Book Review

William A. Phillips, *The Cooperative Neuron: Cellular Foundations of Mental Life*

Oxford: Oxford University Press, 2023. Pp. 384.

William A. Phillips' *The Cooperative Neuron* is a work of rare ambition. Phillips shifts neuroscience's explanatory focus from systems and networks back to the single cell, arguing that the computational complexity required for cognition is built into the architecture of the neocortical pyramidal neuron itself. He frames this as "cellular psychology," a field analogous to molecular biology in scope and transformative potential (Chapter 1). Just as DNA revealed that molecules could convey vast amounts of information—far more than previously thought possible for molecules in general—context-sensitive pyramidal neurons reveal that individual cells can selectively amplify relevant signals through cooperative agreement between input streams. The book's scope is breathtaking. Phillips traces these mechanisms from evolutionary origins in mammalian neocortex through their role in consciousness, attention, and working memory; their malfunction in mental conditions and even the framework's implications for machine learning. He concludes with reflections on knowledge, self-identity, and even life's purposes (!). Thus, the book's ambition is both its strength but also risks stretching the explanatory purchase of contextual cooperation of neurons too thin.

Phillips convincingly argues for cooperation on multiple levels: neurons cooperate with local circuits through context-sensitive gain modulation; apical and basal dendritic compartments cooperate through coincidence detection; and populations cooperate by amplifying coherent signals while suppressing noise. According to him, this complex web of neuronal contextual cooperation explains the reduction of cognitive conflict[1] and production of coherent percepts, thoughts, and actions adapted to current goals and needs. But whether one mechanism can bear such varied explanatory weight still remains to be demonstrated.

Phillips begins by dismantling the point neuron dogma—the assumption that neurons are single-point integrators summing inputs indiscriminately. This "dendritic democracy" implies all synaptic inputs compete equally to determine output. Phillips argues this architecture is computationally blind: it collapses the fundamental distinction between evidence for a proposition and the context determining its relevance (Chapter 2). In its place, he establishes the two-point processor model, grounded in structural asymmetry between basal dendrites (processing feedforward driving inputs) and apical dendrites (processing feedback modulating inputs) (Chapter 3). This segregation allows neurons to amplify relevant signals while preserving what they represent. For Phillips, these biological streams map directly onto the philosophical categories of content and context: "content" is the primary feature the neuron detects via feedforward input, while "context" is the broader cognitive state or top-down expectation that determines the feature's current relevance. However, it is worth noting early on that while Phillips meticulously operationalizes these concepts biologically, the translation from mechanistic "content" to semantic "content" remains a tension that philosophers will need to carefully unpack.

Crucially, this architecture instantiates cooperation at the cellular level (Chapter 3). Neurons seek agreement between input streams, amplifying transmission when feedforward content and feedback context cohere, attenuating it when they conflict. The mechanism is coincidence detection: when apical and basal inputs arrive within approximately 15 milliseconds, the neuron undergoes a phase transition from isolated spiking to high-frequency bursting via back-propagation activated calcium firing. This burst serves as a physical token of validation—a mechanical adjudication of coherence based solely on causal timing. Agreement-seeking reduces incoherence and maximizes coherence of the system's overall activity. Phillips further shows how mental states—wakefulness, sleep, dreaming—depend on neuromodulators regulating apical function, with general anesthetics operating by blocking apical inputs from influencing neuronal output (Chapter 4).

What makes neurons "cooperative"? Phillips formalizes this through Coherent Infomax, a theory developed with statistician Jim Kay over decades (Chapter 8). Standard "infomax" theories propose neural systems maximize total information transmission, preserving sensory input while reducing redundancy. Coherent Infomax differs fundamentally: it maximizes *coherent* information—that is predictably related across multiple input streams—rather than total transmitted information. As Phillips emphasizes, "the difference between these two objectives may seem tiny when expressed in these symbolic terms, but the differences in their functional consequences are huge" (p. 276).

This distinction is formalized through partial information decomposition, which distinguishes how different inputs contribute to output. Rather than detailing the mathematical minutiae of partial information decomposition, it is enough to grasp its functional consequence: Phillips uses this formal analysis to prove that context-sensitive modulation has a distinctive biological signature. The PID framework demonstrates that apical input (context) contributes almost no unique information to the neuron's primary output. Instead, it selectively amplifies the signal already present in the basal input (content)—boosting the transmission of coherent signals while suppressing isolated noise. The two-point neuron physically implements selective amplification, seeking "agreement" between content and context to determine what merits high-salience transmission.

The framework's explanatory reach extends across multiple domains. Phillips argues conscious experience depends critically on context-sensitive pyramidal cells (Chapter 5), with selection between alternative possibilities as a cardinal function—apical amplification determines which among competing percepts, thoughts, or actions becomes part of conscious experience. He relates this to recurrent processing theory, global workspace theory, and integrated information theory, while implicating apical function in selective attention, working memory, and emotional prioritization. The evolutionary story (Chapter 6) traces context-sensitive neurons to six-layered mammalian neocortex, distinguishing them from three-layered reptilian allocortex and suggesting they underlie mammals' distinctive cognitive flexibility. Developmentally, apical dendrites grow and differentiate postnatally with experience-dependent regulation. Phillips demonstrates that apical dysfunction is implicated in diverse mental illnesses (Chapter 7): epileptic loss of consciousness, schizophrenia's disrupted context-sensitivity, anti-NMDA encephalitis, autism spectrum disorders, and foetal alcohol spectrum disorders.

The framework is also considered in regards to machine learning, speculating that implementing principles of contextual cooperation of neurons may prove more efficient than some standard deep learning algorithms (Chapter 8), but this is probably the most tentative part of the book. More interestingly, Phillips' framework exemplifies how biological detail can adjudicate philosophical debates and move them forward.

Consider predictive processing (PP) theories, championed by figures like Karl Friston and Andy Clark. In standard PP models, a central assumption is that top-down predictions "explain away" or suppress matching bottom-up inputs, leaving only prediction errors to propagate forward. Phillips demonstrates that this interpretation is neurobiologically implausible: apical amplification *strengthens* rather than silences feedforward outputs when top-down context agrees with bottom-up input. He reinterprets "error coding" as precision weighting—context modulates gain without corrupting semantic identity. The apical dendrite increases response to matching input rather than silencing it, suggesting prediction highlights rather than erases the world.

This biological intervention bears on broader debates about representation and error detection. Philosophers like Krystyna Bielecka and Marcin Miłkowski have recently argued that naturalized representation requires comparing independent information sources to flag inconsistencies.[2] However, as Mark Bickhard and Romain Brette note, this risks circularity: how does the system know different sources (vision, proprioception) should agree without already possessing semantic knowledge about their shared referent?[3] Phillips' framework suggests a mechanistic answer—physical synchronization of inputs rather than logical comparison of semantic content. When apical and basal streams coincide temporally, the calcium spike validates; when they fail to coincide, the signal attenuates. Whether this solves or merely relocates the semantic presupposition problem remains philosophically contentious, but it provides concrete biophysical detail about how error detection might work without semantic homunculi. This work has been further developed philosophically by Tomáš Marvan and colleagues, who argue apical amplification is a conscious-status awarding mechanism.[4] Marvan's approach exemplifies the call to action this book demands: philosophers taking specific cellular mechanisms and systematically connecting them, piece by piece, to existing philosophical literature on e.g. consciousness, representation or mental content.[4]

Despite its revolutionary potential, the framework still leaves important philosophical details unresolved. Phillips does admirably try to define his terms carefully within information theory and neuroscience. For instance, in Section 1.4, he characterizes "context-sensitivity" as "modifications of the strength or salience of a signal without corrupting the information that the signal transmits" (p. 18), a principle recurring throughout his formal framework. However, he does not connect these technical definitions to philosophical debates about representation, meaning, and intentionality. His operationalizations of "content," "context," and "information" are precisely specified within information-theoretic frameworks but remain unlinked to existing theories in semantics and philosophy of mind. Additionally, questions of constitutive relevance remain. Phillips provides an exhaustive description of entities (calcium channels, NMDA receptors, VIP interneurons) and activities (disinhibition, bursting, apical amplification). But does coincidence detection within a 15-millisecond window *constitute* attention, or merely *enable* it? The specific contributions of, say, calcium spikes versus burst firing versus sustained depolarization to phenomenologically distinct processes like attention, working memory, and cognitive control require further mapping. These are productive gaps inviting mechanistic decomposition work.

Nonetheless, the main philosophical question is what grounds semantic content in this framework. Phillips presupposes neural outputs possess representational identity distinct from signal strength—something that can be preserved or corrupted. But how do coincidence detection patterns and calcium spikes acquire this semantic property? The transition from "selective sensitivity to X" to "representation of X with correctness conditions" remains underspecified. It could be forgiven if it were a strictly neurobiological project. But, given Phillips' explicit ambition to explain "mental life" from cellular foundations, the unaddressed metasemantic foundations are particularly salient. On the plus side, it presents a challenge to philosophers - determining whether the bridge between cellular mechanisms and philosophical theories of mental content can be successfully built.

The Cooperative Neuron is a tour de force restoring the biological neuron to the center of cognitive science. Phillips provides neurobiological foundations for understanding context-sensitivity, coherence-seeking, and mental process coordination. The book offers philosophers concrete cellular nuts and bolts: important details that could be used in grounding debates on content, representation and prediction in biological reality. Substantial work remains, but Phillips has provided a fantastic foundation. For professional philosophers with some understanding and interest in neuroscience, as well as neuroscientists who are more inclined to connect the biological level to mental phenomena, this book is essential reading. It transforms how we should think about its problems, while providing an extensive mechanistic roadmap. The future of philosophy of mind must take the cooperative neuron seriously.

NOTES

1. Danek, A.H. & Flanagin, V.L. (2019). Cognitive conflict and restructuring: The neural basis of two core components of insight. https://pmc.ncbi.nlm.nih.gov/articles/PMC7179339/

2. See Bielecka, K., & Miłkowski, M. (2020). Error Detection and Representational Mechanisms. W J. Smortchkova, K. Dołęga, & T. Schlicht (Red.), *What are Mental Representations?* (pp. 287–313). Oxford University Press. For Miłkowski's recent work on this topic, see: Miłkowski, M. (2025) Interactivism mechanized: bridging the gap between cognition, correspondence, and computation. *Phenom Cogn Sci* . https://doi.org/10.1007/s11097-025-10082-7

3. Bickhard, M. "The Interactivist Model," *Synthese* 166 (2009): 547–591; Brette, R., "Is Coding a Relevant Metaphor for the Brain?" *Behavioral and Brain Sciences* 42 (2019).

4. Marvan, T., Polák, M. Bachmann, T., & Phillips, W.A. "Apical Amplification—A Cellular Mechanism of Conscious Perception?" *Neuroscience of Consciousness* 2021, no. 1:. See also: Marvan, T., & Phillips, W. A. (2024). Cellular mechanisms of cooperative context-sensitive predictive inference. *Current*

Research in Neurobiology, *6*: 100–129.