Philosophical perspectives on the Singularity

Luuk van der Vleuten

780255

14 – 03 – 2014

Department of Philosophy, FGW, Tilburg University

Philosophical Perspectives on the Singularity

Introduction

The technological progress that characterizes our modern world is becoming an increasing part of our lives. Due to scientific research and economic growth, new possibilities of improving our lives are becoming actual. In the fields of artificial intelligence (AI), nanotechnology, biomedicine, computational technologies and genetics new opportunities for applying scientific insights are coming to fruition. This leads to innovative developments in applied science fields such as robotics, augmented or virtual reality, body enhancement, communication technology, medical treatment of patients and psychological therapies. These new forms of technology are likely to have a great impact on our everyday life. Our ways of working, learning and communicating are going to change drastically through implementation of new technologies in our world. Both human longevity and the longevity of information will be challenged and eventually extended. Because the way we interact with the world and with each other is constantly subject to change our way of experiencing the environment is going to change accordingly. Not only does this raise implications for specific fields in the philosophy of mind, but also for the social, cultural and economical sciences. In the very least, a critical look at this technological progress and its merits is required.

In 1987, inventor and futurist Ray Kurzweil wrote *The Age of Intelligent Machines*, which he later adapted to another book called *The Age of Spiritual Machines* (Kurzweil, 1999). These two books were updated by Kurzweil into one book called *The Singularity is Near* (Kurzweil, 2005). In this summative book, Kurzweil describes what awaits the human race as technological development progresses exponentially towards the future. In his view this is a future in which technology changes our everyday life, our scientific research, our way of government and even our place in the universe. According to Kurzweil, the way of being human that we have become used to is going to change radically and very quickly, to the point that we won't recognize ourselves as humans anymore. All this will happen in an event which Kurzweil calls the technological Singularity.

The aim of this essay is not to investigate whether Kurzweil's predictions are accurate or sensible, because this would still be a problem of hypothetical empirical research. I am confident that Kurzweil has his numbers and statistics right, because in addition to his own results regarding the prediction of some sort of event, he also mentions a conservative estimate of this event happening within the near future. This gives his reader a time frame for certain fundamental changes to occur. These are the fundamental changes that are linked to our way of communicating and perceiving the world. Combining Kurzweil's predictions, a timeline of what awaits us on in the future can be constructed.

One specific field of scientific and also philosophical research is the field of artificial intelligence. In the case of Kurzweil's predictions, this specific field is of great importance for the coming of a technological Singularity. In this article I will explore some of the views on machine intelligence developed in the philosophy of mind and put them to use in a critique of Kurzweil's conceptions of realizing artificial intelligence. So my focus is not on the process of technological and scientific progress, but rather what we can say in philosophical terms about machine intelligence and its implications for mankind and the environment we live in. As the difference between what is human and what is machine is disappearing, Kurzweil believes that no matter how much we change, we will always remain human. However, this argument is not clearly explicated and needs elaborating.

First, I will explain briefly the content of Kurzweil's *The Singularity is Near* and the problems that are philosophically interesting to examine. In the next part of this paper I will look at some viewpoints that are prominent in the philosophy of mind regarding machine intelligence. Finally I will use the philosophical conceptions of machine intelligence to shed some light on what Kurzweil is talking about when he's speaking of the technological Singularity. In other words, I will attempt to give an answer to the overall hypothesis of this article: What role can the philosophy of mind play in the discussion on the emergence of artificial intelligence?

Chapter 1: The Singularity is Near

In the near future, the human race will become one with their technology. The rate at which our technology is growing is exponential. Technological evolution is necessarily the next step in the evolution of mankind. Ray Kurzweil defends these claims in his book, *The Singularity is Near*, where he presents and summarizes two decades of research on the progress of scientific development and the growth of technological advancement. In general, his thesis is that exponential growth in technological development and scientific research will result in an event called the Singularity.

This chapter will contain some information on the content of *The Singularity is Near*, explaining Kurzweil's thesis, but especially focusing on machine intelligence as well as enhancement of our own intelligence. At the end I will explicit the philosophical issues and conceptions that are constantly in the background of Kurzweil's texts.

1.1 The law of accelerating returns

Central to Kurzweil's thesis is his concept of the law of accelerating returns that describes the acceleration of the pace of, and the exponential growth in, the products of an evolutionary process. In this definition, the term 'evolution' refers to a more general concept than biological evolution or natural selection as occurs in the global ecosystem. Kurzweil contrasts the exponential growth of the law of accelerating returns to a simple linear growth. He emphasizes the observation that on a smaller scale, any progress will seem like it is of a linear type, but when looking at a large scale, an exponential increase can be attributed to the same line of progress. The law of accelerating returns is Kurzweil's mathematical model to help describe the increasing progress made by scientific research and technology, leading mankind towards the Singularity.

The concept of the law of accelerating returns has developed from Kurzweil's research into empirical data on statistical changes in several different fields and industries, ranging from the amount of cell phone subscribers to the amount of transistors per square inch (Moore's law) to the cost of sequencing one DNA base pair. In *The Singularity is Near*, Kurzweil illustrates his

research into these fields and industries with a couple dozen graphs, most of which are a logarithmic plot. This means that while the plots on the graphs seem linear, they're actually exponential of nature, since the graphs' axes are of logarithmic scale. This way of illustrating the law of accelerating returns is remarkable. It shows how the evolutionary process of the law of accelerating returns works through specific developments such as nano-engineering patents and internet data traffic, but also the massive scale of development of life on earth in general.

The law of accelerating returns is essential to Kurzweil's idea of the technological Singularity. The prospect of a limitless advance of technology and the increase of scientific knowledge depends on wholly on the rate of this development. Without the continuously increasing rate of exponential growth, there will be no time at which technological progress 'booms'. Because the law of accelerating returns is a statistical observation, it is still hypothetical at best. Despite of the vast amount of empirical evidence, there is no guarantee that the law of accelerating returns will keep emerging in the future as it has in the past as Kurzweil predicts. However, as mentioned in the introduction, I will not contest Kurzweil's empirical statistical research, but restrict myself to the philosophical viewpoints connected with his conception of human nature, machine intelligence and the Singularity. In this essay, the law of accelerating returns is accepted as a working hypothesis from this point on. The predictions that Kurzweil provides in line with the law of accelerating returns are at least as hypothetical as the law itself, but there is definitely merit in looking at what kind of future he has in mind for us.

1.2 The Singularity

To make Kurzweil's concept of the technological Singularity clear, I continue by illustrating the exponential growth of technological and scientific progress with a summary of what concrete things the Singularity means to Kurzweil. He divides the whole course of the universe into six epochs, but since they are not relevant for this discussion, I will not elaborate on this categorization.[1]

The power (price-performance, speed, capacity and bandwidth) of information technologies is growing exponentially. Human brain scanning is an exponentially improving

1 Throughout his book, Kurzweil has references for his empirical claims. See Kurzweil, *The Singularity is Near*, pp 497.

technology; both temporal and spatial resolution is doubling every year. Scientists are obtaining the tools necessary for reverse engineering the brain, which basically amounts to decoding the brain's principles of operation. There are already impressive models and simulations of a variety of the brain's several hundred regions. According to Kurzweil, we will have a detailed understanding of how the brain works within twenty years. Following the reverse engineering of the brain and the growth of information technology, we will have enough hardware to emulate human intelligence with supercomputers by the end of this decade. Personal computers with the ability to emulate human intelligence will follow a decade later. Effective software models of human intelligence will be available by the mid-2020s. As we create artificial intelligence that is at the level of human beings, these intelligences will pass the Turing test around the same time, indicating that AI and human intelligence will be indistinguishable from each other.

Human biological intelligence and AI will, as it were, join forces and provide the best of both worlds. Generally speaking, the brain excels at pattern recognition and reasoning (planning, what-if experiments), while machine intelligence has far more processing power and memory capacity than the brain. Kurzweil emphasizes that the most important strength of machine intelligence is their ability to communicate and share knowledge at way higher speeds than human communication through language. The advantage that this gives to machines is that they will have complete freedom of design and architecture. They aren't slowed down by biological limitations such as the relatively slow switching speed of interneuronal connections or a fixed skull size; their performance will be consistent. Once machines achieve the abilities that humans have, like science and engineering, they will have control over their own programming and thus the ability to manipulate themselves.[2]

What follows is the machines taking over the process of the law of accelerating returns. They will result in a feedback loop in which machines are improving their own abilities at increasingly faster rates. By this time, nanotechnology will enable the manipulation of reality at the molecular level. Also, nanotechnology will give rise to nanobots, such as robotic red blood cells. By directly connecting with the body's nervous system, nanobots will be able to extend human experience by creating virtual reality. For humans, this is a crucial point in the technological development of the Singularity, since it is this point where we cease to be human

2 Kurzweil mentions that, by using genetic biotechnological tools, humans are doing the same thing already.

and become transhuman[3]. Another benefit of nanobots would be that they can clean the air from pollution. Once nonbiological intelligence is implementable in the human brain, our own intelligence will grow as fast as machine intelligence. Of course, this means that ultimately, almost all our intelligence will be artificial. Another important aspect of the Singularity is that machine intelligence will be able to understand emotions (or emotional intelligence). By gradually replacing our experience of the real world with that of virtual reality, our environment and even other people will have the ability to appear just like you want them to appear. And you wouldn't even notice the difference.

The next step is looking beyond the Singularity. Kurzweil predicts that intelligence will saturate matter and energy in our vicinity of the universe. By circumventing the supposed limit of the speed of light, our intelligence could stretch literally everywhere. By saturation, he means that all the matter and energy in the system will be put to use of computation of some sort. When we've completely saturated the entire universe with our intelligence, we will supposedly be able to determine our own fate. For Kurzweil, this means that we can somehow alter the laws of the universe and use that freedom for even more computational capacity.

Kurzweil holds the belief that the ultimate goal or purpose of the universe is to gain in intelligence, to saturate all the matter and energy in the universe with intelligence. This is also the goal of evolutionary processes, including both biological evolution and technological evolution. The step (or jump) from biological intelligence to machine intelligence seems to be the biggest and most challenging one, since it occurs right before the Singularity. After machine intelligence achieves human levels of reasoning and cognitive capability without the biological limitations involved, this intelligence will simply take over the process of evolution and automatically follow Kurzweil's prediction of machines enhancing their own intelligence, creating a loop that turns ever faster and faster. By this time, Kurzweil's long term view of how we are going to colonize (or utilize) the universe seems to become the necessary future to which there is no alternative.

---

3 The term 'transhuman' was coined by futurist and science-fiction author FM-2030 who used it as a shorthand for 'transitory human'. His view of a transhuman, though inspired by fiction, is to a high degree comparable to Kurzweil's view, since both look at prostheses, reconstructive surgery and intensive use of telecommunications.

1.3 Philosophical criticism of Kurzweil – Searle's Chinese room argument

Throughout his book, Kurzweil touches upon several philosophically interesting aspects of human nature and machine intelligence. After every chapter in the book, there is a dialogue between several characters, most prominently Ray Kurzweil himself, Molly from the year 2004, George from the year 2048 and Molly from the year 2104. The other characters are scientists and thinkers such as Charles Darwin, Eric Drexler and Sigmund Freud. These dialogues are written as clarification of what has been discussed in the book's chapters, but there are also some philosophically relevant issues that are discussed in the dialogues. To distil Kurzweil's philosophical opinions from the book, I have taken these dialogues into consideration and, while most of it seems to be rhetoric, I will give an account of what philosophical happenings might be going on all throughout his text. Another noteworthy part of the book is where Kurzweil replies to his critics.[4] Out of all the philosophical perspectives on human nature and artificial intelligence, he found John Searle's Chinese room experiment to be the most fitting for discussion in his book (Searle, 1980). So along with Kurzweil's general philosophical opinions, I will particularly look at his notion of the Chinese room experiment in light of Searle's original text.

- Why humans will not be replaced by artificial intelligences: AI learns a lot from its biological heritage, so it's useful to keep biological entities alive~. Also, we're already preserving biological heritage for no useful purpose, Kurzweil says that future enhanced humans and AIs will want the same thing.
- Why humans will not become pets: enhanced humans will be at least as smart as the AI~
- Why human nature will not dissolve: Kurzweil's view on human nature is something like 'being intelligent', not 'being biologically intelligent'
- The problem of teaching AI: it's actually not that hard at all
- Kurzweil compares feeling someone's touch with a computer process – no discernable difference between virtual and actual reality (same will go for AI)
- For Kurzweil, a discussion of consciousness inevitably veers off into conversations about psychology, behavior or neurology. He quotes Leibniz saying that consciousness cannot be found in the brain, would one walk through it as if it were the size of a mill. This

4 In this 56 page long chapter, he discusses several criticisms in detail, most of which are practical, but some are philosophical of nature.

indicates that Kurzweil's concept of an intelligence, whether natural or artificial, is not localized in the brain.

The question of what is so special about human intelligence doesn't even seems to cross Kurzweil's mind at all. Given the impact and the relevant consequences of the Singularity, Kurzweil provides us with a solution that may seem all too simple, but is still worth looking at. When the Singularity hits, the distinction between humans and AIs will disappear over time. Because there will be no difference between humans and AIs, it is not questioned whether or not AI will be conscious. In other words, there will be one single conscious form of entity, with a biological portion of intelligence and a nonbiological portion of intelligence.[5]

Kurzweil says that there exists no objective test that can conclusively determine the presence of consciousness. Any such test would have to have philosophical assumptions built into it. But still, he would not deny that we do have consciousness and moreover, that it's worth something. The respect for consciousness is something that post-Singularity humans and their AIs will have to sustain. Kurzweil's strongest argument is that observers seem to refuse to see consciousness in a system "unless it squirts neurotransmitters, is based on DNA-guided protein synthesis, or has some other specific biologically human attribute."[6] With this reasoning, he already accepts the idea that animals are conscious, along with all other biological organisms. But his point is that limiting consciousness to biological entities seems like an arbitrary boundary. On the other hand, he doesn't want to go as far as saying that if a system appears to behave like a conscious being, it is a conscious being. All that Kurzweil can say about a machine's behavior is that it will definitely appear to be emotional, intelligent and thus human, but nothing more than appearance.[7] But of course, he also mentions that we have no reason to say anything more than that about human beings either. With this line of arguments, Kurzweil's opinion isn't as strong as you could think it would be. A defensible claim of artificial intelligence one-on-one identify with biological intelligence is left out of the discussion. The reason for this

---

5 Kurzweil remarks that in a short amount of time, almost all of human intelligence will be nonbiological, because of the relative limitations of biological intelligence.

6 *The Singularity is Near*, p. 378.

7 The bridge from consciousness to intelligence can be understood clearly and without conceptual problems. Kurzweil talks about consciousness as the human attribute that is worth preserving. It comprises (or at least underlies) cognitive abilities such as creativity, adaptivity and emotion. These features remind us of intelligence in general, which is something artificial intelligence research attempt to emulate in computer programs.

seems to be that Kurzweil does indeed think that artificial intelligence and biological intelligence are radically different on a practical level and this has to have at least some implications for what is special about biological intelligence. While he clearly states that eventually there will be no fundamental distinction between any forms of intelligence, he does seem to realize there is a gap between biological and artificial intelligence.

This gap that Kurzweil hints at is, although in a different manner, expressed by Searle. In the 'Response to criticism' section of his book, Kurzweil replies to a number of criticisms. Most prominently is his response to Searle's criticism stemming from his Chinese room thought experiment. The thought experiment is the following: Suppose I am in a room that is closed off from the outside world. I am given sets of Chinese symbols and sets of rules for manipulating those symbols. With these tools, I can produce output for the experimenters outside the room. For the experimenters, my producing of Chinese answers to their questions is indistinguishable from a native Chinese speaker. The conclusion that Searle draws is that, while doing a computing task (manipulating symbols), I can convince the experimenters that I am a native Chinese speaker, while in fact I don't understand any Chinese at all. Therefore, understanding cannot be reduced to symbol manipulation. In other words, meaning is not reducible to translation; the person in the room replaces one mentioned expression with another and in neither cases uses the expression. Kurzweil quotes a couple of sentences from Searle's text regarding the brain-machine comparison and the notion that consciousness is essentially biological. With these statements, Kurzweil intends to describe Searle himself as a reductionist, saying that he believes brains are essentially biological machines that produce consciousness, but this producing is limited by neurobiological processes. But ultimately, Kurzweil says, Searle's Chinese room argument is an argument against formalism; formal symbol manipulation does nothing to somehow produce understanding. Kurzweil actually accepts this conclusion, but makes the argument that computing isn't the only thing that machines can do. According to Kurzweil, computing processes are capable of being "chaotic, unpredictable, messy, tentative and emergent".[8] And these qualities, among others, are extremely important or even necessary for the emergence of strong AI.

In his original text, Searle discusses a reply from the view of integration of various systems. According to this view, it's not the person inside the Chinese room that has

8 *The Singularity is Near,* pp. 460.

understanding of the Chinese text, but rather the entirety of the Chinese room, meaning Searle himself along with the syntax rules as well as the output rules. Searle's main argument against this reply is that the understanding of English has nothing to do with the understanding of Chinese; the two systems have nothing to do with each other. Partially, this explains why the person inside the Chinese room has no understanding of Chinese, but moreover, it shows that the conjunction of a person plus a set of rules and the execution of those rules does not constitute understanding: translation in the sense of substituting one meaningful expression for another is not understanding. This is however a point where critics of Searle adduce the argument of integration. True integration of the Chinese language system requires a person to link the meaningless squiggles of the Chinese language with representations of what those squiggles refer to, whether internal or external representations. This systems based reply is therefore misunderstood by Searle, because if indeed the English and Chinese subsystems are in no way connected, then there is no unifying system involved and understanding can be on either of the levels independently. It's not simply that the subsystems are thrown together to form a complete system that is able to give meaning to its world, but rather the way how the subsystems are linked that provides this ability of understanding.

Nonetheless, Kurzweil would agree with Searle about rejecting the systems reply, but on different grounds. Kurzweil would say that these representations that are necessary for understanding in an artificial system require the system to be embedded in an environment. It wouldn't matter for this environment to be real or virtual, since we can simply present the computer-brain with the necessary stimuli it needs to be embedded in an environment. The only thing that embeddedness in the real world would imply is embodiment, which means the Chinese room would be somehow implemented in a robot with physical features such as arms, legs and a brain much like ours. But again, these features could also be programmed, in which case the system would appear to itself as if it were embodied, and it wouldn't know it's actually not embodied. The embeddedness would automatically add cognitive abilities such as perception and spatio-temporal awareness, but perhaps also higher-order cognitive functions such as object manipulation or simply moving around in an environment. Searle replies that even then, the homunculus inside the Chinese room would still not understand Chinese. This argument seems to be inescapable for AI researchers, since even if you suppose that the homunculus could see the images through the prosthetic eyes of the robot, it would turn out to be much like the situation in

the classic Cartesian theatre. Of course, Searle's argument is not that strong in the sense that he only tries to criticize the view that a computer program *alone* could account for understanding or explain how the mind works.

In the first part of his original text, Searle mentions that the argument of the Chinese room is an argument against strong AI. And the whole idea of strong AI is that reverse engineering the brain is not necessary to emulate human consciousness in a nonbiological system. In this sense, Kurzweil would indeed agree with Searle, since he puts so much emphasis on reverse engineering the brain and also the supposed collaboration between biological and nonbiological intelligence, to implement the best of both worlds. However, there is a still a major difference between Searle's and Kurzweil's views, since Searle is convinced that "[i]f we had to know how the brain worked to do AI, we wouldn't bother with AI."[9] This is the opposite of what Kurzweil has in mind when he mentions that because we want our AI to be at least as smart as we are, we have to learn from biological intelligence how to do it. Also, in a way that represents our never-ending interest in our biological heritage, the reverse is true, meaning that Kurzweil is convinced that reverse engineering the brain and thus also research into principles of AI will provide us with deep insights into the workings on the mind.

Both Searle's text and Kurzweil's discussion of the Chinese room argument are compelling in their way of dealing with the problem. Coming from Searle's arguments in his article, the question seems to be whether machine intelligence can take over and improve human tasks such as science, while reverse engineering the brain hasn't been completed. In other words, can autonomy, creativity and adaptability be instantiated by purely formal systems? And consequently, will these systems be conscious as a result, as a necessary condition or not conscious at all? For Searle, it's a definite no.

In essence, this is also the conclusion that Kurzweil reaches, when he says that of course formal symbol manipulation isn't enough for understanding to emerge, but computers are (and especially will be) much more capable than what we're familiar with in the present day. Because Kurzweil is convinced that computers will have these capabilities, there is no doubt for him that they will have understanding that is exactly like human understanding.[10] The key difference

---

9 Searle, 1980, Minds Brains and Programs, *The Behavioral and Brain Sciences*, vol. 3, pp. 417-24

10 Kurzweil doesn't say anything about the presumed consciousness of AI, except that we don't know if a perfectly Chinese computer will conscious and, moreover, the Chinese room argument doesn't tell us anything about that.

between Kurzweil and Searle is that Kurzweil doesn't limit intelligence to biological systems. His main argument for this is that, as Searle admits, the brain is a machine and "there is nothing that prevents these biological processes from being reverse engineered and replicated in nonbiological entities."[11] Kurzweil's other arguments are that mere symbol manipulation would not be sufficient for the Chinese room to pass the Turing test. Therefore, Searle's premise that the Chinese room would pass the Turing test already implies that there is understanding somewhere or at some level in the system. Accepting this, the whole rest of the argument becomes obsolete, at least in trying to prove that the person in the Chinese room doesn't understand Chinese.

But of course, for Searle, the premise of passing the Turing test doesn't imply understanding, but at best, it implies that the person inside the room (or the machine, computer) *seems or appears* to understand Chinese. In this sense, it becomes clear that Searle doesn't want to disprove a computer's capability of handling Chinese language, because it certainly can, but rather that there is no sufficient artificial system that understands Chinese. Kurzweil's main response to this conclusion is that it would not be able to convince people of its capacity to speak Chinese with just the symbol manipulation. The striking conclusion is that Searle seems to overestimate the power of symbol manipulation, while Kurzweil, being the radical futurist who has big things in mind for computers and AI specifically as he is, holds a more philosophically conservative opinion in that more human-like mechanisms are needed to simulate intelligence, but that this is well within reach of our scientific research into both the brain and computing. So through his discussion of Searle's Chinese room thought experiment and Kurzweil's comments on this article, we have come to a more refined view of human intelligence, always keeping in mind the importance of our biological heritage to our post-Singularity future.

This is the general outline of the philosophical issues raised in *The Singularity is Near*. Kurzweil emphasizes how we're likely not to understand the Singularity, but that is exactly the nature of the Singularity. And because the role of machine intelligence is crucial, we have to be careful when we consider what the artificial intelligence prospect will bring us. Just now, we have come to some basic understanding of what artificial intelligence is supposed to be. There are some philosophical matters that Kurzweil addresses in his book, but in my opinion he only covers a

11 *The Singularity is Near,* p. 461.

small portion of what is at stake when we look at machine intelligence or the Singularity in general. For this reason, it seems to me that Kurzweil isn't thorough enough in his philosophical considerations. I'm convinced that he does indeed think about these questions a lot, because, for instance, he frequently mentions that we shouldn't be afraid of losing our human nature, but rather, our true human nature will emerge from the Singularity. But these convictions seem to be based on belief, and the underlying concept of artificial intelligence already seems to imply the contrary. The next chapter will dig deeper into Kurzweil's philosophical views on artificial intelligence, as well as introduce some new aspects of his view.

Chapter 2: Philosophical approaches to machine intelligence

Machine intelligence is a necessary requirement for the emergence of the Singularity. To argue about the possibility of the Singularity as a futurist view of our world is still very much like looking into a crystal ball. However, because Kurzweil has such a detailed and thorough description of how this future will unfold, we can conclude that somewhere along the line, artificial intelligence has to happen. So while we can't really question the Singularity's possibility, we can question its legitimacy by looking closer at what artificial intelligence could be and, in Kurzweil's view, would be. In this way, we can gain some insight into the legitimacy of an event such as the Singularity.

2.1 Dretske on the possibility of artificial intelligence

Fred Dretske has argued that "strong" artificial intelligence is not possible. He starts out by making a distinction between two ways of looking at intelligence using an economic analogy. The one way of look at intelligence can be compared to money, which is something that everyone has, but some people have it more than others. The other way of seeing intelligence he compares with wealth, which is something that some people have, in view of the average amount of money. Dretske himself seems to prefer the former view, which is shown by his story on guppy intelligence. He argues that if a guppy swims to the other side of the pool in order to get

food, he is intelligent in the money sense of the word. It is intelligent because it is doing something because it thinks it is a way of getting what it wants and any organism whose behavior is governed by thought is intelligence. This is the basic framework for Dretske's view on intelligence. He doesn't want to think of a Turing Test or a similar comparative test to which artificial systems could be subjected. Rather, Dretske ultimately wants to find out whether or not a machine can have an IQ above zero, which, for him, means that its behavior is governed by its thought. But there is more to intelligence than mere thought, so he goes on to refine his basic 'behavior governed by thought' conception.

Dretske goes on to explain his first addition to the idea of behavior governed by thought. In contrast to reflexive behavior (for example, running away when you see a lion), intelligent behavior must be explicable by thought. While you can think and realize there is a lion in front of you, as well as think that it would be a useful thing to run away, your reflex of running away is not explained by that thought. There is no explanatory relation between the thinking and the doing. The thinking has to explain the doing, which is not the case with reflexes. The next step in Dretske's argument starts off with a hypothetical system programmed to recharge its battery given certain conditions such as the battery being low or there's a recharging facility nearby. When the system has a low battery and it goes to find a recharging facility, the system's behavior is governed by what the systems thinks. Also, the system's behavior is explained by that thought, because that's the way it's programmed. However, what is lacking for intelligent behavior is, according to Dretske, the system does not behave the way it does because of *what* it thinks. In other words, the whole thinking and doing is programmed, which makes it controlled and explicable, but not intelligent. Intelligence is dependent on the content of thought, as opposed to the vehicle of thought. Seeing as most of today's artificial intelligences use preprogrammed routines like the one Dretske describes, he would not see these systems, however intricate and complex they may be, as intelligent.

Coming back to Dretske's argument, he then goes on to add further conditions to his idea of intelligence. The next condition is that of purposefulness. Not only is it necessary for intelligent behavior to be governed by the content of the thought, this content also needs to be purposive. The intelligence of an action depends on whether the content of thought[12] serves some kind of purpose in the required response to some situation. Dretske considers the difference

12 Or, outside of Dretske's distinction between the content and vehicle of thought, simply 'the thought'.

between a plant that changes its color from red to white to increase its chance of survival and exactly such a plant, except this one evolved to change its color from red to white because whiteness attracts pollinators in the summer. While the two plants show exactly the same behavior, the purposes behind their behavior are quite different. Dretske uses this argument to show that explaining behavior requires both internal physical features and external environmental features.

The final condition that Dretske adds to his idea of intelligence is that the governance of thought over behavior has to be learned. To him, part of the explanation of a system's behavior is how that system came to his thoughts and knowledge. He mentions that representations are beliefs, the content of which has to be relevant to the system's behavior. Dretske engages in another thought experiment about a hypothetical artificially created entity named Buster, who lives in a world with other entities called furms. Furms are furry worms that can sting. Prior to encountering a furm, Buster is familiar with the concepts of being furry and being a worm. Naturally, when he first sees a furm, he concludes that there are furry worms. But when Buster approaches a furry worm, he gets stung and never again after that does he dare to come close to a furry worm. This behavior cannot be explained by saying that Buster saw a furm, because he doesn't know what a furm is. It is also not enough to say that Busted avoids furms because they are furry worms. For Dretske, the complete explanation of this behavior is to say that Buster avoids furms because he thinks what he sees is a furry stinging worm and he does not want to get stung. Buster has gained an internal representation of furms as furms (potential stingers).

Because intelligent beings are able to make representations of what they've previously experienced, they can act accordingly. Dretske argues that Buster is intelligent because his behavior is governed by internal representations that depict the object of thought in a behavior relevant way. He acquired these representations by his encounter with furms, so his belief is learned. This seems to be Dretske's key argument in saying that artificial intelligence is impossible. While we can make a Fake Buster that exhibits the same behavior as real Buster, Fake Buster has no explanatory relationship between its representations and the behavior he exhibits. Fake Buster lacks the history that real Buster has.

There are a few issues I have regarding the notion of the 'content of thought' argument. The point of the content of thought as raised by Dretske seems strange to me; if there is a difference between thought with content and thought without content, what could the latter one

mean? What is thought without content? Indeed, Dretske's distinction between the content and vehicle of thought is what constitutes the problem above. What could the vehicle of thought be and why would its content be anything different? An 'empty' thought would not control or trigger behavior any differently than a reflex would. So there is really no such thing as an empty thought or a vehicle of thought, because there is no connection with behavior. I agree with Dretske that it's the content that has effect on a system's behavior, but because thought is always about something, it makes no difference whether you're speaking of thought or the content of thought. In a general way, one could say that a thought always has content.

In Dretske's view, being intelligent comes down to have the right kind of history: a learned history. At the end of his article, he simply states that it is impossible for an artificial system to have such a history. Dretske says that if a system were to have the right kind of history, it is simply no longer artificial. This contrasts with Kurzweil's notion of intelligence, which is that the ability to learn and thus the ability to create the right kind of history is a realistic and even necessary part of current and future AI systems. Kurzweil presents several ways for a program or system to be autonomous, adaptive and creative; machines or applications that will show a great deal of the intelligence that Dretske describes. But because it's a principled matter for Dretske, he would probably continue to argue that artificial history is just not the right kind of history and can never amount to actual intelligence. While most of Dretske's conditions added to the 'behavior governed by thought' idea do seem important to me for the possibility of AI, this final condition of learning seems too strict. Accepting this condition, the only intelligent beings could be biological organisms that took billions of years to evolve into complexity. For Kurzweil, however, this is exactly where technology can, as it were, take a shortcut and circumvent phylogenetic as well as ontogenetic development. Conclusively, both Kurzweil and Dretske would agree that adaptive and creative ability are necessary for the emergence of AI, but they differ on whether or not it is possible to program these cognitive capacities in an artificial system.

While stating at the beginning of his article that intelligence is like money, gradual, Dretske ends up with such a strict definition of intelligence that most human behavior would end up not satisfying his conditions for having intelligence. Some human behavior might be considered intuitive or reflex driven, which may not seem like intelligent behavior, but is actually a bit part of how we would judge an agent to be intelligent or not. He emphasizes that AI

may become quite intelligent and to us, their behavior will look very intelligent and rational. But even then, AI would simply not be intelligent *enough*. In that sense, he seems to end up with a wealth-kind of view on intelligence, rather than the money-view. This shows that indeed intelligence is like money, but some people add limits or have a comparative view of money, and that is when the wealth-view of intelligence arises. Both those views share the same basics (intelligence is like money), but one sees this to be enough and the other finds it necessary to add a limit. To me, it seems that Kurzweil naturally has an open mind towards intelligence and does not limit his view of intelligence to organisms, because he is a defender of strong AI. But he does not engage in the same kind of philosophical process as Dretske does, attempting to disprove the possibility of AI. The most important part of the hypothetical discussion between Dretske's and Kurzweil's views is the emphasis on the learned history of an artificially intelligent agent. I agree with the importance of this aspect of AI, but there are still some problems regarding this learned history. In the next part, these problems as well as some others will be discussed by utilizing Dennett's knowledge on consciousness, specifically in the field of artificial intelligence.

2.2 Dennett's book *Brainchildren*

In 1998, Daniel Dennett published a book of selected essays and articles on the relation between artificial intelligence and philosophy, called *Brainchildren* (1998). In this compilation, Dennett addresses a variety of practical and conceptual problems concerning AI, all of which relevant to the philosophy of mind. Topics include pattern recognition, the frame problem, supposed consciousness of robots and logic. From a selection of these essays, I have taken the points made by both Dennett that are relevant to the problems posed by Kurzweil, Dretske and Searle regarding understanding and intelligence in artificial systems.

In the essay *The Practical Requirements for Making a Conscious Robot*, Dennett shares his experience of working on the Cog project together with the MIT AI lab.[13] Before actually discussing Cog itself, he looks at several reasons for the impossibility of a conscious robot. After quickly dismissing any form of old-fashioned dualism, he turns to the argument that robots are by definition inorganic, and consciousness can only exist in organic brains. In other words, an

13 Cog was a project at the MIT, concerned with making a humanoid robot based on the hypothesis that human-level intelligence requires human-like interaction. See groups.csail.mit.edu/lbr/humanoid-robotics-group/cog.

artificial system cannot comprehend or understand anything, because this is exclusively the property of organic brains. This pretty much reflects the view of Searle, as described in his previously discussed article *Minds, Brains and Programs.*

To counter this argument, Dennett remarks that should there be a contest of making a conscious robot; would it allow the use of artificially constructed polymer organic muscle tissue? If anything, Dennett says, no AI researcher would abide by such a rule; in other words, of course it would be allowed. Another argument discussed by Dennett is the supposed condition of intelligence that it should be born naturally (aptly dubbed 'origin chauvinism' by Dennett). In other words, robots are artifacts, and consciousness abhors an artifact. This may seem like it is simply the same as Searle's argument that understanding can only exist in an organic brain. But there is more to this stance on what constitutes intelligence in a brain. The naturally born brain isn't special because it's biological, but because it has a history. This is more or less what Dretske concluded in his article *Can Intelligence be Artificial?*, when he says that an artificial system has to have a history of learned abilities and knowledge. Dretske defends the position that this history cannot be programmed and must be learned. Moreover, when a history of a system is learned, the system is no longer artificial, according to Dretske.

Dennett seems to have this position into account, but gives it a more practical interpretation. For Dennett, robot infancy is important, not because it provides some kind of "mystic stamp of approval on [the] product"[14], but because it's simply too much work to hard-code every little detail about the world in an artificial system. Such a stage of robotic learning would be meaningful or even necessary for an artificial system to achieve a human-level of intelligence. So while Dennett emphasizes the importance of robot infancy, he dismisses the practical importance of a mystic stamp of history. On the other hand, he recognizes the extra quality that this stamp could mean to people in general. He illustrates this by means of taking a look at the origin of a film. If you use a computer to recreate exactly the pixels in every frame of *Schindler's List*, the result would be exactly the same as the actual film, but the history (or production) of the film would not involve actors, set builders and a director. It would leave out an intrinsic quality of the actual film. But this is no problem for roboticists, because even if they manage to recreate a crude kind of intelligence or understanding, they still win. Comparably, an

14 Dennett, *Brainchildren*, 1998, pp. 175.

animated film can real, or even good. In that same sense, an AI system can have that same quality, even if it's limited compared to biological intelligence.

Looking at Dennett's discussion of these arguments against strong AI, he appears to be optimistic on the possibility of AI. However, he is certainly not nearly as optimistic as Kurzweil . Dennett clearly mentions that it's conceivable that reproducing the speed and compactness of the brain's biochemically engineered processes is not possible in other physical media than the brain. This is the exact opposite of Kurzweil's view, who says that the biochemical firing of neurons is vastly limited compared to the ever increasing speed, capacity and compactness of nonbiological computational processes. So instead of nonbiological computation being AI's biggest weakness, for Kurzweil, it is its greatest virtue. This view is still compatible with Dennett's optimistic 'recommendations' to AI researches, saying that they should not be hesitant to look at biology's evolved solutions and fixes, for instance, artificial polymer muscle tissue. This is something that Kurzweil independently addresses all throughout *The Singularity is Near.*

In another essay from *Brainchildren*, called *When Philosophers Encounter Artificial Intelligence*, Dennett briefly explains a functionalist view of the mind. He calls this the mind as gadget. First of all, the mind as gadget is a refutation of the idea that the mind is something that is governed by "deep" mathematical laws. Nevertheless, the mind as gadget is designed, not by a perfect and ideally rational engineer, but by "natural selection, which is a tinker."[15] Most importantly, the mind as gadget is analyzable in functional terms: ends and means, costs and benefits, elegant solutions on the one hand, and on the other, shortcuts and cheap ad hoc fixes. Dennett argues that while nature is full of these brilliant solutions, philosophers have been hesitant to accept its implications, hammering on their aprioristic methods of investigating the mind. In accordance with Kurzweil, Dennett claims that there is only one way to investigate these evolutionary biological functions, and that is by the empirical mind-set of reverse engineering.

This concept of reverse engineering, which is crucial in Kurzweil's view of artificial intelligence and the Singularity, is further explored by Dennett in his essay *Cognitive Science as Reverse Engineering*, also from *Brainchildren*. In general, Dennett explains, reverse engineering is just what the term implies: the interpretation of an already existing artifact by an analysis of the design considerations that must have governed its creation. Dennet first considers reverse

15 Dennet, *Brainchildren*, 1998, pp. 269.

engineering of human artifacts, such as an electronics company using reverse engineering to figure out how and more importantly why their rival company put together a certain device. Dennett mentions that even though designs are never optimal because designers sometimes put something stupid or pointless in their design, optimality must still be the default assumption. If reverse engineers can't assume any good reason for the implementation of the features they encounter, they have nowhere to begin. But, Dennett remarks, this idea of reverse engineering has one downside, namely that it is top-down. This means that reverse engineers give reasons for certain functions to have been implemented in the system and that this function has been executed optimally by its machinery. This is an over-idealization, especially in the case of the evolutionary history of life. In other words, this is simply not the way Mother Nature designs systems. There are no top-down goals set in the evolutionary process. Problems are not formulated, proposed and confronted with several solutions, one of them being the optimal one. However, again, just because there is a difference in the design processes, it doesn't mean that reverse engineering is less applicable. So while the design processes have been different, their product remains of the same sort and the reverse process of functional analysis is equal regarding both sorts of product. Dennett even goes as far as saying that the whole of biology is the reverse engineering of natural systems.

By focusing on reverse engineering the brain, this view is already a step ahead of Kurzweil, who reduces reverse engineering to simply working out how the brain works in increasingly smaller and thus simpler levels, not looking at what considerations governed its creation. For Kurzweil these considerations are not important; Mother Nature did her job well, and we're going to copy her without looking at why the brain came to be as it is. On the other hand, Kurzweil's opinion is also that AI researches should look at the reverse engineering of the brain to learn from and copy its cognitive functions, not simply doing what Mother Nature herself did. In this way, the reverse engineering of the brain, which Kurzweil values as important for AI, isn't completely bottom-up, because there's also top-down modeling which filters out whatever processes are obsolete for achieving their goals.

This brings back an important topic in the discussion between Kurzweil and Searle, which is the notion that there seems to be a difference in the way the two thinkers look at artificial intelligence concerning its emergence from current technology. Kurzweil often emphasizes the importance of evolutionary algorithms, adaptability and learning capabilities in

AI systems. These cognitive abilities and functions are strongly linked with his conviction that computers are not only formal symbol manipulators, but can also work with chaotic, tentative and probabilistic processes. These processes are really the key behind the workings of adaptability and evolutionary algorithms, according to Kurzweil. But because this part of his argument is so important, Kurzweil may prove to have put himself in a dangerous position, betting all his money on this one option. Still, his progressive look at artificial intelligence systems will still be more useful than Searle's dismissal of even the slightest understanding in such systems.

Dennett seems to take Kurzweil's side on this debate by making an interesting remark about how we will probably look at AI systems in the future. He says that it is very likely that after an AI system has reached the right kind of level of intelligence, we will simply ask it about its internal states, much like we do with other humans, rather than looking at the internal wiring of the system. Its own pronouncements might really be much more informative and certainly easier to access than performing a CT scan on the machine. Also, in contrast to both Kurzweil's and Searle's view, Dennett makes clear that embodiment can indeed make a difference for AI, but not because of any principled reason. Embodiment is of importance, not because it adds something that simulations can't, but because unless you confront the problem of giving a system a body to act in an environment, you might overlook or misconstrue fundamental problems of design.

Kurzweil would probably be convinced of this, and might even be ready to add another practical reason for the importance of embodiment. It might just turn out to be much simpler to give a body to an AI system to let it interact with our real environment, rather than to give it a virtual body in a virtual environment. Why create another reality when we have a perfectly good one right here? And what better place for a robot in infancy to learn about the world than the actual world it wants to learn about? Dennett envisions that robot infancy will be much like human infancy. A human child has a genetically imprinted basic module for language. And before it is even born, it's already exposed to language. This process continues through childhood and, moderately, also in adulthood. A robot or AI system capable of comprehending language would have work on these same principles; using a language acquisition device to learn the ability to understand and speak a language by interacting with a linguistic environment[16].

16 Naom Chomsky's concept of a language acquisition device is the idea that humans are born with the innate facility for acquiring language (not language itself). Chomsky, 1965, *Aspects of the Theory of Syntax*, pp. 32.

Immediately, developmental robot psychology thought experiments come to mind, such as using child directed speech to stimulate a robot's language abilities. These practical concerns for AI research give the debate on the possibility of machine intelligence a different, more practical direction. Instead of looking at what should be possible or impossible for AI researchers to achieve, a view on how practical problems will be overcome, fixed and solved will provide insights into what processes underlie cognitive functioning in general, whether they are natural or artificial.

Despite Kurzweil and Dennett agreeing over the importance or even necessity of robot infancy, they still differ in opinion on the strengths of computational media on which these processes must run. Kurzweil says that computational processes are already a whole lot faster than the biochemical firing of neurons. And this quantitative aspect of computation is growing at an exponential rate. Dennett however, mentions that computational processes are just not fast and compact enough to match the biochemical processes. But Kurzweil provides an empirical statistical study in the actual numbers of the speed, capacity and compactness of computation, while Dennett does not.

In light of the discussion of Dretske's and Dennett's ideas about machine intelligence, we've come to see some more specific requirements for the emergence of artificial intelligence. Robot infancy will be a key factor in the development of these systems. Embodiment is probably the way to go if you want an AI to be able to interact with the world. These features bring forth a very human-like view of an artificially intelligent system. Apparently, the more human-like an AI agent is, the closer it gets to core mental processes such as understanding or consciousness. However, there seems to be an emphasis on the early stages of AI development on the one hand, while on the other hand its eventual goal, in Singularity terms, is left out. The next chapter takes a step back to the Singularity to question the human-like concepts on artificial intelligence presented in this chapter.

Chapter 3: Revisiting the Technological Singularity

After confronting Kurzweil's conception of artificial intelligence with various modern and contemporary views, there is still Kurzweil's broader problem of the technological Singularity and what we can discover about what it's like to be transhuman. Personally, Kurzweil is already trying to live out his dream, taking every opportunity he can to prolong his life, for instance with a large and strict diet consisting of dozens of vitamin pills a day. This is still, however, only the beginning. The changes that Kurzweil describes in *The Singularity is Near* are far more profound, as they are all sublevels of a bigger change: that from biological to nonbiological intelligence. This chapter will consist of a brief outlining of Kurzweil's futurist opinion on human nature, followed by a discussion of an essay David Chalmers wrote on an event such as the Singularity. Finally, a critical view of his opinion in light of the previous chapters will be presented.

3.1 Kurzweil on the Singularity

Kurzweil's opinion concerning the nature of humans is really a complex and far-reaching issue. It concerns a different look on evolutionary processes as well as a speculative belief in the goal of these processes. For Kurzweil, evolution is not limited to biological evolution as it has appeared over the last four billion years. There are different stages of evolution, and biological evolution is merely an intermediate part of the whole evolutionary process of the universe. Kurzweil claims that the goal, or at least the general direction, of evolution is an increase in intelligence. More precisely, it's an increase in intelligence relative to the available matter in the universe. So intelligence isn't just the goal of biological evolution, or even just evolutionary processes in general, but really the goal of the whole universe. Kurzweil is convinced that this is the way to interpret the history of the universe. For that reason, he says that humans will be the main supply of intelligence through the colonization of space. Only two issues stand in our way; firstly, it's still possible an alien race is well ahead of the human race and will beat us in the race of space colonization; and secondly, more importantly, in order to reach to every part in the universe, we will somehow have to figure out how to circumvent the speed of light, should this be possible at all. Kurzweil even mentions a few studies that hint on that possibility.

On the whole, Kurzweil holds the opinion that it is only these final cosmologically scaled problems that we have no way of dealing with yet and pose a genuine threat to the idea of a universe saturated with intelligence. For him, the major part of the development towards, during and after the Singularity is a given. This includes artificial intelligence and enhancement of human intelligence. What might be even more remarkable is Kurzweil's opinion that throughout all of this, we will never cease to be human. In his dialogues between Molly, himself and the scientists, he briefly mentions that the experience of being human will never disappear. This includes humanoid themes such as sense of self, love, beauty and art. Kurzweil says that we will always matter to machines intelligence, since we essentially learn from biological processes how to artificially reproduce them. So even our biological part, which will be insignificantly small according to Kurzweil, will not be lost. But even if it does, Kurzweil still holds that being human is not about being biologically intelligent, but simply being intelligent, regardless of what media this intelligence is instantiated in. Another aspect of the Singularity to keep in mind is that the shift from biological to nonbiological intelligence will happen gradually, though exponentially. In the same way humans have in the past adopted new features to their concept of humanity, so it will be for the shift to nonbiological intelligence as well. These arguments allow him to keep a concept of humanity through and beyond the Singularity. Kurzweil's emphasis on the human-like intelligence that AI will most likely possess, would probably get the benefit of the doubt by people like Dennett or even Searle, but is AI really necessary to be like that for such an event to occur?

3.2 Chalmers on the Singularity

In an essay from 2010, Chalmers describes a clear and logical distinction between the explosion of intelligence (an intelligent machine making an even more intelligent machine) and the exponential growth of computing power (Chalmers, 2010). His view of the Singularity is different from Kurzweil's. It's a broader sense of the Singularity: "A loose sense refers to phenomena whereby ever-more-rapid technological change leads to unpredictable consequences."[17] With this view, Chalmers takes the Singularity seriously, as opposed to the

17 Chalmers, *The Singularity: a philosophical analysis*, 2010, pp. 9

general academic consensus. His discussion of some of the topics put Kurzweil's views in a proper philosophical perspective.

Chalmers mentions various reasons why emulating the brain would not produce AI. Even if the brain is a machine, and we can instantiate the physiological processes from the brain in a synthetic framework, the resulting emulation might still not be AI. This is likely to come from the view of AI researchers, who have been shifting the focus of their research from emulating agents in a virtual environment to embodied robot intelligences.[18] Probably the most important reasoning that Chalmers provides is that the Singularity is not dependent on conscious or humanly intelligent machines. As long as an AI can produce itself a better AI, then that is all you need for the explosion of intelligence. In other words: the AIs that will supposedly be responsible for the Singularity to occur may not be able to pass the Turing test very well at all.

Another topic which Chalmers discusses is how this explosion will begin. What system would the emergence of the first AI rely on? Evolved systems seem to have the upper hand on hardcoded programs or neuron-by-neuron emulated agents. Of course, the most attractive option is a combination of evolutionary algorithms, neural networks and a hardcoded core of some database system, along with the ability to learn or to have a history, as Dretske's paper suggests. Also, this is the most likely option for the follow up of the first AI: a more intelligent machine that improves itself. Again, this system does not require having a human level of intelligence, as long as it can increase its intelligence towards the point which it exceeds human level intelligence.

This is contrary to that which Kurzweil has in mind for artificial intelligence and the Singularity. He sees more potential in the enhancement of the human biological brain than Chalmers; according to Kurzweil, our merging of our bodies with technological components is an essential feature of pre- and post- Singularity life. Kurzweil would deny that the explosion of intelligence from the Singularity is different from an exponential growth in computing power. The view that of the Singularity that Kurzweil presents is in my opinion best described by the word 'transhumanism'. All through the book *The Singularity is* Near, Kurzweil keeps reminding the reader that human biological heritage will always remain important. Chalmers is more inclined to a practical and scientific AI system rather than a human-like system, even though he

---

18 See for instance Pfeifer & Scheier, *Understanding Intelligence*, 2001, pp. 91

agrees that the way to achieve an explosion of intelligence relies on naturalistic processes like evolution and the ability to learn. This distinction divides Kurzweil as an idealist on the one hand, and Chalmers as a realist on the other.

Conclusion

"(…), it could turn out that any conscious robot had to be, if not born, at least the beneficiary of a longish period of infancy." (Dennett, 1998, *Brainchildren*, pp. 157)

In the above chapters, we've gone from a basic logical interpretation of artificial intelligence to a more sophisticated and rational view. Firstly, the discussion between Searle and Kurzweil has shown that logic cannot rule out the possibility of artificial intelligence. Through Dretske and Dennett, the importance of an artificially intelligent system's infancy or history is now viewed as a necessity. Finally, the distinction between Chalmers and Kurzweil points to the conclusion that an AI system need not be human-like or organic, but there is still a lot to learn from biological intelligence. But overall, Kurzweil's opinion remains that to be human is to be intelligent, no matter in what form.

Kurzweil's view that being human should be something like being intelligent, rather than biologically intelligent seems fair to me, since principally, it doesn't really matter in which physical media cognitive functions are instantiated. If rocks could simulate the firing of neurons, then on some kind of scale, it could be functional. This view is similar to that of Chalmers, who explains that the classic canonical view of intelligence is outdated, and that it cannot be measured; and more importantly, that this does not matter for the emergence of artificial intelligence. But Kurzweil always repeats the importance of our biological heritage, which to me is something he insists on, because he doesn't want to step on the toes of those people who are afraid that losing our biology means losing our humanity. While Dennett and Kurzweil have different levels of optimism regarding the possibility of AI, they seem to agree on the fact that a sort of robot infancy and a capability of learning are crucial to the emergence of AI. Searle and Dretske would actually not disagree with this, but they would simply dismiss it as a form of AI.

For them, an artificial intelligence system is artificial because it has no history or way of learning, for instance, a language.

Most of the difficulties that thinkers have with artificial intelligence are based on an anthropomorphizing of artificial intelligence systems. Both optimistic and pessimistic views are marked by this classical view of artificial intelligence. Kurzweil's predictions about artificial life in a post-Singularity world suggest that AI systems will be very human-like to us (or at least able to convince us of their supposed humanity). I believe a more conservative opinion is appropriate, since the development of AI is likely to take on a different direction than Kurzweil predicts. Instead of working towards a completely simulated human agent, the artificial systems that first will appear will have a high level of cognitive ability, while not even able to interact with the outside world in the sophisticated manner that humans do. Of course, Kurzweil is correct in mentioning the various techniques and technology that will add to the advance of AI research, such as neural networks, simulation of the human brain and evolutionary algorithms. These concepts all have their own place in AI research. However, the soundest argument against the emergence of AI (what Chalmers calls a *defeater*) is not of practical nature, but rather something concerning the creators' motivation. It is not difficult to imagine that replicating a human is something that we can already do biologically, so we don't have to research AI for that reason. In most cases, however, it is this anthropomorphic character that seems to form the motivation of AI research.[19] This coincides with the classical canonical view of AI that Chalmers mentions in his paper on the Singularity. So with a change of perspective from the classical interpretation to the practical interpretation on the possibility of artificial intelligence above human level intelligence, an analogous change in the direction of the research is likely to occur.

19 For instance, the MIT Cog project started by Brooks has presented a physically humanoid robot.

List of references

Chalmers, D. J. (2010) The singularity: A philosophical analysis, *Journal of Consciousness Studies* **17**, pp. 7–65.

Dennett, D. C. (1997) *Brainchildren: Essays on Designing Minds.* Cambridge, MA: MIT Press.

Dretske, F. (1993) Can intelligence be artificial?, Philosophical Studies, **71**(2), pp. 201–216.

Kurzweil, R. (2005) *The Singularity is Near: when humans transcend biology*, New York: Viking Penguin.

Pfeifer, R. and Scheier, C. (1999) *Understanding* Intelligence. Cambridge, MA: MIT Press.

Searle, J. R. (1980) Minds, brains, and programs, *Behavioral and Brain Sciences*, **3**(3), pp 417–424.