# Offloading Consciousness: Will Cyborgs Have Selves?

by
Will Wright

A thesis presented to the Honors College of Middle Tennessee State University in partial fulfillment of the requirements for graduation from the University Honors College

Spring 2015

Offloading Consciousness: Will Cyborgs Have Selves?

by
Will Wright

APPROVED:

Dr. Ron Bombardi
Philosophy

Dr. John Pennington
Psychology
Honors Council Representative

Dr. Philip E. Phillips
Associative Dean
University Honors College

## ABSTRACT

As technology persistently alters and improves on our subjective, perceptual tools and the experience they allow, we are forced to ask what exactly it is that makes a minded self. The mind is generated by neuronal systems, physical formal networks in the brain. The resultant phenomenon of a self is, therefore, the necessary result of complex connections and regulatory structures. Artificial replication of these systems will allow us to offload pieces of our functionality onto that of an identical machine. This outsourcing of ability will not only alter what it means to be human but will demonstrate how it is that matter produces the subjective nature of experience. The answer to the puzzle of consciousness lies in the technology which we create to supplant functional parts of it. Artificial forms of intelligence will allow humans to objectively analyze the nature of the mind, simultaneously understanding and reshaping the essence of self.

Table of Contents

# I. INTRODUCTION

Life is an astonishing predicament to be in. What is even more astonishing is that there is something to which life's predicament can appear astonishing. Surprisingly, our brains do not seem at all amazed by the miniscule intricacies and the grand complexities of the universe. It appears, rather, that it is we who are the arbiters of understanding, the begetters of about-ness and the manufacturers of meaning. Despite this ostensibly unique position, however, the nature of this self remains peculiarly uncertain. Most of us *feel as if* we have a firm grasp on what it means to be a self. We *feel as if* our nature, our essential being, is apparent and understood, perhaps not by everyone but definitely by the selves which are making said claim. As we begin a thorough examination of what the nature of self is, the illusion of expertise begins to fade away. At the hands of scientific exploration, the conscious self takes on an antithetical character, one that undermines its foregoing recognition.

The entirety of human endeavor is directed toward the purpose of understanding. Understanding the operations of the world allows a higher capacity to survive; therefore those that do it best appear, if only slightly, more frequently in the genetic pool. This slight edge is enough. Every form of knowledge has this incentive as its origin. Our species has been searching for the answers to life's great problems since before modern cognitive functionality first appeared. Unwittingly, our ancestors started a tradition of searching for some meaning or insight in order to gain a deeper proficiency which we are continuing to this day. The proficiency that understanding endows naturally gives the

owner a leg up in competition for resources. It is evident that Homo sapiens are currently in the lead with this aptitude for knowledge; nonetheless, the zenith is far from reached.

Philosophical reasoning, scientific theories, and religious institutions rise and fall, answers are given and taken away, all in an effort to encapsulate the pursuit towards true acumen. No problem, however, is more mysterious in its intricacies and intimacies than that of our own minds. It is true that we are thinking things, but what is the quality of this thought, what is its purpose and how does it appear to be. This deep-seated, innate desire to uncover the esoteric realities of our world is the driving force of all scientific pursuits. When we begin to focus this notion inward, we encounter what philosophers call the "hard problem" of consciousness.

The hard problem of consciousness, a term originated by philosopher David Chalmers in 1995, can be summarized by the question: "Why is there a subjective feeling attached to groups of sensory information whatsoever?" Essentially it is pointing at the concern for why our notion of self is present. The hard problem can be distinguished from "easy" problems of the mind such as how we are able to control behavior and in what ways do we receive a stimulus, incorporate it, and respond. These are clearly not simple matters, but it is useful to set them against the hulking presence of experience and the hard problem.[1]

Perceptibly, we alone have a privileged access to our internal realities. For many of us, that is as far as our understanding goes. The subjective nature of our minds gives the appearance that we are forever unable to experience what it would be like to be another person, never mind other creatures, such as a bats, with their incredible sonar

[1] Chalmers, David. "Facing Up to the Problem of Consciousness." *Journal of Consciousness Studies*, 1995, 200-219.

capabilities. The advent of technology gives us a novel ability to reach past the veil of ignorance and see objective reality for what it truly is. In this light, the world takes shape in a way which it never has before, leaving even the most intricate and complex entity in the known universe, namely our brains, unraveled and lain bare. In this paper I will present and defend the thesis that artificial intelligence is the microscope with which to view and comprehend the inner working of the mind.

First, it is necessary to frame our problem, to find consciousness in the sea of epistemological difficulties. Once this notion has been thoroughly unpacked, I will demonstrate what science has revealed about the brain and how these advanced perceptions can serve us in unmasking the mind. Next, it will be imperative to trace the history of artificial intelligence and determine where, if at all, a useful connection can be made between the biological intelligence of our minds and the non-biological version striven toward in our machines. Then I will throw the fundamental obligations humans encounter in order to claim mindedness at our supposedly intelligent machinery and see if they have any answers for our most isolated and confidential probings. Finally, I will contend that, through the process of gradually offloading the functionality of our brain onto technology, the mind-body problems which are so very stubborn in philosophy finally will be given a new perspective and perhaps even a few definitive answers.

Technology has consistently set the stage for new lifestyles and greater heights for humanity to achieve. The consequences of our tumultuous time is that the way we know the world today is vastly different than the way people did even a decade ago. The extent of this epistemological shaping rests solely on the shoulders of the technology we have invented. Functions of the brain can be substantiated and duplicated in the external world

in an attempt to see what is genuinely taking place. The implications that this technology has for neuroscience and our understanding of cognitive functions as a whole are currently unfathomable. By offloading and recreating essential brain processes onto artificial intelligence, we will examine what makes us human in an entirely new way.

## II. FINDING CONSCIOUSNESS

As the march of scientific undertaking continues to unveil and expose the reality of the world around us, one bastion sticks out of which we have little comprehension. Consciousness has been a thorn in the philosopher's side since the first sentient being feigned to ask why things were they way they were. The putative sovereignty of consciousness as divorced from the causal nature of physical laws and descriptions continues to poke at and irritate many great minds. At once so intimately familiar and yet, simultaneously, so very foreign, the most common approach to thinking about consciousness is simply not to. By embracing the situation and enjoying the practicality of its function one can blissfully embrace ignorance. For some people, however, this is simply not an appealing solution. A take on Socrates' claim that the unexamined life is not worth living perhaps it can be said that the unexamined mind is not worth minding. Many will take at least a few brief moments once in a while to think about why things are as such. Aforesaid cursory reflection is not an uncommon pastime for the average person. When we begin to delve deeper and unravel the spool of thought, however, the problem soon becomes rather daunting. It is in the face of the many existential dilemmas instigated by inward reflection that the most courageous thinkers take up their sword. Resting on their shoulders, we are able to fathom one of the most complicated, complex, and perplexing things in the universe: our own brains.

Humans naturally want to have answers to problems. It is not often that one will find someone completely content with not knowing an answer to a question presented to him or her. Indeed this is one of the most crucial aspects of human nature. Questions

about the world thrust themselves upon us as we seek answers to the stimuli we encounter. In the case, however, that an explanation is not readily apparent and cannot promptly be uncovered, humans will conveniently provide a narrative of events that, at the moment, seems to be the most proficient answer at describing the circumstances. Utilizing the bits of knowledge we have, a causal chain is linked together so that we may understand and, with any luck, manipulate our environment. So long as the predictive ability of this narrative is better than any others, it does not matter whether it is based on empirical knowledge or otherwise forced onto the world. Legitimate explanations of causally connected events in the world, which, incredibly, are becoming less and less outside our realm of understanding every day, have always been the primary pursuit of any scientifically inclined mind. Unfortunately, the compulsion for answers often overshadows the facts, and we must at all times be cautious that we are not making the data fit with our particular world view.

There is only one way to abolish an entrenched narrative. A great many incongruous manifestations must be shown to be occurring which are unable to be accounted for with the reigning system of beliefs. It is only then that, often begrudgingly, the scientific community will begin to address the concerns which those inexplicable instances create. Most will try in vain first to assimilate these occurrences into the accepted doctrine, striving to salvage the vestiges of their narrative. Once there is shown to be an overwhelming amount of incompatible occurrences, a Kuhnian paradigm shift will occur and throw everything which the scientific community previously subscribed to

up into the air.• Examples of this process can be seen in the tumultuous transitions from the Ptolemaic, geocentric model of the universe to a Copernican understanding in the 1540s and that of phlogiston theory of combustion succumbing to Lavoisier's theory of chemical reactions in 1783. [1]

Scientific revolutions, while often riotous events, make all of humanity their beneficiaries. It is essential to try to remove the outdated scruples as one would a band-aid, quickly and as pain-free as possible. The point here is that in every case where this has occurred, it is always through the tremendous effort of forward thinking scientific minds running on divergent streams of thought. Humans are astonishingly tenacious when it comes to holding on to their perspectives. Undermining the commonly accepted set of ideologies has a historically high chance of getting someone killed. We are fortunate enough to be in a time and a place in which dissention, at least in most respects, is not entirely met with anything other than perhaps derision. Rewriting all the dense volumes of books which the proponents of a doomed theory have slaved over and all the teachers have preached on from their lecterns, however, is not something that most people find comfort in contemplating.

Nevertheless, it is my contention that the prevailing perspective, knowledge and understanding of the subjective experience of our brain are doomed to go the way of Aristotelian mechanics and spontaneous generation. With the wealth of information which neuroscience has brought to light about the way we think and experience our environment, the scientific community can no longer maintain the same hackneyed

---

• Thomas Kuhn detailed the process of paradigm shifts in his book *The Structure of Scientific Revolutions* in 1962. Essentially, this shift is a major upheaval of basic belief structures in the current prevailing theory of science.

[1] Kuhn, Thomas S. *The Structure of Scientific Revolutions*. 2nd ed. Chicago: University of Chicago Press, 1962. Print.

narrative of our consciousness. A great many thinkers, some of whom I will be discussing in this text, have, in my humble opinion, irrefutably demonstrated that this is so; yet there is still a majority of the populous holding to the antiquated belief system of dualism and the immaterial mind it postulates. When the scientific community shifts paradigms it will inescapably take some time for the other groups in society to come to grips with the implications. Once the anomalies and paradoxes inherent within the current concept of consciousness have been thoroughly unveiled and are consequently stacked against the theory, however, established systems of thought will crumble under the weight. Out of its wreckage, an improved and refined understanding of the mind will emerge.

So what is it about David Chalmers' "hard problem of consciousness" that is so difficult for us to come to terms with? It should be relatively easy to figure out because we are all so intimately connected to our own consciousness, or so it would seem. Surely, it is the one thing that we should be the most sure about, right? It seems improbable that we would have superior understanding and predictive powers of how large bodies move in the solar system and yet still be perplexed by something as intimate and essential to our existence as consciousness. The problem began when we did what humans always tend to do: we dictated a narrative on a set of phenomena in order to provide a causal chain and appease our curious nature.

Humans have been pondering why they are pondering for centuries, but it was not until René Descartes, the father of Modern Philosophy, first postulated his dualism that the contemporary scientific discussion of consciousness really began. It was Descartes in his *Discourse on the Method,* first published in 1637, that started us on a path to understanding why and how we think the way we do. In large part, all of the great

thinkers on the subject that came after him have had to respond to his work one way or another. Unfortunately for Descartes, he was almost entirely wrong in his assessment. Yet it must be acknowledged that we are forever indebted to his work and must be grateful for his instigation of the conversation.

Descartes summed up the impression of the mind that had been conclusively confirmed for centuries in his philosophical theory of dualism. Indeed, many did not see the necessity for even beginning to examine inner reality as they supposed the case was so definitively settled. Descartes' dualism postulates that there are two planes of existence: a physical world in which the body, the world, and all other feasible entities reside, and a metaphysical realm which the mind or soul inhabits. It is from this ethereal domain that the true essence of a person is to be found. From this sovereign, isolated location, Descartes believed that the mind was free of material determination in its nature and could never be constrained or indeed ever understood empirically. Not only was this mysterious mind free, but it was also indestructible, at least through earthly means. To reconcile the connection the mind feels with the body, he postulated that through some gland in the brain, our minds interacted with the body, conceivably by pulled levers and pneumatic piping, and thereby governed one's corporeality. From this seat, the master homunculus, a Latin word meaning "little man," would control and manipulate the world utilizing only the physical manifestations of the body. This mind and body, or *res cogitans* and *res extensa* respectively, existed as completely different and distinct types of entities. Yet it was apparent even then that they had an intimate relationship, albeit one which for Descartes was enigmatic and evidently unknowable.[2]

[2] Descartes, Rene. *A Discourse on the Method of Rightly Conducting One's Reason and Seeking Truth in the Sciences*. Waiheke Island: Floating Press, 2009. Print.

The immaterial mind is still, to this day, a pervasive structure of belief. It must be admitted that to the unreflective person this is indeed how it *feels* to be conscious. When we think about our actions and who we are it does seem that I am nicely situated in a place where I can choose to operate actions from my body and command myself around in any faculty this incorporeal self desires. It appears to me that I am entirely in control of what I want to do. Ostensibly, I have the free will to boss my body around and assert my influence, needs, desires, etc. Lamentably, this is an outdated understanding of what is truly happening when we are thinking, and even thinking about our thinking.

Cartesian dualism began to show weakness under the pressure of cases such as the infamous Phineas Gage in 1848. Gage was an American railroad foreman who, when tampering down blasting powder to remove some rocks, accidently created an explosion that caused a crowbar to go flying through his skull, destroying the left frontal lobe of his brain. Miraculously, he survived the traversing of the large iron rod through his head. The problem was that, while Gage lived physically, the Gage people knew and loved did not survive mentally. In other words, who he was as a person was lost in response to his physiological damage. His behavior was altered as a result of his accident. Friends stated that they perceived his intrinsic nature had irrevocably changed. Once a mild-mannered and efficient worker, Gage became much more vulgar and care-free after his accident. Unable to hold down a job, it is now believed that the loss of his pre-frontal cortex caused him to lose a lot of his previously exceptional executive functionality.

The simple fact that physiological changes in the brain can cause permanent damage to the psychological, mental mind of Gage undermines the distinction which dualism posits. Descartes believed his mind to be separate from and indestructible by the

physical world. If the metaphysical, immaterial mind can be altered by a purely physical, material crowbar, then the separation of the mind and body is an incorrect understanding. It would be one thing to claim, perhaps, that the receptive communicative seat of the mind had stopped functioning and therefore the two worlds could no longer interact. However, this is simply not the case. His mental discourse was still present; it was just altered. It is clear that the very essence of Gage was changed with the destruction of a piece of his brain. It is now known that the pre-frontal lobe is the area of the brain which is most responsible for planning tasks, motivation, and other executive functions of the brain. At the time of his accident, the perception that the mind was shared throughout the brain was unthinkable. The lack of functionality that the pitiable railroad foreman suffered from gave the first clues to start unraveling the brain. [3]

Another anomaly that antiquates Descartes' dualism appears with the notorious hole which Gilbert Ryle demonstrated in his book *The Concept of the Mind*, first published in 1949. In this work, Ryle exposes Descartes as having made a category-error when presuming that the mind necessarily must be a separate entity unto itself. For Ryle, the conscious mind is not a distinct, immaterial thing but is instead the categorical name we use to point to a group of events which all culminate to project what we consider to be our minds. The revolutionary part of this stance is that all of these subsystems that combine to make the mind are exclusively physical processes. The mind is indeed immaterial but only in the fact that it does not exist as such. Instead, Ryle indicates, it is the group name of a list of processes in operation within our brain.

[3] Damasio, Antonio R. *Descartes' Error: Emotion, Reason, and the Human Brain*. New York: Putnam, 1994. Print.

To illuminate Descartes' category-error, Ryle uses the example of a foreigner visiting a university for the first time. The visitor is shown all of the buildings, teachers, programs, libraries and sports arenas during her time there. When her tour has finished, however she turns to her guide and asks why she had not seen the University. The prospective student has confused two distinct categories. By believing that the University was itself an actual place instead of being the group name of all of the iterations of things which when combined make up the "University," she has muddled two kinds of things: the concrete and the abstract. This is what Descartes and other dualists have done when they look at the brain and are amazed to see that the mind is not located there. Their own prejudices and ideologies then skew their scientific understanding when they state that it must, therefore, be immaterial and ethereal. They presume its existence and subscribe to an understanding that incorporates the anomaly of its absence.[4]

The implication of these cases testifies to support the thesis that the mind is a physical entity constituted exclusively in the brain. That realization in and of itself was revolutionary. The idea that one's mind was isolated and disembodied can be said to be dubious at best. It is evident nonetheless, that the old, erroneous perception still holds sway in many groups in the world today. Once it has been determined and generally accepted that the mind is a part of a physiological system, a truly scientific enquiry can begin. Descartes put the mind on a pedestal, which was shrouded by its incorporeality. Thinkers since his writing have worked diligently to dismantle this notion. Who we are became situated in the physical realm thanks to their efforts. People slowly have begun to come around to the fact that the mind can be studied, understood, and perhaps even explained completely. No longer were the great thinkers hesitant to postulate a physical

[4] Ryle, Gilbert. *The Concept of Mind*. University of Chicago Press, 2000. Print.

description of what is happening when someone is thinking. This essential transformation, however, was just the beginning of unearthing the true nature of the mind.

## III. Unmasking the Mind

Once we have situated the mind clearly within the brain, it is possible to begin to approach the hard problem of consciousness scientifically. The immateriality of the mind held this pursuit at bay for much too long. It seems indisputable that we enjoy a privileged access to our own minds, at least for the time being. I know what I am thinking and, if I control my outward behavior well enough, can hide my inner feelings from others. At least this is how it seems to me. But who and what is this me to which it seems? A vicious circle begins when one starts to answer, "Well, of course I am 'me', who else could it be?" The trouble is that this internal me is incredibly mysterious not only to everyone else but also in many respects to myself (or should I say itself?). What constitutes this self and what is the manner of influence I have over it, if indeed I do at all? How did it get here and where is it going? All of these questions serve to demonstrate just how mysterious we truly are. In order to unmask consciousness it is necessary to go to the source and start from the beginning.

There is an idea in evolutionary biology that, when traced all the way to the origin of our species, there is a single female who is the most recent direct ancestor of everyone alive. The name that biologists use for her is "Mitochondrial Eve." Clearly the "Eve" is a biblical allusion used to demonstrate that this one genetic relative is, for all intents and purposes, the one who started our line of descendents. The difficulty in setting this divide is that Eve's mother and her offspring had practically identical genetic structures to her, similar to one's parents and oneself, yet they are not the ones said to usher in an entirely

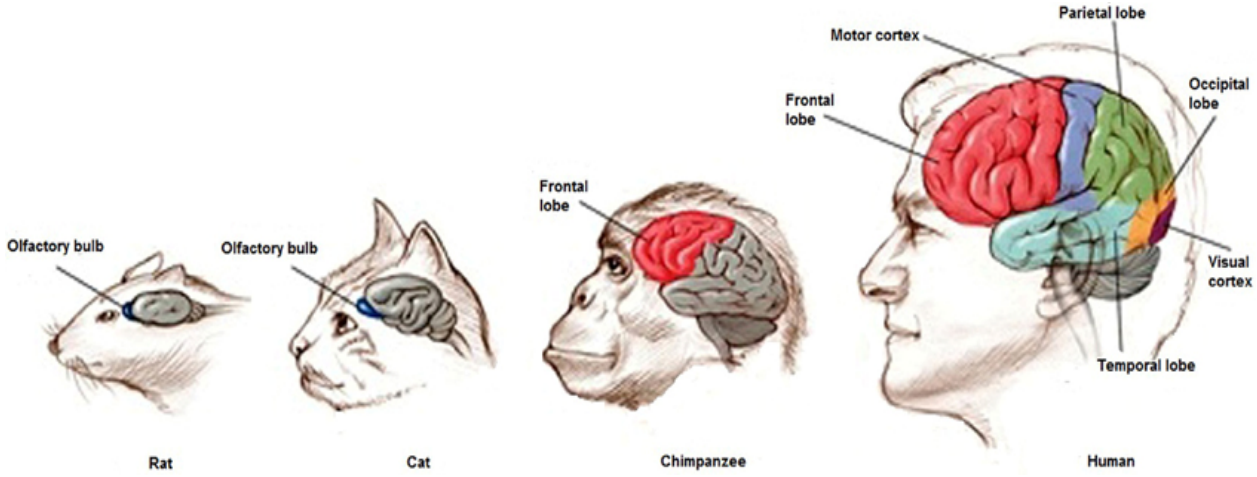
new subset of species. It would appear that this characterization is somewhat arbitrary, a divided line distinction put in only as a part of the reflective nature of history.[1]

The mind has a similarly ambiguous past as "Mitochondrial Eve." If we know that our species today has the ability to be conscious and that our parents, presumably, were sentient beings and their parents had sentience and…etc. At what point does mindedness suddenly appear? Is there a point where a mindless, unconscious being reproduced and *viola* the capacity to consciously think showed up? Or perhaps "thinking" transformed into *real* thinking? Distinctions like these only serve to further obfuscate and undermine our understanding of consciousness. The gradualism in evolution makes it difficult to draw the line which biologists employ with Eve. Instead of a sudden change, gradual reform left a conscious self. This conveys the impression that pieces of mentality, minimalistic bits of mind-like operations, coalesce with more and more competent machinery until one day minds like ours began to enter the world stage. The eukaryotes and bacterium of so many millions of years ago must have had some form of objective, a purpose to which they applied their survival tools. These mindless mechanisms began to incrementally create competences and, inevitably, the comprehension which we enjoy today.

It is evident that Homo sapiens have a unique ability to manipulate the world in response to their needs for survival. Most animals will shape their environment to some extent. Beavers create dams and birds build nests but none could compare to the amazing capability we have at utilizing tools in order to thrive in an environment. We can make tools which help us to make tools; iron sharpening iron so to speak. Without many

[1] Dennett, D. C. *Intuition Pumps and Other Tools for Thinking*. New York: W. W. Norton and Company, 2014. Print.

organic weapons, our ancestors soon found that objects could be used to compensate for the apparent disadvantage. The reason that we were able to dominate in this respect is thanks, in large part, to our gigantic cerebral cortex. Taking up a significant percent of the adult human brain, its purposes include language, memory, reasoning, perception and more.[2] Not only did our brain evolve to be larger in size but it also developed a series of folds and grooves, called gyri and sulci, which were then able to contain even more neural connections while taking up significantly less room. A dramatic increase in the size and efficiency of our cortical structures set our species on a path that would one day give it the thought of being able to artificial replicate itself.




[3] Relative neocortical sizes

The evolutionary advantage of an immense cognitive processor gave us the ability to improve communication, thinking, and thriving in revolutionary new ways. Enough

[2] Kurzweil, Ray. *How to Create a Mind: The Secret of Human Thought Revealed*. New York: Viking, 2012. Print.
[3] Leisman, Gerry, Calixto Machado, Robert Melillo, and Raed Mualem. "Intentionality and 'Free-will' from a Neurodevelopmental Perspective."*Frontiers in Integrative Neuroscience*," 2012.

tiny mechanical entities had consolidated so that we could manipulate objects and symbols, giving us an edge over other species. Our gigantic frontal lobes made connections in the world and solved for the problems plaguing our species' survival. Our mammalian ancestors became skilled at predicting the states which other entities were in. They could, for instance, look at the visage of another creature and gauge their intent by realizing that the manifestations of the internal processes were manifested externally as evidence. Surmising the internal intentionality of a creature based purely on its external indications requires taking what philosopher Daniel Dennett calls the intentional stance.[4] I cannot know what, if any, sort of mental life any other being has but it behooves me to infer that it is as much like mine as possible. When I am hungry I perform this set of actions. That thing is performing an equivalent set of actions. Therefore it is hungry. Undoubtedly, this capacity played a key role in survival. Our neocortex allowed for an incredible amount of hierarchical structuring and learning, millions of neural connections could be made which allow for processes and understanding never before imagined.

Owing to this anatomical, physiological advantage, we developed a place in nature where we could utterly dominate those organisms and entities around us that had significantly less neural possibilities. Some would say that this incredible brain structure is evolution's finest creation. That perspective, however, creates an image of evolution as having some direct goal which it is intending to achieve, as if this form of survival were somehow programmed for or otherwise planned. I believe this is an incorrect understanding of how evolution works. Instead it appears that the occurrences leading up to our brain's development simply gave our ancestors the best possible chances of survival in the environment that they were in at that specific moment. Had some other

[4] Dennett, *Intuition Pumps*, 2014.

evolutionary tool been more successful at the time, then the biological world as we know it would simply not exist. We are indebted to these mind-full relatives and their ingenuity. Darwin's theory of evolution gives us perspective on how we have the brain we do and yet the great mystery, the hard problem, still remains. Our robust brain structure is a product of natural selection but that does not tell us how a mind, an experience-haver, arises from physical, mechanical, and altogether natural events.[5]

It helps first to reflect on how well we know ourselves. For instance, I know that my brain is composed of bundles of neurons and electrical signals and yet it seems like there ought to be more to this story in order for the "I" that I know and love to persist. An approach often used by philosophers in the evaluation of this situation is to create a hypothetical world where there are people like myself (and presumably you, the reader) on one side and on the other, there are zombies.• These philosophical zombies are not the same as those man-eating zombies in Hollywood films. Instead, they are completely harmless, well, at least to the extent that the average person is harmless, for these zombies are practically identical in every aspect to you and me. There is, however, one notable exception: the zombies have no internal, mental conscious life whatsoever. They are functionally indistinguishable from the average person; they talk, walk, and laugh like the average person, all the while completely lacking any internal mental processes. The train is running, but the engineer is asleep at the helm. The question becomes: could you tell the zombie from the conscious human? The point that this hypothetical thought experiment is trying to make is that if the zombies functioned externally the same way as

[5] Dawkins, Richard. *The Blind Watchmaker*. New York: Norton, 1986. Print.

• Robert Krick introduced the term Zombie into the philosophical realm 1974 in a paper entitled "Sentience and Behaviour." Since this time it has become a common thought-experiment for many philosophers of the mind.

the normal minded person, then we would not be able to distinguish between the zombies and the people like us. I cannot be aware of your cognitive phenomena, but instead I must infer your mindedness from its external manifestations by taking an intentional stance. I presume your consciousness by its exterior materializations through the intentional stance.

On the outside it would appear that a mental life is inconsequential so long as the functionality is the same. The trouble is that the zombies would not have access to an internal life in the same way we do. They would, perhaps, be able to mimic the reactions we have toward stimuli such as pain, sadness, etc. but only in the outward manifestations and would never be able to really know or feel what it were like. Their movements would be entirely learned behavior and not truly experienced. It is simple to ape the motions to go through when you stub your toe, for instance. A zombie would only need to watch the act play out a few times to have the reception and feedback response convincingly portrayed. When this sort of stimuli is received, namely a strong electrical signal in a certain region of the brain, the best action to take is to wince and grab your toe. The only difference is that there is not really someone there to experience the sensation. It is vital, therefore, that we find some difference between the zombies and ourselves that can distinguish a mind-haver from the mindless.

Another way to frame this problem is outlined by Richard Rorty in his book *The Mirror of Nature*, published in 1979. He constructed a planet in another galaxy on which there lived people identical to us. The only difference is that "these beings did not know that they had minds; [t]hey had notions like 'wanting to' and 'believing that'… but they had no notion that these signified mental states -states of a peculiar and distinct sort-

quite different from ‘sitting down’ or ‘having a cold’.” Rorty called these beings Antipodeans and demonstrated that they did not refer to experience by any sort of mental talk, such as anger or sadness, but instead spoke in terms of neurology and biochemistry. Instead of saying that they felt pain, for example, they might instead say that their group c nerve fibers were transmitting strong electrical signals. Happiness is spoken of as releases of dopamine instead of a state that the self is in. The philosophers encountering the Antipodeans divide into camps, of course, about whether there is actually a mind experiencing sensations; does a perfect physical description of these events conclude that there is a self feeling or must it eliminate the self altogether.[6]

The Antipodeans, just like the zombies, are completely lacking mental discourse *as we know it*. The caveat of “as we know it” will prove to be decisive here. It is pronounced that these entities have an accurate understanding of events; they can describe what is happening to them, how it is happening and in what manner it is presented. The trouble is that there seems to be no them to which these things can happen or appear. Perhaps, it must be argued, that there is something we are missing. In Cartesian dualism there is the inherent presupposition of a soul, an immaterial self. Descartes, and many others, subscribed to a certain belief structure and forced reality to conform to their points of view. When a mind was not evident in the brain it was no matter because the mind is, in fact, an immaterial unknowable thing but, rest assured, we had it. They were befuddled by their presumptions, confounded by the collision of evidence and ideology. Similarly, we ought to try to disgorge ourselves of all assumptions in order to discover exactly what may separate us from our functionally

[6] Rorty, Richard. *Philosophy and The Mirror of Nature*. Princeton: Princeton University Press, 1979. Print.

identical carbon copies. Zombies and Antipodeans allow us the chance to discern a potential distinction between the minded and the mindless.

Scientists are not born with the knowledge to enquire about the world in the elegant ways that they do. The logical coherence of thought of which society praises scientists comes only after much formulation and turmoil. They are, instead, born with a certain set of instincts, a genetically encoded set of rules which prevail to help them survive as infants. Most people would agree that this is not consciousness. These formal systems are not decided or chosen by the infant but are automated to occur in response to specific stimuli. Lowered blood sugar leads the toddler to cry out for food. Children will instinctively do a Palmar grasp in response to objects being placed in their hands. These techniques for survival are not consciousness as we typically perceive and experience it.

An infant's mental life, therefore, can be likened to the zombie's in the fact that, for all intents and purposes, it could be entirely absent. From these ingrained sets of skills, we soon mature and form new and more complex understandings of the world. We go through Jean Piaget's stages of development, all the while forming more and more accurate perceptions of the world around us. Our brains allow us the unique opportunity to build hierarchical representations of the world. At first, we believe that only what is in front of us exists and soon begin to understand object permanence, which eventually culminates in symbolic thinking and manipulation.

At first, we are a blank slate running the instinctual moves naturally engrained and utilizing the interpretation mechanisms of our bodies to come to better terms with the world; our hands grapple with objects and our fledgling eyes strain to make sense of the blur of colors. During this explorative time our brain is forming brand new synaptic

connections between groups of neurons. Patterns of connections are created which funnel in electrical signals in response to stimuli, sending them down predetermined dendrite paths when familiar with the signals and creating new connections when faced with unfamiliar incitation. Relatively quickly, children assimilate a multitude of experiences under their belts and can make predictions about outcomes based on past experience. Neural pathways are laid out in order that reactions to specific stimuli produce learned behavior.

Imagine, for instance, a young girl who has perpetually connected eating her vegetables with her father allowing her to have a cookie. “Eat your vegetables and you will get a cookie,” he has repeated many times. Her dad has done this enough that the result can accurately be predicted with statistically significant numbers. The young girl’s brain is aware that the cookie is a means to a highly desirable chemical release. She does not comprehend why this causal chain occurs the way an Antipodean might but only that it is as such.

The young girl does not know, or care, that when the exact stimuli of vegetables collide into her receptive tools, they stimulate a certain set of neural pathways that, in accordance with their pre-established pathways, begin firing. The electrical signals travel down these patterned trails and culminate in moving muscles to perform certain tasks in order to create the now reinforced behaviors and reach the predicted goal. The odorants and wavelengths of broccoli are present in the sensory machines, the nose and eyes respectively, which act according to their biologically mechanized duty and send electrical signals to specific areas of the brain. The brain receives this and subsequently moves the hand to the fork and the food into the mouth so that the sweet dopamine will

be released in response to a delicious chocolate chip cookie treat. The brain predicts what will happen by relying on connections which are formed in the cerebral cortex by experiences. If this time her father were to withhold the cookie for some reason then the pattern would be met with failure to achieve its goal. Accordingly, another set of neuronal pathways fire in response to the incorrectly predicted outcome. For most young children the predetermined response to that is to scream.

It is clear that our experiences teach us how to behave in response to certain stimuli. We know, for instance, what it takes to get other people motivated. The behavior that they exhibit is directly linked to their mental processes. To induce the young girl to stop crying, the father knows that either he teaches her what sort of responses will not work through the reinforcement of a copious amount of counter-instances or he capitulates to her efforts and gives over the cookie. Either way, some form of behavior will be reinforced and thereby further entrenched as the right course of action. Learning as youth is not a choice, imitation is the hallmark of species survival. The children are little scientists that look, listen and experiment on the world. Their experimenting comes from emulating the sounds and actions of the people around them but they do not choose to do this. We are ingrained and trained to respond to the environment in specific ways responding to our experiences. Infants are not born with linguistic knowledge or the ability to perform complex tasks. They mirror the efforts of those around them and gradually refine their actions in an effort to produce efficiency. It is strange to see, however that this is the same way that our hypothetical zombies could possibly learn to perform and interact with the world.

In order to free the prevailing intuition of the mind from its entanglement with zombies and Antipodeans we must find where the self becomes a force, when the mind takes control and forces its conscious hand. When does the instinctive imitativeness of our youth become self-determined performance? At what age am I said to be free from the deterministic nature of my past experiences which conditioned and guided me? At what moment do I take the step from being Pavlov's dog to being Pavlov?[7] We do not decide or choose our DNA. Neither do we have the option of which environment to be born into. Furthermore, we cannot dictate the sets of experiences which shape and mold our cognitive structures. It soon becomes apparent that "a living organism at any moment in its life is the unique consequence of a developmental history that results from the interaction of and determination by internal and external forces."[8] Surely, one would be forced to ask, there is something that is missing from the equation? This physical description of who we are is entirely devoid of conscious life and thereby makes us the zombies of the thought experiment.

There is no room in this process for consciousness as we perceive it. Yet I am stuck between the indisputable fact of my having a mental discourse and my inability to locate it in a physical system. Our consciousness is not the determinate factor in our actions but is, rather, a character witness to them. Most would claim that it indeed *feels as if* my actions are the result of my conscious choice but I would caution that appealing from one's experience of the world is a notoriously inaccurate way which to argue. This anecdotal fallacy does not regularly hold much sway in the scientific community. Our perceptions necessitate and confirm a mind but it is only your mind telling you that, and

[7] Ryle, *The Concept of Mind,* 1949.
[8] Lewontin, Richard C. *Biology as Ideology: The Doctrine of DNA*. New York, NY: Harper Perennial, 1992. Print.

it is a notorious liar. We perceive, for example, that the world is full of colors, smells and mass and yet objective science has continually reinforced the idea that the world really consists of wavelengths, colorless odorless chemicals and that atoms are almost entirely comprised of empty space. The emergent manifest image is a phenomenon that results from the efforts of our biological tools. What then is the conscious self to which I cling so tenaciously? Once the list of incompatible anomalies is fully laid out against the prevailing judgment of the mind, only then can we overthrow it and demystify it entirely.

The commonplace perception of the self, when laid out thusly, begins to fall apart. The classical conceptualization of consciousness implies that the environment acts on the body, then some mysterious, immaterial actions occur in our minds, and finally a chosen set of behaviors is produced. Input occurs; some mysterious X of the mind happens; output results. The dilemma commences when pressured to decide whether or not there is a causal, law-like connection in this chain of events. Answering with yes or no, it seems, is incompatible with a mind as we currently have it. If it is a causally determined chain, therefore answering yes, then one does not need the concept of the mind whatsoever. A yes to the causal chain indicates that an input on a person will always lead to a certain set of outputs; thus the inner life, that mystical X, is superfluous to understanding. It becomes an imagined nicety which is inconsequential and powerless in its completely determined position. This is called the theoretician's dilemma of which B.F. Skinner and other behaviorists were proponents. Conversely, if the path previously set out is not a determined, causally linked system, therefore answering no, then there is no way of predicting or anticipating what could come out the other end. There would be no explanatory power in the study of human behavior. There is, through this perspective, no

point in talking about what a person does or normally might do because anything is possible in reaction to a stimulus. In fact, discussion of mental events would be worthless altogether. Essentially, we are left with either eliminating the notion of mental life altogether or completely starting from scratch with our understanding. This is the power of paradigm shifts.[9]

Naturally, the first response to this is to reassert that one does, in fact, have privileged access to one's own mind and it does not matter what sort of evidence is brought against it there is nothing one can know more truthfully and intimately than one's own mind. Our minds reassure us confidently that this is so. However, the perception that we have the ability to garner insight into the reasons or purposes of our actions is undermined at every turn. The knowledge of one's self, what Daniel Dennett coined as auto-phenomenology, is actually one of the worst ways to determine how we are feeling or what we are thinking.[10]

Conversely, hetero-phenomenological descriptions of a person tend to be relatively systematic and accurate. Auto-phenomenological descriptions are difficult typically because they dictate that a person's understanding of the causes of their volition must necessarily be infallible. One could not determine whether another person was having the experiences they claimed regardless of any physical manifestations. If someone were to consistently insist that they were happy while acting as if the opposite were true, our course of action, however, would be to follow the physical manifestations of internal life to circumvent the auto-phenomenological narrative. Furthermore, these internal states are always determined by an outside cause, manifesting from emotional

[9] Flanagan, Owen J. *The Science of the Mind*. Cambridge, Mass.: MIT, 1984.
[10] Dennett, *Intuition Pumps*, 2014.

responses to our environments. The body is unconsciously performing and processing a great many functions all at once. We are forced into certain states and then asked to give an account of how we feel only in the aftermath.[11]

Many psychological theories are substantially based on the premise that the objective observer can know the reason for a person's actions more accurately than that person can. The brain will lie to itself and confabulate reasons to justify actions whose origins it does not know. When we act unconsciously, it is only after the fact that we are suddenly knowledgeable of the impetus. Ironically, we are sometimes the last ones to know the causes of our actions. Studies of bystander effect, for instance, show that as the number of people around someone in crisis goes up, the likelihood that anyone will help exponentially decreases. Yet, when asked afterwards whether or not the number of people had any effect on their decisions, people will unanimously say that it did not. Furthermore, people will often succumb to social pressures in a group in order not to stand out, and yet again when pressed for reasons for their actions will grasp at any reason other than the truth.

Further undermining our antiquated notions of the mind, neuroscientists investigating to what extent our brain will confabulate and provide a narrative of events whose provenance is unknown have devised several clever experiments in order to test the brain. The most interesting cases involve patients with split-brains. A split-brain is a unique case in which a person's *corpus callosum,* the connective structure between the two hemispheres of the brain, is damaged or severed completely. In these experiments, the researchers will show the left eye, which is connected to the right hemisphere of the brain, a certain image and ask the person questions about why she had certain reactions to

[11] Flanagan, *The Science of the Mind,* 1984.

the stimuli in the picture. Generally the picture is stimulating enough to cause a visceral reaction, such as a funny face to cause the patient to laugh. Then, when the left hemisphere, where the language center of the brain is mostly stored, is asked why the reaction happened, it will confabulate and create a reason for the outburst, something like "I just noticed your shoes and they are funny." The left hemisphere was not privy to the stimulus which was had solely in the right side and, when pressed for reasons, fabricated an answer to explain its behavior. The self of the person is completely convinced by the brain's deceit and resolutely subscribes to the reason presented for their reaction.[12] It is evident, though, that she has been tricked. The brain, and therefore the person, does not know the true reason for its actions and so quietly created one to provide the appearance of control. Patients will adamantly state that their physical reaction was due to their own imposed narrative despite all evidence to the contrary.

This is the process which is occurring when we place some mystical driving force on our body. Anomalies continue to subvert our ideas of the mind and demonstrate that the phenomenon of consciousness is instead an emergent image, a witness to events as opposed to their author. It is an effort on the part of our brain to create reasons for actions which are entirely outside of its own control, determined rather by physical processes. It was Benjamin Libet's experiment in the 1970s which, for most, unmasked the mind and revealed it for what it really was. Libet attached EEG electrodes to participants that were in front of a timer. The participants were then instructed to do simple tasks such as move an arm or push a button and to record the timer at the first instant they became aware of an urge or desire to start the motion. What he discovered was that brain activity commenced roughly 300-500 milliseconds before the participant consciously registered

[12] Cushman, Fiery. "What Scientific Concept Would Improve Everybody's Toolkit?" *Edge*.

having initiated the act.[13] This experiment has been repeated several times since with the same results. It can be concluded therefore that our motor cortex is already firing off electrical signals and in motion about a third of a second before we register the will to do something. It is clear that there is a lot of activity occurring in our brain prior to our "conscious decision" to do the action. [14]

In the face of these experiments it becomes rather difficult to maintain the relationship with our consciousness which is still so pervasive in many aspects of society. We like to imagine that we are the captain of our ship, steering it whichever way we so desire. It appears however that we are the ship which has been set in motion through the biological necessity of our genetic structure and the random chance nature of experiences which it encounters. We do not know our destination, indeed it is practically unpredictable, and yet its destination is already determined, out there on an island so many years away. Experiences enter our formal physical systems and push our ship around like strong winds. People crash into and divert our voyage like giant waves. The fact that it is impossible to know what the end will be does not mean that there is anything we can do to change it. Thoughts and intentions do not originate in some supernatural place; rather they simply appear in our consciousness while the traditional understanding of the self bears witness. We cannot have any impulse to stop or change our thoughts other than those which were already going to be there. Our behavior is the result of the physical cause-and-effect relationship which our brain has with the world. Our mind is the result of a formal, deterministic system which operates in accordance to the laws of nature.

---

[13] Harris, Sam. *Free Will*. New York: Free Press, 2012. Print.
[14] Kurzweil, *How to Create a Mind*, 2012.

The great German philosopher, Arthur Schopenhauer, once wrote that "everyone believes himself *a priori* to be perfectly free, even in his individual actions, and thinks that at every moment he can commence another manner of life…but *a posteriori* , through experience, he finds to his astonishment that he is not free but subjected to necessity, that in spite of all his resolutions and reflections he does not change his conduct, and that from the beginning of his life to the end of it, he must carry out the very character which he himself condemns."[15] This perspective does not mean that our semblance of choice, our intentions and reasoning are unimportant. The narrative which the brain provides is itself a part of the causal chain of events. The manifestation of a conscious mind is as useful as our manifest image of reality. It behooves me in my daily practical life to imagine colors out in the world and also to create an autobiography of my events. It is important to understand, however, that this form of a mind is itself functionally dictated. I cannot choose what I choose or exhibit control over my actions as classically conceived but I, or more precisely my brain, can use this information to make better choices and utilize the tools of my biological system. Experiences have trained me to react to information in a certain way. Revealing the truth of my biologically determined nature to my own emergent illusion of a mind could allow it to reshape its actions in a more scientifically informed way. Whether or not it will do that, however, is not my conscious decision to make.

Reevaluating ourselves with the understanding that "unconscious neural events determine our thoughts and actions—and are themselves determined by prior causes of which we are subjectively unaware" is an important concept with which to come to

[15] Schopenhauer, Arthur, and T. Bailey Saunders. *The Wisdom of Life, and Other Essays*. New York & London: M.W. Dunne, 1901.

terms.[16] This re-conceptualization is bound to have resonating implications in our society. When we begin to shrug of the dogma of a ghost haunting our body and instead understand ourselves as products of what we see and do, a great many things can be reevaluated. Everything from how we interact with others, to motivation tools for ourselves to intrinsic societal structures must necessarily be reexamined and reassessed. Instead of punishment for a crime, for example, perhaps with rehabilitation a criminal could, in certain cases, become the best option for people who are merely slaves of their own biology, constantly haunted by a devil in their bloodstream.

Let's imagine that there is a murderer who is brought on trial and found guilty. She admits the act but claims that her past experiences determined that she would one day react to stimuli in this way and therefore she had no control. It is clear that this person is guilty of the action regardless of whether or not she could have controlled it. She simply is the type that would murder in the given situation. What if, however, it was discovered that there is a tumor pressing on her frontal cortex. Prior to her having the tumor, she showed no signs of being a danger to anybody, but afterwards her entire personality changed. Is she still responsible? Her biology created the tumor. Her biology created her mind. Yet, for some reason, we want to prescribe less culpability to this case than the first. The tumor seems to be acting upon the accepted notion of the unrestrained self, the self that had free action before the tumor's imposition. Brains of sociopaths perform differently than the average person and could perhaps, with the help of technological innovation, have their physiology changed to behave in a less dangerous way. If, for instance, a tiny machine were placed in the head of a sociopath, which would release dopamine every time the stimuli of a smile was received instead of when it sees

[16] Harris, *Free Will*, 2012.

terror, the entirety of who that person is, or was going to be, could be replaced and enhanced. The reevaluation of operational ability of the brain necessitates conversations such as this.

Analyzing the mind as a set of physical systems means that it can be fully understood by natural law. While it is still too complex for us to fully predict, a new brain scheme will serve as an upheaval to our lives. When you have an argument with your friend it could be because you are genuinely upset or because you are in a bad mood due to being hungry and having low blood sugar or perhaps a combination of the two. This perspective change certainly reveals consciousness and free will as misconceptions. Instead, it replaces the incompatible perception of the incorporeal mind with a deterministic, biochemical notion of self. It is a difficult thing to initially embrace. Simply having this mindset, however, gives us a better understanding of what is happening and can therefore allow us to "steer a more intelligent course through our lives (while knowing, of course, that we are ultimately being steered)." [17]

This scientific revolution will not come easily. As I previously stated, humans are notorious for holding tenaciously to a narrative that they provide, or more accurately have been provided. Thought experiments, such as the zombies and Antipodeans, challenged us to find distinctions between the purely physical and the purely mental. It is only when we came up empty that we were forced to accept the inevitable, there is no distinction. In this way we can once and for all expel the ghost which has been haunting our machine for so long. The difficulty was intensified due to the fact that I certainly do *feel* as if I am indeed in charge. I am assured by my own brain that I am not the zombie, and yet at every turn it appears that there is no functional division to be made. The manifest image

[17] Harris, *Free Will,* 2012.

of the mind will continue to hold sway and is, as evidenced by our continued existence, a somewhat practical, albeit incorrect, way to view the world. It is crucial that the physical nature of consciousness is unmasked in this way and incorporeal dualism completely expunged in order that a scientific approach to the brain can really get underway. With our mindedness now fully in the camp of the Antipodeans, we can begin descriptions of "mental events" which have as their structure completely physical processes. This reconstitution throws the door wide open for artificial intelligence.

## IV. Illuminating Intelligence

Artificial intelligence has been affectionately corralled as a partner in the struggle to understand the nature of the mind. Before one can reasonably begin to assess the abilities of thinking machines it is necessary to throw off the outmoded notion of the mind and exchange it for a shiny, new perspective. In order that we do not allow our ego-centric perspective of intelligence to bias or taint our work on artificial intelligence, it need be accepted that consciousness is resultant from physical processes. One of the most important tools for rethinking the relationship between the brain and the emergent mind, and thereby addressing the hard problem of consciousness, can be found in computational modules. As technology advances, the opportunity to utilize machines becomes more and more pervasive. If the superior size of our cortex is evolution's finest product then it must be said that the advent of artificial intelligences is the finest product of that product. We are beginning to see mechanized robotics take a pivotal role in society. Our communication and movement in the world only has the proficiency it does thanks to their capabilities. This will undoubtedly have profound implications for how we understand and interact with the world. Even today, we can see that the strides technology has made have irrevocably changed many aspects of our lives. This upheaval is prolific in the transformation it offers. It seems that we are only on the cusp of this development.

The originally intended purpose of artificial intelligence was to engineer machines capable of recreating procedures which would typically require humans in order to complete. Performing equations on paper, or perhaps in the brain, has been replaced with

the calculator, as an illustration. This sentiment accurately sums up the position that proponents of weak psychological artificial intelligence would like to take it. Their stance is that computers are tools to understand the mind. Champions of strong psychological artificial intelligence, on the other hand, take this one step farther and posit that not only can reproduction of human functions help us in gaining perspective but that this advancement will inevitably reveal our own minds as computers, and our computers as minds.[1] I hope to demonstrate that it is only through the unification of the tools in weak artificial intelligence with the brain functions offloaded onto them that strong artificial intelligence will ever be accepted or even entertained by a majority of people. We will be able to utilize machines as a tool to evolve to the next level of humanity, a unity of man and machine. In order to do this, though, scientists first needed to begin to discern the functionality of the human brain and its copious intelligences.

The abilities which the term intelligence conveys are multifaceted. It's unmistakable that intelligence, when correctly applied, confers the aptitude of striving for a goal while learning and reasoning to some, even minimal, extent. By definition, all forms of life have some type of intelligence. Dogs, when trained properly, know to signal to their owners when they need to go outside and ants have a natural hierarchical structure to their colony. Bacteria seek a source of energy and plants lean toward the sun. Machines also have some form of intelligence. A drink machine, for instance, knows to get my soda when I put in a dollar and press a specific button. It follows its programming and completes its task. These are very minimal levels of intelligence. This does not imply any form of mindedness or internal mental structure, only that it has the capacity to handle the tasks presented to it. This, in a fundamental way, is how we assess intelligence

[1] Flanagan, *The Science of the Mind*, 1984.

in each other as well. When I witness others presented with tasks, I engage their levels of intelligence by how they compare with my own or some other standard of ability. If there is ever going to be a truly human level form of intelligence then we will need to get much more complex than a coke machine.

The still fledgling process of inventing artificial intelligence commenced thanks to a computer scientist named Alan Turing. In fact, Turing is considered by most to be the father of artificial intelligence. He was the man who began the discussion of intelligent machines, and dared to dream of them as someday mimicking human levels. It is his test for intelligence, aptly named the Turing test, which still serves as the decisive factor in whether or not we attribute human level intelligence to a machine. What Turing proposed in 1950 with his seminal paper, *Computing Machinery and Intelligence*, was a test which was created to solve the dilemma of thinking machinery. He hypothesized that a machine which could perform functionally identical tasks in the way we did would necessarily have to be assigned human intelligence. The machine had to be indistinguishable from a human and if it was unable to be differentiated from that of its counterpart then there could be no decisive division. Turing's revolutionarily empirical approach to intelligence allowed for a definitive answer. He drew a line in the sand. If a critic were to try and discredit his attempt then they would be compelled to explain what his test did not cover and thereby provide an alternative. [2]

The great theoretician went on to propose a hypothetical, yet mathematically possible, machine and how it would pass his test. The Turing machine serves as a theoretical ancestor of modern artificial intelligences, a "Mitochondrial Eve" of sorts. Essentially, the hypothetical machine is given a potentially infinite strip of paper with 0's

[2] Turing, A. M. "I.—Computing Machinery And Intelligence." *Mind*, 1950. 433-60.

and 1's on it. The machine scans the paper and, upon encountering either a 0 or a 1 responds in accordance with its set of rules, which have been preprogrammed onto it. The executive unit then goes on to interact with the 0's and 1's by either erasing them or replacing them or some other predetermined action. Turing himself proved that for every two specialized Turing machines there exists one which can do the work of the two. This demonstrated that there could be a universal Turing machine which is capable of being programmed to do the work of all other specialized iterations.[3]

The Turing machine's impact as a thinking tool started a scientific understanding of intelligence which continues to resonate today. It is clear though that this sort of intelligence, while useful in our world, does not match the way in which human intelligence operates. The Turing machine adheres explicitly to its set of rules. As we know, however, humans evolve, change, adapt and learn in response to experiences. There was simply no possibility of adaptability in the face of experience in his machine. We change our values and aims in response to new stimuli while the best that Turing machines can do is optimize predetermined performances. Yet, while his machine could not readily adapt, his work has stood the test of time and is used today as a hallmark for levels of intelligence.

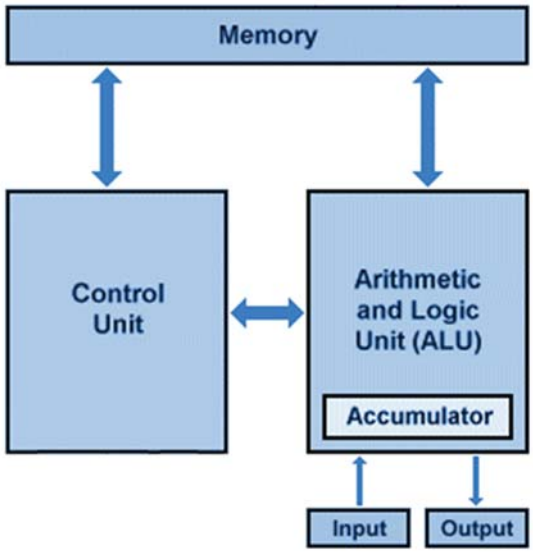
When we traditionally think of artificial intelligence, most of us think of a communicative system that produces human-like actions and feedback. This is the artificial intelligence found in movies such as *Terminator* and *I, Robot.* In order to truly achieve our level of intelligence, machines will need to engage the world the same way we do. Representational states and symbolic manipulation are the certification of human functionality. Letters and other characters on a page can be assembled to produce

[3] Flanagan, *The Science of the Mind*, 1984.

meaning and understanding. If technology were ever going to make the claim at an artificial version of our intelligence which strong psychological exponents suggest, it would have to be able to interpret the world this way. Our language, and therefore our thought process, is emblematic. We manipulate these representational symbols in order to form perceptions and to garner understanding. The symbols which a computer would be manipulating are meaningless to it except in relation to other symbols. The context and essence of the symbols is given to the system by us. This fact mirrors the way in which words are taught to us and only contain meaning when situated in a network of other words and meanings, all serving to reinforce one another. There isn't any reason that if a machine starts out with good programming and sensible input that its conclusion will not be a valid form of associative thought. Correct interpretation throughout will necessitate sensible conclusions in the end.[4]

A computer utilizing symbolic manipulation operates on von Neumann systems. Von Neumann systems have the capacity to receive an input of symbols, check them against the instructions which are stored in a memory unit and produce the corresponding symbols on the other end. Like the Turing machine, von Neumann systems can compute abstract symbols. Plus it can be much more complex than the binary readability that Turing proposed. Due to its capacity to retain a wealth of instructions in a memory unit, the machine can essentially manipulate any number of inputs while continuing to produce the desired output.

[4] Flanagan, *The Science of the Mind*, 1984.




[5]Illustration of a Von Neumann Machine

This sort of conceptualization seems to be much closer to our own faculties. We appear to encounter stimuli, check them against an internal memory in our brain and produce the output which would consistently be produced if the sum total of the stimuli in the environment were perfectly reproduced at another point. The trouble that is encountered, however, is that, while this may be a functionally correct recreation of our intelligence at a specific instance in time, it is not the actual way we operate. Our brains are the most complex things in the universe and in order to fully understand their nature it is not good enough to merely produce the same results. To fully comprehend how the mind emerges and one day recreate its appearance, we will need to be better than that; it is essential to recreate the entirety of the process and not just mimic the output. In this respect, von Neumann systems are close but not quite good enough.

So what exactly is wrong with the von Neumann systems? They appear for all intents and purposes to accurately replicate our mental operation. The problem is that

[5] "Programming Languages." TechnologyUK. January 1, 2001. Accessed March 3, 2015.

"there is no all-purpose memory warehouse in the brain; there are no smallish, domain specific memory stores either" like there is in the von Neumann system.[6] So while the von Neumann architecture correctly mimes the symbolic translation process of the brain, it is still limited by the necessity of having a storehouse of all its memories and actions. There is no corresponding anatomical structure in our heads that serves this function. It appears that having memory function in this way does not allow for any form of learning. A von Neumann machine would be akin to a human infant, forever interpreting the world based on the programs with which it began.

Critics of artificial intelligence generally cite this form of intelligence when analyzing its limitations. Most make the claim that such technology could not possibly be intelligent enough to have a mind due to the fact that their processes are entirely predetermined from the onset. They do not operate in response to the world nor do they have any comprehension. A notable illustration of the understanding lacking in von Neumann systems is expressed by philosopher John Searle in his renowned Chinese room thought-experiment.

The premise of Searle's representation is that the inner life of a computer is comparable to placing an English speaking man in a room with a book containing correlating Chinese symbols and a set of rules written in English. The rules allow the man to receive questions written in Chinese and, using his rules, provide a correct answer also in Chinese by pure association without ever actually understanding the language. A bystander would witness questions going in and correct answers coming out and might, therefore, be tempted to ascribe understanding to this room. Searle cautions us against

[6] Flanagan, *The Science of the Mind*, 1984.

this impression. The man in the room, in Searle's view, neither possesses understanding nor any pertinent form of intelligence with which to further our own understanding. [7]

Searle's Chinese room is a rather ingenious explicative of what a von Neumann computer program is doing. It does receive input in one language and then uses rules in another language to translate the input into the correct form of output. The problem that Searle did not realize is that his thought experiment is also an excellent report of how the brain functions. Our brain does not understand English; its language is entirely composed of electrical signals. A person says "hello" to me, I transmute that into electrical signals which follow a path in my brain inevitably stimulating the motor cortex which moves my mouth and vocal chords in a way to create a vibration which sounds something like "how's it going man?" My brain is not aware that what it just said has another meaning other than that it has traditionally led to successful acquisition of success in pursuing its goals. The notable distinction is that our brains have adaptability; the freedom of plasticity in the face of experience.

The difficulty with both the Chinese room and von Neumann machines as conceptualizations of the brain is that they are purely serial processes with preloaded memory units containing sets of preordained instructions and no ability to adapt. Human brains, on the other hand, often operate in parallel and are acclaimed for their adaptive nature. The man in the room is not given the chance to learn the Chinese characters because he has no teacher.[8] He "knows" that two phrases are associated but not what the context could be. The von Neumann machines are designed for serial tasks and are

[7] Searle, John R. *Minds, Brains, and Science*. Cambridge, Mass.: Harvard University Press, 1984. Print.
[8] Bombardi, Ron. "The Education of Searle's Demon." *Idealistic Studies* 23, no. 1 (1993): 5-18.

completely without plasticity. For instance, von Neumann machines specialize in arithmetic and games such as chess.

A famous example of this came in 1996 when a machine named Deep Blue defeated Garry Kasparov, a world champion chess player, in a one-on-one challenge. The artificial intelligence of Deep Blue had the instructions and tactics of chess programmed into it. This extensive programming was enough to beat the reigning human but it still had troubling limitations. Particularly, it could not learn from its mistakes. If there were to be a flaw in its game play which resulted in losing a vital piece, the human player may be able to exploit it, assured that the machine will always do the same responses. Fortunately for Deep Blue the iterations of possible chess boards are so extensive that this scenario is unlikely to ever play out again. The principle is still problematic for understanding intelligence. Serial processing can be extremely fast but it is a limitation which our brain has overcome.[9]

In order to achieve true human-inspired intelligence it became imperative to develop a machine which could more accurately demonstrate the human capacity of learning and responding to new experiences. To do this, machines would have to replicate the complex structure of connections which form in our brain as a response to practice and training. Memories do not exist in our brains as a file tucked away in a neuron somewhere but rather as a web of interconnected synaptic pathways. When a specific set of qualia—the phenomenal properties of experience—are perceived the brain is going through determined states or dispositions. Memories are not permanent residents

[9] Kurzweil, *How to Create a Mind,* 2012.

of your mind but are produced when neural connections fire a precise way. These neural pathways are laid out as we learn and are refined as we are enlightened by experience. [10]

To fully understand what I mean, try remembering the last time you were reading a book. Can you recall the points which were being made? What about the emotions that the words elicited, the clothes you were wearing or the pages you stopped at? It is hard to recall these specific details. It seems that at any moment "we are conscious of only a tiny fraction of the information our brains process…[a]lthough we continually notice changes in our behavior."[11] While we do not record the specific qualia, our brains do form new connections in response to experiences, gradually refining actions in order to produce the most efficient responses to stimuli. Redundancies, such as driving your car to work, however, do not need to be stored and are quickly forgotten experiences. The environment triggers specifically patterned neural networks in our brains which bring out the semblance of memories into our emergent consciousness.

It is this sort of learning that artificial intelligence must be able to accomplish in order to accurately recreate human intelligence. The brain is a system of formal structures connected together; individual sets of limited mindedness all coalesced to create the mind. The more we progress toward an accurate conceptualization of our brain, the closer we get to a working model which we can then use to explore some of the fundamental questions of its nature. In the end, we will want to construct a realistic account of the operations of the brain. It is evident that Turing and von Neumann systems do not function precisely like our brains. For this reason, it was necessary for cognitive scientists to embrace other theories of the mind than serial models. It was soon discovered that the

[10] Flanagan, *The Science of the Mind*, 1984.
[11] Harris, *Free Will*, 2012.

millions of analog neurons coupled with the synapses they form hold the answers to the paramount questions about human brain functionality and therefore the emergence of a self. A proper artificial intelligence which reflects these neuronal networks and their plasticity is the key to unlocking and unraveling the mysteries of the self. First, however, we had to create a computer that could learn.

## V. Teaching Machines

Neural processing in the brain happens in parallel. Modern day proponents reject the Turing test as sufficient evidence for intelligence on the grounds that it only demonstrates a one track mindedness. While "it is very important how the input-output function is achieved; it is important that the right sorts of things be going on inside the artificial machine" so that we can overcome the "behavioral failures of the classical…machines and [create] machines with a more brain-like architecture."[1] When stimuli crash into our sensory organs they do not come one after the other but are rather bundles of information. Seeing a person entails analyzing color, motion, a detailed outline of the shape and maintaining some amount of awareness about one's environment.

If my brain were to operate in series then I would need to stop typing in order to breathe, then beat my heart, then blink, and finally type another letter. While this could be done in rapid succession, simultaneous perception would be impossible. It is apparent that brains can handle many actions at once without thinking about them whatsoever. A seemingly infinite amount of neurons form synaptic connections so that the world can be processed as it collides with our sensory tools. We have this complex set of neural networks in place in order to maintain functionality in the face of so many stimuli from the environment. If non-biological intelligences are to reach our level, therefore, they must be able to work in parallel. By reverse engineering the brain, scientists put the

[1] Churchland, Paul M., and Patricia Smith Churchland. "Could a Machine Think?" *Scientific American*, 1990, 32-37.

connectionist model found therein to work. Thus, parallel distributive processing was born.

Parallel distributive processing mimics human learning through the creation of dispositional states which are the result of patterns of activation. In order to address the problems with serial processing, it is "the bet of parallel distributive processing researchers that most of cognition—sensory experience, perceptual recognition, possibly scientific theorizing itself—is to be analyzed in terms of dispositions to experience, recognize and classify incoming electrical signals in terms of certain characteristic neural-activation patterns." In other words, if we were to look into someone's mind as they had an experience we would not see a single neuron in one location firing and proclaim that to be the source of the experience. Instead there would be a sequence of patterns reacting to one another, sending signals down synaptic pathways millions at a time. These sequences of patterns create who we are by the way in which they are connected. The patterns they form are entirely unique, as no two people will have their cortical column of neurons connected the same way. When we are in the dispositional states of thinking or acting we are in reality acting out the "characteristic patterns of activation" in our brain which are molded and refined as new experiences are incorporated and new actions are undertaken. [2]

Neuronal connections send electrical signals down channels when prodded by a stimulus. Suddenly, memories begin to surface; previous experiences arise when someone is asked about their childhood home or their best friend. You have no control over these occurrences; instead they float to the surface of your emergent consciousness. Memories are the result of the synaptic connection of millions of neurons in response to

[2] Flanagan, *The Science of the Mind*, 1984.

experiences. When a certain pattern is firing then the brain is in a notifying dispositional state. Dispositions are law-like, predetermined results of a large group of activated neurons which are reacting to stimuli. As MIT neuroscientist Sebastian Seung once put it, "identity lies not in our genes, but in the connections between our brain cells."[3]

Parallel distributive processing works by creating a system and instilling it with the sole purpose of using its machinery to establish more efficient connections. The Hebbian theory in neuroscience describes a similar basic goal for neuronal synaptic formation, stating that neuronal connections adapt in order to provide the most proficient reaction. Parallel distributive processing operates on three levels: input units, hidden middle units and output units. The system starts out giving random answers, feeling its way along. Initially it does not know whether its interpretation of the stimulus will be correct or incorrect. It will posit something like "Am I getting A or B? Perhaps this is B?" The connections which the system has in place are entirely ignorant and untrained, like an infant child grasping and suckling. Then a teacher must be introduced into the system in the form of human guidance or even another identical system that has already learned. We can even give the machine genetic algorithms which mimic evolution and its search for optimal solutions. Either way, the teacher trains the network by responding to its interpretations with either a check or an x, saying "Yes, it is in fact B". Training the network over time by rewarding proper responses and punishing incorrect ones allows the system to determine what the correct strength levels must be for specific stimuli to produce certain actions. The coaching system gives feedback to the efforts of the fledgling program and allows it to formulate answers on its own so that one day it can work efficiently without assistance. Trials show that even if the teacher is relatively

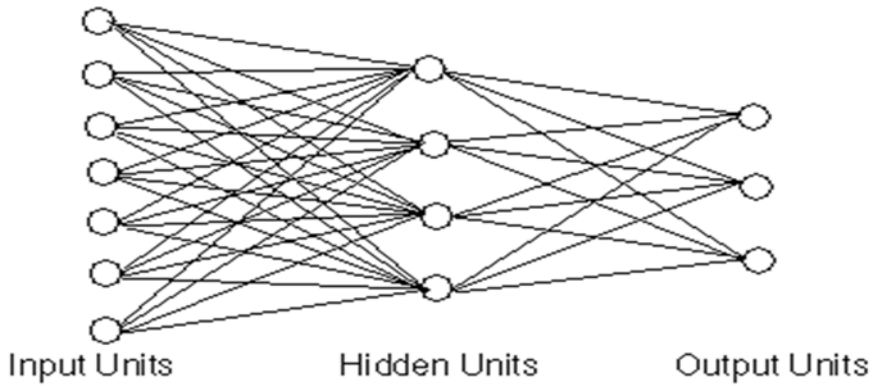
[3] Kurzweil, *How to Create a Mind*, 2012.

inaccurate that the new system will still reach an accuracy level of 100%. The only prerequisite is that the coaching system must be correct at least over half of the time.[4]

An example of this process provided by Owen Flanagan in *The Science of the Mind* will serve to illuminate exactly what this technique looks like. Imagine a system whose goal was to distinguish between mines and rocks buried under the earth. The machine is equipped with sonar producing and detecting equipment in order to gather stimuli from the world but does not know which acoustical pattern is representative of bouncing off a mine versus that of a rock. At first it must amass a significant amount of data and group the responses based on signal strength. Presuming that there is a significant distinction between the signal strength of mines and rocks, the system filters the results through hidden units in order to group activity levels. The system is initially ignorant of what it is trying to do and guesses at which one is which. The teacher then enters into the system. It "reads the network's guesses and it gives the network feedback about how it is doing" in order to "train up the network". The system takes in the critique and, through back propagation, tailors its channels to respond to signal strengths which have a higher percentage of being correct. If, hypothetically speaking, a strong signal means the sonar signal is of a mine, then every time the system guesses that the strong weighted value found in the hidden units indicate the sensing of a mine, the teacher will reward it. The network will send its success back down its connections to more accurately predict a strong signal as a mine. Once it has done this a number of times, the system can now discriminate between rocks and mines. The mine-rock detector learns to take on activation patterns in response to certain stimuli, in this case adopting the disposition of stating whether something is a rock or a mine. Nowhere is there a central

[4] Flanagan, *The Science of the Mind*, 1984.

processing unit with a list of instructions. Instead the parallel distributive process' intelligence comes from the system learning how to respond to its environment over time, changing connections in response to new experiences.[5]




[6]Illustration of parallel distributive processing

Machines employ parallel distributive processing to experience a relationship with the world which is dynamic and malleable. They mature and learn in response to stimuli and can determine what sort of actions are the most efficient for achieving its goals and will favor them over others. The correlation between how these machines learn and how neuroscientists currently understand the way a human learns is indistinguishable. Like parallel distributive processing, humans have no center for storing memory files, can handle multiple stimuli at once, utilize representational, disposition states in order to produce action and mature past our initial programming through the education of a teacher. For now it is clear that artificial intelligence will have to use this structure in order to exceed the limitations of blind programming.

[5] Flanagan, *The Science of the Mind*, 1984.
[6] Garson, James. "Connectionism." Stanford University. May 18, 1997.

This is definitely a complex understanding of our brain; a comprehensive perception that took a lot of smart people many years of work to understand. A fascinating thing about our brains is that once knowledge of the world is hypothesized, it does not take the rest of the community long to comprehend it, comment and further refine it. Parallel distributive processing gives machines the ability to obtain knowledge on their own and reach new innovative conclusions. Optimizing the processing of the wealth of incoming stimuli will lead to breakthroughs which are currently out of reach simply because of the time it would take a human mind to mathematically analyze the trillions of possibilities. The old fashioned artificial intelligence of Turing and Von Neumann could theoretically accomplish this as well but not nearly to the extent of parallel distributive processing. The General Problem Solver, created in 1959 by Herbert Simon, J.C. Shaw and Allen Newell, demonstrated the ability of a program to solve symbolic problems when it constructed a proof for a theorem which renowned mathematicians Bertrand Russell and Alfred Whitehead failed to solve in their 1913 highly influential work *Principia Mathematica.* Even this old fashioned computer demonstrates the ability artificial intelligence has to outperform human minds. [7]

With the competency to learn and accumulate data, machines have already begun to demonstrate their superiority over their biological counterparts. The best example of the application of the analytically modern, adaptive nature of artificial technology occurred in 2011 when an artificially intelligent computer system created by IBM named "Watson" competed in the question and answer game show *Jeopardy*. For the first time, computational ability was matched in a battle of wits against some of the brightest minds to play the game. Competing in *Jeopardy* is not the same as creating a proof or beating a

[7] Kurzweil, *How to Create a Mind*, 2012.

world chess champion. While those are tremendous accomplishments, they pale in comparison to the performance aptitude necessary to execute the requirements of the question and answer game show.

Computers are logic machines; therefore it was not surprising when Deep Blue won at chess, which is perhaps one of the most logical games one can play. At the time of its victory, critics stated that while machines could surpass human logical abilities, the nuances of language and perception would be forever out of technology's grasp. This notion was finally put to rest when Watson beat the two most successful contestants to ever play *Jeopardy* in front of a live studio audience. For those unfamiliar with the game, *Jeopardy* is based on a reversal of the typical question answer format. The host provides the answer and the contestant must give the question to which the provided answer would be correct. The true difficulty comes in the fact that the host's questions, or rather posited answers, are generally some combination of metaphor, pun and double entendre.

Therefore, in order to compete in this game, Watson had to not only learn the English language but also discover the minutia that comes with its mastery. What makes this feat even more fascinating is the fact that Watson's knowledge was not programmed by anybody. Furthermore it could not be connected to the internet and had no physical memory bank while playing. A vast majority of the information Watson had was a result of its meticulous reading of millions of documents, including the entirety of *Wikipedia*. The clues which Watson answered correctly include "A long tiresome speech delivered by a frothy pie topping" to which it correctly responded "Meringue" and "National Teacher Day and Kentucky Derby" to which it again correctly responded with "May". In response to his defeat, Ken Jennings, the longest running winner in *Jeopardy*'s history,

stated that "The computer's techniques for unraveling clues sounded just like mine…I felt convinced that under the hood my brain was doing more or less the same thing." [8]

Watson, without the help of an internet connection or memory banks, had to have the knowledge some other way in order to achieve success. Moreover, it would have been much too slow combing all of them for answers; if it wanted to win it had to do something differently. It is evident to see that machines like Watson can easily pass the Turing test for intelligence and then some. But how does it work? Essentially it is the same sort of processes which occurs in our brain when a question is asked. The brain forms connections between experiences and filters stimuli through them in order to produce the correct physical movements and thereby the answer. Watson used patterns of connections in order to discover answers as well. Clusters of these connections have been conditioned to activate when a signal of a specific strength is detected. The keywords in the clue trigger patterns which connect to produce representational symbols that have been associated with the stimuli.

Opponents of Watson's intelligence argue that all it is doing is blindly running algorithms and statistical analysis which does not constitute a mind. The trouble with this argument is that they are then forced to define a mind outside the mathematical probabilities of association. It seems that the further artificial intelligence encroaches on the human mind, the more that its opponents are pushed into the immaterialist corner with Descartes. What Watson and other artificial intelligences demonstrate is that our minds are transactions of physical properties which can be thought of in terms of statistical analysis and mathematically computation. When I receive a stimulus, my brain weighs all of the different options available, but neurons will only fire if they receive the

[8] Kurzweil, *How to Create a Mind*, 2012.

right strength of signal. Currently the most successful models of the human brain utilize "mathematical techniques that have evolved in the field of artificial intelligence [and have been proven to be] similar to the methods that biology evolved in the form of the neocortex."[9]

As our technology begins to master all the intricacies of mental life, it is natural to begin to imagine the day when full brain simulation in its entirety will be feasible. Proponents of strong psychological artificial intelligence are already starting to see this day on the horizon, but most others are hesitant to make this claim. Digital brains, of course, need not perfectly replicate our biological brains in order to be functionally identical and offer a lot of insight. It must be admitted, however, that even an operationally exact brain would not offer as much understanding as a literal reproduction of the structures in our biological brain. There are two ways to conceive of the imitative progression proposed by strong psychological artificial intelligence proponents. The first is to create a brain which learns through hierarchical pattern recognition and then let the brain grow and mature. The process would be akin to our own; essentially the system would receive a huge supply of stimuli and be forced to resolve and adapt to it. Of course there would need to be a parental teaching figure to provide the necessary feedback. Its propagation loop would allow it to reach adulthood the same way a human would. The other option is to scan the brain of an already formed adult brain and recreate its connections. The difficulty in this would be in downloading all of the cortical structure with all of their microscopic subtleties and then accurately maintaining the precise shapes as it is uploaded onto a non-biological system.

[9] Kurzweil, *How to Create a Mind*, 2012.

The pursuit of a fully simulated human brain has already begun but is still in its early stages. A computer scientist with IBM named Dharmenda Modha has already successfully simulated a piece of our visual cortex. With 1.6 billion virtual neurons and 9 trillion synapses, it only runs about one hundred times slower than the real thing in our heads. Bear in mind that one hundred times slower than such a complex and quick machine like our brains is still incredibly fast. Additionally, The Blue Brain project, started in 2005 by the École Polytechnique Fédérale de Lausanne in Switzerland, is an ambitious project led by Henry Markram with the goal of a full brain simulation by 2023. Naturally, there will need to be a lot of trial and error before these things are perfected but the race for simulation has begun and the finish line is on the horizon. [10]

[10] Kurzweil, *How to Create a Mind*, 2012.

## VI. Will Cyborgs Have Selves?

As we approach an inorganic level of artificial intelligence comparable to our own, it is necessary to discuss the quintessential philosophical question which is bound to arise in its wake. It is imperative to address whether a cyborg, short hand for cybernetic organism, will have a "self" as we perceive it today. If a fully simulated brain does indeed have a synthetic form of human intelligence, must it consequently have a self? When I address this cyborg will it (they?) want to answer to "who are you" as opposed to "what are you" lines of questioning? If not, then why not? While we may not be the arbiter of our experiences, there still remains some emergent image of a conscious self to which we feel intimately connected. A perspective of whom or what one is, no matter how far off, is generally cited as a prerequisite for an understanding of one's self. A self, defined thusly, has historically been reserved for our species alone. To reflect this, we created a hierarchy of consciousness. Animal rights exist as a separate category from the rights of humans and even then there is an order of magnitude based on perceived understanding and cognitive function; adult humans have more rights than infants and primates have more rights than ants. If highly intelligent robots were to enter into this ranking, where would their position be? It is interesting to start evaluating this problem because as the thread pulls, a great many presuppositions begin to fall apart.

A good first place to start is by reflecting again on Daniel Dennett's intentional stance. The intentional stance is something which we are all familiar with. As previously stated, physical manifestations give insight into cognitive processes regardless of whether or not the thing is conscious. The bird is flying away because it is scared of the cat and

sometimes people will say that their car is angry when it does not start. We easily ascribe intentionality to other humans, without acknowledging the presupposition made that they aren't in reality zombies. Without access to another's mental life, we make the assumption that it is there in similar degrees based on their performance of actions in relation to our own when we feel a certain way. We can use this to infer that they are feeling a similar sort of thing.[1]

When we apply this to technology, it gets even more complicated. Deep Blue, the chess playing machine, is often given intentionality in the fact that it "wants" to win and "knows" how to do so. Are these dispositions enough to ascribe the intentionality which we would give to a fellow human (or even to ourselves)? It certainly benefitted Kasparov to imagine Deep Blue as having intentionality regardless if it truly "wanted" or was rather just going through its preordained motions. The challenger ought to predict its moves by making the presupposition that it will make the best move possible for itself. The anticipation of its moves is based on the knowledge that it is aware of the rules and seeks to succeed. By ascribing these states to Deep Blue, we have taken the intentional stance. We cannot see its "mind" anymore than we can see our own but the inference assists us in determining motives of systems which we are incommunicable. One can say, "If I were that machine, I would move the bishop and check the king."

It is of my opinion that the people are intending, and within these people there are parts intending as well, albeit to a much lesser degree. The most basic forms of intention are clearly not conscious and have no sense of self but it is from them that we must derive our own selfhood. These pieces of intentional machines somehow form to create the conglomeration which at once perceives the amalgam of its parts and yet seemingly

[1] Dennett, *Intuition Pumps*, 2014.

stands above and in contrast to them. The parts which form the whole of intentionality are not traditionally given consciousness and yet from their efforts intentionality is gleamed, albeit on a grandeur, composite scale. This gradual progression of intentional states, each stage accompanied by slightly less mindless mechanisms, can be theoretically recreated artificially. The same way that mindfulness is derived from a "Mitochondrial Eve" in our long line of ancestors, so too can mindfulness gradually appear tracing the line from a battery to a cyborg. Surely, the intentional state of a hand calculator is nothing impressive but in combination with a complex network, a multitude of stances all compiled and interlocked, a sense of a self can become an emergent phenomenon.

Critics will claim that because the intentional stance is a tool to predict actions of other entities that it does nothing to address the internal self of the cyborg at question. It does not answer whether or not there would be something to which all of the electrical pulses would appear to, would be experienced and felt by. They have made the same mistake Descartes made of missing all of the trees in an effort to see the forest. They can certainly impersonate a self but perhaps the cyborgs are just robotic versions of the zombies, functionally the same as humans but with only the Antipodean narrative and none of our cherished mental life? A thought experiment may help elucidate the answer.

Let's imagine a man who is sleeping in his bed sometime in the distant future. While in his slumber, a group of scientists scan the person's brain. Then they perfectly reproduce every single aspect of his brain, every connection, and perfectly match the details in a mechanically, fully-automated replica. On the outside the cyborg is identical with the slumbering man but on the inside, instead of a brain, there is a densely compacted set of wires and processors. When the organic man wakes up he is confronted

with an exact replica of himself. It certainly looks like him and can act like him. It even exclaims "I like this new body!" –at which it dawns on him that it is something he too would say if he had a shiny new body. The question is whether or not the cyborg has the phenomena of a conscious self. Is there something inside the cyborg which is truly feeling or is it that there are only dispositional states devoid of *real* meaning? [2]

Another thought experiment will be necessary to further alleviate this stubborn conundrum. Imagine that the same man, in order to get away from his annoying cyborg replica, runs out into the street and is hit by a car. He is unlucky enough to have a small piece of his brain damaged in the accident. Fortunately for him, however, a scientist has just perfect a mechanical simulation of that exact area of the brain which can be transplanted into his head allowing him to regain all of his previous functionality. He thinks the same, acts the same and his friends attest to how familiar he is. Clearly he has not changed. Unfortunately the pitiable man again runs into the street and is hit by a car which regrettably damages a different section of his brain. The other mechanical piece remains intact but another small section is irrevocably damaged. Again, however, the scientist comes to the rescue and replaces the exact area with an artificial substitute designed to perform the exact same function as the damaged part. The grievous man continues to be hit by cars all the while exchanging damaged areas of his brain for mechanized ones with the same functionality. Soon he realizes that his brain is entirely artificial. We are forced to ask at what point in this version of Sorites paradox does the conscious mind disappear? Is replacing one neuron enough to dismantle consciousness or

[2] Kurzweil, *How to Create a Mind*, 2012.

might there be a specific part of the brain that is responsible for the emergent image of the mind? [3]

It is clear that the man, through the artificial restorative procedures on his brain, has become the cyborg which was scanned and made of him in the first hypothetical scenario. The man still maintains all of his functionality and appears to be the same old guy, while hopefully luckier than before his transition. He can still behave the same and even reports first hand that his self has remained intact. His friends are convinced and subscribe to strong artificial intelligence while, on the other hand, many of his acquaintances refuse to accept his self reports and continue to insist that he is dead and that the hollow shell in front of them is offensively impersonating a dead guy they once knew. The cyborg tells them that this hurts his feelings but he gets over it soon enough.

Turing's test would certainly say that these sorts of artificial intelligences ought to be given the status of an intelligent self equal to that of a human. They can talk and report on internal mental events, regardless of whether or not one personally believes they indeed have one. The cyborg's answers are indistinguishable from those of its human counter-part. With the likelihood of the eventuality of partially mechanized people in mind, many skeptics may have to start opening up people's heads to see if they have mechanical parts before being their friends because they cannot be too careful to not be fooled.

Neuroscience has exposed the true nature of the brain as that of a complex set of physical systems. Consciousness emerges out of this network naturally and automatically. If representational states and neuronal connections are enough to produce the manifestation of consciousness, then there is no reason to deny that same attribute to a

[3] Kurzweil, *How to Create a Mind, 2012.*

system which is identically hardwired. If a cyborg cannot be accurately distinguished from the functionality of the average human than it has the right to make a claim at self-hood. When a machine encounters a set of stimuli it will react as it has been trained to do through experience. Once artificial intelligences begin to learn and improve on their own comprehension, it is at this point that a self must be attributed. When I ask who-questions of a cyborg and they respond with whatever conclusion they want to respond with, then I must necessarily proclaim some sort of self is present. “Who are you?” I could request and they may respond with “I am a cyborg!” or perhaps “I’m Steve, who are you?” If, upon continued persistence as to the nature of their self, they reveal all sorts of interesting aspects of self, eventually becoming a little off-put by my persistence, I cannot stamp my feet and proclaim that they are liars for there is no way to demonstrate their belief state is any different from my own. The only argument a critic would be left to repudiate with is that they simply do not want to attribute a self to the cyborg.

Of course, the extent to which our definition of an emergent conscious self is present relies on the complexity of its connections. A toaster would not be said to have any concept of a self in the same way that a new born infant would have no emergent phenomenal mindedness. They are both, rather, simply running the program which has been given to them. The difference of course is that the infant has the capacity to learn and form new connections while the toaster can merely burn my toast. Other sorts of machinery, however, are far more complicated than a toaster and are exponentially more adaptive to their environment in a way that demands selfhood.

Under the tutelage of the best teacher there is, experience, intelligence is formed through associative behavior. Our brains fashion connections between stimuli in the

world in the same manner that a cyborg might, by paying attention and testing hypotheses. When an infant is hurt somewhere on its body its brain senses that and then instinctively produces a wail. The wail serves multiple purposes. First, it beckons the parental figure in order to obtain protection. At the same time, however, the young child watches as the parent holds the hurt spot, perhaps kisses it to make it better. The brain notices this enough times and links up the word “foot” and “pain” to certain stimuli in specifically designated areas in its brain. These words become associated with systems and patterns of receptors which are then manipulated in order to convey internal dispositions and representational states. A narrative emerges as a result of our language capabilities but it is not one that we can dictate or choose.

For cyborgs this would be the case as well. We would not need to decide whether or not consciousness is an emergent phenomenon in their heads, in fact it would be impossible to. It would be impossible only in the same respects that it is impossible for us to determine if anyone else is experiencing consciousness around us. Merely viewing the structure is not enough to discover a mind; if it were, then Descartes would have been satisfied by seeing my mind situated snugly in my brain somewhere pulling levers, a homunculus smiling up at him from his throne in the brain. Instead, to guarantee that a structure produces a self one must *be* the structure and the self. Anything else is mere extrapolation. Taking the intentional stance, we assume that we each are conscious of our experiences based solely on the fact that we can see the causal chain of experiencing play out and can convincingly talk about the process. Our zombies would not be hard pressed in fooling us. The self consists entirely of the ability to convincingly function as if one is bearing witness to their own deeds. If there is no mental discourse happening in the brain

of some entity, so long as it functions indistinguishably from my appearance of self, it can make no difference.

The prevailing perception of consciousness is and will continue to be hard for humans to shrug off. We are notoriously uncomfortable with sharing the title of sentient being with other entities. We must remember that the mind is not some other-worldly gift. The "mind is simply what brains do."[4] If the mind is a result of the brain, and the brain can be shown to be comprised solely of physical, formal systems then it is easy to see that the mind is a resultant property of these systems. It has no real, tangible location but is instead a conceptual tool which our brain uses to try and understand itself. Its notion of self, however, is strikingly inaccurate. As we perfect the impression of ourselves through the looking glass that is artificial intelligence, a new conception will appear. Once we do indeed shrug off the ghost still haunting us from past theories of the mind and fully accept the revolutionary paradigm of the corporeal nature of consciousness, only then will we be able to see that the self is not derivative of the substance of a thing but rather derives from its particular manifestations, functionality and contextual connectivity. Cyborgs will have selves due to the certainty that by learning and reacting, one must be experiencing. The external and internal realities are inescapably linked and must be conceived as such. I am not controlling my body and neither is it controlling me; they are one and the same.

[4] Minsky, Marvin Lee. *The Society of Mind*. New York: Simon and Schuster, 1986.

## VII. Conclusion: Offloading Consciousness

Once the brain is understood as a purely physical, formal system, we can begin an honest pursuit into an operational investigation. Intelligence is the name given to the ability not only to respond judiciously to stimuli in one's environment but also to learn from it and use that information to improve on its interactions. The key word is "judiciously," which implies learning. This is how we grow and how we become agents in the world. It is now time to begin learning properly from the knowledge that we have discovered about our brain.

The brain rewires itself in response to new experiences. We know that our perception of self is a direct result of these experiences. The formally deterministic nature of the systems of our brain can now be translated into computational systems. The computational systems will have to exhibit every form of intelligence of a human and, in so doing, can quickly surpass its capabilities. With any luck, they will want to take up the mantle of understanding which we have carried for so long. This scientific reimagining of what it means to be human has origins in the first hesitant creations of artificial intelligence. While artificial intelligence is only in its fledgling stages, already it has reinvented our understanding of the mind. By creating artificial intelligence that can experience and learn in the same way we do, we can extend ourselves and our abilities, thereby improving many aspects of our lives.

Objectively examining the concept of a mind, tackling the hard problem of consciousness, has proven to be one of the most daunting philosophical tasks. Fortunately, we have been able to create representational systems in the world that can

help alleviate our confusion. Language allowed us to take the first steps in offloading pieces of ourselves onto a physical manifestation. Internal desires and needs could be communicated symbolically and then deliberated upon beyond the limitations of body language. Precision increased and our species began to extend its mind and distribute its inner content beyond the cage of unpredictable, internal mental states.

After language developed, writing became the first step in offloading one's mind, which truly surpassed the limitations of the biological body. The written word has the capacity to make the ideas and thoughts of a person immortal. This fact forever changed our interaction with the world. Suddenly, it was possible to place one's understandings and musings into the world on a medium entirely separated from the body. We no longer needed to maintain ideas as a fleeting mental state, a momentary glimpse at understanding. Instead, knowledge was made tangible and garnered an advanced level of permanence.

Originally, the ancient Greeks were unaccustomed to writing and had to memorize complex arguments and entire plays. They did this with relative ease because they were so adept at pattern recognition. Technology, however, will not be denied. Writing soon permeated their world. There were some even then who saw the potential that technology has in supplanting and improving essential human functions. The most notable of these ancient Greeks was the great philosopher Plato, who stated in his dialogue, entitled *Phaedrus*, that "[the people's] trust in writing, produced by external characters which are no part of themselves, will discourage the use of their own memory within them [and that they] have invented an elixir not of memory, but of reminding..."[1]

[1] Plato. *The Dialogues of Plato*. Trans. and Edited by Benjamin Jowett.New York: Random House, 1937. Print.

Unbeknownst to him, Plato was giving insight into the remarkable ability humans have for integrating technology. A modern day example of Plato's fear can be given when people reflect on how well they knew the phone numbers of their friends and families before cell phones compared to how well they know them now. Before, without the incredibly useful contact section in a cell phone, if you wanted to talk to your friend you had to store that information in your memory, but now it is as simple as typing a name and pressing call. This change is subtle but speaks volumes to the small ways technology infiltrates our world, often without a moment of hesitation from us.

We are intrinsically linked to our personal experiences, an increasingly dominant factor of which is of our own creation. The tools that arise from our remodeling of the world mediate our experiences, acting as a sort of transitory barrier between what there is and how we perceive and understand it. The way we represent and think about the world is inconsistent and fleeting due to our own unflinching remodeling process. Everything from sensory perception to memory to our very forms of communication are changing at an exponentially increasing pace. As a result, so too are our brains. We can use this wealth of mechanized power to understand not only the world and how it operates but also ourselves and how we operate within its context. Artificial development itself is a tool that has given us deeper insight into what we are now and what we are to become. By reverse engineering the brain, we must fine tune our theories of it and subsequently address the philosophical concerns therein.

The artificial intelligence revolution has already begun. The fact that there is a wealth of knowledge at all times just a few clicks away has an incredible amount of impact on how we interact with the world. Already there are generations of people who

will not know what it is like not to have the entirety of human knowledge continually at their fingertips at all times. Technology has inextricably entrenched itself into the most intimate functions of our lives. Non-biological forms of intelligence continue to mediate our daily experiences. Humans have stretched out their internal mental functionality into the world, seamlessly offloading consciousness bits at a time. As technology advances, we offload pieces of our mental system onto the external world. The consequence is that we are no longer required to have extensive neural patterns to solve certain specific problems. A quick Google search does a better job than trillions of synaptic connections could do at retrieving information.

Each mechanized device represents a degree of intelligence, or, more precisely, a piece of our intellectual capacity. When we "outsource parts of our thinking" onto our technologies we are fulfilling a process that began when our earliest ancestors started manipulating the world in an effort to ensure its survival.[2] Simply by allowing cell phones, and other technology, to maintain pieces of information even as seemingly innocuous as photos, contacts, or calendar dates, we are continuing a progression that began when the first modern neocortical structures appeared on the Earth a great many years ago. As we inculcate more functions into non-biological forms, we undertake further the procedure of escaping the limitations of our physiology.

This desire for this transformation stems from the recognition of certain design restraints of our heads. Our brains are vast and extremely sophisticated, yet they still have their physical restrictions. We only have a certain amount of space in our skull to harbor neural networks, but the capacity of artificial intelligence is virtually unlimited. In fact,

[2] Kurzweil, *How to Create a Mind*, 2012.

"digital circuits are 10 million times faster than biological neocortical circuits."[3] Artificially created intelligence will let us overcome our inherent constraints.

An entirely physical description of the brain dictates that one day there will be a sufficiently systematic way of predicting behavior. From there, it is not difficult to fathom a calculated and predicted set of behaviors that we can scientifically discover and be privy to. The more we learn and study the fundamentally human aspects of our brains, the easier it is to understand and reproduce them. A good example of how technology is beginning to probe into quintessential human functions can be seen in the assistance it offers in dating and falling in love. While this externalization is far from being perfected, the fact that technology already has become ingrained in its procedure is testament enough. Online dating sites have mixed results, but many people will testify to the ability that this sort of technology has at matching compatibility. The seemingly random "falling" aspect of falling in love is not something that most people would be comfortable saying a mathematical formula could predict. And yet we know that with an exhaustive understanding of the physical processes of the brain, this is a distinct possibility. Technology could take what is perhaps the most intimate and mysterious function of a person and quantize it into a neat formula. As artificial intelligence continues to shed light, the actualization of our brain states will become progressively more likely.

Offloading bits of consciousness will be the crux for answering those most inexorable and obstinate mind-body problems. The seamless gradualism offered through this transition depicts a process by which the self becomes shared between mechanical and biological media. We already see many people driving into lakes on the command of a global positioning system; but imagine a chip with this capability attached to the brain

[3] Kurzweil, *How to Create a Mind*, 2012.

instead of in your hand. In this case, we would often be unsure about whether the brain had the idea to turn left at the next light or the chip. Either way, we know where to go. The distinction will begin to blur in this way until the man and the machine are inseparable.

Changing a self into an improved, slightly mechanical self will not be jarring. In fact, the upgraded functionality will be—and always has been—welcomed with open arms. Still, opponents will find a way to disparage a separated system of intelligence. Artificial intelligence continually has crossed every line that these people have drawn and yet more lines get placed immediately after. I say that we are better for this as it gives motivation and stimulus to the scientific community at large. Eventually, however, these critics are backed into a Cartesian dualism which they surely did not want. Offloading serves to unify the two supposedly distinct forms of biological and artificial intelligence and resuscitate the hackneyed perception of the self.

We are on the cusp of a technological boom that will undoubtedly lead to a better understanding of how to identify the functionality and purpose of so-called “human” qualities and their qualia. The better we understand the physical processes constituting our natures, the more likely it appears that someday a singularity between human and machine will be reached. Artificial intelligence at this crucial juncture will be indistinguishable from our own form of intelligence. Man and machine, the constitution perpetually unified. The competency demonstrated by technology endeavors to displace and make obsolete our basic human functioning and all of the notions that we concocted in its image.

As we continue to offload basic human functions onto technology, artificial intelligence will advance our pursuit of understanding and aid in our continued survival. Through new mechanization, we can outsource more and more of our functionality and supplant it with the seemingly infinite capacities technology demonstrates. As artificial intelligence advances, its opponents will find less and less room with which to maneuver their dissent. First, it was never beating Turing's test; currently, it is the inability to "truly know" or "really understand" experiences the way we do. As our reproductions rise to each challenge, critics are forced into a conceptualization of Descartes' immaterial mind in which they claim that while a machine can "sort of" know, it is still missing some ethereal factor that it just cannot possibly attain. Again, they assert, that our self is somehow incorporeal and permanently unknowable. It seems that the only way these critics will be dug out of their ideological entrenchment is by the transmutation of an accepted conscious self into a mechanical self, all the while maintaining the essence of its selfhood—seamlessly offloading the entirety of a brain's functionality onto identical machine states. While this is undoubtedly a prospect for the distant future to wrangle with, we can see today that the very roots of our emergent conscious selves are revealed, modified, and redistributed at the hands of artificial intelligences. Continued offloading of essential human functions onto the world will develop an idea of the self that is just beginning to be formulated and take shape. The combination of biological and non-biological intelligent technologies has already begun. As the fledgling process progresses, so too do we. The heights we can reach and the understandings we can reap with the help of artificial intelligences are virtually limitless.

## Bibliography


Bombardi, Ron. “The Education of Searle’s Demon.” *Idealistic Studies* 23, no. 1 (1993): 5-18.

Calvin, William H. *The Cerebral Code Thinking a Thought in the Mosaics of the Mind*. Cambridge, Mass.: MIT, 1996. Print.

Chalmers, David. “Facing Up to the Problem of Consciousness.” *Journal of Consciousness Studies*, 1995, 200-219.

Changeux, Jean, and Alain Connes. *Conversations on Mind, Matter, and Mathematics*. Princeton, New.Jersey: Princeton University Press, 1995. Print.

Churchland, Paul M., and Patricia Smith Churchland. “Could a Machine Think?” *Scientific American*, 1990, 32-37.

Cushman, Fiery. “What Scientific Concept Would Improve Everybody's Toolkit?” *Edge*.

Damasio, Antonio R. *Descartes' Error: Emotion, Reason, and the Human Brain*. New York: Putnam, 1994. Print.

Damasio, Antonio. *Self Comes to Mind: Constructing the Conscious Brain*. New York: Pantheon Books, 2010. Print.

Dawkins, Richard. *The Blind Watchmaker*. New York: Norton, 1986. Print.

Dennett, D. C. *Darwin's Dangerous Idea: Evolution and the Meanings of Life*. New York: Simon & Schuster, 1995. Print.

Dennett, D. C. *Intuition Pumps and Other Tools for Thinking*. New York: W. W. Norton and Company, 2014. Print.

Descartes, Rene. *A Discourse on the Method of Rightly Conducting One's Reason and Seeking Truth in the Sciences*. Waiheke Island: Floating Press, 2009. Print.

Flanagan, Owen J. *The Science of the Mind*. Cambridge, Mass.: MIT, 1984. Print.

Garson, James. "Connectionism." Stanford University. May 18, 1997.

Hacking, Ian. *Representing and Intervening*. New York: Cambridge University Press, 1983. Print.

Harris, Sam. *Free Will*. New York: Free Press, 2012. Print.

Hofstadter, Douglas R., and D. C. Dennett. *The Mind's I: Fantasies and Reflections on Self and Soul*. New York: Basic, 1981. Print.

Kirk, Robert. "Sentience and Behaviour." *Mind*. 1974, 43-60.

Kuhn, Thomas S. *The Structure of Scientific Revolutions*. 2nd ed. Chicago: University of Chicago Press, 1962. Print.

Kurzweil, Ray. *How to Create a Mind: The Secret of Human Thought Revealed*. New York: Viking, 2012. Print.

Leisman, Gerry, Calixto Machado, Robert Melillo, and Raed Mualem. "Intentionality and 'Free-will' from a Neurodevelopmental Perspective."*Frontiers in Integrative Neuroscience*, 2012.

Lewontin, Richard C. *Biology as Ideology: The Doctrine of DNA*. New York, NY: Harper Perennial, 1992. Print.

Minsky, Marvin Lee. *The Society of Mind*. New York: Simon and Schuster, 1986.

Mishlove, Jeffrey. *The Roots of Consciousness: The Classic Encyclopedia of Consciousness Studies: Revised and Expanded*. Rev. ed. New York: Marlowe, 1993. Print.

Nagel, Thomas. "What Is It Like to Be a Bat?" *The Philosophical Review*, 1974, 435.

Plato. *The Dialogues of Plato*. Trans. and Edited by Benjamin Jowett.New York: Random House, 1937. Print.

Priest, Stephen. *Theories of the Mind*. Boston: Houghton Mifflin, 1991. Print

"Programming Languages." TechnologyUK. January 1, 2001. Accessed March 3, 2015.

Putnam, Hilary. *Mind, Language and Reality: Philosophical Papers Volume Two*. New York:Cambridge, 1980. Print.

Rorty, Richard. *Philosophy and The Mirror of Nature*. Princeton: Princeton University Press, 1979. Print.

Ryle, Gilbert. *The Concept of Mind*. University of Chicago Press, 2000. Print.

Searle, John R. *Minds, Brains, and Science*. Cambridge, Mass.: Harvard University Press, 1984. Print.

Schopenhauer, Arthur, and T. Bailey Saunders. *The Wisdom of Life, and Other Essays*. New York & London: M.W. Dunne, 1901.

Turing, A. M. "I.—Computing Machinery And Intelligence." *Mind*, 1950, 433-60.