Philosophy of Artificial Intelligence

Çağatay Yildiz - 2009400096

May 26, 2014

Contents

Chapter 1

Introduction

In this chapter, I will first explain what philosophy means and how it involves in every part of our lives, which I think most scientists do not have an idea of. Then, I will provide my reasons to choose such a topic to study and list resources I made use of.

1.1 Philosophy

Philosophy literally means *love of wisdom.* The word originated from the Ancient Greek [1] where *phileo* meaning "to love" and *sophia* meaning "wisdom". First use of this words dates back to 14th century.

1.1.1 Definition of Philosophy

In a broad sense, one can define philosophy as a thought activity seeking to understand fundamental truths about the people, the world in which we all live, and the relationship between people and universe [2]. Mirriam-Webster's Dictionary defines philosophy as *"Critical examination of the rational grounds of our most fundamental beliefs and logical analysis of the basic concepts employed in the expression of such beliefs"* [3]. One definition or another, philosophical investigations aim at getting deeper and deeper in understanding of both natural and man-made parts of the universe by asking question tirelessly.

1.1.2 Philosophy of Anything

A widely influential philosopher-and mathematician- of 20. century, Alfred North Whitehead noted in one of his books that *"The safest general characterization of the European philosophical tradition is that it consists of a series of footnotes to Plato"* [4]. This quotation may seem annoying at the first seen but an underlying interpretation is that there is no real progress in philosophy. What Plato, or you may say Socrates, had claimed 25 centuries ago are still topics of hot debates. Put it simply, this is the nature of philosophy.

Science, on the other hand, is basically constructed upon continuous development. For instance, Einstein has shown that what Newton had stated year ago is incorrect, or insufficient, to explain how universe works. This is also true for technology, which is just an application of advance in science.

For many scientist, philosophical aspects of whatever their field of study are not within the borders of what they should think of. You cannot see many mathematicians questioning what *a set* in reality represents. A set is just a unit used to represent certain things for most scientists. As Moody puts clearly [5], philosophy is the study of *foundational issues* and *questions* in whatever discourse(scientific, literally, religious and so forth). For mathematics, zero, sets or points are very basic concepts and they are usually taken for granted. Similarly, other fields of science incorporate such foundational questions. The reason why science continuously progresses is rooted in the fact that scientists do not spend much time to such foundational issues. This is where philosophy comes into the play: Questioning basic concepts, assumptions or axioms of the field of study.

1.2 What Philosophy of AI is Interested in

Philosophy of artificial intelligence is a field of study that is concerned with the question whether AI is possible or not. Put it another way, if it is possible to build an intelligent machine that can think is the main topic of interest. In addition, unknowns such as the nature of rationality, the power of human mind and what kind of features a thinking machine should have are investigated [6]. Of course, this list is not exhaustive but all other topics are related to those in some ways. Here are some fundamental issues studied by by artificial intelligence researchers [7]:

- Can machines think? Can they solve any problem in the same way as human-beings do?
- Can a computer have consciousness, emotions, soul or morality?
- Is human brain basically a computer?
- Is it moral for humans to build a machine that can think? What would its possible outcomes be?

1.3 Why to Study Philosophy of AI

Many people consider *science* as the most basic way of understanding universe. In fact, investigating universe is a common interest for philosophy and science. Throughout the history, philosophy has questioned the world and science has came up with answers. Take Ancient Greek or China, for instance. It is not a coincidence that almost all philosophers of that time were science men.

As artificial intelligence is one of the newest fields of study, people studying AI must be in the guidance of philosophical investigations noting that this has always been the case for other disciplines such as economics, psychology, sociology and so forth. I can further claim that there is not many disciplines whose subjects are as related to the philosophical discussions as those of artificial intelligence are. From another perspective, findings in artificial intelligence may give answers to unsolved philosophical problems. For example many are in hopes of clearing up the mystery of human mind thanks to findings in AI. Finally, I believe that artificial intelligence is the hugest step in history to enlighten secrets of universe as well as mankind.

1.4 This Study

1.4.1 Contents

This report contains several chapters. First chapter is dominantly dedicated to the relation between philosophy and artificial intelligence. In the next section, I have dived into the history of AI and explored some developments that have been subjects of philosophical discussions. In the next section, I have examined mind-body problem since the question whether AI is possible or not is quite related to the question whether we have a *mind* or not. In chapters 4 and 5, I tried to go over Turing Test and Chinese Room Argument. Both of them are very famous arguments and influence very much the way philosophical discussions on AI are evolved.

1.4.2 Resources I Used

Here, you see the list of items that I have used during my research.

- Philosophy and Artificial Intelligence by Todd Moody [5]
- Computing Machinery and Intelligence by Alan Turing [8]
- Minds, Brains and Programs by John Searle [9]
- Stanford University's Encyclopaedia of Philosophy [10]
- Internet Encyclopaedia of Philosophy [11]

The first item is a nice introduction book on philosophical aspects in artificial intelligence. I have read all of it except for one chapter. Second and third items are two well-known papers and I have read them as well. Forth and fifth items are two online philosophy encyclopaedias. I have been consulting them for a couple of years while making philosophy readings. During this research, I have read sections that are related to topics included in my report.

Apart from those, I had a look at following resources when necessary and read some parts of them. These helped me enhance my knowledge of certain topics:

- The Philosophy of Artificial Intelligence / Edited by Margaret A. Boden. [12]
- Artificial Intelligence : A Philosophical Introduction [13]
- On Being a Machine [14]
- The Mind and The Machine : Philosophical Aspects of Artificial Intelligence / Editor, S.B. Torrance. [15]

Chapter 2

Some Philosophical Stages in History of AI

In this section, some developments in the history of artificial intelligence are examined. While listing them here, I tried to include those that eventually became a matter of hot philosophical debates and emphasized philosophical aspects as much as possible. Developments in AI are listed very well in Todd Moody's book. So, I combined what's discussed in Moody's book with internet research and provide a nice summary.

- The first clear example of a computer is a **difference engine**, which is just an automatic machine used mainly for calculating logarithms [16].

- After Difference Engine, "Adding Machines" were built. These were machines that can do nothing but addition. Therefore, it is not quite possible to call them "computers". Also, not much philosophical debate had occurred on whether they really make additions or they just seem to do it. At the end of the day, what they did is basically reading their tape and manipulating tape containing the output.

- The first machine that can be identified as "Electronic Computer" was **Colossus**. It was the first programmable computer and built in 1943. By long and numerous computations, Colossus broke Germans codes during World War II.

- Next was ENIAC (1946), or **Electronic Numerical Integrator And Computer**. Just like Colossus, it was a physically giant object, weighing about fifty tones. It was such a speedy device for multiplication that it used to be called "Giant Brain" [17].
  Both Colossus and ENIAC were used to perform lengthy and numerous arithmetical operations. These machines were quite good at making speedy operations. But this is all about quantity of computer's operations, no intelligence involved. Therefore, even though arithmetic has been traditionally associated with intelligence, they cannot be called "intelligent beings" from artificial intelligence standpoint.

- In 1946, John von Neumann, called *father of modern computer*, showed that a computing machine should be able to perform its operations without being wired by hand, which is a property of Colossus and ENIAC. He suggested today's computer architecture, which depends upon a central processing unit, memory and instruction set. This was a milestone in computer science since what we today call programs started to be written thanks to von Neumann architecture.

- In 1948, the first computer that can play chess was built at MIT. As many offer, this is the starting accomplishment of artificial intelligence. Since then, chess has always been a nice example illustrating AI thinking because it is a quite accessible, simple and valid way to exemplify AI thinking.

- The formal starting point of artificial intelligence can be considered as the conference at Dartmouth College in 1956 where John McCarthy initiated the term "artificial intelligence". Some other attendants of the conference were Marvin Minsky, Allen Newell and Herbert Simon, all of whom became leading figures in AI movement. In a year, the programming language LISP was introduced, which was the first high-level language created specifically for research in artificial intelligence.

- In 1965, ELIZA came into the stage. It was the first well-known example of software that can communicate with a human-being in a natural language and taken as a model for future automated psychotherapists. The founder of ELIZA, Joseph Weizenbaum, on the other hand, thought differently and became one of the first people questioning "the religion of science" [18]: *"... while sentimental people argue that God is love, the tough modern man, or at least the tough modern Western man, knows that God is really intelligence. I hope it is very clear that I totally disagree with this position. It is, however, the dogma of a for-the-moment-victorious "religion" that worships intelligence and its embodiment in the computer. This "religion" pronounces an apocalyptic prophecy. According to this prophecy — which certainly has a basis in reality — the earth's people will one day destroy themselves and their gene pool."*

- In 1970, SHRDLU appeared as another fascinating development. It was simply a program that manipulates blocks according to instructions given by a user. What makes SHRDLU different than all other programs was that it can *communicate* to people in English and answer simple questions in "block-world". To do so, SHRDLU has an internal representation of block-world, a parser to understand natural language and a tool mapping instructions given in natural language to block-world, which were all sophisticated during 1970's.
  While some have claimed that SHRDLU *understands* commands, some other have argued that such a limited system cannot have an understanding at all. They continued that people's understanding of actions taken by SHRDLU can be applied to any domain; however, SHRDLU lacks in such generality. Therefore, we cannot talk about understanding, they say [5].

- Few years later, Roger Schank of Yale University, recognizing SHRDLU's limitations, came up with the concept of script, which are used by computer to make inferences about real world [19]. To illustrate, assuming it had necessary scripts and no missing data, a computer can understand if you like a dress in a store and pay some money at the end, you probably buy the dress. He even devoted a book named *Scripts, Plans, Goals, and Understanding: An Inquiry Into Human Knowledge Structures* to his ideas.
  A criticism of Schank's view is that *real* human thinking is so complex and world knowledge of different areas are so interdependent that script approach can never leads to success. Apart from this criticism, some cognitive scientists believed that human knowledge is made of lower-level units that form scripts(=concepts in brain). If this is the case indeed, what we need in artificial intelligence is not scripts but mechanisms to create scripts.

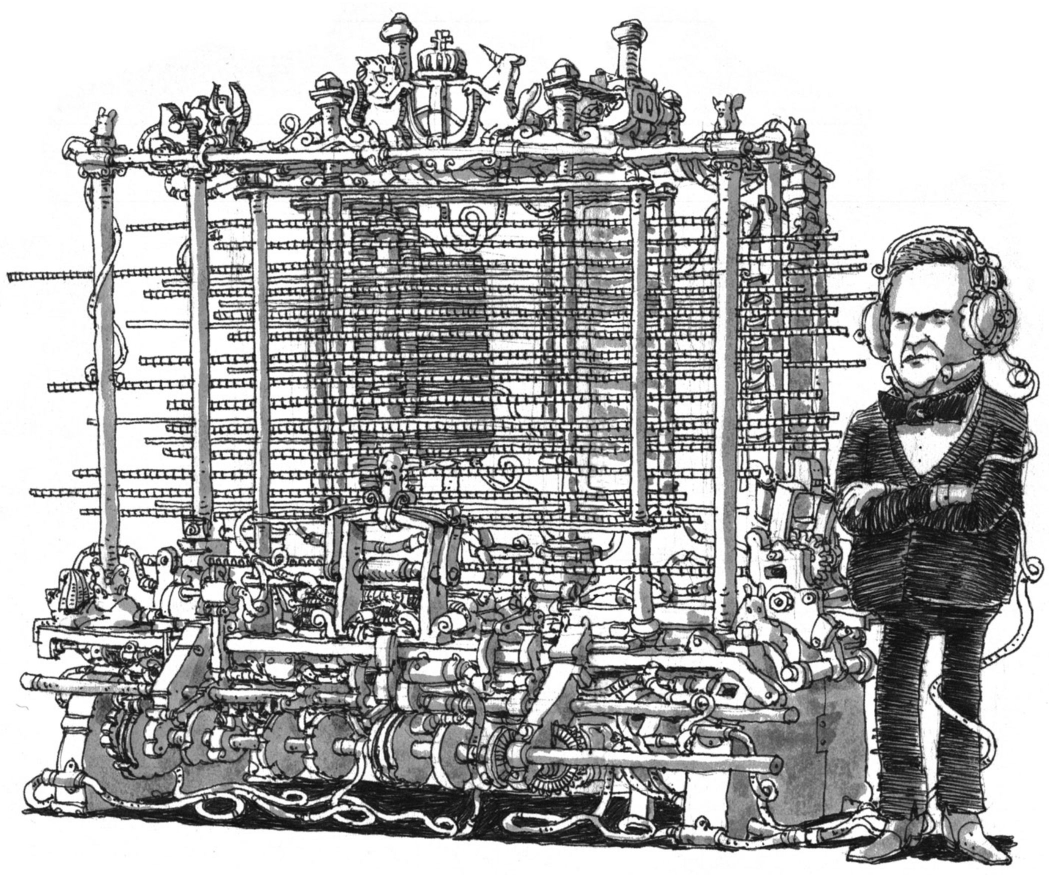

Figure 2.1: An illustration of Charles Babbage and his difference engine [20]

Chapter 3

Mind-Body Problem

3.1 Why to Discuss Mind-Body Problem

Mind-body problem has always been one of the most famous issues in philosophy. Its origins dates back to Ancient Greek and Plato was the first person who systematically argues the problem. The discussion here is simply whether there is an entity called mind that is not physical and what its properties are if it does exist.

What artificial intelligence researchers study is in fact quite related to this problem. On one hand, we have bunch of works in AI that are mainly investigating how human brain works. What's more, computers are devices that look very much like human brain as both have exceptional capabilities. Therefore, findings in AI may serve purposes of philosophers. On the other hand, we have philosophers discussing philosophical aspects of human cognition. They have been trying to identify distinguishing features of human thinking and formalize concepts such as mind, consciousness, soul and so on. At the end, we come up with two disciplines feeding one another.

3.2 Where the Problem is Originated From

Starting from the first man, the mean of expressing the existence of human beings has been the existence of body. If one talks about a person, then there should be a body belonging to that person. This is also true for animals or any other creatures.
The existence of body cannot be reduced to an organ pumping blood to vessels. From the simplest living beings like amoeba to the most complex ones, body is a part of the world of ***perception***. That is, what actions a being can take in the universe is sharply bounded by what it can perceive. A dog can see a flower but it cannot get the joy of all different colors that flower has. This cannot be expected from dogs since they do not have the ability of to *perceive* different colors.

In contrast to all other living beings in the universe, man has the capability of ***thinking***. It can be shortly stated that thinking is the second half of a person's experience, or consciousness. Everything related to thinking starts and ends in so-called *mind* whereas perception is directly resulted from external world stimuli.

Yet, perception and thinking differs in many aspects. First thing first; perception is the product of human body whereas thinking is a much complex process that is traditionally associated with *mind*, which is out of human body. Secondly, one cannot *stop* perceiving at all while not to think is possible. Of course, it is possible to close eyes to stop seeing around; however, this is nothing but avoiding brain from getting input. On the other hand, thinking seems to be very much under conscious control. What and what not to think of is not a tough task. The final distinctness between perception and thinking is that the former can be shared among different individuals but the latter is not. For example, the color of a bird flying in the air can be confirmed by some other people; nevertheless, no one can confirm what you are thinking at the current moment.

In short, the world can be divided into two, an inner mind and an external reality. In fact, this division is inevitable and does make sense. But this is exactly what leads to mind-body problem: Your body is an object in the *external* world. Therefore, your mind, having different properties than your body, cannot simply be your brain, your brain is *external* to your mind. But in this division, what is *you*? Why does your mind belong to you but no other body? What kind of effects do your physical states have on your mental states and vice versa? **What is the relation between your mind and your body?**

3.3 Explanations on the Problem

Platonic Dualism: Plato is the first Western philosopher that has emphasized mind-body problem. In fact, his views on this problem is not a simple set of thoughts but a vital part of his metaphysics. According to him, what we see in the **external world** is only a part of reality and true beings are not just physical objects but eternal Forms, which exist independently of physical world. These objects make up **intelligible world**. To him, Forms are what intellect uses while understanding certain facts in the external world. Since human mind can access knowledge of Forms and Form are not parts of visible world, mind must itself be in the world of abstract objects. So, mind must exist independent of body.
A problem with Plato's view is its lacking in explaining what attaches a mind to a body. In other words, what how *my* mind is mine and yours is yours? His prize student Aristotle did not believe in Platonic Forms that exist independent of visible world. To him, body and soul, or mind, is one and united. However, he also believed that intellect is more than a bodily organ since if it were, it could receive not all Forms but only physical ones, which he thinks is not the case [21].

Cartesian Dualism: Rene Descartes was the one who explains a modern version of dualism. One of two basic methods Descartes has followed while making philosophical investigations was "method of doubt", in which he searches for things that are true themselves, without basing upon some prior knowledge and incorporate no doubt in themselves. At the end, he came up with his very famous saying **I think, therefore I am**. The reasoning here is the following: Everything seems to be possible for one to doubt but one cannot doubt his/her own existence. Doubting require thinking, then there must be someone thinking. Therefore, my existence is guaranteed by me thinking. At the end, my body can be subject of doubting but my mind cannot. Once again, we came up with dualism.
The problem of interaction appears in Cartesian Dualism as well: How can an immaterial mind trigger actions in material body and vice versa? As a solution, Descartes provided that a portion of brain bridged this interaction and called there "pineal gland". But same problem is still at the table: How

come immaterial mind can interact with a material called pineal gland? Proponents of Descartes came up with another explanation, that is, all mind-body interactions is handled by the hand of God. I find it unnecessary to discuss this view here.

Materialism: The opposite view of dualism is materialism. According to this view, there can be no two categories of things. Everything in the world are physical entities and follows rules governed by physics. Therefore, for the materialist, there is nothing called mind; all what human beings done including thinking is the product of bodily actions. There is simply nothing beyond human brain. In fact, materialism is a class of theories and there are many variants of it, which are not discussed here. Critique of materialism has many distinct aspects. From a philosophical perspective, materialism lacks in explaining what is so-called *qualia*, or subjective/conscious experience. Another strong criticism is relatively new and comes from physics. In *The Matter Myth* [22], Paul Davies claims that *"Newton's deterministic machine was replaced by a shadowy and paradoxical conjunction of waves and particles, governed by the laws of chance, rather than the rigid rules of causality. An extension of the quantum theory goes beyond even this; it paints a picture in which solid matter dissolves away, to be replaced by weird excitations and vibrations of invisible field energy."* So, what we call atoms, or matter, may not be "as matter as we think". Therefore, today, materialism is not as strong as once thought.

Chapter 4

Can Machines Think?

Alan Turing's well-known paper titled *Computing Machinery and Intelligence [8]* can be considered as one of the greatest leaps in science since this paper has transformed the way people think of the definition of intelligence, capabilities of computing machines and even the mind-body problem. Therefore, it would be a shame not to go over this paper in my report. In this chapter, I will examine the paper and some objections to it.

4.1 The Imitation Game and Turing Test

As noted at the beginning of the paper, the question "Can machines think?" is quite vague. To discuss machines' thinking capabilities, Turing suggested a structured version of this question. In fact, he created a game called *Imitation Game*. In this game, we have a person, a machine and an interrogator. Each stays in a different room and communication between interrogator and other agents is handled by a teleprinter. The interrogator knows one of them is a machine while the other is a man and refers them as X and Y, not knowing if X is man or machine. By asking whatever question he likes, interrogator tries to identify which agent is machine. The object of machine is to lead the interrogator to mistaken conclusion whereas man tries to help him find the machine.

While many people highly criticize Turing's view, he claimed in contrary that *"... the odds are weighted too heavily against the machine"*. In other words, this task would be much easier if it would be men to be identified. Since men cannot make arithmetical operations as fast and accurate as machines, a simple question can end the game.

By the time this paper was published, fundamentals of current computers were about to be completed, or von Neumann architecture was at the stage. In fact, Turing spent a couple of pages in this paper to explain a structure based on that of von Neumann, which is equivalent of explaining what future computers will look like. When defining the imitation game formally, Turing described the machine that can win the game as "having an adequate storage, suitably increased speed of action and provided with an appropriate programme" [8]. He further claimed that at the end of 20. century, the storage capacity of computers will be about 10^9 units, which he thinks is sufficient to play the imitation game successfully. He also set the limit of success as interrogator's not more than 70 percent right identification at the end of a five minute questioning.

Many variations of this game exist in the literature. A famous and simplified version, called Turing Test, is quite popular. In Turing Test, there is only one player other than the interrogator. The duty of interrogator is to identify what this player is, i.e., a man or a machine. At the later stages of his research, even Turing approached the game as described in this paragraph. The criteria of success has evolved as well. Today, a computer "passes" Turing test if it is identified as human not less often than a real human-being being identified as human-being.

4.2 Objections Replied in the Paper

In this paper, Turing not only unfolded his views but he also handled some possible objections to his path-breaking thoughts. Objections he replied are titled as (1) The Theological Objection; (2) The "Heads in the Sand" Objection; (3) The Mathematical Objection; (4) The Argument from Consciousness; (5) Arguments from Various Disabilities; (6) Lady Lovelace's Objection; (7) Argument from Continuity of the Nervous System; (8) The Argument from Informality of Behavior; and (9) The Argument from Extra-Sensory Perception. Here, I am going to go over some of those as follows:

4.2.1 Theological Objection:

What theologists basically claim is thinking occurs thanks to the soul given by God and soul is granted only to man and woman, no other animal or machine. Although he did not take theological arguments seriously, Turing tried to reply them in theological terms. He questioned the reason why God, considering all He can do, does not unite souls with machines. This possibility simply can never be ruled out. Turing also claimed that he is not very impressed with theological arguments and by exemplifying contradictions between Bible and Galileo on the movements of Sun and Earth, he aimed proving that what is written in Bible is not hundred percent correct.

4.2.2 The Mathematical Objection:

Turing himself was a mathematician and had nice answers to critics from mathematical perspective. Many arguments are based on Gödel's incompleteness theorem and Turing's halting problem, both of which lead to the fact that some questions cannot be answered by the machine in the course of Imitation Game. Turing himself was aware of this and stated that these questions are meaningful only if humans can answer them. That is, if neither machine not man can answer the same question, this very question does not help interrogator to draw conclusions.

4.2.3 Arguments from Various Disabilities:

Without any doubt, one can find things that current machines cannot do. Some examples stated by Turing are to be kind, friendly, have a sense of humour, fall in love, etc. Turing found such claims interesting and believed that people supporting this view probably follow the principle of scientific induction, which has no validity in our case. As an answer to such claims, Turing said that machines that cannot do some of the abilities stated above does not imply those machines lack in intelligence. To him, it would be chauvinistic to expect intelligent machines to have same tastes as we do. Also, he reduced the problem of diverse behaviour to storage capacity limitation. In other words, he anticipated that as storage capacity grows, machines will have many various abilities like those listed above.

4.2.4 Argument from Continuity of the Nervous System:

The nervous system and human mind differs from digital computers in that human beings are not discrete-state machines. A tiny change in neuron impulse may result in huge changes in the action triggered by this impulse. Turing agreed with all these but further claimed that a continuous-state machine can be simulated by a discrete-state machine with very small amount of error. He exemplified that digital computers can perform not much worse than a differential analyser (a continuous-state machine). In other words, interrogator is not expected to realize errors resulted from continuity issue.

4.3 Other Objections

4.3.1 The Turing Test is Too Narrow:

There are many objections saying that Turing Test is not so wide to incorporate all aspects of human mind. According to this view, winning the Imitation Game is something that an intelligent machine can do but there can be other things achieved not by the computer but by human mind. So, machine's abilities are a subset of human mind's capabilities.

A simple answer to this reply is that success in Imitation Game depends on a large variety of abilities. A machine at the stage must have memory, language skills to communicate, huge amount of information and the associations among them and it should also be able to understand rules of games, which is not a simple thing to do even for human-beings. It would be unrealistic to expect that a machine that can play the Imitation Game is unable to perform worse in other tasks. Besides, simply the fact that interrogator asks questions from everyday circumstances shows that machine passing Turing Test has to be able to solve problems in a quite wide variety.

4.3.2 The Turing Test is Too Hard:

There are some people who have claimed that Turing Test is too hard for a computer to pass. They go further and state that there can be no computer that can pass the test if appropriate questions are asked. French is among these people and has delivered quite fascinating questions. He has asserted [23] that human cognition has exceptional properties that can never be replicated by a computer since this computer is required to be able to operate in the exactly same way as low-level structures in out brain operates.

He has given what we call *assertive priming* as an example. Research on priming shows that it is easier for people to identify a series of letters as a word if they are presented this word before. In other words, early exposure to the word makes people pay more attention to the word later. If interrogator is given sufficient data about this research, he can easily distinguish the machine from the person during Turing Test.

The idea here is that there are some aspects of human cognition that are hard to simulate for a computer. The previous example is just a single example of such aspects, researchers probably have not discovered all distinct features of human mind. Therefore, we cannot expect machines to simulate all of them.

Opponents of Turing Test state that this feature of human brain has nothing to do with human intelligence. In other words, testing the machine by this example and looking at whether it has some eyes or arms are essentially the same thing, that is, they are both irrelevant to thinking.

Chapter 5

Chinese Room Argument

In the last chapter, I have examined Alan Turing's attempt to formalize and apply the definition of intelligence. All he was interested in was regarding identifying machines as *intelligent* or not. However, many scientists and philosophers that adhere to his view go even further and broaden the scope of Turing Test. Some even claimed that passing Turing test is equal to having a consciousness, or having ideas, beliefs, thoughts, etc.

This is a quite strong claim. Writings attributing consciousness to agents apart from human beings were not so common back in 1970's; therefore, it is heavily criticized from many perspectives. A simple reply to this argument is to ask whether a computing machine can understand a joke or sarcastic statements. A deeper and well-formulated criticism was done by John Searle in his very well-known paper Minds, Brains and Programs [9]. In his article, Searle propounded a differentiation between two thesis, that are, whether a computer that can pass Turing Test does have a mind or it does not. He called the first definition *strong AI* and the second one *weak AI*. These two terms have been widely discussed ever since.

5.1 Strong and Weak AI

Before going deep into the paper, it would be nice to distinguish what Searle calls "strong" AI from "weak" or "cautious" AI. Here I will be using his terminology as much as possible so that I do not leave the trace of his thoughts.
According to weak AI, computers and programs are simply tools that help us understand human mind. However, strong AI is much more than that. It is claimed by strong AI that appropriately programmed computers are, in fact, minds. In other words, such computers have the ability of understanding. These computers, they say, have cognitive states just like human beings do. Of course, Searle's problem was not with the weak AI thesis but the conclusion that a computer that passes Turing Test essentially have a mind, or strong AI.

5.1.1 Arguments in Favour of Strong AI

Turing Test, or strong AI, is not constructed upon any blurry philosophical discussion. It is quite simple to understand and apply Turing Test. As I have mentioned earlier, diving into philosophical arguments is something that blocks advances in science. In fact, many scientists do not pay attention to how strong Turing's arguments are. They simply try to build a machine that can pass the test.

A more important aspect of Turing Test is that it is very much objective. Rules of the test are quite clear. Causality is just there; interrogator can simply say which answers persuade him to think that machine is not in *this* room but in *that* room. People watching this test can also draw conclusions and these will probably be very similar to those drawn by the interrogator.
In fact, strong AI thesis makes an operational definition of the mind [5]: *A mind is whatever set of functional capabilities that enables a system to behave in ways of characteristics of systems already known to have minds, and those capabilities are detected by the Turing Test.* This definition, just like Turing Test, is very straightforward. It does not take the structure of the *intelligent* system into account. It could be made of billions of neurons and trillions of synapses or millions of tiny circuits constituting greater devices. According to this definition, mind is what a system is capable of and is not in relation with material composition.

Another point is that the definition of intelligence, or understanding, is very much unclear. One way of measuring understanding could be a series of questions-answers. If we talk about a system that understands things, then it is expected to be able to reply questions. Once again, this is a way of measuring understanding and it is indeed an appropriate way. But this is simply what a machine that passes Turing Test does. It takes questions in the form of typed-texts, parses them, *understands* and gives meaningful answers.

5.1.2 An Argument Against Strong AI: The Jukebox Argument

The father of Jukebox Argument is Ned Block [5], a former philosophy professor in MIT. His argument against Turing Test is in fact pretty simple and contains an alternative perspective of this problem.
First, remember that Turing Test occurs in a limited amount of time. In other words, the number of questions asked is finite. It could be a huge number but at the end of the day, it is not infinite. Now, suppose that someone has built a monstrous jukebox that contains all questions that can be asked in English. Along with that, answers to all of these questions are stored in the jukebox. And finally, this jukebox contains a script mapping questions to answers.
When a question is asked by the interrogator, all what jukebox does is to find this question in its storage and return the corresponding answer. This approach may feel uncomfortable for a moment but theoretically it is not problematic at all. In fact, the way many problems in theoretical computer science are solved in a similar way.

Obviously, jukebox does nothing but string matching. There is no understanding involved. This is simple to conclude because sentences are not even parsed. One more remark is that this machine is expected to pass Turing Test. Since it is a man-made machine, its builders can design it in such a way that any kind of question-answer pattern is stored. Therefore, people supporting this argument concludes that Turing Test cannot be a sufficient condition for understanding, or intelligence.

An objection to Jukebox Argument is that jukebox is indeed intelligent but it has a different format, it is built-in. There is no doubt that some intelligence is stored within this jukebox. However, from AI perspective, this kind of intelligence means nothing. It is static and has no capability of inferring new information from already known information. In other words, even if jukebox stores some intelligence, it does not act intelligently.

5.2 Minds, Brains and Programs

5.2.1 Searle's Keynotes

Basic points of Searle's thinking is revealed in the abstract of his paper:

- Human brain *causes* human mind. Put another way, what gives the power of forming causing relations to human mind is actually human brain. This is also saying that brain processes result in **intentionality**, a term he has used so frequently.
- Instantiating a computer program is not a sufficient reason for having intentions. Therefore, if it is somehow showed that computer programs cannot have intentions, then it will be concluded that minds are more than machines. In fact, Chinese room argument is targeted at that phenomena: The simulation of a computer program by a human-being.

A more computer scientist way of referring to these arguments, also used by Searle very often, is as follows [24]:

- **Axiom-1:** Computer programs are formal, or syntactic.
- **Axiom-2:** Minds, however, have mental contents, or semantics.
- **Axiom-3:** Brain causes mind.
- **Conclusion:** The ability to make syntactic operations is not the sufficient condition for having semantic meanings. This is done by human brain for human-beings. Any artificial brain must duplicate such specific causal powers of human brain, which is certainly not just running a simple program.

5.2.2 Chinese Room Argument

Chinese Room Argument can be considered as a thought experiment. It is not quite possible to apply what's described in the argument to the real world but that does not make it less valuable. Since Searle's main goal was to show syntactic operations do not result in understanding, he tried to simulate such operations by himself.

Now, imagine that you are locked in a room. There are only two entrances to the room, one of which serves as where input is brought to the room and one serves as output buffer. In addition to that, suppose that you do not know Chinese. It would be more realistic if you have no idea of Chinese, either written or spoken and you can even cannot distinguish Chinese from any other language such as Japanese.

From the input hole, you get a paper written in Chinese. Things that you cannot identify is written on this paper; letters, figures, marks or whatever. In the room, there is also a huge manual written in English, or any language you know. This manual is composed of instructions telling you to write things, depending on your input paper, to a blank paper.

In fact, what you are doing is a long and tiring process. The instruction manual is huge and new papers keep coming while you are producing outputs. Once again, you do not have any idea of the things you deal with. All you know is how to do it but symbols, marks or whatever you read and write make no sense to you.

Searle took the argument and went further. He considered that you are getting so better at this job that his answers to the questions and the answers given by a native Chinese speaker are simply indistinguishable (As you see, he directly targeted Turing's arguments). However, a native Chinese can understand questions and reply them accordingly while you understand nothing. From a Chinese speaker's perspective, what you do in that room is equal to what a computer does when running code. They both execute formal operations and this job does not incorporate any understanding at all.

A nice summary of Searle's claims would be the following [25]:

1. *If Strong AI is true, then there is a program for Chinese such that if any computing system runs that program, that system thereby comes to understand Chinese.*

2. *I could run a program for Chinese without thereby coming to understand Chinese.*

3. *Therefore Strong AI is false.*

5.2.3 Objections Replied in the Paper

Just like Alan Turing's paper, there are many objections to this work. Following ones are answered in the paper and I am going to examine a couple of them below: (1) The Systems Reply(Berkeley), (2) The Robot Reply(Yale), (3) The Brain Simulator Reply(Berkeley and M.I.T.), (4) The Combination Reply(Berkeley and Stanford), (5) The Other Minds Reply (Yale), (6) The Many Mansions Reply(Berkeley)

- **The Systems Reply(Berkeley)**: A much anticipated objection to Chinese Room Argument agrees with that the man in the room does not understand Chinese. However, this man is just a part of a system, it is the *Central Processing Unit* in the room. The system is somewhat larger; it contains a gigantic book of rules and bunch of paper and pen. Needless to say, they correspond to instruction set and memory in a computer. Conclusion: Not the man but system as a whole understands Chinese.
  A support to System Reply comes from Jack Copeland, a professor of philosophy. He has pointed out in his 2002 paper [26] that just like modules in mind solves some sort of equations that leads a baseball player to catch a ball, the man in the room may not understand Chinese but system as a whole can.
  Searle's response to this reply is pretty simple: Suppose that this man internalizes all the system. That is, he memorizes all instructions and makes syntactic operations in the head. If this man goes out, he can make conversations in Chinese but he still understands nothing. Although he *becomes* the system, he cannot ask for hamburger in a restaurant, as Searle said.

- **The Brain Simulator Reply(Berkeley and M.I.T.):** This reply is kind of interesting and it directly aims at showing that Searle's logic is problematic. According to Brain Simulator Reply, for a moment let's forget about whatever our system does and suppose that it "simulates the actual firings at the synapses of the brain of a native Chinese speaker when he understands stories in Chinese" [9]. Then, we have to conclude that the machine does understand Chinese since "at the level of the synapses there would be no difference between the program of the computer and the program in the brain of a Chinese.
  Before discussing this reply, Searle has noted that this kind of simulation would exactly be the opposite of purposes of strong AI since it is claimed by strong AI that we do not have to know

how brain works to get how mind works. This seems reasonable to me. He, then arrived at his basic objection to the reply and claimed that “even getting this close to the operation of the brain is still insufficient to produce understanding” [9]. To him, brain simulator simulates only the formal structure of neuron firing and lacks in simulating structures producing meaning and mental states. My little research had given no clue of what Searle means by *structures producing meaning and mental states.* I think he is supposed to be much more clear on that and show some evidences for the existence of such structures. Otherwise, one may simply consider mental states as nothing but illusion.

Chapter 6

Further Work

To me, this research was a lot of fun. I wished I had some more time to study more topics. Certain discussions that I had wanted but found no time to study are as follows:

- Connectionist vs Symbolic Models
- Motives, Mechanisms Emotions [27]
- Cognitive Wheels [28]

Bibliography

[1] Definition of philosophy. [Online]. Available: http://philosophy.fsu.edu/Programs/Undergraduate-Program/What-is-Philosophy

[2] Definition of philosophy. [Online]. Available: http://en.wikipedia.org/wiki/Philosophy

[3] Definition of philosophy. [Online]. Available: http://www.merriam-webster.com/dictionary/philosophy

[4] A. N. Whitehead, *Process and Reality.* Simon & Schuster, 1979.

[5] T. C. Moody, *Philosophy of Artificial Intelligence.* Simon & Schuster, 1993.

[6] Philosophy of artificial intelligence on philpapers. [Online]. Available: http://philpapers.org/browse/philosophy-of-artificial-intelligence

[7] Philosophy of ai, wikipedia page. [Online]. Available: http://en.wikipedia.org/wiki/Philosophy_of_artificial_intelligence

[8] A. M. Turing, "Computing machinery and intelligence," *Mind*, vol. 59, no. 236, pp. 433–460, 1950.

[9] J. R. Searle, "Minds, brains and programs," *Behavioral and Brain Sciences*, vol. 3, no. 03, pp. 417–424, 1980.

[10] Stanford encyclopedia of philosophy. [Online]. Available: http://plato.stanford.edu/

[11] Internet encyclopedia of philosophy. [Online]. Available: http://www.iep.utm.edu/

[12] *The philosophy of artificial intelligence / edited by Margaret A. Boden.* Oxford University Press, 1990.

[13] J. Copeland, *Artificial intelligence : a philosophical introduction.* Blackwell, 1993.

[14] A. Narayanan, *On being a machine.* Halsted Press, 1988.

[15] *The Mind and the machine : philosophical aspects of artificial intelligence / editor, S.B. Torrance.* Halsted Press, 1984.

[16] Difference engine. [Online]. Available: http://consc.net/mindpapers/6/all

[17] "Overseas air lines rely on magic brain," *Life Magazine*, p. 45, 8 1937.

[18] Computerized gods. [Online]. Available: http://www.vedicsciences.net/articles/computerized-gods.html

[19] Roger schank's thinking. [Online]. Available: http://www.rogerschank.com/biography.html

[20] Difference engine image. [Online]. Available: http://mattiasa.deviantart.com/art/the-difference-engine-207781965

[21] Dualism. [Online]. Available: http://plato.stanford.edu/entries/dualism/

[22] J. Gribbin and P. Davies, *The Matter Myth: Dramatic Discoveries that Challenge Our Understanding of Physical Reality.* Halsted Press, 1991.

[23] R. French, "Subcognition and the limits of the turing test," *Mind*, vol. 99, no. 53–65, 1990.

[24] Chinese room argument. [Online]. Available: http://www.iep.utm.edu/chineser/

[25] The chinese room argument. [Online]. Available: http://plato.stanford.edu/entries/chinese-room/

[26] J. Copeland, *The Chinese Room from a Logical Point of View.* Oxford University Press, 2002.

[27] A. Sloman, "Motives, mechanisms emotions," *Emotion and Cognition*, vol. 1, no. 3, pp. 217–234, 1987.

[28] D. Dennett, "Cognitive wheels: The frame problem of ai," *Minds, Machines and Evolution*, vol. 1, no. 3, pp. 217–234, 1984.