

Can computers think?

In 1950, Alan Turing wrote a paper that even till the present day, provokes and influences thought about a difficult topic that discusses whether or not we can create something that is capable of thought. I intend to provide a critique of Turing's arguments and show that, whilst I cannot agree with the way in which he attempts to tackle this subject, thought can indeed be represented by artificial processes. However, as we shall see in the forthcoming arguments and as Turing also found, it is difficult to progress such a view in a clear cut manner and without opposition presenting complex discussion at each stage.

Turing was forced to consider the delicacies and the essence of human existence and the mind by the sudden death of his closest friend at a young age, shown in letters to his deceased friend's mother.¹ However, it was not until twenty years later that, whilst working as Deputy Director of the computing laboratory at Manchester University as one of the first to write software programs for the computer there, but during a confused time in his life, he produced the paper *Computing Machinery and Intelligence*.² Whilst he proposes in this paper "to consider the question, 'Can machines think?',"³ he immediately replaces this question with a problem "in terms of a game which we call the 'imitation game'."³ This game "is played with three people, a man (A), a woman (B), and an interrogator (C) who may be of either sex. The interrogator stays in a room apart from the other two. The object of the game for the interrogator is to determine which of the other two is the man and which is the woman."³ Just so we know that he is aware of what would make it a fair test, Turing adds, "In order that tones of voice may not help the interrogator the answers should be written, or better still, typewritten."⁴ He also makes the game more interesting by inserting the following rules, "It is A's object in the game to try and cause C to make the wrong identification,"⁴ and "The object of the game for the third player (B) is to help the interrogator."⁴ He then introduces the purpose of this game and the basis of the rest of his discussion when asking "What will happen when a

¹ Charles Gimon, *Heroes of Cyberspace: Alan Turing* <<http://www.gimonca.com/personal/archive/turing.html>>, 1997

² Andrew Hodges, *Alan Turing: A Short Biography* <<http://www.turing.org.uk/bio/part7.html>>, 1995

³ A. M. Turing, "Computing Machinery and Intelligence," *Mind*, Vol. LIX No. 236, October 1950, p.433

⁴ Turing, *Mind*, p.434

machine takes the part of A in this game?’ Will the interrogator decide wrongly as often [...] These questions replace our original, ‘Can machines think?’.”

This game is introduced as Turing believes it to be an accurate test of thought and one that ignores trivial “disabilities”⁵ such as “[the machine’s] inability to shine in beauty competitions”⁵ or “[a man] losing in a race against an aeroplane.”⁵ There are many opportunities to question the suitability of such a test at this point. Most importantly, it must be asked whether Turing has approached the original question from the right angle. He is substituting the question of whether or not a machine can think, for whether or not one such machine could imitate a human through language. This question brings about two more considerations if it is to be taken seriously. Firstly, we must ask whether or not the ability to use language to succeed in the test would indicate that the creator of the language has the ability to think, which I will discuss later in this piece. Secondly, in this game or test, our perception of whether or not the machine is capable of thought is limited to how well it imitates a being with specifically human thought. Turing is therefore to deny all thought that doesn’t closely resemble that which originates from a human. Whilst Turing does indeed consider such an objection and even deems it “a very strong one,”⁵ he immediately and puzzlingly dismisses it, appealing to the fact that if a machine “can be constructed to play the imitation game satisfactorily, we need not be troubled by this objection.”⁵ It would seem that Turing is willing therefore to dismiss any other machine’s ability to think. However, at this point we must remember Turing’s original purpose which was to find out if it is possible for a machine to think. His test was not designed to test for the ability to think within any machine, but whether one such machine could exist. Turing is willing to pass over the possibility of machine’s thinking in other ways to pass his test which he sees as more stringent. Therefore, for the moment, we must grudgingly leave this objection alone and consider the validity of the test in determining the possibility of a machine that can think.

Turing’s test does however still have its merits. Since the game involves attempting to imitate a man who is trying to make the interrogator guess he is the woman, it requires skill. Turing considered it skill that was a result of thought. Such internal processes required for the game include creativity, understanding, and the ability to not

⁵ Turing, *Mind*, p.435

be trapped by pure deductive reasoning. By the last example I refer to the problem Turing introduces whilst discussing the mathematical objection and describing *Gödel's* Theorem which describes the inability of a logical system to correctly reply to questions such as “Consider the machine specified as follows.... Will this machine ever answer ‘Yes’ to any question?” The dots are to be replaced by a description of some machine [...].”⁶ A machine that could perform all of these tasks would clearly be a very complex one. However, since it is very conceivable that these skills could be programmed straight into a machine, it is difficult to have any certainty of whether even a machine which displayed these skills perfectly could be seen as one that thinks.

Since Turing's paper, there have been many publicised attempts at creating a machine that could pass the test, but most work on the principle of programming the machine with lists of responses in advance to questions. This is quite different from the idea of the machine creating the responses itself. Ned Block, a prominent and current philosopher who often deals with the topics of consciousness and the mind, expresses similar views. He writes, “even if a high budget government initiative produced a program that was good at passing the Turing Test, if the program was just a bundle of tricks [...] with question types all thought of in advance, and canned responses placed in the machine, the machine would not be intelligent.”⁷ This notion is explained in a famous example often called The Chinese Room by John Searle, who often expresses his opposition to the idea that it is possible to create a machine that could think. In his 1980 paper, Searle describes a room in which he is locked in. In this room, he is fed a set of symbols, and a rule book in English. This rule book explains how to, upon receiving certain symbols, reply with other symbols.⁸ However, the symbols that Searle is receiving and responding with are in fact those of Chinese writing, of which Searle has no knowledge of. In his analogy, the symbols Searle is provided with are called ‘the questions,’ the rule book is called ‘the program’ and the symbols he gives back after consulting ‘the program’ are called ‘the responses to the questions.’⁸ He then asks us to

⁶ Turing, *Mind*, p.445

⁷ Ned Block, “The Mind as the Software of the Brain”, in *An Invitation to Cognitive Science*, edited by D. Osherson, L. Gleitman, S. Kosslyn, E. Smith and S. Sternberg, MIT Press, 1995 as reprinted at <<http://www.nyu.edu/gsas/dept/philo/faculty/block/papers/msb.html>>

⁸ John Searle, “Minds, Brains, and Programs,” *The Behavioral and Brain Sciences*, Vol. 3, 1980, as reprinted at <<http://members.aol.com/NeoNoetics/MindsBrainsPrograms.html>>

“Suppose also that after a while I get so good at following the instructions for manipulating the Chinese symbols and the programmers get so good at writing the programs that from the external point of view [...] my answers to the questions are absolutely indistinguishable from those of native Chinese speakers.”⁸ Searle’s aim is to show that just because a machine could pass the Turing Test, it would not understand what it was saying in the same way as a human would, just as Searle doesn’t understand his responses in the same way a Chinese speaker would.

The specific refutations of this analogy and their replies are numerous and are not within the focus of this paper. What are of significant importance are the consequences of what it entails. If it is true that the artificial system following the rule book does not think, then there are two possibilities to explain the difference between human thought and the processes of a system such as in the Chinese Room. The first possibility is that humans possess something that the system does not. This entails that we are built in a different way to the system Searle is describing. This explanation would seem to require a definition of human thought and its constituents, which although is essential to such an argument, is not attempted in depth by Turing in his 1950 paper. Thought is a perplexing subject. If we are to consider its meaning as something objective we run into a lot of difficulty as it is impossible to reason about thought from outside the realm of thought and also to define it as something non-human since we are only aware of it within ourselves, and are only able to reason about it through thought itself. It is somewhat easier to refer to the physical make-up of our brain and the processes within it as our brain states, and all forms of thought and emotion as our mental states. Searle describes a system with brain states (the person receiving the input, and using the rule book to provide an output), but no mental state. He aims to show that no matter how complex the system is, it can never have mental states. He even goes so far as to say that even if you would “Imagine a robot with a brain-shaped computer lodged in its cranial cavity, imagine the computer programmed with all the synapses of a human brain, imagine the whole behavior of the robot is indistinguishable from human behavior, and now think of the whole thing as a unified system and not just as a computer with inputs and outputs,”⁸ the robot would not be able to attach meaning or intentionality to its actions. However, Searle never provides us with an explanation of how the brain is able to create mental

states that are different from its physical states. Without this, and assuming that the mental states actually exist within the physicality of the brain, it seems possible that a machine could be designed that could learn, experience things, and make decisions just like a human and would essentially be thinking like a human.

This enables us to consider the second possible consequence of the Chinese Room analogy. A consequence I am sure that Searle never intended. It is conceivable that humans are on the same level as the Chinese Room and do not actually think in the way Searle infers either. This would entail that human thought is not thought but is the sum of our programming. For example, what we consider as thought processes could be the internal language that is a result of our lines of code, or biological configuration that, to any believers in the Theory of Evolution would agree, had developed during our two billion or so years of evolution. In the same way, it then seems possible to design a machine that had a system of what we could call thought, perhaps as a thought language or a series of maps. In this way, a machine's thought could be very different to that of a human, and it's ability to perform in the Turing Test could be very poor. However, since it has, according to this view, a process of thought in a similar way that humans have, it should not be rejected at all and it is here that it is shown most effectively that the Turing Test is unsuitable, and could even be providing potential developers of machines of artificial intelligence with a goal that leads them down the wrong path entirely.

However, we must not consider this as a conclusion. It is still possible to argue that, even if the neurons in the brain do not actually understand what they are doing, there could still be a mind that provided thought on a different level outside of the material nature of the brain. The first possibility is explained within the theory of dualism which states a complete difference between the mental and the physical. In a book detailing the many issues in the philosophy of mind, it is stated for dualists that "On the material or physical side are such features as how much a person weighs, [...], how their brain operates, and generally those facts about a person studied in the physical and biological sciences. On the mental or psychological side are such features as how a person is feeling, what they are thinking, what they are seeing, and how intelligent they are."⁹

⁹ David Braddon-Mitchell and Frank Jackson, *Philosophy of Mind and Cognition*, Oxford: Blackwell Publishing, 1996, p.3

However, dualism is an unpopular view as it is subject to heavy objections, such as the fact that if one were to look at the events in the brain which link a cause to an action, it would be hard to see where the mental events come into the picture to influence the physical brain events. It is also possible to argue the existence of an immaterial mind, or soul, with a theological belief structure. However, any such argument could always be responded to with the argument that Turing provides, “He [the Almighty] has freedom to confer a soul on an elephant if He sees fit [...] An argument of exactly similar form may be made for the case of machines.”¹⁰

In light of the lack of relevance to the topic this therefore provides, we can focus on a theory of mind that seems to agree with the above hypothesis that the mind, whilst being capable of thought, is physically represented. In my opinion, this is best represented by a form of functionalism. Functionalists would say that mental states are “internal states within us, but we identify and name them by the effect the world has on them, the effect they have on one another, and the effect they have on the world by causing our behaviour.”¹¹ In addition to this, functionalism also contains the view within the concept of Multiple Realizability that our mental states are not limited to our physical design, and that any other states of physical chemistry could realise similar functional roles.¹² It is therefore also possible, that an artificially created state could realise functional roles that are similar to those created in us. If we were indeed to build the parts of a machine in a way that filled the same functional roles as the parts of a human brain, a machine would be created that was capable of thought in the same way humans are.

We have therefore seen just how influential Turing’s paper is and how it relates and stands up to current ideas. Whilst I have shown that the test that Turing presents us with is not a suitable one, his paper outlines the many issues facing the question of whether or not it is possible to build a machine that can think. However, I have shown that thought is not something we should necessarily become chauvinistic about and it seems we should be optimistic about future developments in this field. As Turing remarked over fifty years ago, “We can only see a short distance ahead, but we can see plenty there that

¹⁰ Turing, *Mind*, p.443

¹¹ Braddon-Mitchell and Jackson, *Philosophy of Mind*, p.41

¹² Braddon-Mitchell and Jackson, *Philosophy of Mind*, p.43

needs to be done.”¹³ There is still a lot to be done in terms of developing a structure of thought for a machine, but we are in a position where the road ahead is a lot clearer.

¹³ Turing, *Mind*, p.460

References:

- Block, Ned, “The Mind as the Software of the Brain”, in: *An Invitation to Cognitive Science*, edited by D. Osherson, L. Gleitman, S. Kosslyn, E. Smith and S. Sternberg, MIT Press, 1995, as reprinted at
<<http://www.nyu.edu/gsas/dept/philo/faculty/block/papers/msb.html>>
- Braddon-Mitchell, David and Jackson, Frank, *Philosophy of Mind and Cognition*, Oxford: Blackwell Publishing, 1996
- Gimon, Charles, *Heroes of Cyberspace: Alan Turing*,
<<http://www.gimonca.com/personal/archive/turing.html>>, 1997
- Hodges, Andrew, *Alan Turing: A Short Biography*,
<<http://www.turing.org.uk/bio/part7.html>>, 1995
- Searle, John, “Minds, Brains, and Programs,” *The Behavioral and Brain Sciences*, Vol. 3, 1980, as reprinted at
<<http://members.aol.com/NeoNoetics/MindsBrainsPrograms.html>>
- Turing, A. M., “Computing Machinery and Intelligence,” *Mind*, Vol. LIX No. 236, 1950, pp.433-460