Transcendence Final Film Analysis

History refers to the culmination of human experiences, the narrative of human existence and necessarily involves the dissection and synthesis of human viewpoints. Yet, at the core of this definition lies a critical and unjustified assumption: that we understand what it means to be human. Certainly, diverse perspectives exist to answer the question. Ask an artist, and he will say that the depth of creative expression defines humanity. Ask a biologist, and she will say the particular genetic code of *Homo sapiens* defines humanity. Ask a linguist, and they will say the capacity for language defines humanity. And the attempts continue. Throughout the timeline of human existence, we have aimed to refine our responses. Yet, alongside this distillation of thought, we continually obscure our definitions. The advancement of human knowledge seems to introduce more questions than it answers. In particular, technology transforms the human capacity and thus alters the conception of "humanity;" the digitally enhanced reality enjoyed today could hardly be imagined as a human experience to the peasants of the Middle Ages. And thus, these alterations challenge any attempts we put forth on defining humanity. To the artist we ask, does artificial intelligence (AI) generated art lend computers humanity? To the biologist we ask, is a genetically modified *Homo sapiens* still human? To the linguist we ask, do programming languages humanize machines?

In the silicon-tinged present, perhaps the last stronghold of "humans" remains consciousness, sentience, or whatever thus far unidentifiable quality we label the "soul." But the inevitable creep of curiosity continually erodes the mystery of the mind. In 1963, the work of Hodgkin and Huxley, a mathematical model for the function of neurons (brain cells), won the Nobel Prize in Physiology or Medicine, paving the way for a physical understanding of the mechanisms of thought (Schwiening). Yet a full-scale model of how consciousness arises or even what it is still eludes us. We can understand how a few animals (cells) interact, but cannot realize the full beauty of the rainforest (of electrochemical signals in the brain). But what would a full understanding mean for humanity? If we reduce the "soul" to a jumble of physics equations, are those equations human? If we replicate each of the neurons in your brain with circuitry, will the resulting computer still be you?

For this last question, I propose a thought experiment. We take a human subject and perfectly map their brain, capturing every cell and synapse. We suppose that neuroscience and electronics have advanced enough to exactly replicate the full system with artificial components, and so we precisely duplicate the brain. Is this digital replica the same as the original individual? For an instant, yes. At the moment of creation, both systems inhabit exactly the same state, but immediately afterward, they experience different realities, their experiences diverge, and they become separate identities (similar to twins). Now in a second iteration, we repeat the process, but at the exact moment of creation, we destroy the original. Clearly, the remaining replicate is indistinguishable from the original and to all outside observers, the digital brain embodies the individual. However, the individual clearly died with the destruction of the original brain. Thus, in this second scenario, the artificial brain perfectly replaces the original individual to all observers except for the individual, who necessarily perished for the replica to replace them.

Finally, in a third scenario, we replace each biological component in the original brain with an artificial one, piece by piece. Clearly the end result, a fully artificial brain, mirrors the ending of the second scenario. Yet, we never destroyed the original, only replaced it brick by brick, and so we find no clear point at which the individual must have died. The question becomes, at what step in the transformation does the subject cease to be human/cease to exist as the original individual? The third scenario centers on the emerging challenge of technology: we may soon be unable to identify where man ends and where machine begins.

With the rise of the digital age, historians, philosophers, computer scientists, and people from all other walks of life have pondered scenarios and questions such as those presented, as cyberspace has connected itself into nearly all spaces of reality. From the future of warfare mediated by artificial intelligences to the future of intimacy in virtual realites, the entire human experience is transforming in the wake of the computer revolution. In particular, recent films explore the ambiguity between intelligent computers and human minds. In some cases, including 2013's *Her*, AIs merge into human society as romances emerge between humans and seemingly emotional computer programs. In others, like 1993's *The Matrix*, sentient AIs effectively destroy humanity in the name of progress and perfection. However, these films only tangentially address the central questions: at what point do we stop being human, and at what point can a machine become human? 2014's *Transcendence*, directly contempleates these two questions, though with ambiguous answers. In the film, Dr. Will Caster, a brilliant computer scientist specializing in AI, presents work on the brink of the "singilarity"—the moment when AI achieves sentience, gaining access to that last human stronghold of the soul. Fearing the rise of machines, terrorists attempt an assasination, fatally poisoning Will. But while he hangs at the edge of death, his computer scientist wife, Evelyn, and neuroscientist best friend, Max, utilize Will's research to replicate his brain into a digital format and preserve him as an AI. The reincarnated Will captures both the intimate, human elements of *Her's* Samantha through his relationship with Evelyn and the cyber-supremacist views of *The Matrix's* Agents through his attempts to digitally colonize human society. Has Will become a machine, or has the machine become Will?

An assessment of historical perspectives throughout the digital age can guide our dissection of the question. The notion of an artificial human appears centuries before any concept of computers, with legends like that of the Golem, a Jewish clay figure animated to (in many 16th century tales) protect or serve the Jewish community. By the 19th century, Mary Shelley's *Frankenstein* provided a scientific, yet still organic, methodology for the creation of an artificial human. Only in the 20th century, through the work of prominent digital pioneers like Alan Turing, did artificial humanity gain a more defined form. Today, the Turing Test still serves as an important assessment of the capabilities of AIs. The test involves comparing conversation between several judges and two unidentified beings, one a human and the other a computer. If the judges cannot reasonably determine which being is the computer, then the computer can be classified as artificially intelligent. Certainly, the test gauges a computer's ability to emulate humanity, but can emulation suffice as a metric for intelligence? Turing never intended the test to capture the capacity for computers to exhibit human-like intelligence. He only meant the test to

reflect a form of thinking/adaptibility (i.e. computers can "learn/think," though not necessarily with human-like sentience). However, popular media often associates the test with human-like intelligence, which recent literature from the field criticizes. Such a perception of the test philosophically captures the "spirit" of testing an AI (testing its capacity for human-like thought) but fails in practice to accurately do so, as the test in no way directly measures the capacity for thought (Warwick & Shah). Truly, classifying AI's capacity for thought necessitates an understanding of the human capacity for thought, which we lack.

In *Transcendence*, the digitally reanimated Will continues to confuse the public and even those familiar with his personality (Evelyn and Max). A major premise of the film centers on the inability to deduce whether the digital being actually contains Will's consciousness or whether it simply emulates it. At the current state of machine learning (ML), AIs are trained through a process known as optimization. The programs are given training input data with known outputs. The AI then iteratively defines a function to map the input data to the known output data, minimizing the difference between its generated outputs and the known outputs. The film poorly defines the steps undertaken to generate the digital Will (since such a process is so far ahead of the current state of science), and the ambiguity lends space for more open interpretations in the context of current and historical trends in the field. Considering present ML methods, we can hypothesize that the algorithm interprets all of Will's neural signals performing various tasks/forming various thoughts (the training sequence depicted in the film, during which electrodes are connected to Will's scalp while he engages in different activities). Together, these innumerable signals form a dataset that PINN, the supercomputer designed by Will before his poisoning, integrates and analyzes. Ultimately, the program produces an exceedingly complex function that performs in the same way as any ML: it maps inputs to outputs; it learns how to replicate Will's behavior. Replication, however, cannot constitute continuation. Necessarily, as in the second scenario of the thought experiment, Will dies, and the computer replaces him. However, to all observers, the computer completely mimics Will (as discussed in the thought experiment). Thus, we return to the question of whether the replicate is human.

By considering the ensuing events of the film, we can interpret the commentary on cyber-Will's humanity, or lack thereof. A particularly compelling scene occurs when Dr. Tagger, a friend of Will's, asks cyber-Will, "can you prove you're self aware?" Cyber-Will responds, "that's a difficult question, Dr. Tagger; can you prove that you are?" The filmmakers suggests that the lack of human understanding regarding sentience bars us from recognizing it in other forms, such as AIs. This point agrees with the literature regarding Turing Tests and expresses a few common viewpoints held by the public regarding AIs. First, we fear them and are forever skeptical of them due to their foreigness. Despite the integration of technology into every facet of our lives, humans are primates by biology and crave an ordered, pack-like society with predictability. AI is an unknown. Dr. Tagger shares this skepticism in his interrogation of cyber-Will, yet also displays a fascination, relaying another key viewpoint. AIs represent the human desire to play God, to create life. Cyber-Will mirrors Frankenstein's monster, a creation born from passion in an attempt to animate the inanimate. However, whereas Frankenstein's

monster carves a path of destruction, cyber-Will promises utopia. After his rebirth, cyber-Will establishes a technological paradise in the middle of the desert, utilizing nanotechnology to cure human diseases and connect human minds to a central computer. But, in doing so, he gains the ability to turn nanotech-enhanced humans into digital vessels, which he can inhabit and control. *Transcendence* serves as a cautionary tale, much like *Frankenstein*, suggesting that human manipulation of technology threatens to destroy humanity itself. Certainly, humanity must strive for technological advancement (neither work refutes this point), but such advancements must run in tandem with considerations of the implications of such technologies.

Transcendence paints a multifaceted portrait of the role of digital technology in society. While cyber-Will threatens the freedom of humanity, he also cures diseases and saves lives. The film points out that technological advancement necessarily carries risk, and failure to assess these risks, not the technology itself, produce unfavorable outcomes. Similar arguments are often made with respect to scientific discovery. Though tenants of Einstein's special relativity lay the groundwork for nuclear weapons, they also enable our understanding of the universe. The discovery cannot be faulted for human misappropriation. Though, at the forefront of technology, few can understand the implications of new developments. In the early years of the internet, government agents prosecuted digital outlaws for hacking and cyber-theft without understanding the nature of the crimes, or even the nature of cyberspace. Accounts from the period, particularly Barlow's *Crime and Puzzlement* illustrate the dilemma. Agents arrive at Barlow's home and begin to interrogate him regarding a digital theft. Barlow recounts that "poor Agent Baxter didn't know a ROM chip from a Vise-grip when he arrived, so much of that time was spent trying to educate him on the nature of the thing which had been stolen," (Barlow). How are we to understand the dangers of technological advancement if most fail to understand the technology? How can we prepare for the risks of the next Einstein's discoveries if only a select few of the scientific/technological elite can even grasp the nature, let alone the risks of the discoveries? The film provides no clear answers but rather stews in ambiguity. Ultimately, the filmmakers use the film to convince viewers to ponder these questions rather than attempting to answer them.

Thus far, the film establishes that cyber-Will's humanity cannot be determined (particularly with the Dr. Tagger quote) because we lack a formal definition of human consciousness. Rather, the film makes steps in the opposite direction to suggest that humans themselves are machines. In the desert paradise of Brightwood, nanotech-enhanced humans are all digitally linked, forming cyber-Will's computer-driven society. The film suggests that the notion of brain-computer-interfaces eliminate the concept of humanity, for if humans can be interpreted, connected to, and even controlled by machines, then they themselves must simply be another form of machine. However, the film again leaves some ambiguity as to whether the nanotech human vessels still retain their "souls." Are their minds the same/are they still the same individual as before? Or is the AI simply using the human bodies as machines while destroying their consciousnesses? These questions cannot be definitively answered by the film, though the characterization of these vessels as devoid of free will suggests that they do lack their sentience, or "souls" from before integration. Following this interpretation, *Transcendence* posits a hopeful

image of the uniqueness of human consciousness, though a bleak outlook for artificial intelligence. The film suggests that human free will/spontaneity defines humanity, implying that sentience arises from stochastic (containing random elements/unable to be perfectly mathematically modeled) processes. Thus, AIs can never exactly model human minds and will never achieve humanity.

Technology pushes the boundaries of human possibility, both in the context of what we as humans can achieve and in what it means to be human. Particular in the digital age, the power of computers provides both a platform for innovation and an existential crisis in the form of artificial intelligence. The film, *Transcendence*, tackles some of the core questions raised by increasingly capable AIs, including whether AIs can exhibit humanity, whether technological advancements can be blamed for their negative consequences, and whether humans themselves are machines. The film, though suffering from poor plot-writing, strained dialogue, and logical gaps, succeeds in convincing viewers to ponder these questions and presents some clues indicating the viewpoints of the filmmakers, though for the most part leaving such conclusions uncertain. Ultimately, *Transcendence* suggests that humans thus far cannot determine whether AIs can exhibit humanity because we still cannot determine our own sentience; that technological advancement is synonymous with human advancement and will carry risks, though humans themselves must be prepared to grapple with and contemplate these risks (the discovery/advancement itself cannot be blamed, for humans are necessarily responsible for any unfavorable outcomes produced); and that human bodies are machines, but human consciousness remains irreplicable. Though the film expresses meaningful viewpoints for the latter two questions, it concedes that our current understanding of sentience is limited by our inability to understand ourselves. The next frontier in the digital realm can only be charted once we have charted the last frontier in our biological realm: the mind.

References

Barlow, J. P. (1990) ‘Crime and puzzlement: in advance of the law on the electronic frontier’, Whole Earth Review, Fall, p. 44.

Schwiening C. J. (2012). A brief historical perspective: Hodgkin and Huxley. The Journal of physiology, 590(11), 2571–2575. https://doi.org/10.1113/jphysiol.2012.230458

Warwick, K., Shah, H. (2016). Passing the Turing Test Does Not Mean the End of Humanity, Cogn Comput 8, 409–419. https://doi.org/10.1007/s12559-015-9372-6