

## The Self-Aware Machine:

### The Human Mind versus the Mind of Human Creation

By Joseph Worndl

Since the genesis of robotics, philosophers and engineers alike have considered the possibilities and limitations of artificial intelligence. Can robots think? Can computers ‘feel’? As Phillip K. Dick offers in the title of his 1968 novel, “Do androids dream of electric sheep?” The question of whether or not a machine can be conscious of itself proves fundamental. It is this question which dictates whether or not the mechanical can be viewed alongside the biological. Through an analysis of the nature of consciousness and the essential similarity between that of the engineered and that of the engineer, I will argue that a machine can indeed be self-aware.

To be properly understood, this argument need be understood as a thought-experiment rather than a concrete scientific proposal. I, with some humility, must confess my lack of expertise on the subject of electrical engineering. However, this proves immaterial in a broader sense. Technology and its perceived limitations change quickly. Until recent times, the very nature of this discussion would be inconceivable as it is predicated on the possibilities offered by high-technology. To suggest that a machine cannot be self-aware because technology could never be so sophisticated or advanced would be to say that humans just are not ‘smart’ enough to conceive of such technology. This practical interpretation of the issue, while suitable at an engineers’ meeting, renders a meaningless philosophical debate. As proven countless times from the invention of the rocket to that of the television, society’s view of technology’s possibilities and limitations is entirely relative to its context in the history of human development. Otherwise, the subject of this essay could just as reasonably be whether or not tiny people can survive in a rabbit-eared box and perform for people at scheduled times of the day. In short, my approach

toward defending the possibility of the self-aware machine need be founded on an examination of the essence of consciousness and not on the programming code necessary to realize such a vision.

Before I can support or discount the possibility of the self-awareness of artificial intelligence, I must offer some grounding in the nature of human consciousness. I am intelligent. I am capable of formulating thoughts, forming perceptions of the external, and recognizing my own existence. In case I doubt the last provision, René Descartes offers, “I think therefore I am.” Since I am formulating an argument, I must exist as I am doing something. Since I am thinking, I know that I am thinking; I therefore know I am doing and consequently know that I am.

However, this raises the question: “What animates me?” What allows for me to think thoughts and logically determine that I exist. Neuroscientists have determined that human thoughts and actions are controlled by electrical impulses coming from nerve cells regulated by physiological chemicals. The networks stemming from these components produce thoughts and regulate involuntary functions of the body. To an external observer without expertise in neuroscience, my brain would appear a pile of ‘goo’, without any evidence of intelligent thought. There exists no ostensible link between the physical and the psychological. Yet, I think as I write. You think as you read whether or not you recognize this piece as thoughtful. Ultimately, the electrical impulses moving through me cause this thought, the psychological product of physical workings which can produce the thought “I exist.”

Likewise, engineers regulate the function of machines through a non-organic mechanism. Too, this mechanism channels electrical impulses to its different components. Even the most complex computers of the day rely most basically on a series of ‘switches’ turning various

components ON or OFF, creating a pattern of ONs and OFFs formulating an ‘instruction’. The electrical switches in a mechanical mind (the central processing unit) function similarly to the networks of neurons in the human mind. In working while using this framework, artificial intelligence experts formulate a multitude of ‘thoughts’ for the computer to think. Similar to those of the human mind, the pattern of electrical impulses appears meaningless to the unscientific observer, but real thoughts *could* exist. Ultimately, the electrical impulses through the computer create this thought, the psychological by-product of the computer’s physical workings.

Consequently, in theory if not in practice, engineers could build a computer which could produce the thought “I exist.” Already, we see that engineers have managed to construct computers which can diagnose illnesses and file tax returns, all using variations of basic on/off switches. Based on precedent, it would then appear simple, though without much practical use, for the engineer to construct a machine that can formulate the thought “I exist”, using the system of ONs and OFFs. I could even rightly argue that today’s computers often demonstrate self-awareness, either explicitly or implicitly. The computer on which I compose this piece can determine the cause of its own malfunctions and suggest solutions to them (albeit with varying success). Others can identify when they become overheated and regulate the function of a fan to combat this problem. In truth, devising a self-aware machine would involve a mere further application of existing principles.

However, we must also consider the nature of self-awareness, whether or not it is simply a matter of “I exist.” The artificial intelligence skeptic might counter-argue that no quartzite concoction could ever be truly conscious of its own existence, citing that such a creation could not really experience authentic emotion, demonstrate an empathetic connection between itself

and other entities, or consider broader existential questions. In this respect, true self-awareness would not manifest itself. I would counter-counter-argue that this objection overcomplicates the issue of contention. Consciousness of one's own existence does not hinge upon emotion, empathy, or (much) introspection. The recognition of one's own existence involves a relatively simple thought. As aforementioned, today's computers perform a host of tasks far more complicated than a simple recognition of their own existence. Conversely, a human infant or any insect appears aware of its own existence as it attempts to defend itself when threatened or acts to preserve its life. Essentially, "I exist" is a simple conception, one which a simple form of intelligence could conceive.

Moreover, one may argue that any thought the computer has of self-awareness lacks authenticity because the machine is intelligently designed, reflecting the intentions of its engineer instead of its own free will. However, a Calvinist theologian would assert that humans' internal disposition is predetermined by a universal creator. Accordingly, we, obviously self-aware, exist analogously to intelligent machines. I argue that the ostensible theological implications of this issue prove immaterial. Regardless of its influences, a thought is still a thought. When individuals are influenced by advertising to buy a couch or frequent a buffet, they still possess that inclination. When individuals are taught by a religion to believe a particular doctrine, they still hold that belief. Even though I present imperfect analogies, this can still be seen when an entity is literally "hard-wired" to hold a particular belief such as "I exist." A thought, regardless of its influences, is thought by the thinker. A machine's recognition of its own existence can be initiated by any external influence. Yet, the recognition of its own existence still exists. One is not linked to the other.

Last, a stickler for accuracy in the usage of technological terminology would identify that, in defending my view that computers can formulate thoughts, I have described the function of various software programs in conjunction with the basic machine. True this is; however, this only reflects the fact that engineers engineer these discs separately and users later integrate them as part of the whole unit and sometimes remove them when no longer needed. However, these programs, when included, can be considered part of the whole. While the ability to insert and remove parts of the human does not often occur, the software program forms a part of the whole computer when present, complementing the whole. For the purposes of this piece, there exists no meaningful distinction.

While no perfect analogy exists between the human mind and the mechanical mind, several key similarities appear. Like humans, machines can, if they already do not, recognize their own existence. Like humans, machines are controlled by an internal mechanism. Like humans, machines can formulate thoughts based on the functioning of these mechanisms. In truth, the mind of the human is not particularly different from the mind of human creation. To this basic degree of self-awareness, engineers can mimic themselves in their machines.