

Review of: "Towards Modeling Artificial Consciousness"

Miranda Mowbray¹

¹ University of Bristol

Potential competing interests: No potential competing interests to declare.

This paper gives a model for consciousness. The research summary by the author published in (Shevchenko et al, 2022) gives the main idea of the paper, which is to model consciousness as a nonlinearly connected Artificial Neural Networks. The paper gives a set of equations intended to specify this model. In my opinion, although the research aim is pleasantly ambitious, this paper is not ready for publication.

1. The paper says that in the model, "consciousness is described by a decaying strange attractor that lasts for 3-4 periods." A reason is given for the decaying nature of the strange attractor used in the model, but not for the choice of 3-4 periods rather than some other

figures. Some justification for this should be given.

(The idea that consciousness may be modelled by a strange attractor is not new: see e.g. William Calvin, "The Cerebral Code", 1996, or Andrew Bailey, "William James, Chaos Theory, and Conscious Experience", in *Advances in Psychology* 126: Systems Theories and A Priori Aspects of Perception, 1998.)

2. It is not made clear how the existence of such a strange attractor is implied by the equations that the paper then gives for the model. An explanation of this should be added.

3. Is there perhaps an error in equation (2.7)? This equation is meant to define the nonlinear dependence between different processes which models C2 consciousness, and so is a rather crucial equation for the paper. The equation defines the variable $x_{ki}^{[0]}$, a component of the input signal of the i th process, in terms of itself. Maybe a component of a different process should occur in the right hand side of the equation?

4. One aspect of the model is motivated by a claim that the consciousness of living beings does not arise immediately after birth. The paper cites Dohaene, Lau and Kouider (2017, 5) for this. I have read Dohaene et al's paper several times but have not found that statement in it. Dohaene et al's paper says that human infants have consciousness (both C1 and C2).

5. The sentence of the discussion mentioning deep semantics grounding on semantic fields gives the unfortunate impression that it was introduced just so that the author could cite three semantics papers written by himself. One citation would suffice to support the content of the sentence.

6. The original contribution of the paper is the set of equations describing the model. No calculations, experiments or results are reported. I am not certain that I understand exactly what the equations are intended to specify, but if it is just a general mathematical model of a set of nonlinearly dependent ANNs, I am not sure that this would be enough content to warrant publication. Adding an experimental comparison between an implementation of the model and a comparable ANN

would make the paper more substantial. I do not believe that is possible to test an artificial network for subjectivity. But perhaps there are some other known features of consciousness, that do not require subjectivity and are not directly implied by the equations, which can be shown experimentally to be displayed by an implementation of the model and not by a comparable ANN?