Integrating mammalian and non-mammalian consciousness research: Metamodels and instantiation models of consciousness

My aim is to identify and resolve a tension in contemporary consciousness science. On the one hand, mainstream work on the neural basis of consciousness—driven largely by studies of humans and other mammals—often treats processes in the cortex as crucial, and perhaps even necessary, for conscious experience (Malach, 2022; Michel, 2022). On the other hand, comparative work on non-mammalian animals—especially fish and invertebrates—has amassed increasing behavioral evidence that many such animals are conscious despite lacking a neocortex (e.g. Crook, 2021; Gibbons et al., 2022). Taken at face value, these two streams of research generate an uncomfortable choice: either downgrade the neuroscientific case for cortico-centrism, or downgrade the behavioral case for widespread non-mammalian consciousness. The guiding thought of this paper is that we should resist that forced choice. The methodological aim is not to dissolve the tension by rejecting one side, but to develop a framework in which both kinds of evidence can retain their probative force while genuinely constraining theorizing about consciousness.

Proposals that assign different neural substrates to different "levels" of consciousness can appear to dissolve the tension (Newen & Montemayor, 2023): perhaps subcortical processes suffice for a minimal form of awareness while cortical broadcasting is required for richer, more cognitively integrated consciousness. The paper argues that this move often underdelivers, because it does not answer the real point of friction: *which processes are sufficient for phenomenal consciousness*, and how do we test that sufficiency? Merely noting that there are fast subcortical control loops and thorough cortical integrations does not by itself reconcile evidence for and against non-cortical consciousness. Without a principled bridge between levels and the evidential roles they are meant to play, a two-tier theory risks collapsing into either a cortico-centric view (if the "minimal" level is not genuinely phenomenal) or a non-cortical view (if it is).

Third, two broad reconciliation strategies in the literature—call them generality and distinct realization—each capture something right, but each also faces a serious obstacle when taken alone. The generality strategy seeks a highly abstract functional characterization of the consciousness-enabling architecture: one property F that is realized by the neocortex in mammals but by different anatomical structures elsewhere. This promises unification across species, yet it invites the familiar worry that overly abstract functional conditions become too permissive (the "small network" problem) and too thin to support mechanistic explanation. Conversely, the distinct realization strategy emphasizes that consciousness may be

implemented by very different mechanisms in different taxa, so mammalian neural constraints need not generalize. But this threatens to sever the mutual constraints that make comparative research scientifically fruitful. If mammalian and non-mammalian research proceed in near-independence, it becomes unclear how either can robustly inform the other.

The paper's positive proposal is to combine what is right in both strategies by explicitly developing two linked model types:

1. Metamodels are *taxon-spanning* models stated in abstract functional terms. They aim to capture the shared organizational features of whatever realizes consciousness across the range of conscious animals (or across a large subset of them). Metamodels prioritize unification: they articulate what kinds of roles, interactions, or architectures must be present *somewhere* in a conscious system, even if different species implement those roles with different anatomical parts.
2. Instantiation models are *taxon-specific* mechanistic models. They aim to specify the particular neural components, activities, and organization that realize consciousness in a given group (e.g., mammals, birds, basal vertebrates, arthropods, cephalopods). Instantiation models prioritize explanatory depth: they should support interventionist "what-if-things-had-been-different" counterfactuals and connect consciousness claims to concrete neural mechanisms.

Crucially, the proposal is not merely to have both kinds of models, but to treat them as mutually constraining. Metamodels guide instantiation-model construction by indicating what functional structures to look for when mapping candidate mechanisms in a new taxon. Instantiation models, in turn, constrain and refine metamodels by revealing which features are genuinely shared across taxa and which were mammal-centric artifacts of an initial model. Integration is therefore iterative: theorists should expect a co-evolution of increasingly detailed instantiation models and increasingly taxon-general metamodels, with adjustments driven by familiar scientific virtues—explanatory power, predictive success, and unification.

To demonstrate that this is more than a terminological reshuffle, the paper develops a worked example centered on the relationship between Unlimited Associative Learning (UAL) (Ginsburg & Jablonka, 2019) and Global Neuronal Workspace Theory (GNWT) (Dehaene, 2014). UAL functions as a promising *metamodel*: it characterizes a minimal architecture that supports flexible, open-ended associative learning through interactions among sensory processing, valuation, memory, motor control, and a central associating workspace-like hub. GNWT functions as a paradigmatic *instantiation model* for mammals: it explains conscious access in terms of global broadcasting enabled by long-range cortical connectivity and

workspace dynamics distributed across cortical networks. On the proposed reading, the point is not to force identity between UAL and GNWT, but to show how they can productively relate. UAL provides a taxon-general template—an abstract organizational profile that consciousness-realizing systems should approximate—while GNWT specifies one mammalian way of implementing that template with cortico-thalamic circuitry.

I will illustrate the virtues of this methodology in recourse to a case study regarding comparative neuroanatomical work on basal vertebrates by Zacks and Jablonka (2023). They can be understood as asking: Where, in early fish brains, do we find structures that play the integrating and broadcasting roles specified in the UAL metamodel, and what modifications to the metamodel are required by putative this non-mammalian instantiation of consciousness? Overall, the framework respects mammalian evidence suggesting that particular cortical structures are deeply implicated in human consciousness, while also leaving space for non-mammalian consciousness supported by sophisticated behavior. It does so without making consciousness either trivially easy to realize (pure generality) or radically disunified (pure distinct realization). More broadly, it offers a concrete methodological blueprint for comparative consciousness science: build metamodels that unify across taxa; build instantiation models that explain within taxa; and let each constrain the other in an explicitly iterative research program.

References

Crook, R. J. (2021). Behavioral and neurophysiological evidence suggests affective pain experience in octopus. *iScience*, *24*(3). https://doi.org/10.1016/j.isci.2021.102229

Dehaene, S. (2014). *Consciousness and the brain: Deciphering how the brain codes our thoughts* (p. 336). Viking.

Gibbons, M., Crump, A., Barrett, M., Sarlak, S., Birch, J., & Chittka, L. (2022). Can insects feel pain? A review of the neural and behavioural evidence. In *Advances in Insect Physiology*. Academic Press. https://doi.org/10.1016/bs.aiip.2022.10.001

Ginsburg, S., & Jablonka, E. (2019). *The Evolution of the Sensitive Soul: Learning and the Origins of Consciousness*. The MIT Press. https://doi.org/10.7551/mitpress/11006.001.0001

Malach, R. (2022). The Role of the Prefrontal Cortex in Conscious Perception: The Localist Perspective. *Journal of Consciousness Studies*, *29*(7–8), 93–114. https://doi.org/10.53765/20512201.29.7.093

Michel, M. (2022). Conscious Perception and the Prefrontal Cortex A Review. *Journal of Consciousness Studies*, *29*(7–8), 115–157. https://doi.org/10.53765/20512201.29.7.115

Newen, A., & Montemayor, C. (2023). The ALARM Theory of Consciousness: A Two-Level Theory of Phenomenal Consciousness. *Journal of Consciousness Studies*, *30*(3–4), 84–105. https://doi.org/10.53765/20512201.30.3.084

Zacks, O., & Jablonka, E. (2023). The evolutionary origins of the Global Neuronal Workspace in vertebrates. *Neuroscience of Consciousness*, *2023*(1), niad020. https://doi.org/10.1093/nc/niad020