Review of: "The functional unit of neural circuits and its relations to eventual sentience of artificial intelligence systems"

Chris Rourk

Potential competing interests: No potential competing interests to declare.

The authors them refer to “object neurons” and cite to various papers and texts from as early as 1967. However, recent research casts significant doubt on the idea of an “object neuron,” such as Nagelhus, Anne, et al. "Object-centered population coding in CA1 of the hippocampus." Neuron (2023). This study reports that “Here, we record simultaneously from **large numbers of hippocampal CA1 neurons** to determine how the presence of a salient object in the environment alters single-neuron and neural-population activity of the area. The majority of the cells showed some change in their spatial firing patterns when the object was introduced. At the neural-population level, these changes were systematically organized according to the animal’s distance from the object. **This organization was widely distributed across the cell sample, suggesting that some features of cognitive maps—including object representation—are best understood as emergent properties of neural populations**.” Based on this research, the idea of an “object neuron” has been refuted and it has been shown instead that object representation is an emergent property of neural populations that changes as a function of the spatial firing patterns of those neurons with changes in the location of the object. Furthermore, the hippocampi consolidate information from short- to long-term memory, which means that the objects identified by them are not stored in them. A detailed analysis of recent research is needed to address the tension between the older materials relied on by the authors regarding “object neurons” and recent research that identifies how objects are dynamically identified, stored and recalled from memory by neural populations.

Pg. 11/14, line 21 - “The author emphasizes, though, that consciousness occurs when there is a massive synchronization of neuronal activity and also when repetitive oscillations in neuronal circuits occur.” Epileptic seizures result in “massive synchronization of neuronal activity” and “repetitive oscillations in neuronal circuits,” see, e.g. Ren, Xin, et al. "Connectivity and neuronal synchrony during seizures." Journal of Neuroscience 41.36 (2021): 7623-7635; Uhlhaas, Peter J., and Wolf Singer. "Neural synchrony in brain disorders: relevance for cognitive dysfunctions and pathophysiology." neuron 52.1 (2006): 155-168; Truccolo, Wilson, et al. "Neuronal ensemble synchrony during human focal seizures." Journal of Neuroscience 34.30 (2014): 9927-9944. These and other references show that consciousness ***ceases to occur*** when there is a massive synchronization of neuronal activity and also when repetitive oscillations in neuronal circuits occur. An explanation should be provided of how the works relied on by the authors could possibly be correct in view of this contradictory information, as this appears to be a citation to works that describe a general, metaphysical idea that is unsupported by any research.

Page 12/14, line 14 “If the electromagnetic field, generated in this way, were an inevitable factor triggering the emergence

of self-awareness, then it should be considered by the creators of the rapidly developing artificial intelligence systems." In a digital data processor, a device called an arithmetic logic unit or ALU performs add, subtract, multiply, divide and logic operations on digital data in a series of operations using a large number of discrete digital logic gates, which currently happens at speeds in excess of one billion times per second. Each arithmetic operation is based on the specific inputs to the discrete logic gates of the ALU for each separate operation, and those logic gates would not function properly if the operation of any one of the logic gates was influenced in any manner by a previous operation of the logic gate or other logic gates, or electromagnetic fields from other logic gates or circuit elements. There is no difference between an ALU operation performed for an AI function and the same ALU operation performed for word processing, playing a video or surfing the Internet, and there is no serious argument that desktop computers that perform these functions are self-aware. Furthermore, an "artificial intelligence system" typically uses distributed processors, and the electromagnetic fields of each distributed processor are designed to be isolated from all of the other processors and not to have any influence on other processors. As such, there is no evidence that AI processors create electromagnetic fields that can interact in any functional manner.

The situation is more complex for the brain, where each cell generates local electric field potentials that can interact with nearby cells but not distant cells, due to Debye shielding. The cells include neurons, astrocytes, glial cells, oligodendrocytes, Schwann cells, endothelial cells, blood cells and others, each of which generate electromagnetic fields. There is no evidence that these electromagnetic fields can be integrated in a manner that would explain the integrated experience of phenomenal consciousness. Such fields do not physically interact below the Schwinger limit, which is on the order of 10^{18} V/m or 10^{9} T, well above the levels of electric and magnetic fields in neurons. Instead, they add as independent vector sums and also interact with the ionic charges in other cells. Neurons, like logic gates, must function consistently or the entire neural system would fail to operate in a functional manner. Electrical and magnetic field stimulation of the brain does not create altered states of self-awareness, such as to make someone perceive an object that is not present, it only provides mild therapeutic effects at best, such as for treatment of depression, or pain and neural dysfunction and damage when used improperly. The idea that self-awareness somehow "emerges" from electromagnetic fields is a metaphysical belief without any rationale scientific basis.

It is also unclear how the "object neurons" discussed at the beginning of the paper are supposed to fit into this electromagnetic field picture. An explanation needs to be provided of whether the "object neurons" are supposed to be located physically adjacent to each other so their electromagnetic fields can interact, or if some kind of neural structure exists that is supposed to route and synchronize the electromagnetic fields. Given that recent research indicates that object representations are emergent properties of large numbers of neurons that are physically separate from each other and unable to interact by electromagnetic field interactions, and the absence of any evidence that electric or magnetic field stimulation of the brain can alter perception of objects, it appears that the hypothesis put forward by the paper is untenable.