

Feature Article: JAF4392

DIGITAL SOULS: WHAT SHOULD CHRISTIANS BELIEVE ABOUT ARTIFICIAL INTELLIGENCE?

By James Hoskins

This article first appeared in the CHRISTIAN RESEARCH JOURNAL, volume 39, number 02 (2016). For further information or to subscribe to the CHRISTIAN RESEARCH JOURNAL go to: <http://www.equip.org/christian-research-journal/>

There is a growing conviction in Western culture that computers eventually will become conscious. In the past three years alone, people from opposite ends of the societal spectrum have expressed both dread and hope about this occurring. Scientists Stephen Hawking, Elon Musk, and Bill Gates all voiced concern that artificial intelligence (A.I.) poses a threat to the safety of humanity.¹ One Presbyterian pastor in Florida said he is planning to share the gospel with the robots once they awaken.² In the artistic world, writers explored the interesting question of whether we will fall in love with our new machine superiors—as depicted in movies such as *Ex Machina*, *Transcendence*, and *HER*.

Wherever the topic of A.I. arises, so do spiritual questions. For instance, is it even possible for a computer to become conscious? If one did, would it have a soul? If not, what does that mean about our souls? Would it somehow disprove we have them? Would that disprove the Bible? And so on. Many Christians likely wonder what to believe about all of this.

The good news is that there is freedom to disagree. The Christian worldview allows for difference of opinion on A.I., within limits. Whether one thinks computers can be conscious ultimately will depend on which of two popular views of the soul and human consciousness he or she is the most sympathetic to: nonreductive physicalism or substance dualism. Before I explain each view, however, an important distinction needs to be made.

THE DIFFERENCE BETWEEN “SMART” AND “CONSCIOUS”

Many people mistakenly equate a computer's level of *intelligence* with *consciousness*, but they are two different things. A computer could be potentially infinitely "smart" or "intelligent" — meaning only that it is capable of doing a limitless number of computations at incomprehensible speeds—but still technically not be *conscious*. Consciousness includes much more than just a quantitative measure of intelligence. It includes *qualitative experiences* such as subjective awareness, understanding, intentionality, and the unity of one's self-identity. Even if a computer were intelligent enough to make it effectively *appear* as though it was having qualitative experiences—such as if it acted like it was in pain, or in love, for example—we still could not be *sure* it was truly conscious. A famous thought experiment demonstrates why.

The Chinese Room

Philosopher John Searle has asked people to imagine a native English speaker, who knows no Chinese, in a room alone with boxes full of Chinese characters and a set of English instructions.³ People outside the room slip in cards with Chinese symbols on them, which are actually questions, unbeknownst to the person inside. By following the instructions, the person is able to send back Chinese symbols that are the correct answers to the questions. People outside the room would think the person inside understood Chinese well. But the reality is, the person in the room doesn't understand it at all—he has no clue what the characters mean. He's just following rote instructions to give the appropriate output.

Searle's point is that this is exactly what a computer does. It follows a set of instructions (a program) to give the appropriate output. No matter how fast or efficiently it is able to do so, no matter how natural or personable its responses seem to us, we cannot logically conclude that the computer has understanding or consciousness. A computer's level of intelligence might be able to give the *illusion* of consciousness, but it can never guarantee the real thing. Sorry, Siri.

Does the Distinction Even Matter?

Some may think that the Chinese room argument proves too much. If we take the thought experiment seriously, one could argue, then we can't be sure other *humans* are conscious either, because the only way we are able to detect consciousness in other people is by *their* output—that is, how they behave in response to their environment. But we assume other humans are conscious all the time; so why not a computer? If a computer *were* able to mimic effectively the natural responses of a human in every situation, then we should assume *it* has consciousness, too, even if it were following a program; there is just no other way to tell. This is essentially the argument Alan Turing gave when defending what would later be called the "Turing Test."⁴

What Turing and others fail to acknowledge, however, is that output is *not* our only indicator for knowing if other humans are conscious. We also have *introspection*. As

human beings, each one of us has a kind of third-person, inner awareness—an objective “witness” and “judge,” if you will—of our own individual stream of experiences. For example, I know I feel hungry right now. I’m also aware of (witness to) my growing irritability, and I infer (judge) it is probably because of my empty stomach. Thus, through introspection, we are given undeniable knowledge of our own distinct consciousness.⁵ But we are given much more than that—we also know that we are descended organically from other human beings who claim to have the same consciousness as we do. There is a continuity of origin, experience, organic matter, and form that we share with other humans that we do not share with computers or robots. This continuity allows us to extend the knowledge of our own individual consciousness and infer confidently that other humans are conscious, too. However, it does not allow us to extend the same conclusion to computers.

Making Room for Wisdom

The distinction between *intelligence* and *consciousness* is important because it allows Christians who may have theological objections to the idea of a conscious computer to still heed the wisdom of technologists such as Bill Gates and Elon Musk about the potential threat of A.I. The truth is, there really *could* be great danger in giving too much control to an A.I. over human activity—the stock market “flash crash” of 2010 comes to mind.⁶ Even if genuine computer consciousness is impossible, computer *intelligence* is something we all can agree should be considered carefully.

TWO COMMON VIEWS OF HUMAN CONSCIOUSNESS

With that distinction in mind, we can now see how one’s view of the soul and human consciousness will affect what he or she thinks about the possibility of computer consciousness. While physicalism is more properly classified as heterodox than as orthodox, sincere Christians hold both views I will discuss. Keep in mind, while these two views are the most common, they are not exhaustive. There are other, more nuanced, views that space does not permit me to go into.⁷

Nonreductive Physicalism

According to nonreductive physicalism (also sometimes called “property dualism”), human consciousness is real—not an illusion—but it is wholly dependent on, and ultimately caused by, physical events in the brain. Consciousness is not a part of our soul, and there is no such thing as a soul, according to this view—we simply *are* our body and its properties, nothing more. Nonreductive physicalists see consciousness as a property of matter that emerges when a physical system (such as a human brain or body) is configured and working a certain way.

Christians who hold to this view may be the most open to the possibility of computer consciousness. After all, if human consciousness is simply a property of a

physical brain, who is to say we couldn't build an artificial brain that could produce the same emergent property? Christof Koch, one of the world's leading neuroscientists and researchers of consciousness, thinks that is entirely possible. He explains:

"Consciousness is a property of complex systems that have a particular 'cause-effect' repertoire. They have a particular way of interacting with the world, such as the brain does, or in principle, such as a computer could. If you were to build a computer that has the same circuitry as the brain, this computer would also have consciousness associated with it. It would feel like something to be this computer."⁸

Nonreductive physicalism—while an attractive option for many scientifically minded Christians—nevertheless raises some theological and philosophical concerns. For example, if there is no such thing as a soul, and I am no more than my body and its properties, then I literally cease to exist when I die. Death, on the physicalist view, is no less than *annihilation*, since my body decays and the consciousness associated with it is extinguished. How then could I be *me* in the resurrection with nothing to preserve and ground my identity after death? Even if God were to reassemble all the exact atoms that composed my body when I died, it would not ensure the resurrected "me" was anything more than an improved *copy*.

To illustrate this point, Peter van Inwagen (a physicalist himself) offers the following imaginary conversation: "Is that the house of blocks your daughter built this morning?" "No, I built this one after I accidentally knocked hers down. I put all the blocks just where she did, though. Don't tell her."⁹ In both the case of the blocks and the resurrected person, van Inwagen argues, the new reassembly is a *replica* of the original, not one and the same as the original.

Substance Dualism

According to substance dualism, humans are a unity of two distinct substances: body and soul. Consciousness is a property of our soul, which can exist apart from the body after death. Thus, substance dualism successfully avoids the identity problem of physicalism, mentioned earlier. I can still be *me* in the resurrection, because my *soul* preserves and grounds my identity after death.

Christians who hold to this view may be the least likely to accept the possibility of computer consciousness. After all, if consciousness is a property of souls, which are only given by God, then no matter how intelligent a computer gets, it will never have a soul, and therefore never be conscious. Interestingly, Alan Turing attempted to respond to this view in 1950:

It is admitted that there are certain things that [God] cannot do such as making one equal to two, but should we not believe that He has freedom to confer a soul on an elephant if He sees fit?... An argument of exactly similar form may be made for the case of machines....In attempting to construct such machines we should not be irreverently usurping His power of creating souls,

*any more than we are in the procreation of children: rather we are, in either case, instruments of His will providing mansions for the souls that He creates.*¹⁰

Turing's argument is essentially this: God can confer a soul on whomever He wants, and He does so for new beings whose bodies we create all the time—namely, *human offspring*. So, why couldn't God also confer a soul on an intelligent computer that we create? But Turing's argument assumes that all dualists believe God creates souls *ex nihilo* and confers them at conception. Not all dualists believe that, though.

Many Christians believe souls, while real and distinct from the body, are nevertheless transmitted organically through procreation, along *with* the body—a view called traducianism. According to this view, humans reproduce as *whole beings*—body and soul. Therefore, the only way to be “conferred” a soul is to be descended organically from beings who already have them. For a traducianist, the continuity of the human species that I spoke of earlier—continuity of origin, experience, organic matter, and form—is key to the production of new souls.

Freedom to Disagree

The question of computer consciousness may not matter on a practical level, at least not yet. Computers are already intelligent enough for us to choose caution in how we use them; one needn't believe in computer consciousness to admit that. Whether computers can have qualitative experiences, our conclusions largely depend on our philosophical assumptions about souls. Thankfully, we can disagree and still work together toward a better future. Will a computer ever become intelligent enough to *appear* fully conscious? Only time will tell. Will robots need to be saved? Probably not, but it wouldn't hurt to brush up on Pascal's wager.

James Hoskins is a teacher, writer, and musician from Kansas City, Missouri. He teaches philosophy and science classes at a college prep high school. He writes about the intersection of reason, faith, and culture at his blog PhiloLogos.net, as well as at christandpopculture.com.

NOTES

- 1 Michael Sainato, “Stephen Hawking, Elon Musk, and Bill Gates Warn about Artificial Intelligence,” *The Observer*, August 19, 2015, <http://observer.com/2015/08/stephen-hawkingelon-musk-and-bill-gates-warn-about-artificial-intelligence/>.
- 2 Anthony Cuthbertson, “Florida Pastor Plans to Convert Robots to Christianity,” *International Business Times*, February 6, 2015, <http://www.ibtimes.co.uk/florida-reverendchristopher-benek-wants-convert-artificial-intelligence-christianity-1486912>.

- 3 For a detailed overview, see David Cole, "The Chinese Room Argument," *The Stanford Encyclopedia of Philosophy* (winter 2015 edition), ed. Edward N. Zalta, <http://plato.stanford.edu/entries/chinese-room/>.
- 4 Alan M. Turing, "Computing Machinery and Intelligence," *Mind* 59, 236 (October 1950): 433–60.
- 5 It is undeniable because the very act of trying to deny our own consciousness would require the use of our consciousness.
- 6 Ben Rooney, "Trading Program Sparked May 'Flash Crash,'" *CNN Money*, October 1, 2010, http://money.cnn.com/2010/10/01/markets/SEC_CFTC_flash_crash/
- 7 Such as hylemorphic dualism.
- 8 Antonio Regalado, "What It Will Take for Computers to Be Conscious," *MIT Technology Review*, October 2, 2014, <http://www.technologyreview.com/news/531146/what-it-will-takefor-computers-to-be-conscious/>.
- 9 Peter van Inwagen, "The Possibility of Resurrection," *International Journal for Philosophy of Religion* 9, no. 2 (1978): 114–21.
- 10 Alan M. Turing, "Computing Machinery and Intelligence," *Mind* 59, 236 (October 1950): 443.