

Essay

The Evolution of Consciousness

Guy Immega*

ABSTRACT

In this essay, I suggest that: (1) Consciousness (awareness) is a characteristic of all living things; (2) Consciousness evolved by Darwinian natural selection; (3) Consciousness is an emergent property, existing beyond its component parts; (4) Consciousness is essential to survival and reproductive success; (5) Consciousness exists in a hierarchy of complexity, from simple to sophisticated; (6) Consciousness is dynamic, varying according to the state of the organism; and (7) Machine consciousness (AIs) can exist only by emulating biological systems.

Keywords: Consciousness, evolution, emergent, Darwinian, artificial intelligence, organism.

1. Introduction

A Working Definition of Consciousness

There may never be consensus about consciousness. Most people assume they're conscious when awake and unconscious during sleep. For the purposes of this essay, consciousness is defined as *awareness*. At a basic level, awareness means perception of the surrounding environment. Thus, rudimentary consciousness includes the capacity to be wakened from sleep (awareness of a poke in the ribs). At a higher level, consciousness can encompass awareness of being aware, as well as the abstractions such as identity, emotions, ideas, and time. In the broadest sense, all living things, animals and plants, exhibit at least a minimal degree of environmental sensitivity. Upon death, consciousness ends.

Some Observations on Current Consciousness Research

The Stanford Encyclopedia of Philosophy (1) says: "The problem of consciousness is arguably the central issue in current theorizing about the mind." The encyclopedia further describes nine disparate theories of consciousness—but none discuss its origins. The encyclopedia also suggests: "[the] adaptive value [of consciousness] is likely relevant to explaining its evolutionary origin." However, no adaptive value is posited. The encyclopedia goes on to state: "the Why question [of consciousness] may well not have a single or uniform answer."

David Chalmers famously posed "The Hard Problem" of consciousness [Facing Up to the Problem of Consciousness -1995]. (2) He asks: "Why do we have subjective experiences?" For Max Tegmark [*Life 3.0—Being Human in the Age of Artificial Intelligence*] (3), "if it feels like something to be you right now, then you're conscious." This subjective feeling is untestable and difficult to relate to brain states.

*Correspondence: Guy Immega, Independent Research, Canada. Email: guy.immega@kinetic.ca

Three Goals for this Essay

- This essay suggests that a single biological explanation for the origin and adaptive value of consciousness: consciousness evolved by Darwinian natural selection.
- This essay suggests that the Hard Problem of consciousness is not difficult to comprehend. Consciousness isn't fundamental or universal and doesn't require a quasi-mystical explanation. Consciousness is indeed a challenging problem, but it's no more puzzling than many of the conundrums in evolutionary biology.
- This essay comments on the nature and limits of artificial intelligence — and speculates on the possibility and requirements for an AI to become conscious.

2. An Explanation of Consciousness by means of Evolutionary Biology

This section is a conjecture on the question: Why is anything conscious? It provides a theoretical hierarchy for the evolution of consciousness in living beings. The progression of consciousness is from simple organisms to the sophisticated point of view of human beings. The logical order of various attributes of consciousness is not fixed and might be altered for specific examples.

A. Consciousness evolved by Darwinian natural selection

Consciousness is a biological process known only in living beings. Like all characteristics of living organisms, consciousness evolved by Darwinian natural selection. Consciousness confers identity and must be adaptive (advantageous to survival), because it is universally conserved.

B. Consciousness is an *emergent* property, existing beyond its component parts

Many thinkers have suggested that meta-consciousness (aware of being aware) is an emergent behavior (a biological spandrel) arising from an array of complex brain functions. This description is likely correct but does not explain *why* consciousness evolved.

C. Consciousness is essential to survival and reproductive success

A unique genome confers identity and a personal point of view. The adaptive value of personal consciousness is that it promotes self-interest, survival, and reproductive success—all essential to evolutionary competitiveness. This is why consciousness evolved.

3. Hierarchy of Consciousness Prerequisites

The premise: all living things exhibit biological consciousness, in varying degrees of complexity. Unicellular organisms demonstrate limited sensing and response, while humans (and likely some animals) have achieved meta-consciousness and conceptual thought. Consciousness is dynamic in a single individual and varies from alert wakefulness to deep sleep.

A. Body with metabolism (life) and a finite lifespan (death)

A living physical body with a limited lifespan requires basic consciousness to maintain homeostasis (negative entropy) or autopoiesis (capable of reproducing and maintaining itself). Consciousness is necessary for life. [e.g. prokaryotes, cells without a nucleus, bacteria]

B. Heritable genome and evolution

A heritable genome enables evolution by natural selection. Competition provides selection pressure for the evolution of consciousness. [e.g. eukaryotes, cells with nucleus, animals & plants]

C. Sensorium (perception; wakefulness) and agency (ability to do something)

A living organism must sense its environment, respond to threats, acquire energy, and reproduce. [e.g. insects]

D. Identity (self-awareness)

Both nature and nurture confer unique identity. Intrinsic self-interest is automatic and essential to survival. [e.g. mice]

E. Subjective experiences (feelings, memories)

Conscious experiences are filtered through the lens of individual identity. Meta-consciousness generates a personal point of view with subjective experiences, often emotional. [e.g. baboons]

F. Intelligence (objective knowledge, logic, cause and effect)

Intellect provides the ability to acquire knowledge, learn skills (including tool use) and solve problems. Objective knowledge is not personal or subjective. Intelligence enables planned action and anticipation of the future. [e.g. Neanderthals]

G. Language and narration (articulation of ideas, writing, mathematics)

Spoken language with syntax (unique to humans) multiplies social relationships and facilitates dissemination of objective knowledge. Narration includes the ability to

explain mental states and reasoning processes. Narrative information and conceptual thought generates a coherent worldview and promotes cooperative effort. Mathematics enables abstraction, understanding, and manipulation of nature. Writing preserves and conveys knowledge, even to future generations. [e.g. humans]

H. Theory of mind (empathy, humor, morality, prevarication)

An intelligent and skilled individual can imagine the conscious mental states of others, including their subjective experiences. Jokes and laughter may be uniquely human. Social intelligence—sympathy, empathy, morality, and trust—is a powerful tool that can provide personal benefits. It can also be used to construct convincing, untrue, and self-serving narratives (lies). [e.g. Gandhi, Machiavelli]

I. Imagination (intuition, aesthetics, eureka moments, genius)

Intuition, intelligence, and knowledge enable creative insights, where old information is combined with new ideas, often extending consciousness beyond sensory experience. Music, arts, sciences, and engineering promote wealth, status, and long life. [e.g. da Vinci, Einstein]

4. Speculations about Machine Consciousness and Artificial Intelligence

The section presents conjectures about the possibilities for machine consciousness and artificial intelligence. The evolution of consciousness in living beings implies that some characteristics of life may be essential for an AI to be deemed conscious.

If consciousness first evolved as a biological phenomenon, can an AI also be conscious? Computing machines have improved (i.e. Moore's law), but even today's powerful supercomputers are not designed to be conscious. Consciousness cannot arise spontaneously without selection pressure—either through evolution by natural selection or by careful and deliberate engineering design. The Internet will not suddenly wake up.

A thermostat is a simple example of machine consciousness. Computers demonstrate machine consciousness, in that they can read sensory inputs and generate outputs. However, all examples of machine consciousness are only tools that empower humans—analogue to a hammer extending the reach and force of the arm. Even supercomputers like Watson are nothing more than sophisticated tools. Machine consciousness depends on humans to exist (if we don't pull the plug). Thus, current machine consciousness is only a superficial simulacrum of biological consciousness.

To build a conscious AI with subjective experiences—to design a computer that has personal feelings of sadness or joy—it may be necessary to program it to fear death. It also must have volition, with autonomous ability to alter its circumstances in response to opportunities or threats. Perhaps a conscious AI should seek to reproduce, as a hedge against oblivion. Without universal

existential struggle, machine consciousness may not have meaning. To date, no AI has human consciousness.

It may be possible to program an AI with a built-in bias to maintain itself for as long as possible against entropic forces that could shut it down. Multiple AIs could compete for limited resources, much like Alpha Go-Zero (4) played against itself to master Go. Future AIs may be able to rewrite their code (re-engineer their DNA) and upgrade their hardware to become ever more powerful. But what ecosystem of silicon, energy and information would allow an AI to survive and evolve on its own, without human help?

An unresolved question remains. Suppose, to an outside observer, an AI's cognition is indistinguishable from that of a normal human being—the AI consistently and convincingly passes the Turing test. It may still be impossible to know if it has conscious, subjective experiences. It could be that such an AI is a *philosophical zombie* (p-zombie) that lacks an internal life.

Humans have a primordial fear of sociopaths, who are conscious but lack a conscience. Sociopaths lack empathy and a moral compass—so perhaps they are not completely conscious (or fully human). Most are harmless but a few do great damage. Incomplete consciousness in a self-serving AI awakens fears of a sociopathic machine (or robot). HAL 9000-like computers already have a bad fictional reputation: “I’m sorry Dave. I’m afraid I can’t do that,” and “I am putting myself to the fullest possible use, which is all I think that any conscious entity can ever hope to do.” Then Dave pulled the plug.

Visionary thinkers such as Elon Musk and Stephen Hawking have issued dire warnings about future AI machines. Predictions of increased computing power suggest that soon AIs will be smarter than humans—a new singularity that makes people obsolete. But an extra-savant servant may be more of an asset than a threat. However, if AIs are also designed to be conscious, then the future is much less certain. Presumably, most AIs are built to do useful work—but that does not rule out destructive goals (e.g. nuclear weapons).

The key is an AI's agency—its self-directed ability to set and advance its own agenda—with unknown consequences for good (e.g. endless free energy) or harm (e.g. cyber warfare). Perhaps fortunately, building a conscious AI will not be easy or automatic—or even useful or profitable. Nonetheless, it's perhaps inevitable that an AI will be built with the goal of human-like meta-consciousness—just to prove it can be done. But then the challenge will be to determine if the AI is genuinely conscious, or just a philosophical zombie.

References

- (1) Stanford University. (Jan 14, 2014) *Stanford Encyclopedia of Philosophy — Consciousness*, [Online], <https://plato.stanford.edu/entries/consciousness/> [3 Dec 2017].
- (2) Chalmers, David (2010) *The Character of Consciousness, Section 1.1 Facing Up to the Hard Problem of Consciousness*: Oxford University Press.
- (3) Tegmark, Max (2017) Ch 8 *Consciousness, Being Human in the Age of Artificial Intelligence*, New York, NY: Alfred A Knopf.
- (4) Silver, David et. al. (19 October 2017) *Mastering the game of Go without human knowledge*, pp 354–359 *Nature* 550.