outcome of consistent cross-prediction and representational alignment among temporally active subunits.

Furthermore, the temporal aspect of HTM learning provides an essential

bridge between momentary sensory events and the continuity of experience. Because each HTM neuron can learn multiple temporal contexts via distinct dendritic segments, the network can disambiguate identical inputs based on prior sequence history (Hawkins & Ahmad, 2016). This capacity to represent "what is happening now" in light of "what just happened" underpins the phenomenological flow of time—what William James referred to as the "stream of consciousness."

Importantly, this stream is not merely a sequence of events but a structured narrative shaped by predictive context. The HTM framework enables this narrative cohesion by encoding long-range dependencies and facilitating transitions across representational hierarchies. As higher regions of the HTM hierarchy integrate more abstract, longer-timescale sequences, they provide a scaffold for sustained goals, working memory, and attention—all of which contribute to the coherent organization of conscious contents.

The model also offers insights into how disruptions in prediction and integration might lead to fragmentation of consciousness, as observed in certain neuropathologies. For example, conditions such as schizophrenia have been interpreted through the lens of disrupted predictive coding, where an inability to properly assign precision to prediction errors leads to hallucinations and delusional beliefs (Clark, 2016). Within HTM, a similar failure might manifest as breakdowns in temporal memory, impaired cross-column consistency, or erratic activations across layers—each of which could correspond to disorganized or disjointed conscious states.

Moreover, HTM's approach offers a challenge to traditional views that place the locus of consciousness in a centralized structure such as the thalamocortical loop. Instead, it supports a decentralized but integrated account, where consciousness is distributed but unified through dynamic coordination. This resonates with the predictive processing view, where hierarchical prediction error minimization occurs at all levels of the brain's architecture and where global coherence is an emergent property of nested inferential loops (Friston, 2010).

The flexibility of HTM's architecture also accommodates attentional modulation. Attention, in predictive terms, involves the selective amplification of certain prediction errors while suppressing others. Although HTM does not currently implement an explicit attention mechanism, its sparse coding and competitive activation dynamics effectively result in selective representation. Neurons or columns that best match the contextual input inhibit less relevant alternatives, a process that mirrors attention's role in filtering and prioritizing sensory content (Cui et al., 2017).

This attentional filtering reinforces unity by limiting the contents of consciousness to those most contextually coherent and behaviorally relevant. The sparsity constraint ensures that, at any given moment, only a fraction of neurons are active—those best aligned with ongoing predictions and goals. These

selected representations form a "winning coalition" that constitutes the current conscious scene, providing continuity through time and coherence across modalities.

Ultimately, HTM presents a computational metaphor for the unity of consciousness that avoids central control structures or representational homunculi. Instead, it frames unity as an emergent pattern arising from multiple predictive circuits interacting across time and scale. This architecture supports both the differentiation of content (e.g., visual vs. tactile input) and the binding of these contents into a structured, temporally ordered whole. Such binding does not require a single integrative module but emerges from the recursive, hierarchical, and context-sensitive dynamics of the network itself.

As we consider non-human architectures in subsequent sections—particularly the decentralized nervous system of the octopus—this distributed-but-unified model offers a crucial theoretical tool. It allows for the possibility that systems with distinct structural organizations might still exhibit conscious unity, provided they implement mechanisms for temporal prediction and inter-module coordination. HTM thus serves not only as a model of cortical computation but also as a bridge between neurobiology and the phenomenology of conscious experience.

References

Ahmad, S., & Hawkins, J. (2015). Properties of sparse distributed representations and their application to hierarchical temporal memory [Technical Report, Numenta]. https://numenta.com/resources/white-papers/

Carls-Diamante, S. (2022). Where is it like to be an octopus? *Frontiers in Systems Neuroscience*, *16*, 840022. https://doi.org/10.3389/fnsys.2022.840022

Clark, A. (2016). *Surfing uncertainty: Prediction, action, and the embodied mind*. Oxford University Press.

Cui, Y., Ahmad, S., & Hawkins, J. (2017). The htm spatial pooler—a neocortical algorithm for online sparse distributed coding. *Frontiers in Computational Neuroscience*, *11*, 111. https://doi.org/10.3389/fncom.2017.00111

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127–138.

Hawkins, J., & Ahmad, S. (2016). Why neurons have thousands of synapses, a theory of sequence memory in neocortex. *Frontiers in Neural Circuits*, *10*, 23. https://doi.org/10.3389/fncir.2016.00023

Hawkins, J., Ahmad, S., & Cui, Y. (2017). A theory of how columns in the neocortex enable learning the structure of the world. *Frontiers in Neural Circuits*, *11*, 81. https://doi.org/10.3389/fncir.2017.00081

Hawkins, J., George, D., & Niemasik, J. (2009). Sequence memory for prediction, inference and behaviour. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1521), 1203–1209. https://doi.org/10.1098/rstb.2008.0322

Huffard, C. L. (2013). Cephalopod neurobiology: An introduction for biologists working in other model systems. *Invertebrate Neuroscience*, *13*, 11–18. https://doi.org/10.1007/s10158-013-0147-z

Mather, J. A. (2021). Octopus consciousness: The role of perceptual richness. *NeuroSci*, *2*(3), 276–290. https://doi.org/10.3390/neurosci2030020

Mountcastle, V. B. (1997). The columnar organization of the neocortex. *Brain*, *120*(4), 701–722.