SciVerse ScienceDirect

Procedia - Social and Behavioral Sciences 32 (2012) 456 – 461

Procedia
Social and Behavioral Sciences

4th International Conference of Cognitive Science (ICCS 2011)

Davidson's no-priority thesis in defending the Turing Test

Mohammad Reza Vaez Shahrestani*

Philosophy of Science, Amirkabir University of Technology, Tehran, Iran

Abstract

Turing does not provide an explanation for substituting the original question of his test – i.e., "Can machines think?" with "Can a machine pass the imitation game?" – resulting in an argumentative gap in his main thesis. In this article, I argue that a positive answer to the second question would mean attributing the ability of linguistic interactions to machines; while a positive answer to the original question would mean attributing the ability of thinking to machines. In such a situation, defending the Turing Test requires establishing a relationship between thought and language. In this regard, Davidson's no-priority theory is presented as an approach for defending the test.

Keywords: Turing Test; ability of linguistic interaction; ability of interpretation; thought and language; Davidson's no-priority theory

1. Introduction

Can machines think? Artificial intelligence is seeking to find the answer to this question. During recent decades, the possibility of having intelligent machines has drawn the attention of philosophers as well as experts in the field of artificial intelligence. The Turing Test is one of the well-known theories in this field. Alan Turing (1950), British philosopher and an abstract machines theorist, designed a test to assess machines' intelligence. He began his article with this question: "Can machines think?" According to Turing, in order to understand this question, clear definitions of the terms "machine" and "think" need to be provided. Since it is not possible to define these notions on the basis of their normal uses, Turing believes it is dangerous and inappropriate to use this method to answer the aforementioned question and thus considers the question absurd. In order to clarify the question and use unambiguous words, he immediately replaces it with a new question. The new question is: "Can a machine pass the imitation game?" The imitation game is played with three people, a man, a woman, and an interrogator, who may be of either sex. In this game, the interrogator is kept in a room that is apart and hidden from the other two. The interrogator's role is to identify which one of the two other participants is the man and which one is the woman and if succeeded in doing this, he/she will win the game. To identify the man and the woman, the interrogator can ask each of them some questions. In order to prevent the identification of the two with the help of their tones of voice,

* Corresponding author. Tel.:+98-09137384696; fax:+0-000-000-0000
E-mail address: vaez.dawn1985@gmail.com

1877-0428 © 2011 Published by Elsevier Ltd. Selection and/or peer-review under responsibility of the 4th International Conference of Cognitive Science
Open access under CC BY-NC-ND license. doi:10.1016/j.sbspro.2012.01.069

the answers should be written. The game is played so that when answering the interrogator's questions, the woman helps the interrogator to win the game and the man tries to deceive the interrogator. In the rest of the article, Turing replaces the man's role in the game with a machine and proposes a new question – i.e., "Can a machine pass the imitation game?" – instead of the original one. He does not provide any explanations to justify this replacement and only mentions the ambiguity of the terms "think" and "machine" in the original question for such a substitution. Considering the second question free from ambiguity, he does not give any further explanation for the replacement. In criticism and investigation of Turing Test, lack of a tangible explanation for the substitution of the original question with the new one is considered as an argumentative gap in the test (Turing, 1950). In this article, I attempt to find an acceptable justification for this replacement in order to defend the test. This study uses the well-known theory of Donald Davidson (1917-2005), the contemporary American philosopher, with regard to the relationship between thought and language, which will be briefly explained later. I shall use this theory without evaluating it. The explanations for this philosophical theory can be found in Davidson's works.

2. Turing Test

In October 1950, Alan Turing projected that by the year 2000 it will be possible to build a machine that will have 30% chance of deceiving an interrogator, so that the interrogator would make a mistake distinguishing a human from a machine. (Turing, 1950) In this test, an interrogator is put behind two printers; one printer is connected to a terminal used by a woman and the other printer is connected to a digital computer. The woman and the machine cannot be seen by the interrogator. In order to identify which terminal is connected to the machine and which one is controlled by the woman, the interrogator types some questions for each one. The test is designed so that the woman helps the interrogator to win the game, while the machine is programmed to deceive the interrogator. This test will produce results when the interrogator determines which terminal is connected to the machine and which one in connected to the woman. (Davidson, 1990) If the machine succeeds in deceiving the interrogator and wins the test, it means that the machine can think so claims Turing. This test is one-sided, that is, if the machine fails to deceive the interrogator, no result can be drawn. In the next part I will explain Davison's no-priority theory in order to use it as an approach in defending the Turing Test.

3. Davidson's theory of no-priority

What is the connection between thought and language? In "Thought and Talk", Davidson (1975) poses this question and analyzes and investigates the relationship between thought and language. According to him, the dependence of language on thought is evident because language is used to express thoughts. For instance, when someone says "The candle is out", this statement refers to a candle in the outer world which is out at the time of the utterance and the speaker believes the utterance is true. In fact, by making these sounds, the speaker is uttering words that are only true under such circumstances. In other words, when such belief is absent – a belief in the existence of a candle that is out in the outer world – such sounds would not be produced and this is indicative of the connection between beliefs and language.

On the whole, Davidson is of the conviction that neither language nor thought can be fully explained based on the other one and neither one has conceptual priority over the other. Language and thought are interconnected, that is each one needs the other in order to be understood. In other words, two claims can be made based on this theory:

1. In order to be able to think, we need to have the ability of language interaction with others (i.e. we need to have the ability of interpreting people's words). This principle can also be written this way: "if a creature can think, it needs to have the ability of language interaction".

According to Davidson, a creature cannot have thoughts without being (at least potentially) able to interpret the speeches and viewpoints of others. To him, this conclusion does not require reduction of thought to speech (in a way one may follow in a behaviouristic approach) and thus imputes no conceptual or epistemological priority for language over thought. In addition, we are in need of concepts for thinking because thinking is done through concepts and imaginations. Davidson argues that in order to have a network of concepts, we require the concept of error and in order to have the concept of error, we require language interactions with another person so that we realize the accuracy or inaccuracy of our thoughts. As a result, in order to be able to think, we need to have language interactions with other people; in other words, we need to be able to interpret their speeches (Davidson, 1975).

2. In order to have language interactions with others, we need thinking (It means we need to have a system of beliefs). This principle can be written this way: "if a creature has the ability of language interaction, it must be able to think".

Davidson is of the conviction that a thought is determined by a system of beliefs; in other words, having a thought requires having a background of beliefs. For instance, he says if I consider the thought of going to a certain concert, I need a system of beliefs in my mind so to have this thought. These beliefs include "I will be put to a degree of trouble and expense", "I will be enjoying my favourite music by attending this concert", etc. Therefore, in order to have this thought, I need to have a background of beliefs that determine this thought for me. Meanwhile, having a certain thought does not depend on having a certain belief, that is, I have a thought about going to this concert, but until I decide whether I will go to the concert or not, I won't be having a particular and fixed belief such as "I will go to this concert". Until then, I merely entertain that thought. After proposing the central role of beliefs for having thoughts, Davidson points to the necessity of having a background of beliefs for understanding and interpreting the speeches of others during language interactions. According to this thesis, the person who understands and interprets a sentence in a natural language (say Persian), must have a lot of beliefs and these beliefs must be very similar to those of the person who entertains this thought. For instance, if someone hears the utterance that "this gun is loaded" and aims to interpret it, he must believe that a gun is a weapon and must believe that it is a more or less enduring physical object, etc. Therefore, in order for him to comprehend the sentence, there is perhaps no definite and fixed list of things in which he must believe. However, for the sentence to be understood there must be endless interlocked beliefs. In sum, it can be said that a creature's understanding of an utterance in a language requires having a thought about that utterance, and having such a thought requires having a system of beliefs about it.

On the whole, the fact that the pattern of relations between sentences of a language is very similar to the pattern of relations between thoughts has encouraged the view that taking each pattern as the basic is redundant. If we imagine that thoughts have priority over language, it seems that language has no aim but to express and convey thoughts and this means that language is just a mean for conveying thoughts. If we imagine that language has priority over thought, it would be an attempt to analyze thoughts as speech dispositions, which implies that thoughts at human level are just verbal activities. Therefore, the parallel between the structure of thoughts and structure of sentences provides no argument for primacy of thought or language and it is just a presumption in favour of interdependence of thought and language.

4. No-priority approach in defending the Turing Test

In Turing Test, we are faced with the replacement of "Can machines think?" with "Can a machine pass the imitation game?" If we agree with Turing and attempt to defend the Turing Test, we must be able to provide a reasonable explanation for this replacement. As mentioned above, Turing does not give further reasons for the substitution.

In an article written about the Turing Test, Davidson says: "It is fairly clear that Turing did not believe anything of philosophical importance would be lost by the substitution". (Davidson, 1990, p. 77) Therefore, in Turing's article absence of justification for this substitution is seen as a gap and one needs to fill this gap in order to defend the Turing Test. In addition, we know that Turing's machine (a digital machine with proper programming) needs to deceive the interrogator in recognising the identity so to succeed the test. This is while the relationship between Turing's machine and the interrogator is only possible through written verbal conversations (language interaction) via a printer, i.e. the machine needs to deceive the interrogator through language interaction. In fact, the ability of thinking in Turing's machine in the original questions has turned into the ability of language interaction in the new question. In other words, in order to justify the test, there is a need to establish a relationship between thought and language or in fact we are in need of a view that imputes a systematic relationship between thought and language because if a person does not believe in the relationship between thought and language, it is not possible to acknowledge the value of Turing Test. However, if one manages to argue that the success of the machine in Turing Test constitutes that machines have the ability of language interaction with others –that is to verify the accuracy of the statement that "if Turing machine succeeds the test, it has the ability of language interaction" – we can defend Turing Test based on Davidson's no-priority theory.

The intended approach, which uses the transitivity of materials implication (P → Q, Q → R » P → R) is as follows:

1. If Turing's machine succeeds the test, it has the ability of language interaction. (A statement which we aim to prove.)
2. If a creature has the ability of language interaction, it can think (The second conclusion of Davidson's theory).
3. If Turing's machine succeeds the test, it can think (using the transitivity of materials implication and premise 1 and 2).

In addition, according to the first conclusion, in order for the Turing machine to think, it needs language interaction. If it succeeds the test, based on the statement 1, it has this ability. Therefore, we can at least say that if the machine succeeds in the test, the machine has the ability of language interaction and thus it is not in contradiction with the first conclusion.

As it was stated before, to investigate the accuracy of this approach in defending Turing Test, premise 1 needs to be investigated that is one needs to scrutinize the claim that the success in the test constitutes the ability of language interaction in Turing's machine. It is better to propose the question in this way: is typing the questions by the interrogator and the written answers of the machine for deceiving the interrogator constitutes machines' language interaction with the interrogator? It is obvious that if there were no need to deceive the interrogator and if the machine was programmed so that it was only able to answer the interrogator's questions, the machine's written interaction with the interrogator would not be considered our intended language interaction, which requires thinking. Therefore, what distinguishes this language interaction is the ability of the machine with regard to analysis and identification of the situation for giving answers so that the interrogator does not recognize the interlocutor is a machine. In the rest of this article, I intend to determine when a creature is considered to possess the ability of language interaction, or in other words, what the features of language interaction are.

In order for the machine to succeed in the imitation game, it must be able to deceive the interrogator through written language interaction. This at least requires two abilities:

A. The ability of interpreting interrogator's speeches and words which are provided as the interrogator's questions.

B. The ability of giving dishonest answers to the questions so that the interrogator does not realize or does not suspect they are wrong.

In the article "Thought and talk", Davidson deals with the main feature of language interaction. He believes that we generally assume that the ability of language interaction is to a great extent the ability of speaking with others, but speaking plays only an indirect part in this ability. What Davidson considers necessary for having the ability of language interaction is the idea of an interpreter (the user of the language), that is someone who understands the utterances of others. He is of the conviction that a speaker must have the ability of interpreting and understanding other people's utterances and this is a necessary in language interaction (Davidson, 1975). If we define language interaction as the ability of a creature to interpret the utterances of others and to speak, based on Davidson's view, the ability of interpreting other people's utterances is necessary for the ability of language interaction and thus the ability of talking to others has an indirect and less important role. Based on principles A and B, we can see that a machine needs both these skills in order to succeed the test. It can thus be concluded that if succeeding the test, the machine has the ability of language interaction with the interrogator.

This issue can also be investigated from another perspective. As it was explained before, defending the Turing Test by using Davidson's no-priority theory requires proving the statement that "if the machine succeeds the Turing Test, it has the ability of language interaction". I change this statement on the basis of Davidson's view to "if Turing machine deceives the interrogator, it has the ability of interpreting interrogator's utterances" because deceiving the interrogator by Turing machine implies Turing machine's success in the test and the ability of interpreting others' utterances, on the basis of Davidson's view, is a necessary feature for language interaction. Therefore, such a substitution does not create any problems.

After this substitution, instead of proving the ability of language interaction for the machine in the case of succeeding the test, we need to prove that the machine is in need of interpreting the interrogator's utterances in order to be able to deceive him/her. In order to deceive the interrogator, the machine first needs to translate the interrogators' questions (i.e. his utterances) from a foreign language (interrogator's language) to a familiar language

(machine's language). This is due to the fact that in order to understand the content of the questions and answering them, the machine first needs to translate the sentences typed by the interrogator. Up to this stage, the machine's performance does not constitute interpreting the interrogator's utterances. This process, that is the translation of the sentences from one language to another can easily be performed using a usual dictionary and thus does not require language interaction.

What distinguishes the machine in the Turing Test (if succeeded) from a usual dictionary, which translated one language into another one, is the skill of machine in translating the sentences according to the intended situation; in other words, in order for the machine to deceive the interrogator, in addition to translating the interrogators' utterances, it must be able to determine what the meanings of those sentences in that particular situation are and it must be able to put the translation of the utterance next to each other so that they make sense. This is in fact the ability of interpreting people's utterances. In other words, the meaning of sentences depends on their structure and sentences are not interpreted independently and separately. Moreover, a dictionary can also translate a limited number of words meaningfully, but with the increase in the number of words, the translation would not be coherent and meaningful. It would just be a translation of single independent words and with no attention to the relationship between the sentences. Therefore, it can be said that translation and interpretation are different from one another and the capability of a dictionary is restricted to translation of words and sentences, while the ability of the machine in the Turing Test, in case of success, is simultaneous translation and interpretation.

In "Radical Interpretation", Davidson considers translation a passage from one language to another; that is a relationship between the two languages; however, according to Davidson, interpretation is a passage in the same language realm. He explained this issue as:

> "In the general case, a theory of translation involves three languages: the object language, the subject language, and the metalanguage (the languages from and into which translation proceeds, and the language of the theory, which says what expressions of the subject language translate which expressions of the object language). … And in this general case, we can know which sentences of the subject language translate which sentences of the object language without knowing what any of the sentences of either language mean (in any sense, anyway, that would let someone who understood the theory interpret sentences of the object language). There remains the fact that the method of translation leaves tacit and beyond the reach of theory what we need to know that allows us to interpret our own language." (Davidson, 1973, pp. 129-130).

Therefore, it can be said that in order to understand and interpret the utterances of the interrogator for answering his/her questions, the machine in the Turing test needs a theory of interpretation along with a theory of translation. Thus "translation guide" of the machine provides a sentence in machine's language for any utterance said in the interrogator's language. Then theory of interpretation provides an interpretation for any of these familiar sentences. According to this interpretation and strategy, the proper answer will be selected. Therefore, the ability of interpreting the interrogators' utterances is the result of integrating the theory of translation and theory of interpretation in the machine. Consequently, for the machine to deceive the interrogator, it needs to translate and interpret the utterances of the interrogator and the accuracy of the substituted statement with the original one –that is "if Turing machine deceives the interrogator, it has the ability of interpreting interrogator's utterances, has been proved.

5. Conclusion

It can be said that in order for the machine to deceive the interrogator and thus succeed in the test, it needs to translate and interpret the utterances of the interrogator. Moreover, considering Davidson's theory, such an ability – that is, the ability of interpreting the utterances of people – is the necessary feature of the ability of language interaction. This indicates that the machine has the ability of language interaction. As a result, the original statement, which we intended to show, has been demonstrated. Using the second result of Davidson's theory (principle 2) and based on the explained process, it can be said that the intended machine can think. Thus, one can use Davidson's no-priority theory as an approach for filling the argumentative gap in Turing's article for replacing the original question with the new one.

Acknowledgement

I thank Dr. Mahdi Nasrin from Iranian institute of Philosophy (IRIP), for his invaluable aid. I also thank the anonymous referees who added to the value of this work by their valuable suggestions.

Reference

Davidson, D. (1973). Radical interpretation. In D. Davidson (Ed.), *Inquiries into truth and interpretation* (pp. 125-140). Oxford: Clarendon Press.
Davidson, D. (1975). Thought and talk. In D. Davidson (Ed.), *Inquiries into truth and interpretation.* (pp. 155-170). Oxford: Clarendon Press.
Davidson, D. (1990). Turing's test. In D. Davidson (Ed.), *Problems of rationality.* (pp. 77-86). Oxford: Clarendon Press.
Turing, A. M. (1950). Computing machinery and intelligences. *Mind, 59*, 433-460.