AI & SOCIETY (2020) 35:441–450
https://doi.org/10.1007/s00146-019-00906-x

OPEN FORUM

Classical AI linguistic understanding and the insoluble Cartesian problem

Rodrigo González[1]

Received: 12 June 2019 / Accepted: 8 August 2019 / Published online: 19 August 2019

Abstract
This paper examines an insoluble Cartesian problem for classical AI, namely, how linguistic understanding involves knowledge and awareness of *u's* meaning, a cognitive process that is irreducible to algorithms. As analyzed, Descartes' view about reason and intelligence has paradoxically encouraged certain classical AI researchers to suppose that linguistic understanding suffices for machine intelligence. Several advocates of the Turing Test, for example, assume that linguistic understanding only comprises computational processes which can be recursively decomposed into algorithmic mechanisms. Against this background, in the first section, I explain Descartes' view about language and mind. To show that Turing bites the bullet with his imitation game and in the second section I analyze this method to assess intelligence. Then, in the third section, I elaborate on Schank and Abelsons' Script Applier Mechanism (SAM, hereby), which supposedly casts doubt on Descartes' denial that machines can think. Finally, in the fourth section, I explore a challenge that any algorithmic decomposition of linguistic understanding faces. This challenge, I argue, is the core of the Cartesian problem: knowledge and awareness of meaning require a first-person viewpoint which is irreducible to the decomposition of algorithmic mechanisms.

Keywords Descartes · AI · Meaning · Awareness · Turing test

1 Introduction

In contemporary philosophy, Descartes has been fiercely criticized for posing an insoluble problem, viz., The Mind–Body Problem. For him, two different substances, i.e., mind and body, can be conceived separately; and in virtue of God's supreme powers, He can separate anything which we can conceive separately. Hence, these substances are supposed to be metaphysically different. In particular, Descartes argues that it is conceivable that mental states exist *without* a body, and since at least God can separate mind and body, they are metaphysically different. Despite this conclusion, certain philosophers think that divine intervention is pointless. For them, Descartes, who is regarded as the paradigmatic example of Dualism, leads us into a dilemma regarding mind–body interaction (Crane 2003, p. 230).

Nevertheless, the Mind–Body Problem is not the only factor for which Descartes has been accused of being *the* culprit in posing an insoluble problem. Materialists criticize the French philosopher in view of a new science, viz., Artificial Intelligence (AI hereafter). AI researchers criticize Descartes for holding that the mind is an immaterial soul (Crane 2003, p. 45). The Cartesian argument they aim to counter is as follows: Since the soul is not the body and the body is a machine, the soul cannot be a machine. The rationale for the Cartesian argument is the following: If the mind is not a physical substance, and machines work in virtue of physical mechanisms, like cogs and levers, then machine intelligence should be impossible in principle. Indeed, in human language we produce infinite and flexible outputs (Descartes 1985b, p. 139–140, AT VI, 56–57,[1] CLARKE 2003, p. 166-171). But as I will argue, awareness of meaning underlies our understanding of such outputs. Since such

✉ Rodrigo González
rodgonfer@gmail.com

[1] Department of Philosoph and Center for Cognitive Studies, Faculty of Philosophy and Humanities, Universidad de Chile, Av. Ignacio Carrera Pinto 1025, Ñuñoa, Santiago, Chile

[1] Whenever I quote Descartes, I add the Adams Tannery (AT) convention.

awareness exists from a first-person viewpoint, an insoluble Cartesian problem for machine intelligence arises.

Paradoxically, several classical AI researchers have assumed part of Descartes' view about language and intelligence.[2] For example, *chatbots* are supposed to produce different arrangements of conventional signs so that they answer questions in different ways. This process supposedly implies 'linguistic understanding,' a telltale sign of intelligence and mind. In relation to this assumption, Turing's Test is supposedly the proper method that serves to justify the fact that programmed machines can exhibit linguistic understanding[3] (MOOR 1976 and Copeland 2019) and, consequently, intelligence.

This essay is divided into four sections. In the first section, I discuss the Cartesian metaphysical view according to which the mind is easier to know than the body. Hence, conceiving mind and body separately implies a real distinction. If this is correct, it follows that any machine based upon physical devices, and upon any sort of limited mechanisms, cannot in principle think. In the second section, I then describe how Turing, employing a famous method, viz., the imitation game, attacked this view. But to achieve this end, Turing had to assume part of Descartes' view, namely, he had to endorse that linguistic understanding implies the existence of intelligence and mental states. In the third section, I argue that SAM, i.e., Schank and Abelson's Script Applier Mechanism, has the aim of casting doubt upon Descartes' view. Inspired by Turing, certain classical AI researchers have come to believe that linguistic understanding can be reduced to computations, because these are *purely* mechanical. Finally, in the last section, I explore a challenge that the mechanization of linguistic understanding needs to face. In fact, I concentrate upon showing that linguistic understanding exists from a first-person viewpoint, which in turn leads to an insoluble Cartesian problem for classical AI.

2 The origin of an insoluble problem: the impossibility of mechanical reason

Even though Descartes is one of the major philosophers of our time, his views are quite controversial, both for the layperson and the expert. In fact, his real distinction between *res cogitans* and *res extensa* has provoked a heated debate about the potential mechanization of reason.[4] The *real* distinction in Descartes' Meditations is directly related to the impossibility of mechanizing reason, which Descartes' states as follows:

> First, I know that everything which I clearly and distinctly understand is capable of being created by God so as to correspond exactly with my understanding of it. Hence the fact that I can clearly and distinctly understand one thing apart from the other is enough to make me certain that the two things are distinct, since they are capable of being separated, at least by God. […].
> Thus, simply by knowing that I exist and seeing at the same time that absolutely nothing else belongs to my nature or essence except solely in the fact that I am a thinking thing, I can infer correctly that my essence corresponds solely in the fact that I am a thinking thing. It is true that I may have (or, to anticipate, that I certainly have) a body that is very closely joined to me. But nevertheless, on the one hand I have a clear and distinct idea of myself, in so far as I am only a thinking, non-extended thing; and on the other hand, I have a distinct idea of body, in so far as this is simply an extended, non-thinking thing. And accordingly, it is certain that I am really distinct from my body, and can exist without it. (Descartes 1985a, p. 54, AT VII, 78).

This Cartesian argument, dubbed "Descartes' modal intuition", aims to show that attributes distinguishable in understanding can thereby be distinguished in reality. God's

[2] Here I assume the Cartesian co-extension of the terms reason, intelligence and mind.

[3] Longsworth addresses the issue of whether linguistic understanding is a form of knowledge. In doing so, he states what *desiderata* a theory of linguistic understanding must satisfy. His discussion is directly relevant to this essay, because he examines in what sense understanding an utterance needs one to *know and* be *aware of the meaning that expresses.* He states the point as follows: "[…] those states [of understanding] must be of a sort able to interact with ordinary states of belief, knowledge, etc., and to play the same sort of role as those other states in shaping the subject's consciousness […] What we seek in an account of state-understanding is an account of how such states can play a role in ordinary psychology, how occupying them can impact on the rational development of one's cognitive economy" (Longsworth 2008, p. 51–52). For the sake of argument, I consider the awareness of *u's meaning* as parasitic upon the knowledge required to understand *u*.

[4] See, for example, Marciszewski and Murawski (1995). In their book they assert that the mechanization of reason is paradigmatic in Leibniz's logical calculi and Boole's Algebra of Logic. The relation between logic and the mechanization of reason is pertinent not only for AI, but also for Cognitive Science. Copeland (1993, p. 10), for example, maintains that the philosophy of AI is prior to AI, since Turing wrote "Computing Machinery and Intelligence" in 1950. Dartmouth conference, which gave AI its name, was organized in 1956. Ever since, Minsky stated AI's goal thus: "Artificial Intelligence is the science of making of machines do things that would require the intelligence if done by men" (Copeland 1993, p. 1). I quote this passage in order to show how AI's main goal evolved, from the mechanization of reason, in the 19th century (with Babbage for example), to the making of machines that simulate intelligence, after 1956. Here I mean by 'classical AI' the approach that attempts to making intelligent machines upon the basis of formal rules and representations.

almighty powers are furthermore essential to the argument: He can do what we clearly and distinctly conceive. In this case, Descartes holds that he can conceive a clear and distinct idea of the cogito, as a non-extended substance, and he can conceive a clear and distinct idea of the body, as an extended substance. Since he can conceive the mind, as a non-extended substance, and the body, as an extended substance, the mind must, therefore, be distinct from the body, and so can exist without it. This assumption is doubtless controversial. In fact, Descartes' opponents have countered the argument by claiming that what can be distinctly conceived in understanding need not be the case in reality.[5]

In spite of this possible objection, the French philosopher insists on another crucial difference between mind and body, as only the latter can be divided. His addendum to his Dualism is as follows:

> […] There is a great difference between the mind and the body, inasmuch as the body is by its very nature always divisible, while the mind is utterly indivisible. For when I consider the mind, or myself in so far as I am merely a thinking thing, I am unable to distinguish any parts within myself; I understand myself to be something quite single and complete. (Descartes 1985a p. 59, AT VII, 86).

The above passages illustrate the metaphysical difference between mind and body. The mind, which is intermingled with the body,[6] nonetheless exists unlike any physical thing. It turns out that we are merely thinking things. The mind is single and complete, unlike the divisible body. If this view is correct, and the real distinction is also correct, we can clearly recognize that the mind is not the body because it is impossible to conceive the two as being identical. When we think of an extended thing, nothing mental is conceived, and vice versa. Elsewhere, Descartes argues in relation to the piece of wax, viz., that its essence is to be an extended thing, which is only perceived by the intellect. Hence, the essence of the piece of wax is grasped as a purely mental scrutiny (Descartes 1985a, p. 21, AT VII, 31). As a result, Descartes' Dualism entails the impossibility of endowing a machine, or any physical thing (qua physical), with mental states. The argument is sufficiently clear and so is the dualist approach involved, despite the discussion among scholars on whether Descartes embraces Substance or Property Dualism.[7]

Descartes, who claims that machines cannot think in principle, also provides an empirical argument: Machines cannot produce different linguistic sign arrangements to mean the same thing and, thus, machines cannot use language like humans. This passage, from the Discourse on the Method, explains the Cartesian criteria for the existence of mental states:

> […] If any such machines had the organs and outward shape of a monkey or of some animal that lacks reason, we should have no means of knowing that they did not possess entirely the same nature as these animals; whereas if any such machines bore a resemblance to our bodies and imitated our actions as closely as possible for all practical purposes, we should still have two very certain means of recognizing that they were not real men. The first is that they could not use words, or put together other signs, as we do in order to declare our thoughts to others. For we can certainly conceive of a machine so constructed that it utters words, and even utters words which correspond to bodily actions causing a change in its organs (e.g., if you touch it in one spot it asks what you want of it, if you touch it in another it cries out that you are hurting it, and so on). But it is not conceivable that such a machine should produce different arrangements of words so as to give an appropriately meaningful answer to whatever is said in its presence, as the dullest of men can do. Secondly, even though such machines might do some things as well as we do them, or perhaps even better, they inevitably fail in others, which would reveal that they were acting not through understanding but merely from the disposition of their organs. For whereas reason is a universal instrument which can be used in all kinds of situations, their organs need some particular disposition for each particular action; hence it is for all practical purposes impossible for a machine to have enough different organs to make it act in all the contingencies of life in a way in which our reason makes us act […] (Descartes 1985b, p. 139–140, AT VI, 56–57, my emphasis).

Descartes' argument against machine-like intelligence can be summarized as follows:

1. Machines are physical things.
2. Physical things can only produce finite inflexible linguistic outputs.
3. Machines produce finite inflexible linguistic outputs.
4. If machines only produce finite inflexible linguistic outputs, they cannot think.
5. Machines, qua physical things, only produce finite inflexible linguistic outputs.

[5] See, for example, Hill (2002). This materialist philosopher attacks Descartes' Dualism by stating that what can be conceived need not be the case.

[6] The fact that two things are joined does not entail that they are identical. Descartes emphasizes this point in the sixth Meditation, with the pilot and the ship dis-analogy (Descartes 1985a, p. 56, AT VII, 81).

[7] Take, for instance, Clarke's theory (2003).

Therefore, machines cannot think.

In other words, since machines do not put together signs in a flexible and relevant way, they can only produce finite outputs based upon the disposition of their mechanisms. As a result, machines cannot understand language, because they cannot give appropriately meaningful answers. Thus, Descartes concludes that it is impossible that machines think.

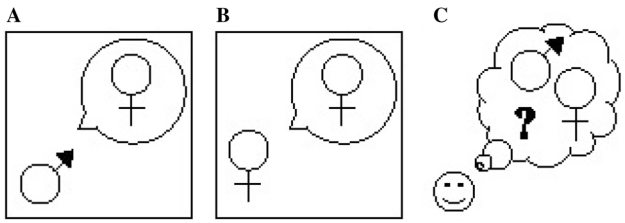
In light of Descartes' view, it is worth pointing out that a machine could *simulate* the use of words, and hence the process of thinking.[8] The French philosopher considers the possibility that a sophisticated machine is created, but this machine could not replicate how words are arranged by genuine reason and intelligence, which are universal instruments. Therefore, even if such a machine managed to use words, it could only count as an ingenious contrivance. In other words, even if a machine exhibited *as-if* intelligence and reason, it would only *imitate* human linguistic behavior, and further, the actions involved in true intelligent behavior.

The Cartesian view, then, has been a crucial challenge for classical AI, at least with regard to the original project as proposed by Babbage and others. The next section, which briefly covers the imitation game, shows how the Turing Test attempts to solve the insoluble Cartesian problem. But, as I argue, Turing bites the bullet here: he believes that if chatbots are appropriately programmed, they would pass the Turing Test and dissolve Descartes' seemingly insoluble problem.

3 Turing: linguistic understanding, a telltale sign of intelligence

Alan Turing faced the insoluble Cartesian problem by addressing a Cartesian question, namely, "Can a machine think?" To this end, he published "Computing Machinery and Intelligence" (Turing 1990), where he explores the issue of whether programmed machines could have mental states. Unlike certain philosophers, Turing introduced a new element of analysis. In fact, his paper dealt with the replacement of the abovementioned Cartesian question. He thought that the replacement was necessary, because the question led to the use of concepts such as "machine" and "intelligence", and eventually to polls about their meaning. Turing's replacement consisted in putting forward a game whose aim was to gather empirical evidence such that programmed machines could deceive human judges, making them believe that programmed machines were humans as well. By doing so, it could be shown that computers may understand language and think. Accordingly, to get rid of the question, Turing proposed the imitation game, which is the core of the so-called Turing Test.

The imitation game has three stages. In the first stage, Turing conceives of a man in room A, a woman in room B, and judges outside the rooms. Each participant plays a specific role: the judges, for instance, need to determine whether they are in the presence of a man or a woman. To attempt to identify the woman or the man, the judges pose questions to them, which are simple and brief. While the woman answers the questions sincerely, the man pretends to be a woman, by answering as though he had female-like intelligence. The questioning, which lasts 5 min or so, focuses upon simple things to determine the sex of the participants in A and B. For example, a typical question is "Do you have short hair?" Consequently, the first stage of the Turing Test can be depicted as follows:

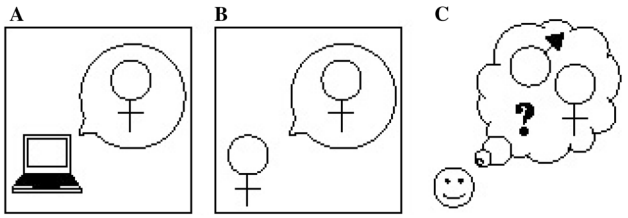


Now, it is worth mentioning an important point about this first stage of the imitation game. If the man can deceive a significant number of the judges, it is said that he has passed the test. What certain commentators have systematically neglected is the importance of identifying the sex of participants (González 2015). Such an identification serves an important purpose. Indeed, it intends to show that a woman's intelligence and her body are metaphysically different and can be separated. Therefore, it is possible to have a man's body and a woman's particular linguistic performance. I deal with this issue below, because it is crucial first to understand Turing's Functionalism, and second to how the British mathematician embraces the Cartesian criteria of intelligence.

After devising the second stage of the imitation game, Turing poses the following question: What if a programmed computer replaced the man in room A, giving the same sort of answers that a woman might give? If intelligence can be separated from its material realization, and from the biological (physical) properties of a woman's brain, then a programmed machine could imitate her linguistic performance. Surely, the programmed machine would need to have enough

[8] I am grateful to an anonymous referee for this journal for raising the issue that AI's main project is the simulation of intelligence. In fact, Bostrom (2014, p. 6) remarks that "[…] on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it". In this essay, I assume that prior to the simulation of intelligence, Leibniz Babbage, Boole and others attempted the mechanization of reason. Take, for example, Babbage's difference and analytic engines (SWADE 2000).

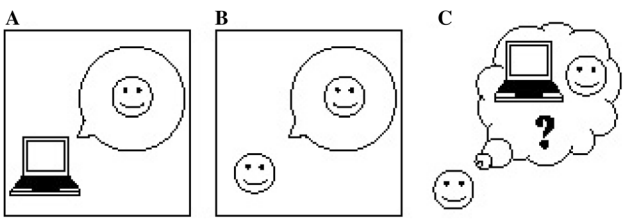
storage capacity and speed to imitate a woman's intelligent linguistic performance (and/or of a man's, as it is clear in the second stage). This second stage, then, can be depicted as follows:



But a caveat is necessary here. In relation to the first and second stages of the game, it simply does not seem to be the case that a gender interpretation is possible. On the one hand, a man does not surpass a woman's intelligence and vice versa. This shows that, despite some interpretations, Turing did not favor a gender revolutionary project. This is worth mentioning, as some interpretations, which have focused upon gender (Genova 1994; Lassègue 1996), claim that Turing had a political agenda in "Computing Machinery and Intelligence." However, it is unlikely that Turing had had such a political project in mind. Indeed, his aim, at least what he explicitly declares in 1950, is to provide a philosophical justification for the plausible success of classical AI.

A point that supports this view is that Turing himself (2004b) simplified his test. He likely believed that the identification of the participants' sex may lead to misinterpretations regarding the purpose of the imitation game. This simplification gave rise to the standard version of the game, which is as follows: A programmed computer is in room A, while a person whose sex is irrelevant is in room B. The judges assess whether they are in the presence of a person or a programmed machine. The person answers the questions sincerely, whereas the programmed computer pretends to be a person. Based on this version of the imitation game, Turing made a prediction: By the year 2000, an average interrogator will not have more than 70 percent chance to make the right identification (Turing 1990, p. 49). Why? Because the machine's understanding of language seems to bear evidence of the existence of mental states, as any Cartesian would be forced to accept.

The third stage (the standard version of the imitation game) can be depicted as follows:



Here I resume the analysis of a point made above. The imitation game is consistent with a functionalist approach to the mind. Incidentally, this approach is deeply unbiological. The capacity to have mental states is regarded by Turing as an unbiological functional property, i.e., as a function that does not require a realization in any specific physical material, like the brain.[9] Turing in turn holds that a computer can be realized in different materials, like a Turing Machine (Turing 1936).

Functionalism, which is later called "Machine Functionalism" (Putnam 1967, 1973), holds that intelligence is a merely *computable* function which does not depend upon any material realization. Indeed, Turing's Functionalism is compatible with Cartesian Dualism in the sense that both hold that mind and intelligence can be separated from the brain (cf. González 2011). Accordingly, while Descartes' view holds that it is conceivable that the mind can be separated from the body, Turing's view holds that intelligence can be separated from the brain through computer programming.

All these suppositions sparked a wave of skepticism about the possibilities of AI to succeed. For example, John Searle (1980) has been particularly critical in this regard because both classical AI and Cognitive Science are profoundly unbiological. Searle's main criticism shows that Turing's view supposes that a programmed machine can be realized in any material, which is the opposite of real cognition and intelligence.[10] Searle's argument against Machine Functionalism is as follows:

> […] indeed strong AI only makes sense given the dualistic assumption that, where the mind is concerned, the brain doesn't matter. In strong AI (and in functionalism, as well) what matters are programs, and programs are independent of their realization in machines; indeed, as far as AI is concerned, the same program could be realized by an electronic machine, a Cartesian mental substance, or a Hegelian world spirit. (Searle 1980, p. 86).

The Searlean criticisms against classical AI are also directed toward Schank and Abelson's theory. Both hold that SAM, their computer program, simulates an understanding of language. In fact, SAM presupposes the Cartesian view: If any agent or machine understands language, then they clearly think.

[9] Turing is clear about his un-biological functionalism when he states: "if now some particular machine can be described as a brain we have only to program our digital computer to imitate it and it will also be a brain" (2004a, p. 112).

[10] I will return to this criticism and to the issue of linguistic understanding in the final section.

4 SAM: AI's classical paradigm of linguistic understanding

Despite the large number of criticisms against the Turing Test, many commentators regard it as a valid method for assessing machine intelligence. To debunk the critics, they claim that the test is simple and effective for gathering empirical evidence that bears evidence of machine intelligence. In fact, the test justifies the fact that programmed computers understand language and, thus, that they have intelligence and mental states. The supposition that underlies the game is that any programmed machine that understands language, therefore, thinks.

Both Schank and Abelson (1977), two AI researchers, are advocates of Turing's test. The program designed by them, SAM (Script Applier Mechanism), was inspired by a general theory of understanding. On Schank and Abelson's view, people understand the situations in which they are in by interacting with an organized knowledge database. Any process of linguistic understanding requires general knowledge, which allows a person to interpret other people's behavior based upon certain shared and standardized needs and methods (for example, when a person goes to bed). However, the interpretation is also a function of specific kinds of knowledge, which makes possible the explanation and participation in daily life scenarios (for example, why tickets are shown at cinemas). But what is the connection between the knowledge database and the understanding of a new story?

According to Schank and Abelson, new stories are understandable in function of certain scripts, or certain standardized sequences of events, all from the viewpoint of an actor. Consider, for example, these stories:

1. Flor came in the pub and ordered a beer. She paid the bill and left.
2. Flor did not go to the party. Abe did not leave any tip, as usual. Police were stopping drunk drivers.
3. Flor fancied buying an umbrella. She found a flyer of a sale on the bus. Umbrellas were sold out.
4. Flor was at the supermarket. At the till she realized that she had lost her wallet.

While story 1 made reference to a clear script [PUB], which typically involves a sequence of events (desiring a drink, sitting at the bar, ordering a drink, asking for the bill, and so on), all of which make sense, story 2, seems to lack a clear script, because it links events which are not necessarily related. For this reason, story 2 has little sense and, without extra information, it turns out to be very hard to understand. By contrast, story 3 overlaps two scripts [BUS] and [SALE], which allow us to understand what the story is about. Likewise, story 4 is also connected to two scripts [SUPERMARKET] and [SHOPPING]. Despite being very brief, story 4 is clear and meaningful.

Even if complications arise, such as interruptions and interference in the sequence of events, matching scripts allow us to take part of daily life situations and to understand stories without boring details. Given the fact that scripts are internalized, people can anticipate future events and can even infer information that has not been explicitly stated. According to Schank and Abelson, implicit information given as answers to questions is a telltale sign of linguistic understanding.[11] For example, for those who understand stories 3 and 4, Flor's frustration is anticipated, that is, her pity about not buying an umbrella, and her shame about losing her wallet.

Nevertheless, according to Schank and Abelson, people understand stories like programmed machines, that is, as the result of the manipulation and processing of scripts. In particular, both remark that SAM and the human being understand stories as follows:

> Understanding then, is a process by which people match what they see and hear to pre-stored groupings of actions that they have already experienced. New information is understood in terms of old information. By this view, man is seen as a processor that only understands what it has previously understood. Our script-based program, SAM, works this way […] A human understander comes equipped with thousands of scripts. He uses these scripts almost without thinking (Schank and Abelson 1977, p. 67, my emphasis).

Schank and Abelson agree with Turing in that intelligence is the result of computing, or information processing. For all of them, mental states, such as those involved in understanding, can be recursively decomposed into mechanical operations, such as script fetching, retrieving, and matching. Moreover, the fact that a programmed machine answers questions about stories offers compelling evidence about the existence of mental states. This is what Searle dubs "Strong Artificial Intelligence", an approach which he intends to debunk with his famous and controversial Chinese Room (1980, 1990). Such a thought experiment shows that it is possible to manipulate Chinese symbols and understand no Chinese at all. The moral then is that syntax, or computer programs, is insufficient for semantics and mental content.

[11] In Sect. 1, I defined linguistic understanding in terms of knowledge and awareness of certain meaningful expression. Still, I note that such knowledge and awareness can also be applied to matching stories and scripts. In fact, although the state-understanding process of stories is more complex than that of expressions, in both cases understanding requires S's knowledge and awareness.

Consequently, "understanding stories" offers little if any compelling evidence for the existence of mental states and intelligence. As I examine below, running programs that make machines answer by script matching does not indicate true linguistic understanding.

5 Mechanized linguistic understanding: psychologically plausible?

Despite what some people may think about the anti-Cartesianism of Cognitive Science, Descartes and Turing seem to be part of a symbiotic relationship. Although Descartes denies that any machine can think in principle, Turing bites the bullet by supporting Machine Functionalism. As noted, this view holds that intelligence is unbiological and can be separated from the brain. Furthermore, on Turing's view, linguistic understanding through information processing bears evidence of the existence of mental states. As I have analyzed above, Turing sought to demystify Descartes' insoluble argument against classical AI, i.e., the impossibility of machine intelligence in principle. To debunk this Cartesian myth, Turing firstly assume that theory T is true, like in a negative thought experiment (Brown 2007). That is, we must assume that linguistic understanding suffices for machine intelligence. But such an assumption has misguided classical AI researchers into believing that the use of words bears evidence of intelligence.[12]

In this section, I concentrate on whether it is psychologically plausible to mechanize linguistic understanding. By "psychological plausibility"[13] I mean the possession of psychologically relevant conscious experiences that can support how understanding can be mechanized, and whether such experiences can be integrated to our cognition. For example, it is not psychologically plausible to believe that we are only molecules. To the contrary, we have psychologically relevant conscious experiences that show that we are far more complex than mere molecules. Molecules, unlike us, cannot have a first-person viewpoint, an issue I shall address below. Similarly, the psychological plausibility of being like a bat is impossible. Desiring to be like a bat, as Nagel examines (1974), leads to imaginations in which we, humans, behave as though we were bats, in a would-be batman behavior. To sum up: My argument here is that appealing to human psychological experiences can establish whether linguistic understanding can be mechanized.

Mechanizing linguistic understanding is not psychologically plausible. Two reasons seem to be sufficient for drawing this conclusion. In the first place, the very notion of linguistic understanding is still controversial, as it requires someone who possesses both knowledge and an awareness of meaning. No one can truly hold that they understand *u* if they have neither knowledge nor awareness of *u's* meaning. We know from an inner perspective, and thus from a first-person viewpoint, when we understand *u*. Searle is emphatic on this point, for example:

> In many of these discussions one finds a lot of fancy footwork about the word 'understanding.' My critics point out that there are many degrees of understanding; that 'understanding' is not a simple two-place predicate; that there are even different kinds and levels of understanding, and often the law of excluded middle does not even apply in a straightforward way to statements of the form 'x understands y' […].

To all these points I want to say: of course, of course. But they have nothing to do with the points at issue. There are clear cases in which 'understanding' literally applies and clear cases in which it does not apply; and these two sorts of cases are all I need for this argument (Searle 1980, p. 418–419).

The second reason is that we can argue on the basis of the Chinese Room thought experiment that no mental states are caused by algorithmic processing and, thus, by mere syntax. Implementing algorithms does not imply the possession of conscious mental states, which is precisely the point at issue in the Chinese Room. Despite the multiple objections to Searle's argument, it seems to be clear that a computer program's capacity to linguistically manipulate symbols is simply insufficient for the realization of semantic understanding (and for intentionality). Moreover, given the symbolic forms and the rules that allow us to manipulate such symbols, nothing interesting arises from a psychological viewpoint. This point reminds us of the above stated first reason, which I now elaborate more fully below.

To support these two reasons, we might consider the question of what the essence of computing is. There is nothing we can imagine in relation to what-it-is-like-to-be-a-programmed-computer (Block 1995, p. 270). If we run an algorithm, nothing psychologically interesting occurs from the point of view consciousness, and alternatively, from a first-person viewpoint. If this is correct, then there is no what-it-is-like-to-be-SAM. Furthermore, SAM would count as an as-if understander of expressions and stories, but the program does not *explain* the psychological process by which someone knows and is aware, from the

[12] Take, for instance, Colby's program (1975), PARRY. Such a program is the simulation of a paranoid person. All the emphasis is on PARRY's psychotic linguistic behavior, which is reflected in the answers given by the program.

[13] Here I mean "psychology" in a broad sense, that is, the scientific study of cognitive intelligence and behavior. Moreover, by "psychologically plausible" I mean the possibility of imagining consciousness experiences, and how these experiences are supposed to be integrated to our cognitive intelligence and behavior.

first-person viewpoint, whether such expressions or stories have meaning.[14]

Searle fully understands English, as a native English speaker. He partially understands French and nothing of Chinese. In the three cases, the understanding process has an inner perspective: Searle knows and is aware of *u's* meaning. The connection between understanding and semantics is obvious: the key point of Searle's Chinese Room is that syntax is neither sufficient nor constitutive of semantics. For example, Searle, as a native English speaker, knows what the word 'insofar' means. Secondly, Searle's French proficiency allows him to understand that "Un demi" refers to a draught beer, for example. Finally, someone unable to speak Chinese, like him, is incapable of knowing and being aware of the fact that soda is "Sūdă" in the Chinese language.

However, it is necessary to introduce a final example of computing that does not include any psychological experience. Euclid's famous algorithm will serve to this end (Penrose 1999, p. 41–44). The point I stress here is that algorithms, or programs, work in the same way as SAM, namely, by running a program with finite algorithmic steps. Still more, the automatism of these steps does not lead to any significant psychological experience.

According to the Euclidean algorithm, to determine the greatest common divisor between two numbers, these rules, which are expressed in finite steps, must be followed. The last one is recursive:

(i) Divide the greater number (N) by the lesser number (M), taking from the result the remainder (R);
(ii) If $R=0$, *halt*;
(iii) If $R\neq 0$, then divide M by R, and go to step 1.

For example, to determine the common maximum divisor of numbers 99 and 15, we have to do this:

N	M	#Divisions	R
99	15	6	9
15	9	1	6
9	6	1	3
6	3	2	0

Once the finite algorithmic steps are followed, it is possible to establish that the greatest common divisor of 99 and 15 is 3. In this example, it is crucial that no psychological experience in relation to mathematics exists. In fact, the algorithm could be followed by someone ignorant of mathematics, as also by an expert. Neither would have an interesting psychological experience associated with running algorithms, and for this reason no conscious mental states are caused in any sense. That is, no agent who follows Euclid's rules for the algorithm would have a conscious psychological experience associated. The same, mutatis mutandis, occurs with linguistic understanding, as Searle's Chinese Room shows. No mental states are caused after running the program, inside the room.

This kind of example shows that we cannot get rid of the insoluble Cartesian problem. Descartes seems to be correct when he holds that language and mind cannot be mechanized. The rationale is not what Descartes imagines though. Mechanisms are finite and inflexible, whereas reason is a flexible universal tool that can be applied to any sort of problem. In turn, my view is that either reason and language cannot be mechanized or, at least, the above arguments do not support their mechanization, since nothing psychologically interesting occurs after running an algorithm. Hence, AI cannot teach us anything about real intelligent cognition by running chatbots, e.g., even if these succeed in deceiving us.

In summary, my argument about the insoluble Cartesian problem for classical AI is as follows:

1. If S linguistically understands u, S achieves knowledge and awareness of *u's* meaning.
2. If S achieves knowledge and awareness of *u's* meaning, S does so from a first-person viewpoint.
3. If S linguistically understands u, S does so from a first-person viewpoint.
4. If programmed machines linguistically understand u, they must do so from a first-person viewpoint.
5. Programmed machines have no first-person viewpoint.

Therefore, programmed machines cannot have linguistic understanding of u.

There is a hypothetical syllogism between premises 1 and 2, which supports 3 as a preliminary conclusion. Then, there is a *modus tollens* between 4 and 5. The outcome is that programmed machines have no linguistic understanding of u, which shows the core of the insoluble Cartesian problem for classical AI.

6 Conclusion

Traditionally, Cartesian metaphysics has been regarded as a theoretical foundation that favors Dualism, that is to say, the view according to which two different substances in the world exist: *res extensa* and *res cogitans.* The former is

[14] Whether Searle endorses a Cartesian view when he holds that intentional mental states have conditions of satisfaction, which are *known* by an agent (Cf. Searle 1983, p. 64) is indeed debatable. Also, I examine elsewhere the Systems Reply to the Chinese Room argument, which involves a mechanism that cannot be internalized: the agent's introspection, which is fundamental to run the thought experiment (González 2012). All these points deserve more discussion in another essay.

related to the divisibility of tridimensional things and the latter to the mind, as a non-extended single complete substance. However, Cartesian metaphysics involves a second part, which is directly related to whether machines can think, on the one hand, and to whether reason can be mechanized, on the other hand. As I have examined in this essay, Descartes holds that we are thinking things as opposed to extended things. However, what most commentators have neglected is that mechanizing reason seems impossible if physically extended things are finite and inflexible. Descartes argues that reason is flexible and unlimited, and that rational beings can in turn produce an unlimited arrangements of words. In contrast, programmed machines, qua physical things, cannot produce unlimited flexibly relevant linguistic outputs.

I also examined how classical AI holds that the mind can be mechanized, while Descartes holds that the mind, as a non-extended single complete substance, cannot be mechanized in principle. Evidence of this impossibility is that linguistic understanding of *u* cannot be reduced to mere machine-like behavior. On the contrary, linguistic understanding of *u* leads to an agent's knowledge and awareness of *u's* meaning, that is, to the assessment of understanding from a first-person viewpoint. The Cartesian example of the wax shows that we judge the wax as a purely mental comprehension. By the same token, we judge that we understand *u* from a first-person viewpoint. The latter is what I have dubbed the "insoluble Cartesian problem", because any attempt to reduce linguistic understanding of *u* to mere mechanisms ultimately fails.

In view of the insoluble Cartesian problem, I have provided two examples that show how linguistic understanding is assessed from a first-person viewpoint, and how this assessment is impossible to reduce to mechanisms. If this is correct, the Turing test, which gathers empirical evidence about linguistic understanding from a third-person viewpoint, is not the proper method to assess the existence of intelligence and mental states. What I have remarked upon in regards to this problem is that AI researchers who support Turing's method have partially assumed the Cartesian view according to which the use of linguistic signs is a telltale sign of intelligence. And here I have shown the paradox that follows, namely, classical AI researchers have embraced part of the Cartesian view to refute Descartes himself. But, if the first-person viewpoint cannot be reduced to anything objective, say, to a third-person viewpoint, the insoluble Cartesian problem arises, pace supporters of classical AI. With their enthusiasm, though, they have simply overlooked that the mind, a phenomenon with a first-person viewpoint, is neither reducible to algorithms nor to any sort of machine-like behavior.

References

Block N (1995) The mind as software of the brain. In: Heil J (ed) Philosophy of mind: a guide and anthology. OUP, Oxford, pp 267–274

Bostrom N (2014) Superintelligence: paths, dangers, strategies. OUP, Oxford

Brown JR (2007) Counter thought experiments. R Inst Philos Suppl 61(82):155–177

Clarke D (2003) Descartes's theory of mind. Clarendon Press, Oxford

Colby K (1975) Artificial paranoia. Pergamon Press, New York

Copeland J (1993) Artificial intelligence: a philosophical introduction. Blackwell, Oxford

Copeland J (2019) "The Church-Turing Thesis". Available at: http://plato.stanford.edu/entries/church-turing/. Accessed on 22.06.2019

Crane T (2003) The mechanical mind: a philosophical introduction to minds, machines and mental representation. Routledge, London

Descartes R (1985a) Meditations on first philosophy. In: Cottingham J, Stoothoff R, Murdoch D (eds) The philosophical writings of descartes, vol II. Cambridge University Press, New York, pp 1–62

Descartes R (1985b) Discourse on the method. In: Cottingham J, Stoothoff R, Murdoch D (eds) The philosophical writings of descartes, vol I. Cambridge University Press, New York, pp 109–151

Genova J (1994) Turing's sexual guessing game. Social Epistemology 8(4):313–326

González R (2011) Descartes, las Intuiciones Modales y la IA. In: Revista Alpha, vol 32, pp 181–198

González R (2012) La pieza china: un experimento mental con sesgo cartesiano. In: Revista Chilena de Neuropsicología, vol 7, pp 1–6

González R (2015) ¿Importa la determinación del sexo en el Test de Turing? In: Revista de Filosofía Aurora, vol 27 (January–April), pp 277–295

Hill C (2002) Imaginability, conceivability, and the mind-body problem. In: Chalmers D (ed) Philosophy of mind: classical and contemporary readings. OUP, Oxford, pp 334–341

Lassègue J (1996) What kind of turing test did turing have in mind? Tekhnema 3:37–58

Longsworth G (2008) Linguistic understanding and knowledge. Noûs 42(1):50–79

Marciszewski W, Murawski R (1995) Mechanization of reasoning in a historical perspective. Rodopi, Amsterdam/Atlanta

Moor J (1976) "An Analysis of the Turing test". In: Philosophical Studies 30, pp 249–57. Reprinted in Shieber S (ed.) The turing test: verbal behaviour as the hallmark of intelligence. MIT Press, Cambridge, pp 297–306

Nagel T (1974) What is it like to be a bat? Philos Rev 83:435–450

Penrose R (1999) The emperor´s new mind. Oxford University Press, Oxford

Putnam H (1967) "Psychological predicates." In: Capitan W, Merril D, Art (ed.) Mind, and Religion. Pittsburgh: University of Pittsburgh Press. Reprinted in Heil J (ed) Philosophy of mind: a guide and anthology. Oxford University Press, Oxford, pp 160–167

Putnam H (1973) "The nature of mental states," originally published as "Psychological Predicates." In: Capitan W, Merril D, Art (ed.) Mind, and religion. Pittsburgh: University of Pittsburgh Press. Reprinted in Chalmers D (ed) Philosophy of mind: classical and contemporary readings. Oxford University Press, New York, pp 73–79

Schank RC, Abelson RP (1977) Scripts, Plans, Goals, and Understanding. Hillsdale, NJ: Erlbaum

Searle J (1980) Minds, brains and programs. Behav Brain Sci 3:417–424

Searle J (1983) Intentionality: an essay in the philosophy of mind. Cambridge University Press, Cambridge
Searle J (1990) "Is the brain's mind a computer program?". In: Scientific American, pp 20–25
Swade D (2000) The difference engine: charles babbage and the quest to build the first computer. Penguin, London
Turing A (1936) "On computable numbers, with an application to the *Entscheidungsproblem*". In: Proceedings of the London Mathematical Society, series 2, v. 42, pp 231–65 (with corrections in v. 43), pp 544–546
Turing A (1990) "Computing intelligence and machinery" In: Mind LIX, n. 2236, pp 433–60, Oct. 1950. Reprinted in: Boden, M. (ed) The Philosophy of Artificial Intelligence. OUP, Oxford, pp 40–66
Turing A (2004a) "Can Digital Computers Think?" An interview in BBC, 15 may 1951. Reference of Turing archives: B.5. Reprinted in Shieber S (ed) The turing test: verbal behavior as the hallmark of intelligence. MIT Press, Cambridge, pp 111–116
Turing A (2004b) "Intelligent machinery, a heretical theory", inedited manuscripts of a conference in "51 Society", Manchester, England. Reference of Turing archives: B.4. Reprinted in Shieber S (ed) The turing test: verbal behavior as the hallmark of intelligence. MIT Press, Cambridge, pp 105–109