Review of: "The Case for Conscious Experience Being in Individual Neurons"

Christer Johansson[1]

1 University of Bergen

Potential competing interests: No potential competing interests to declare.

This review is for v.1. Version two is expanded, which is also one of our recommendations from v1.

Review of *The Case for Conscious Experience Being in Individual Neurons*, by Jonathan Edwards and Pavel G. Somov.

Per Olav Folgerø & Christer Johansson

University of Bergen, Norway.

The article presents an interesting thesis that individual neurons are conscious. Consciousness is defined as the ability to select and integrate information, and the neuron is the only place in the brain that performs this function. The integration of information is thus seen as a necessary condition for consciousness. If this is also a sufficient condition for consciousness is debatable.

The thesis *per se* has, if correct, wide-ranging neurobiological implications. It is easy to accept the premise that the dendritic tree is an organizing matrix. The dendritic tree, receiving thousands of inputs, which are being processed and broadcasted in a network of next stage neurons, is a basic principle in neurobiology. How can consciousness be strictly related to the single neuron? There is strong evidence that cortical neurons continuously compute differences between predicted and actual information (prediction error), by comparing ascending and descending information (Barrett & Kyle Simmons, 2015; Mumford, 1992; Ramstead et al., 2018). This counteracts surprises (i.e., increases of free energy and entropy).

We miss a discussion on Dahaene and Changeux's recent contributions (e.g., Dahaene & Changeux, 2011; Changeux et al., 2021) introducing a neurobiological basis for the so-called global neural workspace (GNW) of consciousness. Their main postulate is that "conscious access corresponds to global information availability": "(a)mong the innumerable neuronal structures available in the brain, the theory privileges a subset of cortical pyramidal cells with long-range excitatory axons, particularly dense in prefrontal, temporoparietal, and cingulate regions, that, together with the relevant thalamocortical loops, forms the *global neuronal workspace*" (Changeux, 2017, p. 9).

The argument that conscious experience is localized to individual neurons requires a more thorough review of the background literature. The original idea was proposed by Edwards (2005). Here, meaning and conscious experience arise from binding information together and the neuron is where this takes place.

The Integrated Information Theory of Consciousness (IIT)[1] proposes a neural information-based account that needs to be differentiated from both the single neuron account and the global neuronal workspace (GNW). It states “that consciousness is a ‘structure’ in the brain formed by a specific type of neuronal connectivity that is active for as long as a certain experience, such as looking at an image, is occurring. This structure is thought to be found in the posterior cortex, at the back of the brain” [2].

The current manuscript needs to go through the background literature and address the question about how to test the theory, and how the different theories can be differentiated in their predictions. In our opinion, such testing will not need to be definite but rather consist of a start for many small steps that will gradually build the theory through gradual refinement. Presenting a paradigmatical, radical, and interesting thesis requires a solid presentation of the recent evidence. A discussion should include a comparison between the *one-neuron theory*, the *GNW* theory, and the *integrated information theory* (IIT) [3].

Building up the arguments could benefit from addressing computational modelling of neurons. One such model, that is congruent with ideas presented in the article, are the so-called Kohonen-networks (Kohonen 1982). In this model, neurons learn to categorize input in competition with other neurons and using heuristics such as “conscience”, which makes simulated neurons less prone to take on all inputs. The first such trained neuron that responds to an input represents the class of the input it reacts to. This may be thought of as that neuron being conscious about its input and having preferences. The activity of such simulated neurons evolves as they are exposed to more input in competition (and collaboration) with other neurons. The input that these neurons send on can become the input to new layers of neurons, who in turn pick the combinations they like to respond to. Such models can be surprisingly powerful.

Robert Hecht-Nielsen (Hecht-Nielsen 2007, Solari et al. 2008) developed a biologically inspired theory called *Confabulation Theory*, which also demonstrates how elaborate narratives can be spun from limited initial seeds or triggers. The narratives are similar to the narratives that generative AI such as ChatGPT can spin. The idea of a conscious neuron is thus not so surprising. Is it the same kind of consciousness that we associate with human consciousness? The idea could be elaborated at different scale levels. The neurons are conscious of the input that is relevant at their scale level. A neuron integrates information and sends a signal if the current combination is something it likes. Other neurons may listen and give feedback (reinforcement or inhibition) and decide if the pattern it receives can excite that neuron enough to send a signal. This is like the Pandemonium Theory (Selfridge, 1959) that predates modern neural networks, and illustrates how a theory can be stepwise elaborated (in this case by mechanisms for learning and adaptation). Each unit is only conscious of its own input. If we move up a level, or three, we may similarly view individuals as conscious of neuronal activity that similarly reach some threshold. Exactly how that happens is not explicit. If we assume that individual neurons are responsible for our consciousness, it is certainly not the same neuron all the time. If we go further up a level, the individual may be organized in groups with their own agency and goals. A group similarly shows agency at its own level and constitutes an integration and selection of the conscious activities of its individuals (i.e., if all the individuals are unconscious that will not contribute to the group consciousness). Similarly, the individuals may have other activities and agendas than the group, which achieves an agency that is different from its individuals. The group starts to live its own life. Groups in turn may affect the animacy and agency of higher order organizations.

Much of the argumentation hinges on definitions of what consciousness is. The presented article implies that there is nothing magical about consciousness, as it always arises in a situation where there is *information* (patterned and/or random), and *integration* and *selection* of information. A discussion on entropy and information could thus be relevant. The neuronal process could be influenced by *free energy* (Yanagizawa 2021) available to neurons, including what could be labeled random activation. Random activation is important to achieve a cascade of activation (Pallbo 1997a,b). A detailed theory about the relation between consciousness, awareness, perception, and cognition is likely to lead to testable predictions. As we understand it, all these concepts are, for now, left vague. It is an attractive step to ground them all in a concept based on the integration and non-linear selection of information, which suggests that consciousness is everywhere to some degree and at different scale levels, where each scale level separates the underlying layers. We simply do not need to be aware of all the processes in order to perform our own integration of the information that is available to us. Imagine that this is also true for a neuron.

References

Changeux, J.-P, Goulas, A., & Hilgetag, C.C. 2021. A Connectomic Hypothesis for the Hominization of the Brain,*Cerebral Cortex*, 31(5), pp. 2425-2449 http://doi.org/10.1093/cercor/bhaa365

Dehaene, S. & Changeux, JP. 2011. Experimental and theoretical approaches to conscious processing.*Neuron.* 70(2), pp.200-27. doi: http://doi.org/10.1016/j.neuron.2011.03.018. PMID: 2152160.

Edwards, J.C.W. 2005. Is Consciousness Only A Property Of Individual Cells? *Journal of Consciousness Studies*,12(4-5), pp. 60-76

Hecht-Nielsen R. 2007. Confabulation theory. Springer-Verlag

Kohonen, T. 1982. Self-Organized Formation of Topologically Correct Feature Maps.*Biological Cybernetics.* 43(1), pp. 59–69. *doi:10.1007/bf00337288*

Mumford, M. 1992. On the computational architecture of the neocortex II. The role of cortico-cortical loops,*Biol. Cybern.* 66, pp. 241-251

Pallbo R. 1997a. A one step motion detection circuitry.*Bio Systems*, *40*(1-2), 141–148. https://doi.org/10.1016/0303-2647(96)01640-1

Pallbo, R. 1997b. Mind in motion, Ph.D. thesis, Lund University Cognitive Studies, Lund, Sweden. https://lup.lub.lu.se/search/publication/29566

Ramstead, M. J. D., Badcock, P. B., & Friston, K. J. 2018. Answering Schrödinger's question: A free-energy formulation, *Physics of Life Reviews*, 24, pp. 1-16

Selfridge, O.G. 1959. Pandemonium: A paradigm for learning." In D. V. Blake and A. M. Uttley, editors,*Proceedings of the*

Symposium on Mechanisation of Thought Processes, pages 511–529, London, UK.

Solari, S., Smith, A., Minnett,R. & Hecht-Nielsen, R. 2008. Confabulation Theory, *Physics of Life Reviews*, 5(2), pp. 106-120, ISSN 1571-0645, https://doi.org/10.1016/j.plrev.2008.03.003 (https://www.sciencedirect.com/science/article/pii/S1571064508000122)

Yanagisawa, H. 2021. Free-Energy Model of Emotion Potential: Modeling Arousal Potential as Information Content Induced by Complexity and Novelty. *Front. Comput. Neurosci. 15*, 1–13. https://doi.org/10.3389/fncom.2021.698252.

[1] https://iep.utm.edu/integrated-information-theory-of-consciousness/

[2] Decades-long bet on consciousness ends — and it's philosopher 1, neuroscientist 0 (nature.com)

[3] Consciousness theory slammed as 'pseudoscience' — sparking uproar (nature.com)