Articles

The Turing Test, or a Misuse of Language when Ascribing Mental Qualities to Machines

Józef Bremer, Mariusz Flasiński

Abstract In this paper we discuss the views on the Turing test of four influential thinkers who belong to the tradition of analytic philosophy: Ludwig Wittgenstein, Noam Chomsky, Hilary Putnam and John Searle. Based on various beliefs about philosophical and/or linguistic matters, they arrive at different assessments of both the significance and suitability of the imitation game for the development of cognitive science and AI models. Nevertheless, they share a rejection of the idea that one can treat Turing test as a test for "machine thinking." This seems to stem from a concern for the proper use of language —one that is a fundamental methodological commitment of analytic philosophy.

Keywords analytic philosophy; Artificial Intelligence; misuse of language; philosophy of mind; Turing test

✍ Józef Bremer, Jesuit University Ignatianum in Krakow, ul. Kopernika 26, 31-501 Kraków, Poland
✉ jozef.bremer@ignatianum.edu.pl 0000-0001-9664-8896
✍ Mariusz Flasiński, Department of Information Technology Systems, Jagiellonian University, ul. St. Łojasiewicza 4, Kraków 30-348 ✉ mariusz.flasinski@uj.edu.pl 0000-0002-0910-4928

Forum Philosophicum 27 (2022) no. 1, 6–25 Subm. 2 December 2021 Acc. 17 February 2022
ISSN 1426-1898 e-ISSN 2353-7043 DOI:10.35765/forphil.2022.2701.01

1. Introduction

The imitation game known as the Turing test (TT), proposed by Alan Mathison Turing (Turing 1950), is one of the most important thought experiments in the history of Artificial Intelligence (AI). In its standard version, the "interrogator," playing the imitation game, aims to distinguish a person from a machine (i.e., a computer) on the basis of a conversation conducted simultaneously with both of them. The test is an important element of the debate about any supposed equivalence between machine-based information processing and human thinking (see Rapaport 2006; Sterret 2020; Proudfoot 2020; Copeland et al. 2017; Levesque 2017; Damassino and Novelli 2020; Neufeld and Finnestad 2020). On the one hand, the imitation game would seem to constitute a valuable test for monitoring progress both in the applied domain of man-machine communication and in the important sub-domain of artificial intelligence that concerns Natural Language Processing (NLP) (Russell and Norvig 2020; Flasiński 2019). On the other, TT has provoked some heated disputes and controversies in the philosophy of mind, cognitive science, AI and popular science forums. In general, this multi-threaded debate has tended to issue from a variety of epistemic assumptions, assumed theories of language, and presumed models pertaining to the philosophy of mind. However, it seems that a misinterpretation and/or misuse of certain key mental concepts is the most important driver of these controversies.

The aim of the present paper is not to contribute to the discussion about TT by taking up a position (along with its putative justification) in respect of any specific threads in the debate. Instead, we would like to consider the following issue: "What, as a matter of fact, is tested by TT?"—or, more specifically, "Is TT a test for (machine/computer) thinking?" We shall therefore focus on analysis of the notion of thinking, since this concept seems to be key here. In fact, it seems to us that an improper use, in a linguistic sense, of this notion is an important source of the aforementioned controversies.

Although the emphasis on language in analytic philosophy is currently not as strong as it was right after the linguistic turn at the beginning of this movement, a concern for the "proper"—whatever this may mean—use of language, with analyzing reasons for conceptual confusion, and with avoiding the misuse of words, still seem to rank amongst the methodological commitments associated with what we now think of as the analytic approach to philosophical matters. We shall analyze the views on the Turing test of four influential thinkers belonging to this tradition (Martinich and Sosa 2001): Ludwig Wittgenstein, Noam Chomsky, Hilary Putnam and John Searle. Wittgenstein has greatly influenced both the philosophy of mind and the foundations of AI (Shanker 1998). Meanwhile, Chomsky arguably counts

as the most influential linguist of the twentieth century and, moreover, his influence on computer science (above all as regards formal language theory, programming language theory, compiler design theory and syntactic pattern recognition) cannot be underestimated. Putnam introduced functionalism, this being one of the most important concepts of what is known as strong AI, but then went on to change his views about the possibility of the latter, based on linguistic considerations. Finally, Searle's Chinese Room thought experiment seems to be as much-famed in the realm of AI as TT itself. Although these four thinkers differ from one another considerably as far as their views on philosophical and linguistic matters are concerned, their objections to TT as a test for "machine thinking" all result from much the same concern for the proper use of language—one characteristic of analytic philosophy. At least, that is what we ourselves aim to show in the present paper.

2. Wittgenstein

Turing begins his famous paper (Turing 1950) with the sentence "I propose to consider the question 'Can machines think?'," while in the context of Wittgenstein's philosophical explorations, the question "What is thinking (thought)?" is one of the most important.

In the *Tractatus Logico-Philosophicus* (TLP) (Wittgenstein 1922), a definition of thought is presented via two theses: "The logical picture of the facts is thought" (TLP 3) and "The thought is the significant proposition" (TLP 4). The idea that a significant proposition pictures a state of affairs (or atomic fact) (TLP 3) is known as the "picture theory of language." Although the later Wittgenstein refuted and replaced this theory with a use theory of meaning, this picture theory is now employed in computer science. For example, it is a fundamental paradigm (together with Chomskyian generative grammar) of syntactic pattern recognition (Flasiński 2019).

The goal of the *Tractatus* is to draw a limit "not to thinking, but to the expression of thinking" within language. Language is treated here as the "totality of propositions" (TLP 4.001). Therefore, Wittgenstein analyzes thoughts not as mental items, but only to the extent that they are represented by logical propositions. This logico-syntactic approach to thoughts ("thinking as operating with signs") is typical for the early Wittgenstein. Turing attended Wittgenstein's lectures on the foundations of mathematics, and was influenced by them (Floyd 2017). He sent him a copy of his famous paper on the *Entscheidungsproblem* (Turing 1936) containing a definition of computation in the form of a machine "operating with signs" of just the kind we know today as a Turing machine.

Wittgenstein's views on what thinking (or thought) amounts to certainly changed radically in the period between TLP and *Philosophical Investigations* (Wittgenstein 1953). Let us therefore now consider how the question "What is thinking?" figured in his approach subsequent to the publication of the *Tractatus*.

2.1. What is thinking?

Reflecting the use theory of meaning mentioned above, the later Wittgenstein is interested in the grammar of using mental verbs, not in questions of philosophical ontology or psychology (Bremer 2019). To this end, in *Philosophische Grammatik* (Wittgenstein 1974), written during the years 1931–1934, he introduced the concept of grammar thus: "Grammar describes the use of words in the language. So it has somewhat the same relation to the language as the description of a game, the rules of a game, have to the game." Wittgenstein "analyzes" our forms of expression (see Wittgenstein 1953, §90), and furnishes "grammatical remarks" that concern the ways in which we use words (and the concepts underlying them). He is only concerned with facts pertaining to language use. Public criteria for the correct application of a word are necessary if that word is to be employed correctly. The meaning of mentalistic concepts (e.g., remembering, thinking, knowing) should not be given in terms of internal, private mental states, as in the Cartesian model of the mind, but should be publicly and intersubjectively verifiable. As he puts it in the following two remarks (Wittgenstein 1953):

> And it is this inner process that one means by the word "remembering."—The impression that we wanted to deny something arises from our setting our faces against the picture of the "inner process." What we deny is that the picture of the inner process gives us the correct idea of the use of the word "to remember." We say that this picture with its ramifications stands in the way of our seeing the use of the word as it is (§305).

> An "inner process" stands in need of outward criteria (§580).

Thus, it seems that he could be considered a weak logical behaviorist.

In line with the methodology of ordinary language philosophy, Wittgenstein's grammatical analyses of the concept of thinking are, for the most part, of a piecemeal nature, being largely embedded in the varied manifestations that we encounter in daily life of that which philosophy and philosophical psychology seek to explore. He believes that the linguistic use of "thinking" is as confused as the use of psychological verbs

generally (Wittgenstein 1967, 1980a). Thinking can be said to be a process or an activity of the mind, but not in the same sense as talking is an activity of a person or writing an activity of the hand (Wittgenstein 1974). The similarity of "thinking" to "speaking" is merely apparent. (This observation proves relevant if and when we find ourselves inclined to treat the imitation test as a "thinking" test.) Unlike with a spoken sequence of words, it is not usually possible to indicate the "beginning" or "end" of a sequence of thoughts, and neither does a sequence of thoughts inevitably develop according to a similar chronology: "I cannot say, e.g., that this or that phase of the process occurred in this time segment. So I cannot describe the thinking process as I can describe the speaking itself, for instance. That is why one can't very well call thinking a process" (Wittgenstein 1980a, §266).

Corporeal activities, such as the circulation of the blood, breathing, digestion, etc., are different from *sui generis* mental activities such as thinking, wanting, or feeling. Construing thinking as a hidden process that proceeds covertly can result in confusion (Wittgenstein 1980b). Moreover, if we deny that thinking is an immaterial (incorporeal) activity, then it is not that such a denial comes from some familiarity or acquaintance on our part with immaterial processes more generally, empowering us to conclude that thinking cannot be found among these (Wittgenstein 1953). In sum, thinking is by no means an obvious or definitively clear concept. Instead, while long viewed as a reliable unity by philosophers, it turns out to be highly sensitive to the varied ways in which it can figure in everyday contexts of language use (e.g., we can "think of something", "think something up", "think ahead", etc.).

2.2. "'A machine thinks (perceives, wishes)' seems somehow nonsensical"

One can therefore hardly be surprised by Wittgenstein's forceful assertion in the *Blue Book* when he writes:

> "Is it possible for a machine to think?" (whether the action of this machine can be described and predicted by the laws of physics or possibly, only by laws of a different kind applying to the behaviour of organisms). And the trouble which is expressed in this question is not really that we don't know a machine which could do the job. The question is not analogous to that which someone might have asked a hundred years ago: "Can a machine liquefy gas?" The trouble is rather that the sentence, "A machine thinks (perceives, wishes)" seems somehow nonsensical. It is as though we had asked "Has the number 3 a colour?". (Wittgenstein 1958, 47)

This seems to directly challenge the understanding of the imitation game as a test for "machine thinking." The grammatical analysis resulting in the conclusion that attributing mental predicates to machines amounts to a conceptual confusion is made in the *Investigations*, and can be summarized as follows (§359–360): (a) We are only justified in ascribing the mental predicates "calculating," "thinking," "feeling" and "knowing" to human beings, or to a being appropriately similar to these ("We only say of a human being and what is like one that it thinks"); (b) Machines are neither human, nor appropriately and sufficiently similar to human beings ("Could a machine think?—Could it be in pain?—Well, is the human body to be called such a machine? It surely comes as close as possible to being such a machine."); (c) *Ergo*, we cannot ascribe mental predicates to machines ("But a machine surely cannot think!"). Wittgenstein then asks "Is that an empirical statement?"—to which he answers "No." The assertion in question is, rather, the outcome of a grammatical analysis. Thus, we can either rename what we mean by "think," or rename what we mean by "machine."

3. Chomsky

In general, Chomsky's views in linguistics and the philosophy of mind run contrary to those of Wittgenstein. His mentalist theory of an innate language faculty centered around the notion of Universal Grammar (UG) (Chomsky 1965) stands in opposition to behaviorist theories in psychology and linguistics (Chomsky 1965, 2019). As he puts it, "the behaviorist position is not an arguable matter. It is simply an expression of lack of interest in theory and explanation" (Chomsky 1965, 193). His theory is consistent with rationalism in philosophy, and Chomsky, as a self-declared Cartesian (Chomsky 1966), states that "[w]e should, so it appears, think of knowledge of language as a certain state of mind/brain, a relatively stable element in transitory mental states once it is attained; furthermore as a state of some distinguishable faculty of the mind—the language faculty—with its specific properties, structure and organisation, one module of the mind" (Chomsky 1986, 12–3).

Chomsky's linguistic grammars are—from a mathematical point of view—generative grammars/systems (abstract rewriting systems) (Flasiński 2016). Although they are now subject to criticism in linguistics, we should note that programming languages are defined with the help of the Chomskyian-grammars-based Backus-Naur Form (BNF), and any compiler[1] design

1. A compiler is a computer program that translates computer code written by a programmer in a certain programming language into code that is "understood" by a computer and therefore executable by the latter.

tutorial will begin by introducing such grammars, as they (and the corresponding formal automata) furnish the basis for any compiler construction. Chomskyian grammar is the fundamental paradigm for the constructing of syntactic pattern recognition systems, such as are used in, amongst others, computer vision, bioinformatics, the analysis of signals in medicine (ECG, EEG, etc.), seismology and radar engineering (meteorology, air-traffic control, etc.), and Natural Language Processing (Flasiński 2019). Indeed, this theory has turned out to be more useful than any other for the development of computer science methods and the designing of IT systems.

Chomsky was also the originator of essentialism in linguistics. Linguistic essentialism is interested in defining universal principles for describing the characteristics of languages. He differentiates between E-language (Extensional, External), which is external to the mind, and I-language (Individual, Internal, Intentional), which is the linguistic knowledge of the individual human being internalized in their own mind/brain. He states that "for H to know L is for H to have a certain I-language. The statements of the grammar are statements of the theory of mind about the I-language, hence structures of the brain formulated at a certain level of abstraction from mechanisms." (Chomsky 1986, 23). At the same time, he rejects E-language as an object of study in linguistics. Obviously, this view stands opposed to those according to which language is considered a social, externally-observed phenomenon—as with, for instance, ordinary language philosophy.

Where Chomsky's "semantic internalism" is concerned, the idea that we should treat I-language as the proper object of linguistic research counts as a basic axiom. In general, according to this approach, meanings are held to be generated by the mind and intrinsic to it (Chomsky 1986). We shall address the issues raised by semantic internalism when we come to present Putnam's views in the area of the philosophy of language.

3.1. (External) imitation of thinking is not modelling of (internal) thinking capacities

Chomsky's skepticism toward the Turing test as a test for thinking follows from his rationalist (anti-behaviorist) and essentialist views. As he writes:

> Questions about computational–representational properties of the brain are interesting and seem important; and simulation might advance theoretical understanding. But success in the imitation game in itself tells us nothing about these matters. (Chomsky 2008, 105)

Chomsky differentiates between simulation/modelling and imitation/duplication. Let us first look at the example of simulation/modelling:

> A completely separate issue is whether simulation might teach us something about the process simulated; whether a chess-playing program, for example, might teach us something about human thought. In the latter case, the topic is very badly chosen, in my opinion, but in principle simulation certainly can provide much insight. That much was well understood centuries ago, though the classical discussion did not fall into the errors of the modern revival. When Jacques de Vaucanson amazed observers with his remarkable contrivances, he and his audience were concerned to understand the animate systems he was modelling. His clockwork duck, for example, was intended to be a model of the actual digestion of a duck, not a facsimile that might fool his audience, the neuropsychologist John Marshall points out in a recent study. That is the purpose of simulation generally in the natural sciences. (Chomsky 1993, 30)

Modelling (of the scientific sort) allows us to better understand, visualize, quantify, etc., a particular aspect/feature/part of the world. If the goal of modelling is to understand better, then what we usually do is define a conceptual (functional) model of a subject that can be used to help us understand or simulate the subject that this model represents. In this context, a (computer) simulation will be the process of modelling performed on a computer. Let us now see what Chomsky has to say about imitation/duplication:

> Let's try an analog [to the Turing test]. We breathe.... We could get a machine that duplicates that completely by some crazy mechanism. Would the machine be breathing? Well, no ... Is it a good model of humans? Well, that we'd look at and see if it teaches us anything about humans. If it does, it's a good model of humans. If it doesn't teach us anything about humans, send it to Hume's flames. (Chomsky 1993, 90)

> Well, you go back to the Turing test, notice that it's not an attempt to explain and understand anything about thinking. (Chomsky 2019, 3)

Imitation is merely replication of someone else's behavior, and duplication is just copying something. This is why the imitation game "doesn't teach us anything about humans"—or, strictly speaking, about human thinking. As both a rationalist and an essentialist, Chomsky holds that cases of

(external) imitation/duplication of thinking are not tantamount to any form of modelling/simulation of (internal) thinking capacities.

3.2. The inappropriateness of ascribing intentional attributions to (non-human) objects

Chomsky is not in principle opposed to the Turing test *per se*:

> The dual significance of the enterprise [i.e., the TT project]—constructing better machines, gaining insight into human intelligence—should no longer be in doubt, if it ever was.… Of the two "useful lines of research" that Turing contemplated, one—improvement of the capacities of machines—is uncontroversial, and if his imitation game stimulates such research, well and good. The second line of research—investigating "the intellectual capacities of a man"—is a more complex affair. (Chomsky 2008, 103–4)

At the same time, apart from the objections concerning "the second line of research," discussed here in the previous section, he—similarly to Wittgenstein—raises objections to attributing mental predicates to machines. First of all, he claims that Turing was not in fact looking to discuss whether machines can think:

> He [Turing] begins by proposing "to consider the question, 'Can machines think?'," but went on to explain that he would not address this question because he believed it "to be too meaningless to deserve discussion" … Perhaps he agreed with Wittgenstein that "We can only say of a human being and what is like one that it thinks." (Chomsky 2008, 104)

> So he [Turing] is not going to discuss it, because the notion thinking is so vague and amorphous that you can't give a response in the manner in which you might in, say, physics or even biology. (Chomsky 2019, 3)

As with Wittgenstein, Chomsky objects to the ascribing of intentional attributions, like thinking, to objects other than people (or anything not appropriately or sufficiently similar to the latter):

> it is idle to ask whether legs take walks or brains plan vacations; or whether robots can murder, act honorably, or worry about the future. Our modes of thought and expression attribute such actions and states to persons, or what we might regard as similar enough to persons. (Chomsky 2008, 104)

> we may say that people think, not their brains, though their brains provide the mechanisms of thought. As noted, it is a great leap, which often gives rise to pointless questions, to pass from common sense intentional attributions to people, to such attributions to parts of people, and then to other objects. (Chomsky 2008, 106)

Finally, like Wittgenstein, Chomsky claims that in talking about "machine thinking" what we are discussing is a linguistic issue:

> It's just like asking, "Does my brain think?" That's not the way we talk English, but if you want to change the language you could say it. The same is true about this breathing device or about machines thinking and so on. (Chomsky 1993, 91)

4. Putnam

If we are to present the philosophical views of Hilary Putnam, then we are obliged—as with Wittgenstein—to consider two distinct phases of this thinker's development: the earlier Putnam and the later Putnam. For computer scientists and cognitivists, the former is recognized primarily as the founder of "machine state functionalism" (Putnam 1960). According to this theory, which is based on an analogy between the mind and the Turing machine, mental states are connected by causal relations in a way analogous to that in which formal automaton states are connected via the transition function (of an automaton). Functionalism has strongly impacted artificial intelligence, and numerous versions of it have been developed by such influential thinkers as Jerry Fodor, Daniel Dennett, Zenon Pylyshyn and others (Flasiński 2016). However, Putnam himself abandoned this (radical) view in the late 1980s (Putnam 1988). Ultimately, he decided to introduce a modified version of functionalism called liberal functionalism (Putnam 2012, 2016). According to this theory, mentality is a collection of functional capacities that are interrelated. These can be described using the vocabulary of computer science (as abilities to compute), psychology, or neurology, and even by means of intentional idioms. (Although the brain can be treated as a computer, the mind is more than the brain.) The transactions of the organism with its environment are to be taken into account when we describe the mind. This last principle follows from his externalist and anti-individualist views.

Semantic externalism is one of Putnam's most important contributions to the philosophy of language (Putnam 1973, 1975). Contrary to Chomsky's semantic internalism, Putnam claims that "'meanings' just ain't in the

head!". Rather, they are determined (in part) by factors that are external to us—something interestingly illustrated by the Twin Earth thought experiment in his paper "The Meaning of 'Meaning'" (Putnam 1975). This reflects the approach adopted in (semantic) model theory (Tarski 1944), in that it uses truth conditions to model the interpretation of the sentences of a given language.[2] Putnam says the following: "Truth involves some sort of correspondence relation between words or thought-signs and external things and sets of things. I shall call this perspective the externalist perspective" (Putnam 1981, 49). Although he altered his functionalist views many times, he remained a semantic externalist until the end of his life.

4.1. The externalist argument against TT

Putnam's early functionalism (Putnam 1960) gave strong support to the idea of the imitation game as a test for "machine thinking." In order to realize just how radical this view was, it is worth reading some excerpts from the paper in question (Putnam 1964, 676–8):

> Throughout this paper I have stressed the possibility that a robot and a human may have the same "psychology"—that is, they may obey the same psychological laws.... In general, such laws, like all scientific laws, will involve abstractions—terms more or less remote from direct behavioral observation. Examples of such terms have already been given: repression, inhibitory potential, preference, sensation, belief. If the human brain is simply a neural net with a certain program ... then a robot whose "brain" was a similar net, only constructed of flip-flops rather than of neurons, would have exactly the same psychology as a human. I have referred to this problem as the problem of the "civil rights of robots" because that is what it may become, and much faster than any of us now expect. Given the ever-accelerating rate of both technological and social change, it is entirely possible that robots will one day exist, and argue "we are alive; we are conscious!" In that event, what are today only philosophical prejudices of a traditional anthropocentric and mentalistic kind would all too likely develop into conservative political attitudes.

This form of his early functionalism was then criticized by Putnam himself in the period when he was introducing semantic externalism: e.g., in his paper "Philosophy and our Mental Life," where he reflected that

2. (Semantic) model theory is the area of mathematical logic that studies the relationship between (formal) languages (treated as collections of sentences) and their interpretations (models).

> in previous papers, I have argued that (1) a whole human being is a Turing machine, and (2) the psychological states of a human being are Turing machine states or disjunctions of Turing machine states. In this section I want to argue that this point of view was essentially wrong, and that I was too much in the grip of the reductionist outlook." (Putnam 1975, 298)

At the time of his rejection of this early functionalism, Putnam raised objections regarding the Turing test. These also stemmed from the principles of semantic externalism, and especially from the belief that Putnam, years later, would refer to as being "the heart of [his] semantic externalism": "(…) our words don't have meanings just by going through our heads; only as a being related to a world and to other people in certain ways do we have thoughts with content at all" (Putnam and Peruzzo 2015, 215). Thus, we can only speak about thinking/thought if words have taken on meanings by virtue of being related to the external world. Putnam presented his arguments against the imitation game as a test for "machine thinking" in his seminal work "Brains in a Vat" (Putnam 1981, 13):

> Suppose, for example, that I am in the Turing situation (playing the "Imitation Game," in Turing's terminology) and my partner is actually a machine. Suppose this machine is able to win the game ("passes" the test). Imagine the machine to be programmed to produce beautiful responses in English to statements, questions, remarks, etc. in English, but that it has no sense organs (other than the hookup to my electric typewriter), and no motor organs (other than the electric typewriter). (As far as I can make out, Turing does not assume that the possession of either sense organs or motor organs is necessary for consciousness or intelligence.) … What should we say about such a machine? To me, it seems evident that we cannot and should not attribute reference to such a device. It is true that the machine can discourse beautifully about, say, the scenery in New England. But it could not recognize an apple tree or an apple, a mountain or a cow, a field or a steeple, if it were in front of one.

4.2. The Turing test: "Syntactic play" is not "intelligent discourse"

The conclusions that issue from the considerations pertaining to the imitation game in Putnam's seminal work (Putnam 1981) are of a linguistic nature, and are similar to those of Wittgenstein and Chomsky presented above. Putnam, as a semantic externalist, claims that some form of correspondence between words and external things is required when we converse about the real world. He formulates this requirement with the help of the basic terms of the Sellarsian approach to the use theory of meaning

(Sellars 1954; Bremer 1997). Sellars distinguishes three dimensions of usage: language entry transitions (from perception to language), language exit transitions (from language to actions), and intralinguistic transitions (within language—i.e., moving between sentences).[3] According to Putnam:

> What we have is a device for producing sentences in response to sentences. But none of these sentences is at all connected to the real world.... What produces the illusion of reference, meaning, intelligence, etc., here is the fact that there is a convention of representation which we have under which the machine's discourse refers to apples, steeples, New England, etc.... But we are able to perceive, handle, deal with apples and fields. Our talk of apples and fields is intimately connected with our nonverbal transactions with apples and fields. There are "language entry rules" which take us from experiences of apples to such utterances as "I see an apple," and "language exit rules" which take us from decisions expressed in linguistic form ("I am going to buy some apples") to actions other than speaking. (Putnam 1981, 14)

On his view, the imitation game corresponds to a case of mere "syntactic play," since only intralinguistic transitions (rules) occur in the test. Finally, he claims that such syntactic play is not intelligent discourse (albeit that it does resemble such discourse):

> Lacking either language entry rules or language exit rules, there is no reason to regard the conversation of the machine (or of the two machines, in the case we envisaged of two machines playing the Imitation Game with each other) as more than syntactic play. Syntactic play that resembles intelligent discourse, to be sure ... The point that is relevant for our discussion is that there is nothing in Turing's Test to rule out a machine which is programmed to do nothing but play the Imitation Game. (Putnam 1981, 14)

5. Searle

As a biological naturalist, Searle (Searle 1983, 2008) has rejected Putnam's functionalism (i.e., Turing machine functionalism) ever since the latter's initial introduction into the philosophy of mind: "I reject any form of behaviorism or functionalism, including Turing machine functionalism, that ends up by denying the specifically mental properties of mental phenomena" (Searle 1983, viii).

3. Sellars' transitions are sometimes referred to as rules.

Searle thinks that both behaviorists and functionalists reject mental phenomena as they fear not eliminating them will lead to the undesirable outcome that is dualism and its associated insoluble mind-body problem (Searle 1983). He proposes biological naturalism as an antidote to such dualism. Pursuing his considerations further, he describes biological naturalism as follows:

> On my view mental phenomena are biologically based: they are both caused by the operations of the brain and realized in the structure of the brain. On this view, consciousness and Intentionality are as much a part of human biology as digestion or the circulation of the blood. It is an objective fact about the world that it contains certain systems, viz., brains, with *subjective* mental states, and it is a *physical* fact about such systems that they have *mental* features. The correct solution to the "mind-body problem" lies not in denying the reality of mental phenomena, but in properly appreciating their biological nature. (Searle 1983, ix)

Two fundamental issues make up the primary subject matter of Searle's philosophy of mind: intentionality and consciousness (Regner 2002). The concept of intentionality is defined by him as the "property of many mental states and events by which they are directed at or about or of objects and states of affairs in the world" (Searle 1983). He distinguishes between intentionality and consciousness thus: "Many conscious states are not intentional, e.g., a sudden sense of elation, and many Intentional states are not conscious, e.g., I have many beliefs that I am not thinking about at present and I may never have thought of" (Searle 1983, 2). Elsewhere he defines consciousness in the following terms:

> Here is a definition: consciousness consists of those states of feelings, sentience, or awareness that typically begin when we wake from a dreamless sleep and continue throughout the day until those feelings stop, until we go to sleep again, go into a coma, die, or otherwise become "unconscious." (Searle 2008, 141)

Searle considers consciousness a basic concept of the philosophy of mind: "The reason for emphasizing consciousness in an account of the mind is that it is the central mental notion. In one way or another, all other mental notions—such as intentionality, subjectivity, mental causation, intelligence, etc.—can only be fully understood as mental by way of their relations to consciousness" (Searle 1992). As we shall see, both of these basic concepts

of his theory of mind—i.e., intentionality and consciousness—are used by him to criticize the Turing test.

5.1. The Chinese Room thought experiment

The basic idea of the Chinese Room thought experiment (Searle 1980) can be summarized as follows: Searle (knowing no Chinese) is locked in a room containing baskets full of Chinese symbols (the data base) and a book of instructions, called the rule book, for manipulating the symbols (the program). People outside the room hand in small bunches of symbols which are questions in Chinese (the input). Searle manipulates the symbols according to the rule book and, in response, hands back more small bunches of Chinese symbols which are correct answers to the questions (the output).

Searle uses this thought experiment to criticize the imitation game as a test for machine thinking/understanding. As he puts it:

> Now suppose that the rule book is written in such a way that my "answers" to the "questions" are indistinguishable from those of a native Chinese speaker.... I satisfy the Turing test for understanding Chinese. All the same, I am totally ignorant of Chinese. And there is no way I could come to understand Chinese in the system as described, since there is no way that I can learn the meanings of any of the symbols. Like a computer, I manipulate symbols, but I attach no meaning to the symbols. (Searle 1990, 26)

Searle's critique of TT facilitated by the Chinese Room thought experiment is based on three main propositions (Searle 1980, 1992, 2010), which we list below.

1. Humans have intentionality and consciousness ("intrinsic intentionality is a phenomenon that humans and certain other animals have as part of their biological nature" (Searle 1992)), whereas computer programs do not. Searle claims to have "demonstrated years ago with the so-called Chinese Room Argument that the implementation of the computer program is not by itself sufficient for consciousness or intentionality" (Searle 2010, 17). As to the question of whether "something [could] think, understand, and so on solely in virtue of being a computer with the right sort of program," he answers "No" and goes on to explain that "the formal symbol manipulations by themselves don't have any intentionality; they are quite meaningless." (Searle 1980, 428)

2. Syntactic operations (like machine computation) do not yield semantics. Searle claims that "computation is defined purely formally or syntactically, whereas minds have actual mental or semantic contents, and you cannot get from the syntactical to the semantic just by having the syntactical operations and nothing else." (Searle 2007, 173)

3. The simulation of natural phenomena on a computer is not the same as the occurrence of these phenomena for real.[4] Searle explains this proposition as follows:
   Simulation is not duplication. You can simulate the cognitive processes of the human mind as you can simulate rain storms, five alarm fires, digestion, or anything else that you can describe precisely. But it is just as ridiculous to think that a system that had a simulation of consciousness and other mental processes thereby had the mental processes as it would be to think that the simulation of digestion on a computer could thereby actually digest beer and pizza.... The point, however, is that any such artificial machine would have to be able to duplicate, and not merely simulate, the causal powers of the original biological machine. An artificial heart does not merely simulate pumping, it actually pumps. It actually causes the pumping of blood. And an artificial brain would have to do something more than simulate consciousness, it would have to be able to produce consciousness. It would have to cause consciousness. (Searle 2008, 68–72)

These objections follow from Searle's views pertaining to both philosophical and linguistic matters, and are different from those of Wittgenstein, Chomsky and Putnam. However, he also says something that is similar to the positions of those three thinkers. This concerns the idea that ascribing mental qualities to computers involves a misuse of language.

5.2. Conceptual confusions resulting from the construal of metaphor as literal meaning

Searle is known for his significant contributions to the philosophy of language and mind. As a philosopher of language belonging to the tradition

4. Searle uses the term duplication in a different meaning from Chomsky (see Section 3). For the latter, to duplicate means to imitate: i.e., to replicate the behavior of somebody/something in an imperfect way, or merely in a certain respect. Therefore, cases of computer simulation/modelling will be more than instances of computer/machine duplication. For Searle, to duplicate means to "produce" something for real. Hence, Searle considers duplication to amount to something more than a computer simulation.

of analytic philosophy, and just like Wittgenstein, Chomsky and Putnam, he is attentive to the proper use of language. His differentiating between "intrinsic intentionality" and "as-if intentionality" (Searle 1992; 1983) constitutes one of the most important distinctions in cognitive science and AI aimed at avoiding conceptual confusion:

> It is very convenient to use the jargon of intentionality for talking about systems that do not have it, but that behave as if they did. I say about my thermostat that it perceives changes in the temperature; I say of my carburetor that it knows when to enrich the mixture … it is important to emphasize that these attributions are psychologically irrelevant, because they do not imply the presence of any mental phenomena. This intentionality described in all of these cases is purely "as-if." … I am just stipulating that by "intrinsic intentionality" I mean the real thing as opposed to the mere appearance of the thing ("as-if"), and as opposed to derived forms of intentionality such sentences, pictures, etc. … any attempt to deny the distinction between intrinsic and as-if intentionality faces a general reductio ad absurdum. If you deny the distinction, it turns out that everything in the universe has intentionality.… The price of denying the distinction between intrinsic and as-if intentionality, in short, is absurdity, because it makes everything in the universe mental.… There is nothing harmful, misleading, or philosophically mistaken about "as-if" metaphorical ascriptions. The only mistake is to take them literally. (Searle 1992, 79–82)

Searle raises similar objections with respect to the misuse of the notions of intelligence and artificial intelligence:

> Similar remarks apply to the notion of "intelligence." There is a perfectly good sense in which my present computer is much smarter, that is, more "intelligent," than the computer I had ten years ago. But, I take it, there is no psychological reality to that sense, at all. On the other hand, when we say that humans are more intelligent than some other species such as dogs, we are talking about a certain psychological reality. Questions like this become important when you ask whether or not you could build an intelligent machine.… The notion of "artificial intelligence" has for decades suffered from a failure on the part of its users to distinguish between creating a simulation of real intelligence artificially and creating real intelligence artificially. (Searle 2007, 173)

6. Conclusion

On the one hand, the Turing test is, undoubtedly, an important idea that continues to stimulate research into AI methods, especially in the area of

Natural Language Processing. On the other hand, as we have mentioned in our Introduction, TT as a test for the "thinking of machines" raises heated disputes and controversies in the philosophy of mind, cognitive science and AI. Partially, such disputes result from a variety of epistemic assumptions, assumed theories of language, and presumed models in the philosophy of mind. However, in our opinion, the improper use ("misuse", as early analytic philosophers would say) of the notion of thinking seems to be the primary source of the controversies.

In order to substantiate this position, we discussed the views on TT of four thinkers who have influenced AI significantly, these being Ludwig Wittgenstein, Noam Chomsky, Hilary Putnam and John Searle. We also presented their positions on philosophical and linguistic matters insofar as these appear relevant to the assessment of both the significance and the suitability of the imitation game for the development of cognitive science and AI models. These stances differ markedly. Chomsky's essentialism and innate language faculty theory run counter to Wittgenstein's weak logical behaviorism and commitment to ordinary language philosophy. Putnam's semantic externalism is opposed to Chomsky's semantic internalism. Searle, as a biological naturalist, rejects any form of functionalism, and therefore Putnam's (machine state) functionalism and liberal functionalism, too. The relations obtaining between these positions are summarized diagrammatically in Fig. 1.

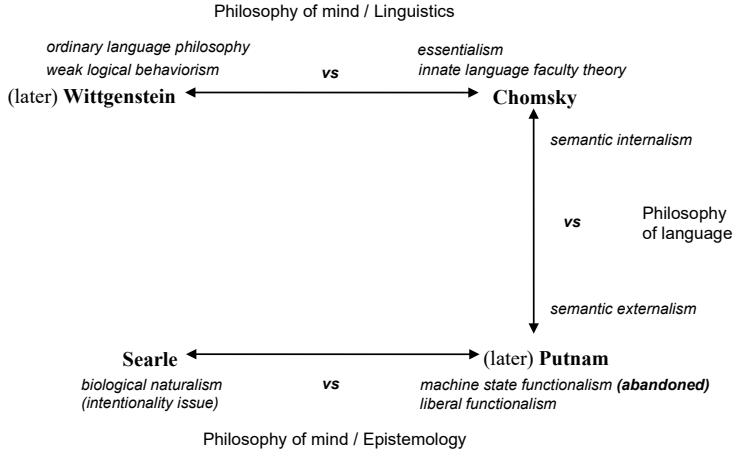

Figure 1. Diagram of selected philosophical/linguistic views (Wittgenstein, Chomsky, Putnam, Searle)

The various assessments of the imitation game given by Wittgenstein, Chomsky, Putnam and Searle, as presented in the preceding sections here,

result from quite different philosophical/linguistic sources. If we posed for these thinkers the question "Why is TT not a test for machine (computer) thinking?", Chomsky could answer "because TT does not model thinking," Putnam could say "because there is no reference to the real world," and Searle could claim that it is "because syntactic operations do not yield semantics" (or "because there is neither intentionality nor consciousness in machines"). Going further, if we were to ask them "So, what is tested by TT?", Chomsky could answer that it is "the possibility of external imitation of thinking by a computer," Putnam that it is "the possibility of imitation of intelligent discourse," and Searle that it is "the possibility of a (virtual) simulation of mental processes."

However, there is still something common to these thinkers where the Turing test is concerned: all of them raise objections to attributing mental predicates to machines/computers. As we have seen in the preceding sections, talk about "machine thinking" is, on Wittgenstein's view, nonsensical, while for Chomsky it is improper (in that only human beings think), for Putnam wrong (as only human beings have thoughts with content), and for Searle mistaken (as it can only be used metaphorically). They consider such an attribution a misuse of language, where this verdict follows from methodological commitments associated with the analytic approach to doing philosophy.

As stated in the Introduction, the aim of this paper has not been to contribute to the discussion surrounding TT by taking up a position in respect of any thread of the debate. Having followed the latter for many years, it seems to us that some controversies may be clarified —at least to some extent—by considering whether talk of "machine thinking" makes sense from a linguistic point of view. We have therefore decided to analyze views about "machine thinking" in TT-connected contexts as these relate to four influential analytical philosophers. We share their (common) view that ascribing mental qualities to machines is a misuse of language. It seems that it could well prove beneficial to participants in the TT debate to take careful note of their justifications for such a stance. And even if one does not agree with that position, and still wants to use the term "machine thinking," it is surely worth considering—and carefully and precisely analyzing—what thinking means in that context.

Bibliography

Bremer, Józef. 1997. *Rekategorisierung statt Reduktion. Zu Willfrid Sellars' Philosophie des Geistes*. Göttingen: Vandenhoeck & Ruprecht.

——. 2019. "Wittgenstein über seine eigenen Emotionen in den Denkbewegungen." In *Wittgensteins Denkbewegungen (Tagebücher 1930–1932/1936–1937) aus interdisciplinärer Sicht*, edited by Ilse Somavilla, Carl Humphries and Bożena Sieradzka-Baziur, 211–31. Innsbruck: StudienVerlag.

Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge: MIT Press.

——. 1966. *Cartesian Linguistics*. New York: Harper and Row.

——. 1986. *Knowledge of Language: Its Nature, Origin and Use*. Westport, CT: Praeger.

——. 1993. *Knowledge and Thought*. Wakefield, RI: Moyer Bell.

——. 2008. "Turing on the 'Imitation Game'." In *Parsing the Turing Test: Philosophical and Methodological Issues in the Quest for the Thinking Computer*, edited by Robert Epstein, Gary Roberts and Grace Beber, 103–6. Berlin: Springer.

——. 2019. *The UCLA Lectures (April 29 – May 2, 2019)*. edited by R. Freidin. Los Angeles: UCLA.

Copeland, Jack, Jonathan Bowen, Mark Sprevak, and Robin Wilson, eds. 2017. *The Turing Guide*. Oxford: Oxford University Press.

Damassino, Nicola, and Nicholas Novelli., eds. 2020. "Special Issue: Rethinking, Reworking and Revolutionising the Turing Test." *Minds and Machines* 30 (4).

Flasiński, Mariusz. 2016. *Introduction to Artificial Intelligence*. Cham, Switzerland: Springer.

——. 2019. *Syntactic Pattern Recognition*. Hackensack, NJ: World Scientific.

Floyd, Juliet. 2017. "Turing on "Common Sense": Cambridge Resonances." In *Philosophical Explorations of the Legacy of Alan Turing*, edited by Juliet Floyd and Alisa Bokulich, 103–49. Cham, Switzerland: Springer.

——. 2000b. "The Turing Test: the first 50 years." *Trends in Cognitive Sciences* 4 (3): 115–22. https://doi.org/10.1016/s1364-6613(00)01453-4.

Levesque, Hector J. 2017. *Commonsense, the Turing Test, and the Quest for Real AI*. Cambridge: MIT Press.

Martinich, A.P., and David Sosa, eds. 2001. *A Companion to Analytic Philosophy*. Oxford: Blackwell.

Neufeld, Eric, and Sonje Finnestad. 2020. "In defense of the Turing test." *AI and Society* 35: 819–27. https://doi.org/10.1007/s00146-020-00946-8.

Proudfoot, Diane. 2020. "Rethinking Turing's Test and the Philosophical Implications." *Minds & Machines* 30 (4): 487–512. https://doi.org/10.1007/s11023-020-09534-7.

Putnam, Hilary. 1960. "Minds and machines." In *Dimensions of Mind: A Symposium*, edited by Sidney Hook, 138–64. New York: New York University Press.

——. 1964. "Robots: Machines or artificially created life." *Journal of Philosophy* 61 (21): 668–91. https://doi.org/10.2307/2023045.

——. 1973. "Meaning and reference." *Journal of Philosophy* 70 (19): 699–711. https://doi.org/10.2307/2025079.

——. 1975. *Mind, Language and Reality. Philosophical Papers*. Vol. 2. Cambridge: Cambridge University Press.

——. 1981. "Brains in a Vat." In *Reason, Truth and History*, 1–21. Cambridge: Cambridge University Press.

——. 1988. *Representation and Reality*. Cambridge: MIT Press.

——. 2012. *Philosophy in an Age of Science: Physics, Mathematics, and Skepticism*. edited by Mario De Caro and David Macarthur. Cambridge, MA: Harvard University Press.

——. 2016. *Naturalism, Realism, and Normativity*. edited by Mario De Caro. Cambridge, MA: Harvard University Press.

Putnam, Hilary, and Léo Peruzzo. 2015. "Mind, body and world in the philosophy of Hilary Putnam (Interview with Hilary Putnam)." *Trans/Form/Ação* 38 (2): 211–6. https://doi.org/10.1590/S0101-31732015000200011.

Rapaport, William J. 2006. "Turing Test." In *Encyclopedia of Language and Linguistics*, edited by Keith Brown, 151–9. Oxford: Elsevier.

Regner, Jan. 2002. "Mind in the World." *Forum Philosophicum* 7: 193–204. https://doi.org/10.35765/forphil.2002.0701.10.

Russell, Stuart, and Peter Norvig. 2020. *Artificial Intelligence: A Modern Approach*. 4 ed. Hoboken: Pearson.

Searle, John R. 1980. "Minds, brains, and programs." *Behavioral and Brain Sciences* 3 (3): 417–57. https://doi.org/10.1017/S0140525X00005756.

——. 1983. *Intentionality. An essay in the philosophy of mind*. Cambridge: Cambridge University Press.

——. 1990. "Is the Brain's Mind a Computer Program?" *Scientific American*, January 1990, 26–31.

——. 1992. *The Rediscovery of the Mind*. Cambridge: MIT Press.

——. 2007. "Dualism revisited." *Journal of Physiology-Paris* 101 (4–6): 169–78. https://doi.org/10.1016/j.jphysparis.2007.11.003.

——. 2008. *Philosophy in a New Century*. Cambridge: Cambridge University Press.

——. 2010. "Why Dualism (and Materialism) Fail to Account for Consciousness." In *Questioning Nineteenth Century Assumptions about Knowledge*, edited by Richard E. Lee, 5–48. New York: SUNY Press.

Sellars, Wilfrid. 1954. "Some reflections on language games." *Philosophy of Science* 21 (3): 204–28. https://doi.org/10.1086/287344.

Shanker, Stuart G. 1998. *Wittgenstein's Remarks on the Foundations of AI*. London: Routledge.

Sterrett, Susan G. 2020. "The Genius of the 'Original Imitation Game' Test." *Minds & Machines* 20: 469–86. https://doi.org/10.1007/s11023-020-09543-6.

Tarski, Alfred. 1944. *Logic, Semantics, Metamathematics: Papers from 1923 to 1938*, edited by J.H. Woodger. Oxford: Oxford University Press.

Turing, A. M. 1936. "On computable numbers, with an application to the Entscheidungsproblem." *Proceedings of the London Mathematical Society, Series 2* 42 (1): 230–65.

——. 1950. "Computing Machinery and Intelligence." *Mind* 59 (236): 433–60. https://doi.org/10.1093/mind/LIX.236.433.

Wittgenstein, Ludwig. 1922. *Tractatus Logico-Philosophicus*, translated by C. K. Ogden and F.P. Ramsey. London: Kegan Paul, Trench, Trubner.

——. 1953. *Philosophical Investigations/Philosophische Untersuchungen*. Translated by G.E.M. Anscombe. 4 ed., edited by G.E.M. Anscombe and R. Rhees. Oxford: Basil Blackwell.

——. 1958. *Preliminary Studies for the "Philosophical Investigations." Generally Known as The Blue and Brown Books*. edited by R. Rhees. Oxford: Basil Blackwell.

——. 1967. *Zettel*. Translated by G.E. M. Anscombe. edited by G.E.M. Anscombe and G.H. Wright. Oxford: Basil Blackwell.

——. 1974. *Philosophical Grammar*. Translated by A. Kenny. edited by R. Rhees. Oxford: Basil Blackwell.

——. 1980a. *Remarks on the Philosophy of Psychology/Bemerkungen über die Philosophie der Psychologie*. Translated by C. G. Luckhardt and M.A.E. Aue. Vol. 2. edited by G.E.M. Anscombe and Heikki Nyman. Oxford: Basil Blackwell.

——. 1980b. *Remarks on the Philosophy of Psychology/Bemerkungen über die Philosophie der Psychologie*. Translated by G.E.M. Anscombe. Vol. 1. edited by G.E.M. Anscombe and G.H. Wright. Oxford: Basil Blackwell.