Trying to find a new criterion thinking machine

Marián, Ambrozy

College of International Business
ISM Slovakia
Prešov, Slovakia
e-mail: ambrozy.marian@gmail.com

ABSTRACT

The present paper is an attempt to start a discussion with the opponents of the T-test mainly in the form of the Chinese room paradox. The paper also analyzes the objections postulated by J. R. Searle, and it basically accepts his arguments. We acknowledge that intentionality cannot be detected by observation. Therefore, we try to postulate a new criterion for determining an intelligent machine. Such criterion should prevent bypassing semantics via syntax which is the point Searle challenges. We refuse to postulate more demanding criteria. We believe there exists a simpler criterion. It is not our ambition to examine whether machines can or cannot think. We summarize ideas opposing the idea of an intelligent machine. We only try to present such criterion for determining whether a machine can think, which could resist the objections of J. R. Searle.

Keywords

T-test, thinking machine, Chinese room paradox

1. INTRODUCTION

The question whether machines can think has accompanied cybernetics from its start as a science. The word cybernetics was derived etymologically from Plato´s work and its original meaning is the ability to govern. It is known that some Marxist philosophers considered it a pseudoscience. Even such a prominent figure as academician Kolmogorov denounced cybernetics for a certain period of time. Some Calvinist philosophers also spoke negatively of cybernetics stating that it is against divine predestination. We assume that the above arguments would make the existence of some fields of physics, e.g., quantum mechanics and deterministic chaos, impossible. Today nobody doubts that this scientific discipline has firmly established itself in the scientific world. Philosophical problems of cybernetics are connected with philosophical questions of logic, programming languages, computer ethics, etc. One might even ask a question whether there are any areas of philosophy of cybernetics and informatics which do not have a parallel in other areas of philosophy. One of the philosophical problems is question whether machines can think.

2. T – TEST

"Turing´s machine as a theoretical computer model together with Church – Turing thesis have become the basic pillar of emerging digital computer technology" (Rompotl 2012, p. 34). In 1950, Alan M. Turing published his article *Computing Machinery and Intelligence*, in which he asked a very serious question: "Can machines think?" This question is connected with two phenomena: the Turing machine and the Turing test. They are not one and the same thing. Although Turing´s article might seem a positive contribution to the scientific problem, it actually includes more questions than answers. "The truth is the original question "Can machines think?" is formulated in a very vague way" (Tvrdý 2014, p. 26). Despite this the article together with Quine´s text *Two Dogmas of Empiricism* rank among the most influential philosophical texts of the second half of the 20th century.

Turing had studied the topic for quite a long time. In the late 1930s, he wrote his article *On Computable Numbers* in which he introduced an idealized machine performing simple pre-programmed operations. In his article Turing suggests constructing the Turing´s machine equipped with a finite number of internal states and with a potentially never-ending tape divided into cells. The machine is able to read, rewrite and delete symbols (Tvrdý 2014, p. 21). This machine reacts in accordance with pre-defined instructions and it either leaves the symbols in their original form or it rewrites or deletes them. Turing asks whether it is possible to use such a machine for solving the problem of the famous Hilbert´s programme, the so called *Entscheidungproblem* or decision problem. The programme would prove that each formalized mathematical proposition can be matched with a formal process which gives the proposition its truth value. "Hilbert and Ackermann in their *Grundzüge der theoretischen Logik* [1928, § 11] which was the first modern textbook of this scientific discipline, define the *Entscheidungsproblem* as the main problem of then mathematical logic" (Kolman 2008, p. 527). The core of the problem rests "in the fact that for the said deductive theory we need to find a general method enabling us to decide whether a proposition formulated within the theory can or cannot be proven by this theory" (Tarski 1966, p. 139). Turing demonstrated that neither his idealized machine is able to find such process. The question of truthfulness becomes even more complex in the so called fictitious worlds, fictitious objects and their predicates (Martinich 2007, pp. 152 – 159).

The first version of the Turing test is simple. It is a game with three players who have no physical contact. One of the players is a man, the second one is a woman and the sex of the third player is indifferent. The only contact the players have is in the form of text. The third player has just one task and that is to find out which of the players is a man and which is a woman. Both these players try to confuse the third player. "Turing believes that a computer would be intelligent if it could apply successfully the strategy of the man in this imitation game" (Páleš 1994, p. 15). In another version of the game the third player plays against a real computer which takes over the roles of the man and the woman players whereas the purpose of the game and its rules remain the same. "If we succeeded in programming a computer so that it could play the imitation game as the man, we would have to admit that it can think and that it has something which makes people intelligent" (Liptáková, Ambrozy 2015, p. 169). According to Turing if the computer is intelligent, it can perfectly imitate the man or the woman, and confuse the third player whose task in the game is to determine the sex of the other players.

Turing tried to develop a reliable test which could find out whether machines can think. As stated above he tried to reduce the notions to their elementary forms without turning away from the original context of the question. Simultaneously, he tried to define the notions of *machine* and *think*. Considering the above mentioned, we can see that Turing understands the thinking of a machine as an ability to play the role of the man or the woman in communication with a human. Turing considers the fact that a machine can fully replace people in the said roles with a sufficient proof of thinking. However, the development of the situation after replacing a human by a machine remains a cardinal question.

Experts on Turing´s work have found some discrepancies in the above mentioned versions of the tests. It depends on the translation and interpretation of the famous Turing´s essay whether the English word *man* is translated into other languages as a male or a human. In such case the game assumes a gender dimension and the third player must determine the sex of the other players which, according to some interpreters of the Turing´s text (Genova 1994), has consequences based on which gender is a social construct and not a biological fact.

The outputs from Turing´s proposals are clear. Turing proposes to replace human thinking by a computer which answers the questions as a meaningfully communicating human being. According to Turing in order to detect the player´s sex or human or machine identity of the communicating player, the thinking machine must perfectly imitate a human or one of the sexes. In Turing´s opinion a successful game comparable to human communication is a criterion of thinking. The computer, however, needs competences from the fields of linguistics and psychology and it must recognize a natural human language. There are also other versions of the Turing test. One of them was presented by Hingston (2010). "The new judging system can be used to support a "reverse Turing Test" for bots, in which bots are evaluated on their ability to identify the other players as human or bot" (Hingston 2010, 348).

There are several objections according to which a machine cannot think. In our opinion, however, their authors have not submitted sufficient evidence proving their standpoint. Turing himself attempted to refute such objections in his legendary paper *Computing Machinery and Intelligence.* Turing tried to refute 9 possible objections to thinking machines. Let us introduce the most interesting of them.

One of the interesting comments comes from the way the Gödel´s incompleteness theorem was interpreted. The consequence of Gödel´s statements does not rest only in the impossibility to elaborate Hilbert´s programme in mathematics (Zlatoš 1995). There are tasks which cannot be performed by a computer and they cannot be performed by a human either. Laws based on formal logic and mathematics apply also to human inference. In Turing´s opinion, the second Gödel´s statement can be applied to the said problem in a way enabling deriving a statement that there are activities the machine will never be able to perform. According to Turing, there are tasks of the imitation game the machines cannot do no matter how much time they are given for completing the tasks. Turing does not agree with the assertion that the said proposition does not concern the human mind.

There are also theological arguments against the claim that machines could think. Most Christian and Muslim theologians do not support the idea of thinking machines because if they did, they would have to admit that machines have a soul. "In some streams of the philosophy of the 20th century there are assertions that religion cannot produce justified cognitive assertions" (Karaba 2017, p. 192). For Alan Turing the said objection is not worth commenting on. He claims that from a logical point of view it is not impossible for machines to have souls and, therefore, it does not prevent God´s intervention (regardless of the type of religion). Similarly, Turing rejects irrelevant objections resulting from ontological superiority of the man over the machine.

Neurologist Geoffrey Jefferson also disagreed with the claim that machines can think. He said that a machine is not able to write a poem or a song, it cannot feel emotions, it cannot be happy or sad, etc. Contemporary advancement of computing technologies is a sufficient proof eliminating the claim that machines cannot create a work of art. Producing art with the help of programming can be used as an example – compare (Kerlow 2004) or (Spiridonova 2015).

Ady Lovelace was one of those who claimed that a computer is an analytical machine which cannot create anything new. He believed that machines do only what they are told to do. It is true that the above mentioned can be said about many machines. "Typewriters, copy machines, cameras and tape recorders, e. g., are machines, which produce signs. Are they semiotics machines? If semiosis is required, a copy machine can certainly not be called a semiotic machine although it may be said to produce signs" (Nöth 2003, p. 84). Turing disputed this opinion, nevertheless, he did not present any relevant counterarguments.

One of the arguments against the T-test rests in the idea that it is possible to cheat in the game and pretend to be an intelligent being. Let us take, for example, a simultaneous game of chess with two chess grandmasters who do not know about each other´s presence, whereas the player copies the moves of the two rivals. Robert French thinks that the T-test focuses on human intelligence only, i.e., it is interconnected with human realia to a large extent. French says there is too much focus on human body. French also uses the analogy of Nordic seagulls as an illustrative example. If philosophers discussed the meaning of flying based on the extent to which a candidate for a flying machine resembles a seagull, the criteria for defining a flying machine would be incorrect. In our opinion French used a completely wrong analogy.

Richard Purtill claims that the T-test is actually a discussion between a programmer and a human (Purtill 1971). It is certainly not possible to agree with this assertion without reservations. Purtill assumes that people are going to construct thinking computers in the future. However, it will not be so soon and the T-test will not be sufficient for testing their intelligence. Purtill´s arguments do not sound convincing as he also admits that brains are physical entities of enormous complexity and with analogically functioning concept.

Michael Scriven assumes that a computer cannot think because it does not have consciousness. In order to have consciousness it needs life. However, there is no satisfactory definition of life. The attempt of Teilhard de Chardin to define life by means of centrocomplexity has never been generally accepted. (Galleni 2011) Turing says that we associate consciousness with a person. Here the problem becomes more complicated because it is very difficult to define a person. For example, Peter Singer believes that some animals are also people. On the other hand, however, he says that a human before the age of one month cannot be considered a person. Similarly, the definition of consciousness is vague, too. There is a theory of consciousness of higher order according to which the consciousness of higher order is the thought reflection itself.

Philosopher Blay Whitby considers the T-test obsolete. In his opinion intelligence should not be tested on an imitation of a human. People should try to better understand cognitive processes instead. Contrary to this, Stevan Harnad and some other scientists try to push the boundaries of the T-test and make the computer imitate a human in everything a human is capable of doing (Harnad 1989).

3. OBJECTION OF J. R. SEARLE

The objection of J. R. Searle is so well known that it is not necessary to analyze it in detail. Apparently, the person in the room does not speak Chinese. To be able to communicate the person gets a book of instructions. It shows him/her how to use Chinese characters to make meaningful sentences. The person´s only communication output are the Chinese characters he/she shows. The person is able to identify the characters only on the basis of their shape. Searle concludes that the person´s knowledge of syntax helps bypass understanding on the semantic level. An outside observer, however, mistakenly believes the person can speak Chinese. Similarly, a computer understands syntax and can thus potentially succeed in the T-test. It cannot, however, understand a language on the semantic level.

The Searle´s objection is based on differences between semantics and syntax and on by-passing the semantic grasping by using the knowledge of syntax. Semantics plays a significant role in informatics, too. "The main role of semantics is to predict the outcome of program execution. The semantics can be viewed as a function which maps syntactic elements to the semantic domains" (Steingartner and Novitzká 2015, p. 252).

There are several objections to the findings of Chinese room experiment, however, Searle disputed them. The system objection says that the whole system can speak Chinese, i.e., it is not only the person in the room but it is the person and the syntactic books of instructions. Another objection speaks of hypothetical existence of a robot similar to a human placed in the Chinese room. In case the robot is equipped with sensomotoric organs, it is able to perceive and move, then in Searle´s opinion the basis of the problem remains the same because the robot uses the characters in accordance with the rules of syntax. There is another argument claiming that simulation of the brain is actually a simulation of the nerve impulses and this way it is possible to get the answers to Chinese questions.

William Rapaport has presented a more serious objection. He assumes that in the above thought experiment understanding is a matter of syntax only seemingly. He thinks that there is a connection between semantics and syntax and his opinion is built on a physicalist basis. He claims that "semantic and syntactic features of language expressions are produced by the same configurations of neurons of the central nervous system and thus there cannot be any major difference between them." (Tvrdý 2014, p. 109) To support his statement Rapaport presents some empirical facts. Two blind and deaf women Helen Keller and Olga Skorochodova learned to use their languages and they both became writers. Helen Keller wrote her autobiography in which she gave a detailed description of her process of learning. (Keller 1954) Rapaport assumes that they learned their languages on the level of syntax not semantics. If this was the case, the Chinese room argument would lose its justification.

4. PROPOSED SOLUTION

There is no counterargument excluding the thinking of a machine. Therefore, we try to present a criterion which would be certainly sufficient for stating that a particular computer can think. It is not our aim to oppose Turing´s criterion for determining whether a machine can think. We are not trying to refute his assertion and we are not saying his criterion is incorrect or insufficient. We are not opponents of the T-test. On the other hand, we are aware of the fact that there are numerous objections to Turing´s criterion. Our solution is an attempt to come with a stronger and more robust solution whereas it is not based on rejection or falsification of the T-test.

"´Classical´ computational model of mind might fail to capture crucial aspects of cognition even though Turing´s account of computation underlies it, and there´s a sense in which ´Turing machines can do anything´" (Fodor 2005, p. 25). It is not our task to say whether the T-test is or is not a sufficient criterion for computer intelligence. We see no serious reasons why the T-test could not manage the task given by Turing. However, there are still many defenders of the Searle´s Chinese room paradox and there are still many attempts to falsify the T-test as a suitable criterion. Therefore, we have tried to find our own criterion the machine has to meet in order to be declared a thinking machine. We try to present a formulation which can withstand all hitherto objections, mainly the Chinese room argument.

The machine must work independently without the possibility to bypass semantics via syntax. Computers use a wide spectrum of logical systems including non-classical logics, for example, linear logic (Mihalyi and Novitzka 2013). We refuse arguments which speak of dependence of semantics on syntax as unjustified. There are no substantial reasons for such dependence. Experts studying syntax and semantics do not agree with it either. Even in case of deaf and blind people the coining of words is connected with objects and it is not a product of syntactic operations. A thinking machine should produce truly inventive operations. For example, imagine a computer which is not connected to the Internet and which contains no pre-programmed questions. If such computer starts to generate questions which do not result from its memory chains and which the computer, under given circumstances, does not normally generate itself, we can speak of an intelligent computer. By questions we mean existential questions which only a thinking being can postulate. However, it remains a question whether such circumstances will be possible.

Computers can authentically and autonomously ask questions of the said type only in case they grasp the semantic part of words. We believe that meeting the above criterion is a sufficient provision of intelligence. If a device with artificial intelligence could meet the criterion, we believe we could consider it a thinking machine.

It is not our aim to say whether machines can ever meet this criterion. We do not touch upon the topic of the consciousness of machines either. Our only result is an attempt to postulate a new objective criterion for affirmation of computer intelligence. We admit there is a possibility that there exists another minimal and more easily performable criterion which has not been presented yet. Our attempt introduces a solution which should withstand objections to hitherto submitted criteria and mainly objections to the T-test.

5. ACKNOWLEDGEMENT

This paper was written as part of the tasks of the ‘Selected philosophical and ethical consequences of theoretical physics’, supported by institutional grants IG-KSV-01/2016/2.1.5.

REFERENCES

[1] J. Rompotl, "Historie kybernetiky a umělé inteligence v kontextu filosofie vědy", Plzeň, pp. 51, 2012.
[2] F. Tvrdý, “Turingův test”, Praha, pp. 213, 2014.
[3] V. Kolman, “Filosofie čísla”, Praha, pp. 672, 2008.
[4] A. Tarski, “Úvod do logiky a metodologie deduktivních věd”, Praha, pp. 248, 1966.
[5] A. Martinich, A. Stroll, “Much Ado about nonexistence: Fiction and reference”, Plymouth, pp. 147, 2007.
[6] E. Páleš, “Sapfo parafrázovač slovenčiny počítačový nástroj na modelovanie v jazykovede” Bratislava, pp. 308, 1994.
[7] M. Liptáková, M. Ambrozy, “The selected connection between intencionality of mind and informatics”, *Scientific Conference on Informatics, 2015 IEEE 13th International,* New York, pp. 168 – 170, 2015.
[7] J. Genova, “Turing´s Sexual Guessing Game”, *Social Epistemology 8 (4)*, pp. 341 – 343, 1994.
[8] P. Hingston, “A new design for a Turing Test for bots”, *Computational Inteligence and Games IEEE (CIG ´10)*, Copenhagen, pp. 345 – 350, 2010.
[9] P. Zlatoš, “Ani matematika si nemôže byť istá sama sebou”, Bratislava, pp. 208, 1995.
[10] M. Karaba, “Aplikácia vybraných vedeckých metodológií na otázky náboženského poznania”, *Filozofia 72 (3)*, pp. 192 – 203, 2017.
[11] I. V. Kerlow, “The art of 3D: computer animation and effects”, New York, pp. 451, 2004.
[12] A. Spiridonova, “Light experiments in russian avangard art”, *2nd International Multidisciplinary Scientific Conference on Social Sciences and Arts SGEM2015, Book 4, Vol. 1*, Sofia, pp. 3 – 8, 2015.
[13] W. Nöth, “Semiotic machines*”, S.E.E.D. Journal (Semiotics, Evolution, Energy, and Development) 3 (3)*, pp. 81 – 99, 2003.
[14] R. L. Purtill, “Beating the imitation game”, *Mind, 80 (318)*, pp. 290-294, 1971.
[15] L. Galleni, “Teilhard de Chardin and the Latin School of Evolution : Complexity, Moving towards and Equilibriums of Nature”, *Pensamiento 67 (254)*, pp. 689–708, 2011.
[16] S. Harnad, “Minds, machines and Searle”, Journal of Experimental & Theoretical Artificial Intelligence, 1 (1), pp. 5 – 25, 1989.
[17] W. Steingartner, V. Novitzká, “A new approach to semantics of procedures in categorical terms”, *Scientific Conference on Informatics, 2015 IEEE 13th International*, New York, pp. 252 – 257, 2015.
[18] H. Keller, “The story of my life”, New York, pp. 382, 1954.
[19] J. Fodor, “Reply to Steven Pinker ‘So how does the mind work?’”, *Mind & Language, 20 (1)*, pp. 25 – 23, 2005.
[20] D. Mihályi, V. Novitzká, “What about linear logic in computer science”, *Acta Polytechnica Hungarica, 10 (4)*, pp. 147-160, 2013.