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

Louis Irwin[1]

1 University of Texas at El Paso

Potential competing interests: No potential competing interests to declare.

Over the years, neuroscientists have generally come to believe that all cognitive processes, including learning, memory, emotion, perception, and consciousness, are processes in populations of neurons (Jablonka and Ginsburg 2022; Limbacher and Legenstein 2020).   However, some neurons have been found in the medial temporal lobe of the human brain that do respond selectively to specific images, even when presented in alternative versions, like familiar objects in different orientations or the same celebrity in different clothes (Quiroga *et al.* 2007). These cells, designated ‘concept neurons’ (Koch 2018), are sparse but not totally unique; an average of about eight different cells responding to a specific ‘concept’ (Quiroga *et al.* 2008).

The proposal by Edwards and Somov that consciousness can be instantiated by individual neurons appears to be similar to the notion of ‘concept neurons’ and potentially could point to useful refinements in our thinking about potential neural correlates of consciousness. Unfortunately, a number of weaknesses in their article prevent it from achieving those refinements.

The greatest weakness is the failure of the authors to define the boundaries of a conscious percept that a single neuron can code for. Let us assume for the sake of argument a simple single percept: the image of a static landscape consisting of a tree in the foreground with plant-covered hills and blue sky with a few clouds in the distance.

What are the boundaries of this image that are relegated to a single neuron? Does the activation of a single cell evoke the entire landscape, or do different elements of the image require different cells: one for the tree, another for the hills, a third for the blue sky and a fourth for the clouds? If a single neuron responds to light reflected from the tree, will the stimulus from any tree activate that neuron, or will different neurons be required to respond to light from aspens, oaks, or pines? Will a single neuron respond to any cloud shape, or will every variation of cloud shape require a unique neuron?

Since the authors emphasize the neuron as the most likely integrating processor in the nervous system, I am assuming they believe that the conscious image of the landscape is determined by integrating inputs into the dendritic tree from the various elements (tree, sky, clouds), and the unique attributes of each element (oak or pine tree, blue or grey sky, thin or fluffy clouds, etc.) But that gets us back to the question of the boundaries of conscious perception attributable to a single neuron. Since the number of variable landscapes to which a single brain could be exposed over a lifetime is very large, a correspondingly large number of neurons would need to be devoted to landscapes alone. Logically then, every unique image at any point in time would have to activate a different neuron. This would appear to require well over the 86 billion neurons in the human brain (Herculano-Houzel 2017).

One way around this problem is to assume that any given neuron not devoted to encoding a long-term memory could be repurposed to respond to different images at different points in time. Perhaps this is the assumption the authors are making. But how does the brain select which neuron to repurpose for which novel input, and which neuron is to be left alone as a permanent repository for a single (remembered) percept? Given that the mechanisms for encoding memory, much less consciousness, are not known, these question can't be answered within the framework of anyone's current understanding. Fair enough. But for a notion as novel as the one proposed by these authors to merit credibility, the problems posed above need to be addressable within the framework of their theory as well as or better than alternative theories can address them, which, I contend, they cannot.

In his classic treatise on *Mechanisms of Memory,* E. Roy John (1967) posed the question of whether memory is a *thing in a place* or a *process in a population.* This was at a time when the possibility was seriously considered that learning might require the elaboration of specific informational molecules — peptides, proteins, or nucleic acids unique for specific elements of experience (Irwin 2021). This article appears to propose that consciousness, like memory, is *a thing in a place*, with single neurons rather than molecules expressing the minimal units of consciousness.

The alternative to the notion that binding of the disparate elements of a conscious percept occurs within single neurons is the more conventional current explanation that consciousness is not *a thing in a place* but is, in fact, *a process in a population.* The population of necessity is a multi-neuronal network of some sort – the iconic concept of which is Donald Hebb's (1949) 'phase sequence' in which 'cell assemblies' of virtually infinite variety are bound together to compose any dynamic perception which any brain could experience over any time span. But this alternative is clearly rejected by these authors, who devote an entire section to arguing against the involvement of any kind of neural networks in the manifestation of consciousness.

On the other hand, if the simplest percept has to be generated by a neural network, where does this leave 'concept neurons'? Could the notion advanced by Andrews and Somov at least resonate to some degree with evidence that complex stimuli can evoke activity in specific neurons? I believe so. Clearly single neurons do integrate hundreds to thousands of inputs through their dendritic branches, as the authors correctly point out. Binding through integration of whatever elements each input represents clearly does occur at the level of individual neurons. But it would take a virtually infinite number of individual neurons to instantiate the virtually infinite variation of sensory inputs that a brain must process over its lifetime. Evidence for the reality of complex information resident in cell assemblies is strong (Butler *et al.* 2019; de Sousa *et al.* 2019; Hofer *et al.* 2009; Takamiya *et al.* 2019). This doesn't rule out the possibility that the activation of specific neurons within a given cell assembly could contribute to consciousness of specific percepts. In other words, the existence of 'concept neurons' and the reality of processing through neural networks could both be true.

Perhaps this is the way that Edwards and Somov actually rationalize the simultaneous demonstrable involvement of millions of neurons in any perceptual experience with their view of how consciousness can be said to emerge at the unitary neuronal level. If so they do not make this view clear. On the contrary, their article seems to dismiss such a compromise. If that is the case, the burden falls on them to answer "at least six impossible things to believe before breakfast" that skeptics of this view of consciousness pose, and leaving it simply at "All that we can perhaps say is that it makes sense to us" doesn't move us toward greater scientific understanding of the mechanisms of consciousness.

In one respect, Andrews and Somov pinpoint a critical issue that no theory of consciousness has yet resolved: What is the 'perceiver' of consciousness? What mechanism is attending to the neural activities that are mediating consciousness? Stated another way, what process in analyzing the 'conscious' neurons while ignoring the neurons active in the unconscious state? The authors' suggestion that quantum activity in microtubules provides the meta-analysis that gives rise to consciousness falls way short of being convincing for three reasons. First, all neurons have microtubules, so what differentiates the 'conscious' microtubules from those playing no role in consciousness? Secondly, there is no credible evidence that microtubules directly affect the electrophysiological activity of a neuron. Thirdly, all the questions enumerated earlier about the boundary limits of a single percept for a single neuron would pertain as well to a single microtubule. (Parenthetically, the authors curiously make no reference to Roger Penrose (1994), the highest-visibility scientist championing quantum processing in microtubules as the key to consciousness.)

Other weaknesses in this article deserving mention include the following:

A great deal of argument is devoted to the question of how localized and/or redundant neurons are that evoke a sense of self, without making clear the connection between this point and the one in the title.

Some references are too obscure to mean anything to many readers. "a walk on Hampstead Heath" and "that yinzz (as they say in Western PA)" are examples.

Some passages are near or totally incoherent, at least to me. Examples include: "Science is business, another way of getting bananas", "Any theory that deals with *individual indivisibles* . . . really has to operate at the level of such indivisibles in physics", and "How much is what a neuron feels its interpretation of input and how much is the interpretation already in the input?"

The sum of all these weaknesses reluctantly leads me to rate this article in its present form (v.2) at the 2 star level. By no means do I wish to imply that the novel concept argued by these authors is undeserving of consideration. I commend Andrews and Somov for their extensive efforts over an extended period to put such a highly unconventional idea into the public domain. I hope they find value in this and the other critiques that can induce them to make a stronger, more precise, and less problematic argument, or a more credible revised formulation, of their interesting idea.

References

Butler C. W., Wilson Y. M., Mills S. A., Gunnersen J. M., and Murphy M. (2019) Evidence that a defined population of neurons in lateral amygdala is directly involved in auditory fear learning and memory. *Neurobiol Learn Mem*, 168: 107139.

de Sousa A. F., Cowansage K. K., Zutshi I., Cardozo L. M., Yoo E. J., Leutgeb S., and Mayford M. (2019) Optogenetic reactivation of memory ensembles in the retrosplenial cortex induces systems consolidation. *Proc Natl Acad Sci U S A*, 116: 8576-8581.

Hebb D. O. (1949) The Organization of Behavior: A Neuropsychological Theory. John Wiley, New York.

Herculano-Houzel S. (2017) The Human Advantage: How Our Brains Became Remarkable. MIT Press, Cambridge, MA.

Hofer S. B., Mrsic-Flogel T. D., Bonhoeffer T., and Hubener M. (2009) Experience leaves a lasting structural trace in cortical circuits. *Nature*, 457: 313-7.

Irwin L. N. (2021) Scotophobin: Darkness at the Dawn of the Search for Memory Molecules. Xlibris, Bloomington, IN.

Jablonka E., and Ginsburg S. (2022) Learning and the evolution of conscious agents.*Biosemiotics*, https://doi.org/10.1007/s12304-022-09501-y.

John E. R. (1967) Mechanisms of memory. Academic Press, New York

Koch C. (2018) What Is Consciousness?*Sci Am*, 318: 60-64.

Limbacher T., and Legenstein R. (2020) Emergence of stable synaptic clusters on dendrites through synaptic rewiring. *Front. Computl. Neurosci.*, 14.

Penrose R. (1994) Shadows of the Mind: A Search for the Missing Science of Consciousness. Oxford University Press, New York.

Quiroga R. Q., Kreiman G., Koch C., and Fried I. (2008) Sparse but not 'grandmother-cell' coding in the medial temporal lobe. *Trends Cogn Sci*, 12: 87-91.

Quiroga R. Q., Reddy L., Koch C., and Fried I. (2007) Decoding visual inputs from multiple neurons in the human temporal lobe. *J Neurophysiol*, 98: 1997-2007.

Takamiya S., Yuki S., Hirokawa J., Manabe H., and Sakurai Y. (2019) Dynamics of memory engrams.*Neurosci Res*: doi.10.1016/j.neures.2019.03.005.