Review of: "Towards Modeling Artificial Consciousness"

Stanley Sutton, Jr.

Potential competing interests: No potential competing interests to declare.

The paper is titled “Towards Modeling Artificial Consciousness”, which leaves wide open what the content and contribution of the paper might be. Thus the title could be more specific.

The abstract is very brief, simply "A new synergetic approach to consciousness modeling is proposed, which takes into account recent advancements in neuroscience, information technologies, and philosophy.” The conclusions are similarly brief. Both are accurate enough as far as they go, but they do not convey anything of the substance of the paper, which they ought to do.

The model proposed in the paper appears to be well founded on about a dozen and a half key references that have appeared in the past several years. However, what is novel about the proposed model (as opposed to just new) is not stated. (Unfortunately, I am not qualified to give an opinion about it’s novelty.)

As the author notes, prior work by others has observed that consciousness in humans develops over some time after birth and is evidently founded upon already trained neural networks. (Although not addressed in the paper, this would seem to be even more completely the case with consciousness in animals.) The model that is described in this paper addresses the emergence of consciousness atop such a preexisting foundation. Thus, a realization of the model should not, by itself, be expected to give rise to consciousness. (Neither should broadly similar computational structures that lack the requisite substrate.) Still, modeling of this sort is essential to the development of artificial consciousness based on neural networks, and a well founded and novel model like this would be a worthy research contribution.

I am a computer scientist with an interest in artificial consciousness as it may emerge, not in neural networks, but in organized systems of human activity. I appreciated the concise review of recent perspectives on human consciousness that was provided at the beginning of the paper. The mathematical modeling that comprises the substance of the paper does not speak directly to the issues with which I am mainly concerned, but it does give me a more sophisticated way to think about how consciousness may emerge on the basis of neural networks, and it may provide concepts that I can leverage in other contexts.