Is Consciousness Computable?

Subhash Kak*

ABSTRACT

This commentary reviews different scientific positions for and against consciousness being a computable property. The role that quantum mechanics may play in this question is also investigated. It is argued that the view which assigns consciousness a separate category is consistent with both quantum mechanics and certain results in cognitive science. It is further argued that computability of consciousness implies the solution to the halting problem which is computationally impossible.

Key Words: cognition, machine consciousness, learning models, quantum Zeno effect

DOI Number: 10.14704/nq.2019.17.05.2359 **NeuroQuantology 2019; 17(05):71-75**

Introduction

The case may be made that since the brain is a machine that is conscious then other machines with appropriate architecture should also exhibit consciousness (Baars, 1988; Baars, 1997). This is the physicalist position that consciousness is either an emergent property or an epiphenomenon. It is also assumed that computers can capture the abstract causal organization of other systems and thus of the brain (Chella *et al.*, 2008; Gamez, 2008). Panpsychism, the view that mind is to be found everywhere, is another position that has recently become popular in academic circles (Koch, 2014), but it is too extravagant in associating mind even with a straw or a rock. Broadly speaking, both physicalism and panpsychism associate mind with matter, although they do it in different ways.

From a computational perspective, it is counter-intuitive that a small subset of activity in the brain in terms of abstract thought is able to capture the workings of the physical world. This happens even though brain function is accompanied by the reorganization of its very structures during learning and function (Kak, 2016), which goes beyond the Turing machine model on which the von Neumann architecture of digital computers is broadly based.

Although learning cognitive tasks may require attention and concentration, once the learning is complete they can be performed literally automatically. The learning apparently leads to training of neural networks that convert the computational problem into one of classification and recognition (Kak, 2011). It should thus be possible to replicate the processing in a machine of tasks that involve computations in the neural circuitry of the brain. But raw awareness appears to be a phenomenon that is different from cognitive processing (Kak *et al.*, 2016).

The design of a conscious machine faces formidable scientific and engineering obstacles and so one must begin with small steps. Architectures that copy models of brain function have been investigated (e.g. Holland *et al.*, 2003; Haikonen, 2012). These architectures include distributive agents and the global workspace theory (GWT) (Baar, 1988; Shanahan, 2006). In the GWT separate parallel processes compete to place their information in the global workspace whose contents are broadcast to

Corresponding author: Subhash Kak
Address: School of Electrical and Computer Engineering, Oklahoma State University, Stillwater, OK 74078, USA
e-mail ✉ subhash.kak@okstate.edu
Tel: + 405-744-6096

Relevant conflicts of interest/financial disclosures: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
Received: 14 May 2019; **Accepted:** 24 May 2019

a multitude of receiving processes. Since globally broadcast messages can evoke actions in receiving processes throughout the network, the global workspace may be used to exercise executive control to perform voluntary actions. This roughly mimics some brain functions but it doesn't address the problem of consciousness.

Postulating a physical substrate of consciousness to be at the basis of complex patterns of activity has been called the integrated information theory (IIT) (Tononi, 2004; Tononi *et al.*, 2016). It amounts to correlations in physical processes that cannot by themselves be the origin of awareness. To see a parallel situation, complex causally connected behavioral patterns can also be seen in social networks, but they are properly analyzed within the framework of an ecological system without recourse to the concept of consciousness.

In the standard neuroscience view, mind emerges from the interoperation of the various modules in the brain and its behavior must be completely described by the corresponding brain function. However, no specific neural correlate of consciousness has been found (Zeki, 2003). Counterintuitive characteristics of the mind may be ascribable to underlying quantum processes (Tarlaci, 2010), but although quantum mechanics might indeed play a role in brain processes, there is no reason to assume that it throws any light on the phenomenon of consciousness (Kak, 2000).

The case for and against a unitary representation of the world and whether self-awareness is computable was discussed by Penrose in his well-known books (Penrose, 1989; Penrose, 1994). The arguments in neuroscience literature are classical and this brings up the question of the fundamental difference between classical and quantum models (Kak, 2013a; Kak, 2013b). According to the orthodox Copenhagen Interpretation of quantum mechanics consciousness is a category different from matter and dual to physical reality (Bohr, 1958; Schrödinger, 1967), which may be shown as in Fig. 1.

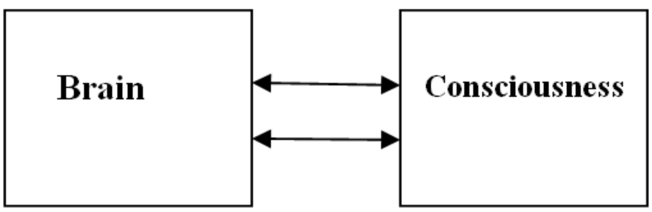

Figure 1. Consciousness as an independent category.

We apprehend reality in our consciousness and not directly in terms of space, time and matter. Consciousness is the doorway that shows us the world and makes self-knowledge possible and it is the source of creativity, although it is constrained by the habits and limitations of the mind (Freeman, 1999). The quality of the manifestation of consciousness in a natural system depends on structure and different modes of processing. Consciousness and the material world complement each other and consciousness may influence material evolution as in the Quantum Zeno effect (Misra and Sudarshan, 1974; Kak, 2017).

Mind And Computability

The statement that consciousness is the ground on which experience is evoked may appear to be a principle that is no different from that of gravity being the same force that works on objects variously depending upon their location. This does not rule out the possibility that the structure of the mind makes it impossible to know reality completely (Gautam and Kak, 2013; Kak, 2014a; Kak, 2014b; Kak, 2015)

We think of ourselves as being outside of the physical world. Even our conceptions of the universe are as if we are not a part of it, and in the words of Schrödinger: "We do not belong to this material world that science constructs for us. We are not in it; we are outside. We are only spectators. The reason why we believe that we are in it, that we belong to the picture, is that our bodies are in the picture. Our bodies belong to it." (Schrödinger, 1967)

If this sense of being outside of the physical world is true, it would be impossible to emulate it by hardware and processing that is within the world. It also follows that it will not be a computational property of the physical elements that comprise the system. Since consciousness is global, it must be non-algorithmic (Penrose, 1989).

Can consciousness be defined as the capacity to know with certainty that one is conscious? (Caplain, 1995; Reason, 2016) This appears to be a circular definition and it hinges on hard-to-define concepts such as "knowledge" and "belief". Consider HM, the guy who lost all new memory with the bilateral resection of medial temporal lobe and his ability to hold beliefs and knowledge was greatly impaired (Squire, 2009), yet, without doubt, he had all hallmarks of consciousness.

We can look for the non-computability of consciousness from its parallel to the unsolvability of the halting problem. Let us define "consciousness" as some privileged state of the mind that makes its processes halt (we don't bother to specify it beyond this description) and its contents registered (which is what we imply by awareness). Humans can get into the state of "awareness" at any time, which means that the earlier computation has halted, and this is irrespective of the initial state of the immediately preceding process. The exceptions to this are if a person is sleeping or unconscious as in coma. But such halting to arbitrary input is impossible from a computability point of view (Davis, 1958). Therefore, it follows that consciousness is not computable.

The consideration of information (or entropy) in physical theory, which is commonly done in many branches of physics, implies an unstated postulation of consciousness (Kak, 2018). Information cannot be reduced to local operations by any reductionist program. It requires the use of signs derived from global properties and the capacity to make choices which, in turn, implies agency. Such agency will be consistent with physical law only if does not involve the expenditure of energy. There are different ways interaction between matter and consciousness may be considered and these include the questions of information and that of dynamics.

The Interaction Problem

Let us consider the interaction problem in the framework of quantum theory. The measurement operation divides the physical universe into two parts: the first part is the system being observed, and the second part is the human observing agent, together with the instruments.

Philosophically, the experimenter must be separated from the system being observed. In the orthodox Copenhagen Interpretation (CI) (Bohr, 1958; von Neumann, 1955), a hypothetical interface called the Heisenberg cut (or the von Neumann cut) is assumed between quantum events and the observer's information. Below the cut everything is governed by the quantum wave function, whereas above the cut one must use classical description. The cut might appear to be arbitrary, but it is merely a way to separate parts of the system, which must be done in a consistent manner.

The CI considers the question of interaction between mental states and the wave function by taking the wave function to have an epistemological reality, that is, it represents the experimenter's knowledge of the system, and upon observation there is a change in this knowledge. Operationally, it is a dualist position, where there is a fundamental split between observers and objects. The placement of the cut between the subject and the object may be done in many different ways that depends on the nature of the interaction between the two.

In the ontic view of the wave function, there is no collapse, and the interaction is seen through the lens of decoherence, which occurs when states interact with the environment producing entanglement. Decoherence causes the system to make transition from a pure state to a mixture of states that the observer is able to measure (Zurek, 2003; Kak, 2014c). The process of decoherence in no way negates the CI picture, for it merely shifts the cut away in such a way that the system under observation and the measurement apparatus are on the same side.

By placing the Heisenberg cut away from both the system and the measurement apparatus, we can speak of interaction between different systems. But it raises other questions: Since the entire universe may be taken to be a quantum system, the question of how this whole system splits into independent subsystems arises. It would seem that the splitting into subsystems is itself an observational choice, rather than fundamental. This splitting serves about the same function as the Heisenberg cut of CI. From the perspective of the mind, the ontic view is troublesome for its own states are determined by transformational operations that rule out agency. But to assume a mechanistic behavior for the mind is to oppose the inconvertible fact of free will.

The experimenter is not describing reality ontologically; rather, he is obtaining knowledge about it and this depends on the nature of his interaction with the system. The knowledge informs his mind and consideration of this information creates a sense of overarching knowledge.

If the physical world and consciousness did not interact then evolution of the universe will remain random. While the psychological part of the psychophysical parallelism notion implies that there exists no specific correlate of consciousness in the brain (as it cannot have a physical basis), the quantum Zeno effect provides a mechanism on how observation can influence dynamics sidestepping the question of the ontological position of the observer.

Discussion

Cognitive capacities are computational but their assignment to the autobiographical self is a process that is associated with awareness and memories. This assignment occurs with consciousness as a singular phenomenon.

Consciousness cannot intervene in physical law but it can change the probabilities in the evolution of quantum processes, and it does so without changing the dynamics, which provides an explanation of how consciousness can be reconciled with the physical law. If the phenomenon of consciousness is contingent on a recursive and self-organizing structure that constitutes the unity of the organism, then we know that current machines will come up short in attaining it.

Consciousness appears to be a privileged state of the mind that makes brain processes halt and its contents registered irrespective of the initial state of the immediately preceding process. But such halting to arbitrary input is impossible from a computability point of view. Therefore, it follows that consciousness is not computable.

References

Baars B. A Cognitive Theory of Consciousness, Cambridge, MA: Cambridge University Press, 1988.

Baars B. In the Theater of Consciousness, New York, NY: Oxford University Press, 1997.

Bohr N. Atomic Physics and Human Knowledge. Wiley, 1958.

Caplain G. Is consciousness a computational property? Informatica 1995; 19: 615-619.

Chella A, Frixione M, Gaglio S. A cognitive architecture for robot self-consciousness. Artificial Intelligence in Medicine 2008; 44: 147-154.

Davis M. Computability and Unsolvability. McGraw-Hill, 1958.

Freeman W. How Brains make up their Minds, London, UK: Phoenix, 1999.

Gamez D. Progress in machine consciousness. Consciousness and Cognition 2008; 17: 887–910.

Gautam A and Kak S. Symbols, meaning, and origins of mind. Biosemiotics 2013; 6: 301-309.

Haikonen P. Consciousness and Robot Sentience, Singapore: World Scientific, 2012.

Holland O and Goodman R. Robots with internal models. In O. Holland (ed.), Machine Consciousness. Exeter: Imprint Academic, 2003.

Kak A, Gautam A, Kak S. A three-layered model for consciousness states. NeuroQuantology 2016; 14: 166-174.

Kak S. Active agents, intelligence, and quantum computing. Information Sciences 2000; 128:1-17

Kak S. Information and learning in neural networks. NeuroQuantology 2011; 9: 393-401.

Kak S. Biological memories and agents as quantum collectives. NeuroQuantology 2013a; 11: 391-398.

Kak S. Probability constraints and the classical/quantum divide. NeuroQuantology 2013b; 11: 600-606.

Kak S. From the no-signaling theorem to veiled non-locality. NeuroQuantology 2014a; 12: 12-20.

Kak S. Observability and computability in physics. Quantum Matter 2014b; 3: 172-176.

Kak S. Measurement complexity and oracle quantum computing. NeuroQuantology 2014c; 12: 374-381.

Kak S. Veiled nonlocality and quantum Darwinism. NeuroQuantology 2015; 13: 10 – 19.

Kak S. Communication languages and agents in biological systems. In Biocommunication: Sign-Mediated Interactions between Cells and Organisms, R. Gordon and J. Seckbach (editors). World Scientific Publishing, London, 2016. pp. 203-226.

Kak S. The limits to machine consciousness 2017. arXiv:1707.06257

Kak S. Epistemic view of quantum communication. In Quantum Foundations, Probability and Information, A. Khrennikov and B. Toni (editors). Springer-Verlag International, 2018; pp. 119-128.

Koch C. The Quest for Consciousness: A Neurobiological Approach, Pasadena, CA: Roberts & Company Publishers, 2004.

Koch C. In consciousness universal? Scientific American, January 2014.

Misra B. and E.C.G. Sudarshan, The Zeno's paradox in quantum theory. Journal of Mathematical Physics 1977; 18: 756–763.

Penrose R. The Emperor's New Mind. Penguin Books, 1989.

Penrose R. Shadows of the Mind. Oxford University Press, 1994.

Reason. Consciousness is not a physically provable property, Journal of Mind and Behavior 2016; 37: 31-46

Schrödinger, E. What is Life?and Mind and Matter. Cambridge University Press, Cambridge, 1967

Shanahan M. A cognitive architecture that combines internal simulation with a global workspace. Consciousness and Cognition 2006; 15: 443–449.

Squire LR. The legacy of patient H.M. for neuroscience. Neuron 2009; 61: 6-9.

Tarlaci S. Why we need quantum physics for cognitive neuroscience. NeuroQuantology 2010; 8: 66-76

Tononi G. An information integration theory of consciousness. BMC Neurosci 2004; 5: 42.

Tononi G, Boly M, Massimini M, Koch C. Integrated information theory: from consciousness to its physical substrate. Nature Reviews Neuroscience 2016; 17: 450–461.

von Neumann J. The Mathematical Foundations of Quantum Mechanics. Princeton: Princeton University Press, 1955.

Zeki S. The disunity of consciousness. Trends Cogn Sci 2003; 7: 214-218.

Zurek WH. Decoherence, einselection, and the quantum origins of the classical. Rev Mod Phys 2003; 75: 715-775.