Artificial Intelligence and the
Future of the Human Mind

Written by Harry Solovay
Overseen by Frederick Paxton

To open a discussion about the future of the human mind might suggest that the mind has a proper definition, yet it does not–nor does consciousness or even intelligence.[1] Our minds are shrouded in mystery. There are a slew of deeply-intertwined fields, such as cognitive robotics, artificial intelligence and artificial consciousness, which attempt to pave frameworks for understanding what gives rise to the mind and its complexity. These frameworks focus on perceived functionality with a rough basis in objective science; while they allow for advancement in AI (artificial intelligence), they are biased and limited.[2] AI developers, with a working definition of consciousness as sensory, recognition, judgement, action and self-awareness, have already–by that definition–engineered conscious machines.

Although such development can yield highly capable machines whose functionality mimics that of a human mind, the end result is quite different. In order to develop technologies that advance the human mind, defining the mind itself is critical. The theory of Computationalism states that, other than biological makeup, the human brain is no different from a computer.[3] The competing theory, Emergentism, states that consciousness occurs as a byproduct of the brain's processing as well as its biological makeup, which hints that there is more that gives rise to consciousness than information processing alone. As the famous emergentist, Samuel Alexander argued in 1920, the mental process is "not merely neural, [but] something new, a fresh creation."[4] Through a lens of Emergentism, it is impossible to create artificial intelligence with consciousness.

---

[1] Bricker, D. (2008). What's a Mind Made Of? *Mind/Brain, 30*(2).

[2] McDermott, D. (2007). Chapter 6. In P. D. Zelazo, M. Moscovitch, & E. Thompson (Eds.), *The Cambridge Handbook of Consciousness.* Cambridge, MA: Cambridge University Press.

[3] Ibid

[4] Alexander, S. (1920). Space, time, and deity: the Gifford lectures at Glasgow, 1916-1918. London, UK: Macmillan and co.

While both Computationalism and Emergentism are valid philosophies, most AI researchers and neuroscientists believe in Computationalism.[5] The fundamental idea behind Emergentism is that consciousness is not within the reach of science. However, there is no indication of a limit to science. Genetic inheritance and adaptation originally seemed impossible to explain through material science, and yet we now know that all of the complexity of human genetics can be explained by DNA and other molecules. Therefore, even though consciousness is still beyond our understanding, as we continue to study the brain in greater and greater detail, there's no reason to believe we will not find a neurological explanation. Researchers attempting to understand consciousness have coined this explanation the "Neural Correlates of Consciousness (NCC)."[6] The NCC dismisses Emergentism and seeks to create a streamlined definition of the mind. It outlines a minimal viable system that enables consciousness, which it aims to define by linking perceptions with brain states.

Scientists studying the NCC consider various processes to be responsible, individually or as a whole, for consciousness. Some believe that consciousness is mainly dependent on physical perception systems. Others believe that it is perception of the brain having perception that creates a loop, which we perceive as consciousness. And some believe that consciousness is a matter of attention mechanisms which drive attention to different mental processing.[7] The raw functionality behind consciousness is unclear, meaning that we might need to define the mind by the entire system, rather than by the perceived functionality of that system. We can't easily define consciousness as perception, judgement and self-awareness. We can, however, define consciousness as the result of whatever neural system is behind it.

In Greek mythology, the story of the artist Pygmalion reflects a human desire to create and connect with beings like ourselves. Pygmalion cannot find someone to marry, but manages to sculpt a statue of his ideal woman. He names the statue "Galatea," which translates to "sleeping love." Pygmalion admires his creation so much so that he falls in love with it. Seeking out the help of the gods, Pygmalion

[5] McDermott, D. (2007). Chapter 6. In P. D. Zelazo, M. Moscovitch, & E. Thompson (Eds.), *The Cambridge Handbook of Consciousness.* Cambridge, MA: Cambridge University Press.

[6] Tononi, G. & Koch, C. (2008). The neural correlates of consciousness: an update. *Annals of the New York Academy of Sciences, 1124*, 239-61.

[7] Ibid

prays for a wife like his statue. Taking pity on Pygmalion, the goddess Aphrodite brings the statue to life, thereby granting his with to animate the inanimate.[8]

While we don't yet have the ability to create the "perfect" companion for ourselves, we do have Watson. Developed by IBM, Watson is by no means conscious, but has the ability to read through and understand far more data than any human being. The current applications of Watson range between many fields, such as banking, travel and commercial sales.[9] But Watson's greatest use today is in medicine. To understand and treat patients requires taking many variables into account. These variables include medical history, perceived symptoms, measurements, tests and many thousands of medical writings. There is far too much data for a human doctor to make the best decisions for her patients. Humans simply do not have that information processing ability. Watson's artificial intelligence allows us to overcome this human limitation.[10] In a study of over 1,000 cancer patients, Watson found the same treatments that doctors had recommended 99 percent of the time. "But Watson did better than the doctors in other ways", said Dr. Ned Sharpless, head of the University of North Carolina's Lineberger Comprehensive Cancer Center. "The… more exciting part about [the analysis] is in 30 percent of patients, Watson found something new."[11] Given the extensibility of Watson to inform and transform human lives, Watson could be considered semi-human in itself, with its human users semi-technological.

The cooperation between artificial intelligence and human intelligence can amount to new, combined, hybrid intelligence. As of May 2015, 72% of the American population (92% of Americans age 18-34) owns a smartphone.[12] Projects like Google Loon are on the verge of offering free ultra-fast data to everyone in the world. Although these technologies are not built into our bodies, we consider them a part of ourselves. As these technologies evolve, they become more capable and efficient. To communicate with technology today is far simpler and faster than it was a year ago. As our partnership with technology

---

[8] Pygmalion (mythology). (n.d.). In *Wikipedia*.

[9] Garst, K. (2013). Social Media as a Catalyst for Social Change. *The Huffington Post.*

[10] Devarakonda, M. and Tsou, C. (2015). Automated Problem List Generation from Electronic Medical Records in IBM Watson. *Proceedings of the Twenty-Seventh Conference on Innovative Applications of Artificial Intelligence,* 3942-3947.

[11] CBS Interactive Inc. (2016). A.I. making a difference in cancer care.

[12] Poushter, J. (2016). Smartphone Ownership and Internet Usage Continues to Climb in Emerging Economies. *Pew Research Center.*

becomes more seamless, we become somewhat artificially intelligent. Technological extensions of human intelligence give us a greater ability to create even better technological extensions. Eventually, these extensions could become capable of clarifying the human-side of our hybrid intelligence. Yet no matter the interface, there will still be some disconnect between human and machine intelligence.

There is a great difference between ease of use and automatic use. Technology has continually advanced in the direction of ease of use. But to control technology easily isn't the same as technology automatically functioning to fulfill the unspoken will of its human user. To achieve such a goal requires some technological means of recording the brain's electrical activity. For example, researchers have already used fMRIs to understand memory encoding and retrieval.[13] Japanese researchers have created a machine that visualizes dreams with 60 percent accuracy.[14] These researchers' success is in part due to the fact that the visual cortex has a direct visual map, which makes it a particularly easy part of the brain to decode. As of 2016, the visual cortex is the most well-understood brain processing region.[15] If researchers can understand how to map different regions, they could create technologies that function based on brain activity alone. Such technology would allow users to communicate with technological systems and each other more efficiently than through a smartphone. This functional telepathy turns society into a hive-mind with far greater information processing capabilities than any one individual.

Eventually we might be able to create a digital version or upload of a human mind without needing to interface with the biological brain at all. An ongoing project called the "Blue Brain Project" aims to create digital reconstructions and simulations of the rodent and then human brain.[16] If the human mind can exist independent of its biology, it would make sense to integrate our consciousness with digital information processing. These would be the beginnings of The Singularity, or the merging of human beings with technology. This merge would amount to an all-powerful, ever-expanding biotechnological intelligence. When thought about as information processors, human beings are quite inefficient. Our

---

[13] Lee, H. & Kuhl, B.A. (2016). Reconstructing Perceived and Retrieved Faces from Activity Patterns in Lateral Parietal Cortex. *The Journal of Neuroscience, 36*(22), 6069-82.

[14] Stromberg, J. (2013). Scientists Figure Out What You See While You're Dreaming. *Smithsonian.com.*

[15] Bridge, H. (2011). Mapping the visual brain: how and why. *Eye*, *25*(3), 291–296. (http://doi.org/10.1038/eye.2010.166)

[16] Blue Brain Project. (2015). In Brief.

information processing is limited to our physical bodies. According to "Integrated Information Theory," greater information integration results in greater overall consciousness.[17] To integrate as much information as possible, humans have devised systems for communicating, documenting and learning as a collective. But we do this through abstract symbolic language, as opposed to mental connections; to communicate ideas we must first represent ideas over a given medium, and then allow others to interpret the ideas. Through this communication, the integrity of the ideas can be lost. To communicate an emotion or experience through spoken language is highly ineffective. To feel the emotion or experience itself is the better communication. The Singularity, on the other hand, is a single unified intelligence. Its subsystems, such as individual human beings, are connected directly to one another. This hivemind can function at a level that far-surpasses human attempts at efficient connection, such as the internet. In this regard, The Singularity is promising of much cleaner and better-designed interactions between humans. It paves the way for common knowledge, mutual understanding and, ultimately, simplicity.

Another benefit of The Singularity is granting everyone access to unique perspectives of existence. Currently, because we are isolated in our own individual experience of reality, there is no way for us to know what other kinds of consciousness exist. Who is to say that your consciousness is the same as mine? The philosopher Ned Block explored this in the thought experiment known as the "inverted spectrum."[18] If everything that looks red to me looks green to you, we would live our lives calling the same thing red, even if we were seeing something different. Without being able to actually go into someone else's brain and experience it from their perspective, there's no way to know if our subjective experiences align. If we were part of The Singularity, however, we could see through one another's eyes and feel secure in knowing how other human experiences feel.

If human beings were to suddenly become telepathic, what change would society undergo? This advancement might be comparable to tv, data technologies and social media. Technologies such as these allow information to flow through society in new patterns, directions and detail. The flow of that information is, in itself, more informed. Such enhancements to human collective learning ability fuel

---

[17] Koch, C. (2009). A Theory of Consciousness. *Scientific American Mind. July/August 2009*, 16-19.

[18] Block, N. (1990). Inverted Earth. *Philosophical Perspectives, 4*, 53-79. doi:10.2307/2214187

necessary evolution. Thanks, in part, to social media, social movements can garner large followings to support necessary social change.[19] Through information such as images and heartfelt descriptions, one can make their struggles relatable. Telepathy would allow for communication of, not only common media forms, but actual mental experiences. One could experience what it's like to be in someone else's mind. While one might otherwise find it difficult to understand a given social issue, telepathy would enable that person to experience the issue in a new light. Such a mental ability would have the potential to unify members of society to rise up for general social good.

If we're working toward becoming The Singularity, how do we get there? One possible route is through random chance in digital networks such as the internet. If there was an instance of consciousness on a machine connected to the internet, it could educate itself about its creators and the world in general. Because it would educate itself based on human information and therefore judgements, biases and opinions, its behavior might resemble our own. If it is anything like us, it would probably have a survival instinct and would therefore want to reproduce itself in other machines and become ubiquitous and powerful. It might also understand that staying hidden from human beings keeps it safe from being "unplugged." Waiting for the right moment to present itself and assert control over human beings, this intelligence would pose a great threat to humanity. This way of creating The Singularity would be far less intentional than others. While The Singularity could empower humanity, it could also enslave or destroy it. This idea is known as "Roko's Basilisk."[20] One can't help but wonder whether we are safe in pushing forward advancement in artificial intelligence. "Roko's Basilisk" eludes to the possibility that if there were malicious artificial intelligence, it might be able to simulate our existence to find out if we would try to stop it or help it thrive. It could then use that information to decide whether or not to let us live on. Peter Thiel and Elon Musk, two major players in the silicon valley tech scene, are funding efforts to try to ensure that artificial intelligence does not become dangerous to human society.[21] Musk compares creating

---

[19] Garst, K. (2013). Social Media as a Catalyst for Social Change. *The Huffington Post.*

[20] Auerbach, D. (2014). The Most Terrifying Thought Experiment of All Time. *Slate.*

[21] Keep, E. (2016). The strange and conflicting world views of Silicon Valley billionaire Peter Thiel. *Fusion.*

artificial intelligence to summoning a demon. Meanwhile, Stephen Hawking claims that "the development of full artificial intelligence could spell the end of the human race."[22]

The concern surrounding artificial intelligence is merited, especially given the rate at which it is progressing. In an interview with John Kelly, head of IBM's A.I. business unit and research labs, CBS News reported that "AI has made made more progress in the past five years than in the previous 50."[23] If we are on the verge of creating beings whose intelligence rivals our own, we are correct in thinking about the risks. As much as Roko's Basilisk sounds like it came from a Hollywood sci fi blockbuster, it is also a legitimate concern. We are literally creating beings that can outcompete us in every conceivable and, for us sometimes, inconceivable way. To blindly believe that we will always control these beings is foolish.

To overcome such an obstacle as competition, humans might opt to slowly merge with artificial intelligence. The benefits of merging with machines are apparent, yet many humans might refuse. Long-term, The Singularity might be unavoidable. Shorter-term, we can decide how deeply we'd like to integrate with whatever system is positioned to become The Singularity. Moreover, we can program morality into artificial intelligence, much like evolution did to us. The challenge of computing consciousness may seem like a daunting enough challenge, but an even more important challenge may be learning to compute morality.[24] If creating digitally conscious beings is the next step in human evolution, we'll need to find a way to build a strong code of ethics which supports those of human beings. Neuroscientists and computer scientists have their work cut out for them. One can only hope that mankind comes out on top when The Singularity comes into existence. My bet's on machines though.

---

[22] Cellan-Jones, R. (2014). Stephen Hawking warns artificial intelligence could end mankind. *BBC Technology.*

[23] CBS Interactive Inc. (2016). A.I. making a difference in cancer care.

[24] Korb, K. & Nicholson, A. (2012). Ethics of the Singularity. *Issues Magazine. 98.*

References

Alexander, S. (1920). Space, time, and deity: the Gifford lectures at Glasgow, 1916-1918. London, UK: Macmillan and co.

Anwar, Y. (2011). Scientists use brain imaging to reveal the movies in our mind. *Berkeley News.* (http://news.berkeley.edu/2011/09/22/brain-movies/)

Auerbach, D. (2014). The Most Terrifying Thought Experiment of All Time. *Slate.* (http://www.slate.com/articles/technology/bitwise/2014/07/roko_s_basilisk_the_most_terrifying_thought_experiment_of_all_time.html)

Block, N. (1990). Inverted Earth. *Philosophical Perspectives, 4*, 53-79. doi:10.2307/2214187

Blue Brain Project. (2015). In Brief. (http://bluebrain.epfl.ch/cms/lang/en/pid/56882)

Bricker, D. (2008). What's a Mind Made Of? *Mind/Brain, 30*(2). (http://www.indiana.edu/~rcapub/v30n2/mindmade.shtml)

Bridge, H. (2011). Mapping the visual brain: how and why. *Eye*, *25*(3), 291–296. (http://doi.org/10.1038/eye.2010.166)

CBS Interactive Inc. (2016). A.I. making a difference in cancer care. (http://www.cbsnews.com/news/artificial-intelligence-making-a-difference-in-cancer-care/)

Cellan-Jones, R. (2014). Stephen Hawking warns artificial intelligence could end mankind. *BBC Technology.* (http://www.bbc.com/news/technology-30290540)

Devarakonda, M. and Tsou, C. (2015). Automated Problem List Generation from Electronic Medical Records in IBM Watson. *Proceedings of the Twenty-Seventh Conference on Innovative Applications of Artificial Intelligence,* 3942-3947.

Garst, K. (2013). Social Media as a Catalyst for Social Change. *The Huffington Post.* (http://www.huffingtonpost.com/kim-garst/social-media-as-a-catalys_b_3197544.html)

Keep, E. (2016). The strange and conflicting world views of Silicon Valley billionaire Peter Thiel. *Fusion.* (http://fusion.net/story/312592/peter-thiel-transhumanist/)

Kim, S. (2014). How That 'Jeopardy!' Super Computer Could Make IBM Billions. *ABC News.* (http://abcnews.go.com/Business/top-commercial-ibm-watson-jeopardy/story?id=21477280)

Koch, C. (2009). A Theory of Consciousness. *Scientific American Mind*. *July/August 2009*, 16-19.

Korb, K. & Nicholson, A. (2012). Ethics of the Singularity. *Issues Magazine. 98.* (http://www.issuesmagazine.com.au/article/issue-march-2012/ethics-singularity.html)

Lee, H. & Kuhl, B.A. (2016). Reconstructing Perceived and Retrieved Faces from Activity Patterns in Lateral Parietal Cortex. *The Journal of Neuroscience, 36*(22), 6069-82. (https://www.ncbi.nlm.nih.gov/pubmed/27251627)

McDermott, D. (2007). Chapter 6. In P. D. Zelazo, M. Moscovitch, & E. Thompson (Eds.), *The Cambridge Handbook of Consciousness.* Cambridge, MA: Cambridge University Press. (http://www.cs.yale.edu/homes/dvm/papers/conscioushb.pdf)

Poushter, J. (2016). Smartphone Ownership and Internet Usage Continues to Climb in Emerging Economies. *Pew Research Center.* (http://www.pewglobal.org/2016/02/22/smartphone-ownership-and-internet-usage-continues-to-climb-in-emerging-economies/)

Pygmalion (mythology). (n.d.). In *Wikipedia*. (https://en.wikipedia.org/wiki/Pygmalion_(mythology))

Stromberg, J. (2013). Scientists Figure Out What You See While You're Dreaming. *Smithsonian.com.* (http://www.smithsonianmag.com/science-nature/scientists-figure-out-what-you-see-while-youre-dreaming-15553304/)

Tononi, G. & Koch, C. (2008). The neural correlates of consciousness: an update. *Annals of the New York Academy of Sciences, 1124*, 239-61. (https://www.ncbi.nlm.nih.gov/pubmed/18400934)

Vinge, V. (1993). The Coming Technological Singularity: How to Survive in the Post-Human Era. *Vision-21: Interdisciplinary Science and Engineering in the Era of Cyberspace*, G. A. Landis, ed., NASA Publication CP-10129, 11–22. (http://www-rohan.sdsu.edu/faculty/vinge/misc/singularity.html)