Review of: "The Evolution of Consciousness Theories"

Igor Aleksander[1]

1 Imperial College London

Potential competing interests: No potential competing interests to declare.

Certainly, the field of consciousness theories does not follow a set of coherent headings. So Ashkan Farhadi's attempt to classify published theories is to be welcomed by researchers in the consciousness domain. The focus of my comments here is to address what the paper does not do. First, the provision of a list of existing efforts is useful, but it largely focuses on cognitive/algorithmic selections (e.g., Global Workspace Theory) of the field with less emphasis on the dynamic/neural side expressed in artificial neural networks, where properties due to the dynamic stabilities of neural networks play a strong role.

While the discussion of Dehaene's (1998) neural version of Baars' (1988) Global Workspace Theory is welcome, the way that the task of a neural structure might give rise to consciousness does not enter this article, and neural dynamics hardly feature. Indeed, missing is a most important contribution by Damasio (2021) on the definition and role of *feeling* and *knowledge* based on the activity of the nervous system. This distinguishes consciousness from basic reactive intelligence in objects such as bacteria. Also pertinent to the author's own discussion of awareness, missing is Damasio's concept of awareness as a gateway to switches in consciousness.

But above all, missing is the recognition of the rise of*Machine Consciousness* as a method for the discussion of the nature of biologically conscious entities through an analysis of their machine models. (See, for example, Reggia (2013)). Reggia points to two reasons for the value of this approach. First, machine models approach consciousness in a precise and executable way. The result of this is that the concept of a conscious machine can be studied in a defined machine way rather than attempting to analyse complex biological mechanisms. This has theoretical benefits. It leads to conditions that are necessary for a *machine* to be classified as being conscious and then bridges to biological entities while still retaining the character of precision**.** Second, given this clarity, conditions of consciousness can be implemented in machines. They could be said to be M-conscious (for Machine) and have an explicative relationship to living entities. I have tried to summarise ideas about machine consciousness in (Aleksander, 2012 and 2022). The language of this description is undoubtedly neural and in the domain of neural automata theory. This seems essential to an explanation of the transition between perceptive neural mechanisms and accessible conscious inner states (that is, experience). This is done by recognising that the inner state of an interconnected neural network, the states of which are forced by perceptual mechanisms (we have called this 'iconic learning'), has the characteristics of 'mind' with a mind/body-like relationship to the physical network.

To summarise briefly, the points I am making are that Farhadi's description of consciousness theorems is useful but restricted to algorithmic descriptions. I propose logical descriptions drawn from automata and dynamic neural network

theory. This benefits from the power of dynamic neural creation of world- and self-experience that is accessible. Damasio's relationship of knowing to feeling is important but not approached …. I may be totally wrong, but it seems that knowing and feeling may be seen to lurk in and emerge from dynamic neural models rather than algorithmic models of the Global Workspace or cognitive concepts of attention and decision making. This does not negate that the material covered by Farhadi is an excellent compendium of algorithmic descriptions of *conscious effects*, but not always indicative of how these come about in a machine, particularly the human one.