Minds and Machines (2023) 33:1–31
https://doi.org/10.1007/s11023-022-09616-8

The Turing Test is a Thought Experiment

Bernardo Gonçalves[1]

Received: 6 July 2022 / Accepted: 10 November 2022 / Published online: 27 November 2022

Abstract
The Turing test has been studied and run as a controlled experiment and found to be underspecified and poorly designed. On the other hand, it has been defended and still attracts interest as a test for true artificial intelligence (AI). Scientists and philosophers regret the test's current status, acknowledging that the situation is at odds with the intellectual standards of Turing's works. This article refers to this as the Turing Test Dilemma, following the observation that the test has been under discussion for over seventy years and still is widely seen as either too bad or too good to be a valuable experiment for AI. An argument that solves the dilemma is presented, which relies on reconstructing the Turing test as a thought experiment in the modern scientific tradition. It is argued that Turing's exposition of the imitation game satisfies Mach's characterization of the basic method of thought experiments and that Turing's uses of his test satisfy Popper's conception of the critical and heuristic uses of thought experiments and Kuhn's association of thought experiments to conceptual change. It is emphasized how Turing methodically varied the imitation game design to address specific challenges posed to him by other thinkers and how his test illustrates a property of the phenomenon of intelligence and suggests a hypothesis on machine learning. This reconstruction of the Turing test provides a rapprochement to the conflicting views on its value in the literature.

Keywords Alan Turing · Turing test · Thought experiment · Epistemology · Philosophy of Science · Conceptual foundations of AI and Machine Learning

1 The Turing Test Dilemma

Alan Turing opened his seminal paper by proposing to replace the question 'can machines think?,' which he deemed 'too meaningless to deserve discussion' (1950, p. 442). The new question, considered to have a 'more accurate form,' would be based on what Turing called the 'imitation game,' and later in the same text,

✉ Bernardo Gonçalves
begoncalves@usp.br

[1] Polytechnic School, University of São Paulo, São Paulo, Brazil

his 'test.'[1] Essentially, according to different interpretations of the various versions of the test, the machine must be able to imitate stereotypes of a woman, a man, or a human, beside a true representative of the kind, to deceive a human interrogator about its true nature. The new question is whether the interrogator, at a distance and having no physical contact whatsoever, would be able to distinguish the machine from the genuine individual through a conversation game. If not, the machine must be considered intelligent.

However, details about this new question and the exact settings for its evaluation (duration, number of test runs, scoring protocol, characterization of the players and interrogators) slip through Turing's 1950 text in a sequence of variations that defies interpretation. Two different versions have been identified (Sterrett, 2000; Traiger, 2000) and have been referred to as the 'Original Imitation Game' (read from pp. 433–434 in Turing, 1950) and the 'Standard Turing Test' (read from p. 442). There is significant disagreement on how the two passages should be read. Some authors acknowledged the presence of a 'gender test' in the first passage (Genova, 1994; Hayes & Ford, 1995). Others considered it to serve as a scoring protocol for a non-gendered test read from the second passage (Copeland, 2004, p. 436; Proudfoot, 2013, p. 395). Others have disregarded any form of gender imitation and read from the second passage instead, a standalone Standard Turing Test (Moor, 1976; Dennett, 2006 [1984]; Piccinini, 2000; Moor, 2001; Shieber, 2007), which turns out to be the most popularized version of the test. According to it, the game tests the machine's capability of giving sensible answers to questions, both complex and simple, indistinguishably from a human in an unrestricted conversation game conducted by the interrogator.

This article refers to the 'imitation game,' 'Turing's test(s),' and the 'Turing test' without committing to a specific passage from Turing's texts; instead, it considers a conflation of several passages that will be examined in due course (Sect. 3). Beyond existing disputes about which Turing test is best for artificial intelligence (AI), this article will characterize influential positions on the 'Turing Test Dilemma,' which asks whether the Turing test is a valuable experiment for AI. Preliminaries are presented (Sect. 1.1), and two main positions are identified (Sects. 1.2, 1.3) relative to the two horns of this dilemma.

1.1 Preliminaries: The Practical Turing Tests

First, it is worth noting that there have been several attempts at running the test as a controlled experiment. Two such ventures received much attention—the 1991 and the 2014 editions of the Loebner Prize Competition held in Boston (Epstein, 1992)

[1] Turing wrote about the 'imitation game' centrally and extensively throughout his text (1950), but apparently retired the term thereafter. He referred to '[his] test' four times—three times in pp. 446–447 and once on p. 454. He also referred to it as an 'experiment'—once on p. 436, twice on p. 455, and twice again on p. 457—and used the term 'viva voce' (p. 446). Later, Turing referred to a 'viva-voce examination' (2004 [1951a], p. 484) and multiple times to '[his] test' (2004 [1952]), including 'one of my imitation tests' (p. 503).

and at the Royal Society in London (Warwick & Shah, 2015). The top-ranked program in 1991 was the PC Therapist, developed by a psychology graduate turned computer programmer. It was inspired by ELIZA, a trick developed in the mid-1960s by Weizenbaum (1966) to imitate a 'person-centered' (Rogerian) therapist by regurgitating the patient's own words and phrases in a simulation of understanding. The top-ranked program in 2014 was Eugene Goostman, the simulation of a Ukrainian boy that claimed to be restricted by his acquired culture at 13 and his use of English as a second language.

Based on his first-person experience with the former, Shieber (1994) provided a rigorous and comprehensive analysis of the problem of implementing a practical Turing test. To have its coherence preserved, Shieber remarked, the Turing test could not be restricted in its domain (it must be open for any conversation topic) or task (it must be open for any question). Shieber recommended that practical Turing tests should not be run until the standard of AI gets close to the high standards required by the test.[2] However, the Loebner Prize has been continued all the same. The 2014 edition was organized by Kevin Warwick and Huma Shah, who had been experimenting with practical implementations of the Turing test for several years (2016). Warwick and Shah (2015) announced the Eugene Goostman program as being 'the first to pass the Turing test.' They argued that their 2014 implementation of the Turing test was unrestricted 'as set out by Alan Turing.' Having received criticism from Vardi (2014), Shah and Warwick (2015) presented evidence that the acclaimed program does seem indistinguishable from humans in conversation. Yet, Vardi's rejoinder ran: '[t]he details of this 2014 Turing Test experiment only reinforces my judgment that the Turing Test says little about machine intelligence' (*ibid.*).

1.2 The Negative Answer: The Turing Test is Too Flawed to Be a Valuable Experiment for AI

Given the relatively good performances of obviously unintelligent machines in practical Turing tests, the scientific community seems to have mostly opted to dismiss the test, which would have been revealed to be 'just a game' (Vardi, 2014) or 'highly gameable' (Marcus et al., 2016). This had been discussed by an earlier influential address given by Hayes and Ford (1995), who declared to have tried to 'take Turing seriously.' They acknowledged that Turing's test 'has been with AI since its inception, and has always partly defined the field.' Further, they recollected, '[s]ome AI pioneers seriously adopted it as a long-range goal, and some long-standing research programs are still guided by it.' They suggested that scientists abandon the goal of constructing a 'mechanical transvestite.' They also referred to the practical Turing tests, which would have shown that the test has plenty of ambiguities, flaws, and

[2] Shieber (1994) made an informative analogy with the Kremer Prize for human-powered flying inspired by the designs of da Vinci. A cash prize is an appropriate incentive in this case because 'the task is just beyond the edge of current technology,' Shieber observed (p. 74). He noted that 'limited tests are better addressed in the near term by engineering (building bigger springs) than science (discovering the airfoil)' (p. 77) and suggested that there still is a substantial scientific gap to be filled in AI.

gaps in its design. Further, it would be a biased and even circular test, the standard of which would be elusive, and it would be unable to detect anything. Accordingly, the Turing test should be rejected and moved 'from the textbooks to the history books' (Hayes & Ford, 1995). Bringsjord et al. (2001) emphasized that attempts to build computational systems able to pass restricted versions of the test have devolved into shallow symbol manipulation designed to fool people and concluded that 'the problem is fundamental: the structure of the [test] is such as to cultivate tricksters.' In summary, '[c]onsidering the importance Turing's Imitation Game has assumed,' Drew McDermott wrote in (2014), 'it is a pity he was not clearer about what the game was exactly.'

In general, critics of the Turing test answer 'no' to the Turing Test Dilemma. According to them, it is unfortunate that the test is an underspecified and poorly designed experiment. However, that position must face the first horn of the dilemma: it is at odds with the intellectual standards of Turing's works (Newman, 1955). Further, if the test is so bad, why has it been defended and attracted so much interest? Would that be due to Turing's credentials alone?

1.3 The Positive Answer: The Turing Test is Too Good to Be Abandoned as an Experiment for AI

The Turing test has been defended before and since its early 1990s practical implementations, primarily by AI philosophers. Moor (1976) was the first to emphasize the generality of the test and to advocate its use in unrestricted experiments (pp. 249–250). Dennett (2006 [1984]) noted that the test comes from a long philosophical tradition ('[p]erhaps he was inspired by Descartes,' p. 297) and observed that it is general enough to subsume several specific intellectual tasks at once. Dennett argued that 'the Turing test, conceived as he conceived it, is (as he thought) plenty strong enough as a test of thinking,' and provoked: 'I defy anyone to improve upon it' (p. 297). He argued that it is a convenient sufficient condition (a 'quick probe,' p. 298) for confirming the presence of a human-level AI. After the first practical Turing tests, Dennett (2006 [1997]) regretted that the Turing test 'requires too much Disney and not enough science' and that it 'is too difficult for the real world' (p. 315). Copeland (2000) rejoined: '[i]t is often claimed that Turing was insufficiently specific in his description of his test' (p. 530). 'A machine emulates the brain,' Copeland clarified, 'if it plays the imitation game successfully come what may, with no field of human endeavour barred, and for any length of time commensurate with the human lifespan.' Concerning the difficulties of implementing such an unrestricted experiment, he suggested that the solution lies in sampling: '[a]ny test short enough to be practicable is but a sampling of this ongoing situation.' Shieber (2007) presented a statistical-proof scheme to substantiate the inferential status of the Turing test as a sufficient condition for intelligence, and arguably it could be adapted along the lines suggested by Copeland. However, despite the availability of such an elegant mathematical device, according to Turing, the test must rely on the judgment of 'an

average interrogator' (1950, p. 442) or of 'a jury, who should not be expert about machines' (2004 [1952], p. 495), and such judgments can be flawed.[3]

In general, supporters of the Turing test answer 'yes' to the Turing Test Dilemma. They hold that, in its original (unrestricted) form, the test is not comparable with the restricted practical Turing tests run so far and is too good to be abandoned as an AI experiment. However, this leads to the second horn of the dilemma: if the test cannot be supplanted, will the success of AI science depend on the chances of average human interrogators against increasingly elaborate, yet still unintelligent, chatbots in unrestricted tests? In any case, does running repeated unrestricted Turing tests bring value to AI?

Altogether, taking the dilemma by any one of the two horns, no simple and general explanation of the Turing test seems available to deal with the other horn.

2 Argument Sketch

This article argues that the Turing Test Dilemma can be solved by reconstructing the test as a thought experiment in the modern scientific tradition. No study of the Turing test appears to have ever reconstructed it as a thought experiment.

A core criticism of the test's value as an AI experiment is that Turing would not have specified exact settings for implementing it, whose design would turn out to be poor and imprecise. This view is evidenced, for instance, by the existence of two widely acknowledged and yet heterogeneous readings of the test: the Original Imitation Game and the Standard Turing Test. However, this article will argue that Turing's presentation of his test (Sect. 3) satisfies what Ernst Mach called 'the basic method of thought experiments' (Sect. 4), characterized by a continuous variation of experimental conditions (1976 [1897]).[4] 'By astute handling of this procedure,' Mach observed, 'we may reach cases that at first blush seem rather different, that is to generalisation of the point of view.' Showing that Turing's presentation of his test satisfies Mach's observations establishes that the Turing test can be understood as a thought experiment in the modern scientific tradition, had Turing been aware of it or not.[5] Accordingly, the critique that the test is an underspecified and flawed experiment can be rebutted by showing the rich methodological structure in Turing's exposition of his imitation game and test.

Also in support of understanding Turing's proposal within the scientific tradition, this article will reconstruct the Turing test as a thought experiment serving

[3] The question of whether Turing, the mathematician, would suggest a subjective criterion for justifying an intelligence claim will be addressed later (Sect. 6.1).

[4] Mach is often acknowledged as the thinker who established the use of the term *Gedankenexperiment* ('thought experiment') in the modern scientific tradition.

[5] Further research may explore the historical and analytical roots of Turing's familiarity with thought experiments. Floyd (2017) identifies the intellectual origins of Turing's analysis of computability with the Cambridge tradition of 'common sense,' with which Turing particularly engaged during his formative years at Cambridge in the early 1930s. 'Logic was approached,' she writes, 'not first and foremost axiomatically, but practically and in thought experiments' (p. 106).

both critical and heuristic uses (Sects. 5, 6). Popper (2002 [1959]) presented a discussion of 'apologetical,' 'critical' and 'heuristic' uses of 'imaginary experiments' (pp. 465–466). Popper found in Galileo's criticism of Aristotle's theory of motion in the context of his polemic with peripatetic philosophers the paradigmatic case of the *critical* use of thought experiments. Similarly, Turing's critical use of his test addressed and posed severe problems to opposing theories of intelligence presented to Turing by his intellectual opponents in the context of controversy. In particular, seeking conceptual change on the meaning of the words 'machine' and 'think,' Turing tried to expose a paradox in a theory of intelligence that tied logical kind to physical kind, which had been presented to him by a contender as will be shown later. This satisfies Thomas Kuhn's conception of the function of thought experiments (1977 [1964]). Popper also pointed out Einstein's experiment of the accelerated lift as a paradigmatic case of the *heuristic* use of thought experiments as 'it illustrates the local equivalence of acceleration and gravity, and it suggests that light rays in a gravitational field may proceed on curved paths.' According to Popper, therefore, the heuristic use illustrates a property of the studied phenomenon and suggests a related hypothesis. The reconstruction of Turing's heuristic use of his test will conform to that scheme. The Turing test illustrates that the perception of intelligence is emotional, and it suggests the hypothesis that a learning machine may be created simple and educated naturally, without reboots or special coaching, to play the imitation game well.

The reconstruction of Turing's critical and heuristic uses of his test will emphasize how it increases understanding of the question 'can machines think?' and prepares for related practical experiments. Attention will be drawn to how the imitation game accomplishes its epistemic goals through its design and not by its execution.[6] Overall, the reconstruction of the test as a thought experiment will provide a rapprochement to the conflicting views on the value of the Turing test for AI and can ultimately end the Turing Test Dilemma as a two-horned issue.

This argument sketch summarizes this article's contributions to advancing a crucial debate on the conceptual foundations of AI and machine learning. The remainder presents the complete argument in detail. The key points will be revisited at the end (Sect. 7).

[6] A similar prospect has been suggested by Kuhn (1977 [1964]) in his analysis of Galileo's thought experiment appearing at the start of 'The First Day' in the *Dialogue Concerning the Two Chief World Systems*: '[f]or his purpose in this part of the *Dialogue*, it is quite sufficient that we may suppose these things [viz., uniformly accelerated motion and equal instantaneous velocities of the bodies at the bottom of their fall] to be the case' (pp. 251–252). Sorensen (1992) also singled out this as a distinguished property of thought experiments compared to practical experiments. He defined thought experiments as 'experiments that purport to deal with their questions by contemplation of their design rather than by execution' (p. 6). More recently, Stuart (2018) developed a related view of the power of thought experiments in establishing not necessarily new (justified) knowledge but understanding. Stuart leveraged two decades of results in the epistemology of understanding. Further work may build upon the contributions of this article to extend the analysis of Turing's imitation tests in connection with the most recent literature on thought experiments in science and philosophy (e.g., Stuart et al., 2018).

3 Turing's Presentation of His Test

Turing's presentation of his test will be studied by emphasizing how he varies the conditions of his test (Sect. 3.1). Then the methodological structure of Turing's exposition will be outlined (Sect. 3.2).

3.1 Turing's Variation of the Conditions of His Test

To replace the question (*Q*) 'can machines think?,' Turing introduced his imitation game:

> The new form of the problem can be described in terms of a game which we call the 'imitation game'. It is played with three people, a man (A), a woman (B), and an interrogator (C) who may be of either sex. The interrogator stays in a room apart from the other two. The object of the game for the interrogator is to determine which of the other two is the man and which is the woman. He knows them by labels X and Y, and at the end of the game he says either 'X is A and Y is B' or 'X is B and Y is A'. [...] It is A's object in the game to try and cause C to make the wrong identification. [...] The object of the game for the third player (B) is to help the interrogator. [...]
> We now ask the question, 'What will happen when a machine takes the part of A in this game?' Will the interrogator decide wrongly as often when the game is played like this as he does when the game is played between a man and a woman? These questions replace our original, 'Can machines think?' (Turing, 1950, pp. 433–434)

The substitute question (*Q′*), therefore, was based on Turing's imitation game and test. It has been referred to, as mentioned, as the Original Imitation Game. Turing then illustrated a few queries that the interrogator could make and suggested that all communication between the interrogator and the participants should be teletyped to neutralize signals such as tone of voice. Structurally, Turing's first presentation of the game relates two variants: a baseline *man-imitates-woman* game and a *machine-imitates-woman* game. Results of the latter are supposed to be compared with the results of the former. In commenting on practical Turing tests, Copeland (2004) argued that this comparison of results is a scoring protocol (p. 436). However, this misses the point that the comparison performs a conceptual function. It reminds the reader of a common-sense truism—namely, that a man can imitate stereotypes associated with women despite their biological difference.

Turing proceeded to discuss strengths and weaknesses of the new problem and which machines would be concerned in the game. Having introduced digital computers as the kind of machine allowed to take part in the game, he paused and revisited the new problem:

> There are already a number of digital computers in working order, and it may be asked, 'Why not try the experiment straight away? It would be easy to satisfy the conditions of the game. A number of interrogators could be

> used, and statistics compiled to show how often the right identification was given.' The short answer is that we are not asking whether all digital computers would do well in the game nor whether the computers at present available would do well, but whether there are imaginable computers which would do well. (Turing, 1950, p. 436)

This new formulation can be identified as Q'': are there 'imaginable computers' that could perform well in the imitation game? This reference to an imaginary experiment should not pass by unnoticed. Turing promised to present that question 'in a different light later,' and proceeded to explain a key scientific property of the new digital computers: their universality. He had given a conceptual description of the digital computer as a discrete-state machine. He then used the imitation game to illustrate his point once again:

> Given the table corresponding to [any] discrete state machine it is possible to predict what it will do. [...T]he digital computer could mimic [its] behaviour. The imitation game could then be played with the machine in question (as B) and the mimicking digital computer (as A) and the interrogator would be unable to distinguish them. (Turing, 1950, p. 441)

Turing further remarked that '[t]his special property of digital computers, that they can mimic any discrete-state machine, is described by saying that they are *universal* machines' (p. 441, no emphasis added). Turing thus used this *machine-imitates-machine* variant of the game to suggest that physical kinds could, in principle, have their logical behavior imitated, as long as the imitating agent was properly qualified for universal computation.

In yet another variation, Turing considered 'again the point raised at the end of §3' (Q''), which he had promised. Now, having explained the science and technology of digital computers and their universality property, he posited:

> It was suggested tentatively that the question, 'Can machines think?' should be replaced by [question Q'', which is also] equivalent to this, 'Let us fix our attention on one particular digital computer C. Is it true that by modifying this computer to have an adequate storage, suitably increasing its speed of action, and providing it with an appropriate programme, C can be made to play satisfactorily the part of A in the imitation game, the part of B being taken by a man?' (Turing, 1950, p. 442)

This version of the test (Q''') reinstates the A/B/C-player structure in a *machine-imitates-man* game. Turing's reference to 'man' has been generally read as masculine generics. This is the case of the Standard Turing Test, which reads in Turing's passage a *machine-imitates-human* game and discards the baseline man-imitates-woman game as an implicit scoring protocol in question Q'''. The present reconstruction of the Turing test as a thought experiment can end the exegetical problem of whether Turing meant an ungendered human, as will be shown later (Sect. 5.3). In any case, Turing's literal use will be followed for simplicity, and this version will be referred to as a 'machine-imitates-man' game.

Once having considered 'the ground to have been cleared' (p. 442), Turing revisited 'the original form of the problem' (Q) and 'the more accurate form of the question' (Q'''):

> I believe that in about fifty years' time it will be possible to programme computers, with a storage capacity of about 10^9, to make them play the imitation game so well that an average interrogator will not have more than 70 per cent, chance of making the right identification after five minutes of questioning. (Turing, 1950, p. 442)

Turing thus guesses an answer to yet another question (Q''''): can a machine of gigabit-storage capacity be programmed to deceive an average interrogator in 30% of the times that it plays the imitation game for 5 min?

In his text (1950), Turing presented research steps that 'should be taken now if the experiment [question Q''''] is to be successful' (p. 455). Therefore, contrary to the view of some commentators that Q'''' is a prediction, and thus it could not rule the test, Turing did suggest that it is a valid version of 'the experiment.' Warwick and Shah (2015) sought to implement the conditions of Q'''' very closely and claimed that the 'Eugene Goostman' chatbot satisfied it. So, thinkers that answer positively to the Turing Test Dilemma either diverge from Turing's original proposal or should not reprobate the claim.

At the end of his text (1950), Turing was unsure about which intellectual field was best to address in a test for machine intelligence. He referred to machines eventually competing with men 'in all purely intellectual fields' and asked (p. 460): '[b]ut which are the best ones to start with?' He pondered that even this 'is a difficult decision' and added: '[m]any people think that a very abstract activity, like the playing of chess, would be best. It can also be maintained that it is best to provide the machine with the best sense organs that money can buy, and then teach it to understand and speak English.' In (2004 [1948]), Turing had discussed kinds of intelligence task to be explored in machine intelligence research (pp. 420–421) and even described an imitation test based on the game of chess, referring to it as 'a rather idealized form of an experiment I have actually done' (p. 431). In (2004 [1951a], 2004 [1952]), he presented yet other versions of his test, having even acknowledged the existence of several 'imitation tests' (cf. Note 1). Altogether, Turing presented various imitation tests not only *throughout* his 1950 text, but also *before* and *after* it.

3.2 The Case–Control Methodological Structure of Turing's Various Conditions and Questions

To replace the original question, Q, Turing posed in his 1950 paper various empirical questions, Q' to Q''''. These are based on different game variants through varying players A and B while keeping C fixed: (i) man–woman, (ii) machine–woman, (iii) machine–machine and (iv) machine–man and (v) machine as A in the absence of B. The questions can be generalized as follows:

Question $Q^{\star}$: could player A imitate intellectual stereotypes associated with player B's type successfully (well enough to deceive player C), despite the physical differences between A and B's types?

Turing's varied conditions establish two levels of case–control structure. At the intra-game level, A plays the case, and B plays the control. At the inter-game level, the case–control structure alternates as follows. Note that question $Q^{\star}$ is open concerning the *machine–woman* and the *machine–man* versions of the game, both of which set the case; however, the same question is settled concerning the *man–woman* and the *machine–machine* variants of the game, which set the control. Beyond Turing's rhetorical use of the man–woman variant of the game, it is well known that a man (A) can possibly imitate gender stereotypes associated with a woman (B) successfully, despite their physical difference. Further, regarding the machine–machine variant, it is also known that a digital computer (A), because of its universality property proven by Turing (1936), can successfully imitate any discrete-state machine (B), despite their physical difference.

We may now proceed to analyze Turing's presentation of his test against the backdrop of classical conceptions of thought experiments in the philosophy of science literature.

4 Turing's Use of the Basic Method of Thought experiments

Turing's presentation of his test satisfies Mach's conception of the basic method of thought experiments, which is variation, continuously if possible. Ernst Mach's characterization of the method will be presented (Sect. 4.1) and then compared with Turing's use of it in his exposition of his imitation tests (Sect. 4.2).

4.1 Mach's Characterization of the Method

Throughout his text, Mach developed sharp observations and insights on thought experiments, which he grounded in countless examples from the history of modern physics, mathematics, and commonsense experience. On the method, he wrote:

> [T]he basic method of thought experiments, as with physical experiments, is that of variation. By varying the conditions (continuously if possible), the scope of ideas (expectations) tied to them is extended: by modifying and specializing the conditions we modify and specialize the ideas, making them more determinate, and the two processes alternate.[7] (Mach, 1976 [1897], p. 139)

[7] Mach's specific words in German are (no emphasis added): 'Wie man sieht, ist die Grundmethode des Gedankenexperimentes, ebenso wie jene des physischen Experimentes, die Methode der *Variation*. Durch wenn möglich kontinuierliche Variation der Umstände wird das Geltungsbereich einer an dieselben geknüpften Vorstellung (Erwartung) erweitert; durch Modifikation und Spezialisierung der ersteren wird die Vorstellung modifiziert, spezialisiert, bestimmter gestaltet; und diese beiden Prozesse wechseln.'

It is important to note the mutually reinforcing connection he suggested between extending 'the scope of ideas (expectations)' and 'varying the conditions,'[8] where variation means 'modifying and specializing,' continuously 'if possible.' Mach illustrated his point through an account of the process of discovery of universal gravitation. Preceding the passage above, he wrote:

> A stone falls to the ground. Increase the stone's distance from the earth, and it would go against the grain to expect that this continuous increase would lead to some discontinuity. Even at lunar distance the stone will not suddenly lose its tendency to fall. Moreover, big stones fall like small ones: the moon tends to fall to the earth. Our ideas would lose the requisite determination if one body were attracted to the other but not the reverse, thus the attraction is mutual and remains so with unequal bodies, for *the cases merge into one another continuously*. Not only logical elements are at play here: logically, discontinuities are quite conceivable, but it is highly improbable that their existence would not have betrayed itself by some experience. Besides, we prefer the point of view that causes less mental exertion, so long as it is compatible with experience. (Mach, 1976 [1897], pp. 138–139, emphasis added)

By 'logical,' Mach means conceptual, and by 'continuous,' he means fluid and extendable. The fall's distance and the stones' size are the experimental conditions, which are continuously varied in the physicist's mind and eventually stretched to the celestial scale. Reciprocally, the concept of a celestial body, such as the earth and the moon, becomes interchangeable with the concept of a stone, and quite unequal stones can then become mutually attracted. The scope of ideas (expectations) tied to the conditions of the fall of stones is extended simultaneously to the conditions themselves. The cases merge into one another continuously: a conceptual integration is established, connecting near-earth bodies to celestial bodies under a unified physical concept.

In the above example, as in most of Mach's examples, the experimental conditions comprise physical quantities, which makes 'continuous' variation coincide with spanning a real-valued domain. However, a close reading of Mach's entire argument, developed in fourteen numbered analytical steps, suggests that his conception of 'conditions' and their variation, 'continuously if possible,' is broad rather than narrow. That is, although Mach mostly referred to quantitative ideas, he meant the physicist's conceptual representation of sense experience rather than the instantiation of a mathematical model with numerical initial and boundary conditions on physical quantities such as distances, angles, and particle densities. This will be illustrated in what follows.

Mach resumed his account of universal gravitation as a remarkable conceptual integration achieved using the method of continuous variation. He referred to

[8] The German words *Vorstellungen* and *Umstände* persisted throughout Mach's original text and were translated to English as 'ideas' and 'conditions'. Alternatively, they could be translated as 'mental images' and 'circumstances.' Note that Mach's seminal text on thought experiments comes as a chapter of his book *Knowledge and Error: Sketches on the Psychology of Enquiry*.

Galileo as a master of this kind of thought experiment and discussed three of his thought experiments, including the one on free-falling bodies:

> If a body of greater weight had the property of falling faster, a combination between a light and a heavy body, would, though heavier still, have to fall more slowly because retarded by the lighter component. The assumed rule is thus untenable because self-contradictory.[9] (Mach, 1976 [1897], p. 139)

Note that properties such as having a 'greater weight,' 'falling faster' and 'fall more slowly' correspond to what Mach calls 'quantitative ideas:'

> Planned quantitative experiment yields many details, but our *quantitative ideas* educated by experiment gain their surest support if we relate them to [unintentionally and instinctively gained] raw experiences. Thus, Stevin adapts his *quantitative ideas* about inclined planes to that experience about the gravity of bodies by means of exemplary thought experiments, and Galileo does likewise with [his quantitative] *ideas* concerning free fall. (Mach, 1976 [1897], p. 141, emphasis added)

Also mentioning Stevin's experiments, Mach connected Galileo's experiments on free fall and inclined planes. In either case, Stevin's or Galileo's, Mach suggests, there is no sharp line distinguishing their thought experiments on inclined planes, on the one hand, and those on the gravity of bodies and falling bodies, on the other hand.[10] Although the inclined plane could be seen as a different setting compared to free-falling bodies, Mach notes that in Galileo's thought, they are the same setting continuously varied, which is done by *modification* and *specialization*. By the method of variation, 'the cases merge into one another continuously,' that is, they are conceptually integrated.

We may now proceed to see how this works in Turing's imitation tests.

4.2 Turing's Use of the Method of Continuous Variation

A reconstruction of Turing's imitation tests as an application of Mach's method of the variation of conditions (continuously, if possible) must show how, in Turing's perspective, the various imitation tests merge into one another continuously. First,

[9] Mach neglects that the contradiction can only be empirically verified by assuming the movement takes place in a vacuum, despite that Aristotle's view of motion echoed by Simplicio in the *Two New Sciences* applies to the fall of bodies in media. For a detailed discussion, see Norton (1996, p. 20). None of this compromises Mach's analysis of thought experiments.

[10] Since Mach, there have been substantial accounts of how Galileo connected his law of free fall with his experiments on inclined planes, transferring empirical evidence from the latter to the former. Based on his Galilean studies, Koyré (1953) states: 'It is well known with what extreme ingenuity, being unable to perform direct measurements, Galileo substitutes for the free fall the motion on an inclined plane on one hand, and that of the pendulum on the other' (p. 224). After the controversies with Stillman Drake on whether Galileo could have ever validated some of his empirical claims, Naylor's (1974) study came to confirm Koyré's point on both historiographical and empirical grounds. This indicates the depth of Mach's insight that variation is the fundamental method of thought experiments and that these 'lie at the basis of science and consciously aim at widening experience' (1976 [1897], pp. 135–136).

it is necessary to introduce Turing's view of creation and evolution, particularly his view of humans, animals, and living beings as machines. Turing's view builds, among other sources, on the organic-machine metaphors of Edwin T. Brewster's *Natural Wonders Every Child Should Know* (1912), which he read in childhood (Hodges, 2012, p. 11).

In (2004 [1948]), Turing dedicated a section (Sect. 3) of his text to describe 'Varieties of machinery.' He observed that '[a]ll machinery can be regarded as continuous, but when it is possible to regard it as discrete it is usually best to do so' (p. 412). A brain, he noted, 'is probably' a 'continuous controlling' machine, but given the digital nature of neural impulses, it 'is very similar to much discrete machinery' (p. 412). Defining the possible states of a machine as a discrete set instead of a continuous set can be convenient for controlling purposes since a 'reasonably accurate knowledge of the state at one moment yields reasonably accurate knowledge any number of steps later' (1950, p. 440). In another section (Sect. 6), 'Man as Machine,' Turing construed the differences between 'man' and human-made 'machine' in terms of a continuum:

> A great positive reason for believing in the possibility of making thinking machinery is the fact that it is possible to make machinery to imitate any small part of a man. (Turing, 2004 [1948], p. 420)

Because 'any small part' could be imitated, he imagined:

> One way of setting about our task of building a 'thinking machine' would be to take a man as a whole and to try to replace all the parts of him by machinery. He would include television cameras, microphones, loudspeakers, wheels and 'handling servo-mechanisms' as well as some sort of 'electronic brain.' (Turing, 2004 [1948], p. 420)

He dismissed such a method as 'altogether too slow and impracticable.'

Turing viewed human intelligence in continuity with animal intelligence, as indicated by his formulation of the 'Heads in the Sand' objection to the possibility of machine intelligence: '[w]e like to believe that Man is in some subtle way superior to the rest of creation' (p. 444). This was an elaboration of objection '(a)' from (2004 [1948]), which referred to an 'unwillingness to admit the possibility that mankind can have any rivals in intellectual power' (p. 410). Turing, a reader of Samuel Butler,[11] considered human-made machines as a species. In ca. mid-1951, he started new foundational research on the genesis and development of organic forms.[12] (In

[11] Butler's 'The Book of the Machines' appears in the bibliography of Turing's 1950 paper. A full-fledged study of the Turing-Butler connections is a topic of future work.

[12] Turing's (1952) study revealed what is called today Turing 'structures' or 'patterns.' Never been observed in nature at the time, they were later experimentally verified and by now have been observed in objects ranging from biological tissues to sand dunes, also appearing at the atomic level (Fuseya et al., 2021).

terms of today's concepts of hardware and software, one could say that he started research on organic hardware formation.[13])

For Turing, the differences in intelligence power among all 'species,' human-made machines included, were contingent products of evolution, whether natural or artificial. Until 1950, however, he addressed the problem of intelligence in terms of programming digital computers, which was the technology that he contributed to developing and was nearly available for use at the time.[14] Along these lines, Turing suggested machine intelligence could be achieved by making a learning program to simulate a child's mind and subjecting it to 'an appropriate course of education,' in analogy with evolution:

> There is an obvious connection between this process and evolution, by the identifications
> Structure of the child machine = Hereditary material
> Changes of the child machine = Mutations
> Natural selection = Judgment of the experimenter
> One may hope, however, that this process will be more expeditious than evolution. The survival of the fittest is a slow method for measuring advantages. The experimenter, by the exercise of intelligence, should be able to speed it up. (Turing, 1950, p. 456).

Turing believed machine intelligence could be progressively developed and eventually achieved by subjecting machines to artificial evolution. The development of the machine's intelligence would depend on the experimenter's intelligence. Although he saw species evolution, whether natural or artificial, physical or cultural, in continuity (i.e., conceptually integrated), his imitation tests presented in (1950) considered hardware fixed (Sect. 3) to focus on software instead (Sect. 7). Thus by machine evolution, in 1950, he meant cognitive and cultural evolution.

This clears the ground for shedding light on Turing's use of the method of continuous variation in the design of his imitation tests. There is a core experimental setup based on players A, B, and C and their goals in the imitation game. In Turing's view, all the players are machines in either organic or inorganic, discrete or continuous controlling form. Other types and subtypes apply: woman and man are subtypes of human, which is a subtype of organic and continuous machine. At the same time, differential analyzer and digital computer with its associated learning program are subtypes, respectively, of continuous and discrete non-organic machine. The fundamental question Turing asks (question $Q^{\star}$) is whether the intellectual and cultural performances associated with the types, namely their related stereotypes, could be

[13] To be precise, Turing's computability theory conceived of 'a fluidity between hardware, software and data' (Floyd, 2017, p. 104), with different elements being explored in different studies.

[14] Turing's interest in referring to digital computers in his argument can also be attributed to their popularization by the media in the context of a public controversy on their meaning and significance. Turing, as shown by Floyd (2017), held a strong and overarching interest in connecting advanced technical knowledge with the common language of 'the man in the street.'

imitated, thus empirically showing that the types can be softly transposed.[15] Note that for any arbitrarily chosen type, say, a 'woman,' further specific subtypes can be continuously conceived and considered as varied conditions of the imitation game: women having property p, women having property $p' \subset p$, and so on. Further, for any two arbitrarily chosen types (say, a 'machine' and a 'man'), a new type can be conceived, whether as a specialization or a modification (cf. Turing's thought experimentation on imitating 'any small part of a man'). The existence of such an evolving continuum of levels and types relates to the fact that concepts are fluid entities.[16] This analysis shows how Mach's characterization of the method of continuous variation applies to Turing's imitation tests, or how, in Mach's sense, Turing's variation of conditions aims to make 'the cases merge into one another continuously.'

Further boundary conditions can be varied (continuously if possible): the game's duration, the number of trials, B's actual presence in the game,[17] and the machine's hardware and software capacities. The question across the various versions of the game can be posed this way: how does C's perception of A's performance change as the game's conditions are (continuously) varied? Will it change if gendered verbal behavior, as a subtype of human verbal behavior, is required? Will it change if the game's duration is reduced? Will it change if the machine's hardware is increased and/or its learning program is modified? For Turing, there is no conceptual discontinuity at all among the various conditions that can be chosen for instantiating his thought experiment.

Mach emphasized that 'the basic experimental method of variation' is found within 'man' himself, who collects experiences 'by observing changes,' above all the changes 'he can influence through his own intervention and deliberate movements.' Mach described the playful, instinctive experiments of a child, such as being surprised by their mirror image or shadow in sunlight, and added:

> If the adult temporarily loses these treasures so that he must as it were discover them afresh, the explanation is that his social upbringing narrows his circle of interests and confines him to it while at the same time he acquires a large number of ready opinions, not to say prejudices, that he supposes not to be in need of examination. (Mach, 1976 [1897], p. 134)

[15] Sterrett (2000) opened this perspective on Turing's imitation game (cf. Sects. 5.2, 6.2). As Floyd (2017) noted, Turing was concerned with *types*—'the delimited, surveyable ordering and organizing of objects, concepts, terms, logical particles, definitions, proofs, procedures and algorithms into surveyable wholes' (p. 104)—and their connections with common sense.

[16] Floyd (2017) opened this perspective on Turing's thought, writing: '[h]e saw the difference in levels and types as a complex series of systematizations sensitive to everyday "phraseology" and common sense, not a divide of principle. This was because he always saw "types" or "levels" as lying on an evolving continuum, shaped by practical aspects, the user end, and mathematics' (p. 142). This aspect of Turing's thought, Floyd suggests, can be attributed to the Cambridge tradition of "common sense." '

[17] Mach referred to the exclusion of certain conditions as 'mentally diminishing to zero one or several conditions that quantitatively affect the result, so that the remaining factors alone must be taken as of influence' (1976 [1897], p. 140). The quantitative result in Turing's case is the interrogator's success rate in making the correct identification of the players.

Clearly, Turing did not lose these treasures. Based on Mach's analysis, the fact that Turing's thought experiment involves cultural issues does not make it unscientific. For Mach, '[t]here is no sharp dividing line between instinctive and thought-guided experiments' (p. 134). This is also in line with Floyd's sensible observation (2017) that Turing used common sense as a scientific tool.

At least rhetorically, Turing did not consider that his thought experiment dispensed with physical experiment. He stated that '[t]he only really satisfactory support that can be given' for his positive answer to question Q'''' 'will be that provided by waiting for the end of the century and then doing the experiment described' in the question (1950, p. 455). Once again, this is consistent with Mach's analysis:

> If a thought experiment is without definite issue, that is[,] when the idea of certain conditions leads to no certain and unambiguous expectation of a result, we tend to turn to guessing, at any rate for the period between thought and physical experiment, that is[,] we tentatively assume an approximately sufficient condition for a result. This guessing is not unscientific, but a natural process that can be illustrated by historical examples. (Mach, 1976 [1897], p. 141)

Mach further noted that '[t]he method of letting people guess the outcome of an experimental arrangement has didactic value too' (p. 142).

In light of Mach's analysis, Turing's exposition of his various imitation tests should not be confused with loose rhetoric. Rather than being sloppy, the presentation of his thought experiments can now be understood as methodical. The various questions that Turing asked offered an empirical basis for discussing the original question (can machines think?) under varied limiting conditions. The design of his imitation game was deliberately flexible to address conceptual problems. This observation liberates AI scientists from Turing's specific rhetoric to design, even if Turing-inspired, meaningful, practical experiments. As Mach emphasized, 'thought experiment often precedes and prepares physical experiments' (Mach, 1976 [1897], p. 136).

We may now proceed to gain further depth into Turing's uses of his imitation tests and examine what specific conceptual problems they address.

5 Turing's Critical Use of His Test

As is often the case with thought experiments, Turing proposed his test in the context of intellectual controversy (Gonçalves, 2022). The significance of the newly existing digital computers was under dispute in post-war England. In 1949, Turing was exposed to strong reactions against his view that machines can think.

Fellow of the Royal Society (FRS) and professor of neurosurgery at the University of Manchester, Geoffrey Jefferson (1886–1961) became Turing's primary intellectual opponent. In his Lister Oration (1949), Jefferson presented a reductionist view of intelligence, characterized as an emergent property of the animal

nervous system. The nervous impulse, he argued, is not a purely electrical phenomenon but also a chemical one that depends on the continuity of specific physical quantities. Further, as will be shown in Sect. 5.2. Jefferson himself used a thought experiment to suggest that gendered behavior is causally related to the physiology of sex hormones. Jefferson's critique of the possibility of machine intelligence was so powerful and comprehensive that it subsumed the objections of other thinkers. For instance, he posited that 'it is cogent argument against the machine that it can answer only problems given to it, and furthermore, that the method it employs is one prearranged by its operator' (p. 1109). This objection was originally championed by Douglas Hartree (1897–1958), FRS and professor of mathematical physics at the University of Cambridge.[18] Moreover, Jefferson cited René Descartes (p. 1106) and suggested that speech is the distinguishing mark of human intelligence compared to other kinds of animal intelligence (pp. 1109–1110). This also covers the objection formulated by Michael Polanyi (1913–1976), FRS and professor of social studies at the University of Manchester. Polanyi had presented to Turing a Gödelian argument (cf. Blum, 2010), which later developed into Polanyi's general theory of knowledge (1974). Essentially, according to it, humans can solve problems that machines cannot. Turing was, until then, using the game of chess as a testbed for machine intelligence.[19] However, Polanyi dismissed it as unimpressive (1974): '[a] routine game of chess can be played automatically by a machine, and indeed, all arts can be performed automatically to the extent to which the rules of the art can be specified' (p. 261). Jefferson's appeal to speech as the hallmark of human intelligence subsumed Polanyi's argument.

It will be shown that Turing's thought experiment attacks those opposing theories of human intelligence through its varied design. It exemplifies what Popper called the critical use of thought experiments. Moreover, it does so by satisfying Popper's methodological rule for 'the use of imaginary experiments in critical argumentation' (2002 [1959]), which is to say that '*the idealizations made must be concessions to the opponent, or at least acceptable to the opponent*' (p. 466, no emphasis added).

5.1 The Function of the Machine–Machine Variant of the Imitation Game

In his Lister Oration (1949), Jefferson argued that the physiology of the nervous system is based on continuous physical quantities. Therefore, it would be incommensurable with the activity of a digital computer, which, as Turing himself explained, is a discrete-state machine. This is a core element of Jefferson's argument. According

[18] Hartree expressed it in his inaugural lecture at Cambridge University (1947, p. 21) in ca. November 1946, and later in his *Calculating Instruments and Machines* (1949, p. 70), then attributing it to Ada Lovelace. Turing responded to it first in his lecture to the London Mathematical Society (2004 [1947], p. 392), and again in his formulation of 'Lady Lovelace's Objection' (1950, p. 450). The debate would reappear in their radio broadcasts in May 1951 (Jones, 2004).

[19] In (1974, p. 261), Polanyi referred to Turing's 1949 argument based on machine chess in the Manchester 'Mind and the Computing Machine' seminar (cf. also Mays, 2000).

to it, thinking is an emergent property that belongs exclusively to the animal nervous system. Therefore it could not be reproduced by computing machines.

Turing executed the critical use of his test against Jefferson's argument, which he formulated as 'the argument from continuity in the nervous system' (1950, pp. 451–452). He acknowledged that '[t]he nervous system is certainly not a discrete-state machine,' for '[a] small error in the information about the size of a nervous impulse impinging on a neuron, may make a large difference to the size of the outgoing impulse.' However, Turing pondered that this does not mean that a discrete-state system cannot mimic the behavior of the nervous system. He argued that the imitation game neutralizes such a difference. He presented an example in which player C asks the other players—A is a digital computer and B is a differential analyzer (a simpler continuous system)—to give the value of a transcendental number such as π. The digital computer could imitate the differential analyzer by choosing at random from a probability distribution between values that approximate the correct answer (say, 3.1416). More generally, the discrete-state machine can use any technique to approximate the continuous-state machine's behavior, and yet an external observer (the interrogator) may not be able to distinguish which is which.

Turing used his test to criticize the argument that a digital computer, as a discrete system, could not imitate human thinking, which is produced by the (continuous) nervous system.

5.2 The Function of Player B and the Man–Woman Variant of the Imitation Game

Wolfe Mays (1912–2005), who was a contemporary of Turing at the University of Manchester and another opponent of his views (Mays, 2001), guessed that a specific source for Turing's imitation game was Twenty Questions (1952, p. 148), a radio parlor game that Turing had made casual reference to in his own writing (1950, p. 457). Despite never relating Turing's imitation game with Twenty Questions, Hodges (2012 [1983]) noted that Turing played the latter with friends during a summer holiday and even 'developed a theory of how to choose the next question so as to maximise the expected weight of evidence of the answer' (p. 389). In the game, the players must identify an entity by asking up to twenty yes-no questions. The only clue that can be provided is whether the item was of animal, vegetable, or mineral nature, which highlights the game's focus on ontological categories. Sterrett (2020) found that since the early 1950s, there have been TV shows whose structure is similar to Turing's imitation game. Inspired by parlor games, the Turing test suits Mach's point that thought experiments are sourced in quasi-sensory information such as combinations of memories of sense elements (1976 [1897], p. 137).

However, why did Turing address the problem of sexual guessing specifically? Sterrett (2000) argued that player A needs to think reactively to avoid giving ingrained responses that would reveal their true kind, and gender is such an intrinsic property of an individual. She remarked that 'cross-gendering is not essential to the test; some other aspect of human life might well serve in constructing a test that requires such self-conscious critique of one's ingrained responses' (p. 470). Sterrett's insight captures in a fundamental way the intellectual skill required from

player A across the various conditions presented by Turing's imitation tests: to turn into what he/it is not. Nevertheless, this question remains: among other possible properties that Turing could have chosen (as Aristotelian differentia for the human genus) in his use of the method of variation, why did he choose gender in particular?

The capability to think through gender had a specific role in Turing's critical use of his thought experiment. Jefferson presented a critique of the artificial behavior of 'modern automata' (1949, p. 1107). He referred to the then famous electromechanical tortoises of the cybernetician Grey Walter and, in doing so, offered Turing an imaginary experiment:

> [...It] should be possible to construct a simple animal such as a tortoise (as Grey Walter ingeniously proposed) that would show by its movements that it disliked bright lights, cold, and damp, and be apparently frightened by loud noises, moving towards or away from such stimuli as its receptors were capable of responding to. In a favourable situation the behaviour of such a toy could appear to be very lifelike – so much so that a good demonstrator might cause the credulous to exclaim 'This is indeed a tortoise.' I imagine, however, that another tortoise would quickly find it a puzzling companion and a disappointing mate. (Jefferson, 1949)

It can be argued that a key function of Turing's 1950 imitation tests is to criticize this thought experiment on automata and gender, which they partly reconstruct. Jefferson brought forward the image of a genuine individual of a kind, which is placed side by side with the artificial one so that the latter's artificiality is emphasized. The function of the genuine individual is to reveal the artificiality of the imposter. That explains Turing's introduction of a control player (B), which only appears as a structural element in the 1950 variants of Turing's imitation tests. In the (2004 [1948], 2004 [1951a], 2004) tests, the machine plays directly against the judge with no control player around. With Popper's rule in mind, the control player can be explained as a concession to Jefferson.

Jefferson referred to 'sex hormones' as a distinctive feature of the intelligent behavior of 'animals' and 'men,' as opposed to 'modern automata' (1949, p. 1107). He remarked that 'neither animals nor men can be explained by studying nervous mechanics in isolation, so complicated are they by endocrines, so coloured is thought by emotion.' He then added: '[s]ex hormones introduce peculiarities of behaviour often as inexplicable as they are impressive' (p. 1107). In effect, Jefferson suggested that machines could not exhibit enough peculiarities of behavior to imitate the actions of animals or 'men' because they are not moved by sex hormones. A machine would give itself away and be found to be 'a puzzling companion and a disappointing mate.' In a further passage,[20] Jefferson stated that he would not agree that 'machine equals brain' until a machine could, among other things, 'be warmed by flattery' and 'be charmed by sex' (p. 1110).

[20] That passage was quoted in full by Turing (1950) in his discussion of the fourth objection (argument from consciousness), which he explicitly attributed to Jefferson (pp. 445–446).

In summary, Jefferson substantiated his argument that human intelligence is an exclusive product of the physiology of the animal nervous system with the thesis that gendered behavior is a causal product of male and female sex hormones. For Turing to meet Jefferson's challenge and conceive a machine that could be convincingly human-like, as opposed to a puzzling companion and a disappointing mate, it would have to be able to learn and successfully imitate gender. The function of player B and the man–woman control variant of Turing's imitation game was to establish, through the simple common sense of a parlor game, that gender stereotypes can be learned and imitated despite the players' physiological differences. Turing thus established from the start of his 1950 text that question $Q^{\star}$ (cf. Sect. 3.2) can be meaningful from a logical point of view (it is not a conceptual paradox) and, therefore, open for empirical study. In other words, rather than serving as a scoring protocol to Q''', Q' serves a rhetorical purpose within the critical function of the Turing test.

Further, the man–woman game tries to expose the existence of a conceptual paradox within Jefferson's theory that physical kind determines logical kind—if a man can imitate intellectual stereotypes associated with a woman despite their physical differences, why could a machine not imitate a woman, a man, or, more broadly, a human? That satisfies Kuhn's characterization of the function of thought experiments (1977 [1964]), for Turing proposed a conceptual change on the traditional concepts of machine and intelligence at the time,[21] which Jefferson had articulated in scholarly form using his background in neurophysiology.

The machine–woman case variant of the game reinstates the question of the learning and imitation of gender stereotypes as a challenging special case of question $Q^{\star}$.

5.3 The Function of Conversation as the Intelligence Task Addressed by the Imitation Game

Since his wartime service from 1941 to late 1949, Turing considered the game of chess as his chosen intelligence task to illustrate, develop and test machine intelligence. In 1948, he discussed a tradeoff between convenient and impressive intellectual fields for exploring machine intelligence. Regarding language, and having discussed 'various games e.g. chess,' Turing (2004 [1948]) wrote : '[o]f the above possible fields the learning of languages would be the most impressive, since it is the most human of these activities' (p. 421).[22] However, he pondered, that field seems 'to depend rather too much on sense organs and locomotion to be feasible.' In the end, he kept his choice for chess and described a chess-based imitation game (p. 431).

[21] The Oxford English Dictionary's definition of 'machine' in the early 1950s (cf. Mays, 1952, p. 149) implies that machine behavior was synonymous with unintelligent behavior, and intelligence was considered an intrinsic property of humankind.

[22] As mentioned, Floyd (2017) contextualizes Turing's interest in the common use of language.

Eventually, as mentioned, Turing's use of chess to test for machine intelligence was directly challenged by Polanyi and indirectly challenged by Jefferson. From 1949 to 1950, Turing changed his option and built his thought experiment in the form of a conversation game. Unlike chess, which is governed by definite rules, good performance in conversation cannot be easily specified. Therefore, Turing's 1950 choice for 'the learning of languages' as the intellectual field addressed in his test can be best understood as yet another concession to Jefferson and, in this case, to Polanyi as well.

Now, note that the machine–man case variant of the game is designed to test the machine's capability of *language learning*, which is Turing's specific uptake of the required skill (language use and understanding). If Turing's various imitation tests are understood as part of his continuously varied thought experiment (Sects. 3, 4), the exegetical problem of whether Turing meant masculine generics in the machine–man game vanishes. That is because gendered language learning, as a challenging special case of natural language learning, had already been implied as a required skill by the machine-woman game.

6 Turing's Heuristic Use of His Test

Turing considered his imitation game as a means to distinguish true language learning from parrot-fashion learning. He addressed this issue also in his response to Jefferson's demand that a thinking machine should be able to create a sonnet on its own (1949, p. 1110). Turing thus presented this example of an exchange between his imaginary machine and player C, the human interrogator, who questions the machine about a sonnet that it has written:

> Probably he [Jefferson] would be quite willing to accept the imitation game as a test. The game (with the player B omitted) is frequently used in practice under the name of *viva voce* to discover whether some one really understands something or has 'learnt it parrot fashion'. Let us listen in to a part of such a *viva voce*:
> Interrogator: In the first line of your sonnet which reads 'Shall I compare thee to a summer's day', would not 'a spring day' do as well or better?
> Witness: It wouldn't scan.
> Interrogator: How about 'a winter's day'. That would scan all right.
> Witness: Yes, but nobody wants to be compared to a winter's day.
> Interrogator: Would you say Mr. Pickwick reminded you of Christmas?
> Witness: In a way.
> Interrogator: Yet Christmas is a winter's day, and I do not think Mr. Pickwick would mind the comparison.
> Witness: I don't think you're serious. By a winter's day one means a typical winter's day, rather than a special one like Christmas.
> And so on. What would Professor Jefferson say if the sonnet-writing machine was able to answer like this in the *viva voce*? I do not know whether he would regard the machine as 'merely artificially signalling' these answers, but if the

> answers were as satisfactory and sustained as in the above passage I do not think he would describe it as 'an easy contrivance'. (Turing, 1950, pp. 446–447)

To understand the heuristic function of Turing's test in the Popperian sense, it is important to emphasize what Turing's imaginary sonnet-writing machine *illustrates* (Sect. 6.1) and what it *suggests* (Sect. 6.2).

6.1 The Turing Test Illustrates a Property of the Phenomenon of Intelligence

Turing presented a standard of intelligent behavior that he thought could be produced by a machine. He believed that the imaginary machine's performance was so 'satisfactory and sustained' that it would stress Jefferson's aprioristic claim that, whatever a machine could do, it would be nothing but a result of shallow symbol manipulation. The practical Turing tests (Sect. 1.1) have shown that Jefferson's point still stands. Whether Turing may have underestimated the power of modern mechanical parrots will be discussed later (Sect. 6.2).

In any case, it is worth noting Turing's manifest uncertainty on how the machine's performance, which he took to be suggestive of true language understanding, would be perceived by Jefferson (perhaps as a mere artifice). Turing had noted (2004 [1948]) that some of the objections to the possibility of machine intelligence were 'purely emotional' (p. 411); therefore, the justification of an intelligence claim could not rest on logic alone. This is an important point illustrated by the heuristic function of the imitation game. The game encodes Turing's insight that explaining 'the cause and effect' of mechanical intelligence makes it unimpressive and seem 'a sort of unimaginative donkey-work' that is unworthy to be called thinking (2004 [1952], p. 500). For that reason, the imitation game has been designed to be a blind experiment centered on behavior rather than on internal states: '[u]sually if one maintains that a machine can do one of these things, and describes the kind of method that the machine could use,' Turing remarked in (1950), 'one will not make much of an impression' (pp. 449–450). It was instead 'the actual production of the machines,' Turing had guessed in (2004 [1948]), that 'would probably have some effect' (p. 411). This explains Turing's use of an imaginary (machine) experiment at a time when he was still waiting for the Manchester Automatic Digital Machine to be available for his first preliminary experiments (Lavington, 2012, p. 99).

Proudfoot (2013) identified in two of Turing's works (2004 [1948], 2004 [1952]) his view that the perception of intelligence is emotional,[23] which she developed into a response-dependence theory of intelligence. This means that a machine can be said to be intelligent if it appears intelligent to 'a normal subject' in certain 'specified conditions' of observation (Proudfoot, 2013, p. 404). In fact, Proudfoot argued (2017), '*the Turing test does not test machine behaviour*'

[23] Referring to a preliminary experiment with machine chess, Turing remarked that '[p]laying against such a machine gives a definite feeling that one is pitting one's wits against something alive' (2004 [1948], p. 412).

(p. 303, no emphasis added). 'Instead,' she wrote, 'it tests the observer's reaction to the machine.' This pushes the Turing test closer to psychometrics and farther from AI. Response dependence can be illustrated through other secondary-quality concepts. For example, a color can be perceived similarly by people who are not colorblind in adequate lighting conditions. This, of course, does not preclude a physics of color, which reifies color as a (response-independent) primary quality concept. However, Proudfoot commits to a notion of 'global response-dependence' (Pettit, 1991, p. 588). This imputes to Turing the view that intelligence is a socially constructed concept whose verifiability rests on the intersubjective judgment of human interrogators. If an unintelligent chatbot fools humans under the specified conditions, the chatbot can be claimed intelligent. Proudfoot takes 'the concept of colour' as being 'very different from the concept of electromagnetic radiation, even though electromagnetic radiation is the physical basis of colour.' 'Likewise,' Proudfoot concludes (2017), 'if intelligence is a response-dependent concept, the concept of intelligence is very different from the concept of computation, *even if* brain processes (implementing computations) form the physical basis of "thinking" behaviour' (p. 305, no emphasis added). Essentially, Proudfoot commits to anti-physicalism: she rejects the reification of the physical concepts of color and intelligence as primary-quality concepts.

Turing, however, did refer to intelligence as a dispositional physical property grounded in material computational power. In (2004 [1948]), he referred to the 'intellectual power' of humankind and other animal species (p. 410) and the 'intellectual power' that the 'isolated man' cannot develop given his limited possibilities for learning (p. 431). In (1950), he referred to 'the power of thinking' (p. 444); and in (2004 [1952]), he said that 'an intelligent human mind' could learn how to learn (p. 497). Turing's physical concept of intelligence and its connection to the Turing test has been explained by his colleague Donald Michie as follows:

> Turing's belief about intelligence was that the *PROPENSITY* is *INNATE*, but the *ACTUALITY* has to be *BUILT*. For him the crux was the brain's ability to make sense of its inputs, that is to *understand* them. And how would we tell whether we had succeeded? To assess degrees of machine understanding he was later to propose what is celebrated today as the *Turing Test*. (Michie, 2002, no emphasis added)

This oral source suggests that Turing did consider intelligence a physical concept and his test a sort of experiment for machine intelligence.

Nevertheless, Turing's experience with Jefferson and others showed that actual intelligence (on the computer, as in the brain) was not enough to *justify* a machine intelligence claim. Especially in the early 1950s, when the traditional concept of intelligence was tied to humans, justifying machine intelligence in terms of inner computational structures would make a circular argument. Instead, machine intelligence had to be demonstrated by addressing language use and understanding—a skill that indisputably belonged to human intelligence—so that it could be *perceived*. Illustrating this is the first part of the heuristic function of the Turing test.

Now, if Turing relied on his test to assess machine understanding, did he overestimate the capacity of human interrogators to unmask mechanical parrots?

6.2 The Turing Test Suggests a Hypothesis on Machine Learning

Human-like chatbots can be based on a combination of psychological tricks and ad hoc schemes to store and retrieve human-built, semi-structured content pulled from the Internet.[24] From a conceptual point of view, machines of this kind can be understood as sophisticated mechanical parrots. For a related example, Sterrett (2020) described how IBM researchers built the unintelligent Watson system to outstrip humans in the popular *Jeopardy!* game by using Internet-based content and exploiting the a priori known structure of the game (pp. 473–474).

Turing reprobated the use of 'the man inside the machine' stratagems that characterizes the top-ranked machines that competed in practical Turing tests thus far. In (2004 [c. 1951b]) he posited that the machine learning processes that he envisioned 'could probably be hastened by a suitable selection of the experiences to which [the machine] was subjected' (p. 473). 'But here,' Turing warned, 'we have to be careful.' 'It would be quite easy,' he continued, 'to arrange the experiences in such a way that they automatically caused the structure of the machine to build up into a previously intended form.' This, he adverted, 'would obviously be a gross form of cheating, almost on a par with having a man inside the machine.' In other words, Turing ruled out from his test machines that are specially conditioned to pass it, just like IBM Watson was specially conditioned for *Jeopardy!*.

For Turing, of course, a machine 'having a man inside' could never be an existence proof of machine intelligence. On the other hand, mechanical parrots disregarded, he considered that the conversation performance of his imaginary sonnet-writing machine could hardly have been produced unless it had truly learned about British Christmas traditions, characters in Charles Dickens' novel, the use of sarcasm, and so on. For Turing, such a performance would be best explained by assuming a true learning and understanding of the English language and the related culture, just as is assumed in viva voce examinations.

Yet, how many examinations should be enough for an existence proof? Turing said:

> It is clearly possible to produce a machine which would give a very good account of itself for any range of tests, if the machine were made sufficiently elaborate. However, this again would hardly be considered an adequate proof. Such a machine would give itself away by making the same sort of mistake over and over again, and being quite unable to correct itself, or to be corrected by argument from outside. If the machine were able in some way to 'learn by experience' it would be much more impressive. (Turing, 2004 [c. 1951b], p. 473)

This passage could be read as supporting the positive answer to the Turing Test Dilemma: Turing believed that unrestricted tests would eventually unmask elaborate yet unintelligent machines. However, is running repeated unrestricted tests

[24] For a survey on how AI applications can exploit human-built Internet resources, see Hovy et al. (2013).

on unintelligent machines valuable for AI? Shieber (1994) noted that unrestricted Turing tests—precisely for being unrestricted—could not support scientific progress in AI. Therefore, seeing the Turing test as a practical experiment reduces its value to its confirmatory power. However, this pushes the test nearer to the psychometric issues related to the judgment of average human interrogators and farther from AI research.

Now, the interpretation of the Turing test as a thought experiment in the modern scientific tradition presents another reading of the above passage, which observes what Turing *suggested*: even elaborate machines could not qualify as 'an adequate proof' of human-level machine intelligence if they could not learn from experience to correct themselves or be corrected without reboots. In fact, Turing held a specific view of what an existence proof would be (1950, pp. 455–459): to raise a simple learning machine through an adapted process of language and culture education that should be analogous to the one that a human child goes through, until the machine could, without reboots or special coaching, play the imitation game well.[25] The second part of the heuristic function of the Turing test is to suggest that this is possible,[26] as developed next.

Turing's concern was not the design of a practical experiment whose confirmatory power would be robust against *false* positives. It was instead the proposal of an empirical criterion for justifying an existence proof of machine intelligence in the presence of *true* positives. As Shieber observed more recently (2016), the Turing test 'works exceptionally well as a *conceptual* sufficient condition for attributing intelligence to a machine, which was, after all, its original purpose' (p. 95, emphasis added).

Yet, why would the playful imitation game be such an adequate proof of the revolutionary possibility of intelligent machinery? If Michie was correct that the test was meant to assess 'machine understanding,' how can Turing's focus on deception be explained?[27]

First, it is worth recalling the question $Q^{\star}$ that can be generalized from Turing's presentation of his test (Sect. 3): could player A imitate intellectual stereotypes associated with player B's type successfully (well enough to deceive player C), despite the physical differences between A and B's types?

In fact, given that the perception of intelligence involves emotion (Sect. 6.1), deception, or the capability to manipulate the states of mind of another agent, must be addressed as an intrinsic meta-task in any experiment related to $Q^{\star}$. The Turing test, therefore, prepares for related practical experiments addressing deception in AI. As Mach remarked, 'thought experiment often precedes and prepares physical experiments' (1976 [1897], p. 136).

[25] Sterrett (2012) presented what appears to be the most substantial account of Turing's views on 'child machines.'

[26] 'These are possibilities of the near future,' Turing wrote in (1950), 'rather than Utopian dreams' (p. 449).

[27] Turing said in (2004 [1952]), e.g., that the machine 'would be permitted all sorts of tricks so as to appear more man-like' (p. 495), and 'it would have to do quite a bit of acting' (p. 503).

Proudfoot (2011) has urged AI scientists to acknowledge the value of the Turing test as a practical experiment, and her position must face the second horn of the Turing Test Dilemma. However, this article's reconstruction of the Turing test as a thought experiment preserves a deflationary view of her argument, which shows how the Turing test introduced the idea that deception can be and should be explored and controlled for in AI experiments.

Sterrett (2020) contributed an analysis that does justice to Turing's distinction between, on the one hand, the perception of intelligence as grounded in deception in the context of a game and, on other hand, intelligence itself as grounded in learning. Sterrett explained how the Turing test addresses deception through a comparative analysis of popular parlor games. 'The game context,' she remarked, 'provides means to hone in on the part of language performances that have to do with being reflective and resourceful, i.e., not "machine-like" ' (p. 471). The intellectual abilities required by impersonation, Sterrett highlighted by citing a passage in Ryle's *The Concept of Mind* (2000 [1949], p. 33), are perhaps most clearly pronounced in the performance of a clown. Observing Turing's background in espionage, the performance of an intelligence agent may also be considered. Deception can be hard even for a sophisticated mechanical parrot to simulate if not resorting to special coaching by the human programmer 'inside' it.

The distinction between true machine education and special coaching appears in Turing's guidelines on how the machine should be programmed. He addressed that distinction through his heuristic execution of the imitation game. He observed that the imitation of *human fallibility* is necessary for deceiving a human observer. He illustrated human fallibility, first, in the form of incapacity for sonnet-writing, and second, in the form of an arithmetic mistake:

> Q Please write me a sonnet on the subject of the Forth Bridge.
> A Count me out on this one. I never could write poetry.
> Q Add 34957 to 70764
> A (Pause about 30 seconds and then give as answer) 105621.
> (Turing, 1950, p. 434)

Now, the key point here is to note how human fallibility appears in Turing's vision of machine intelligence:

> Another important result of preparing our machine for its part in the imitation game by a process of teaching and learning is that 'human fallibility' is likely to be [mimicked] in a rather natural way, *i.e.*, without special 'coaching'. [...] Processes that are learnt do not produce a hundred per cent. certainty of result; if they did they could not be unlearnt. (Turing, 1950, p. 459, no emphasis added)

In effect, the coherence of the Turing test rests in that the machine's capability to deceive the human interrogator about its true kind must be a corollary of its own learning from experience.

The second part of the heuristic function of the Turing test is to suggest the hypothesis that a learning machine may be created simple and educated naturally, without reboots or special coaching, to play the imitation game well.

7 Conclusion

This article has shown that the Turing test can be best understood as a thought experiment in the modern scientific tradition. First, it has shown that underlying Turing's 1950 presentation of various imitation tests (Sect. 3.1), there is a rich methodological structure (Sect. 3.2), which conforms to what Mach characterized as the basic method of thought experiments, consisting of a continuous variation of experimental conditions (Sect. 4).

Second, this article has presented a reconstruction of Turing's thought experiment that satisfies Popper's conception of the critical and the heuristic uses of imaginary experiments. That reconstruction has emphasized how the Turing test increases understanding of the question 'can machines think?' and prepares for related practical experiments. This provides a rapprochement to the conflicting views on the value of the Turing test for AI and can ultimately put an end to the Turing Test Dilemma as a two-horned issue.

Specifically, this article has shown how Turing's methodic variation of his test design consists of a critical use of the test against the view that physical kind determines logical kind (Sect. 5). The various forms of the test, rather than being a result of imprecision and bad design choices, as suggested in the secondary literature, can be seen instead as concessions to Turing's intellectual opponents. This conforms to Popper's rule for using imaginary experiments in critical argumentation and puts an end to the first horn of the dilemma. Turing's imitation tests addressed the following opposing theories of intelligence presented to Turing:

(1) Human-level intelligence is an exclusive product of the physiology of the animal nervous system, and gendered behavior is a causal product of male and female sex hormones (Jefferson).
(2) A machine can only do what it has been instructed to do (Lovelace–Hartree).
(3) A given art can be performed automatically only to the extent that its rules can be specified, as in the game of chess (Polanyi).

In particular, this article has shown that, seeking conceptual change, Turing used his imitation game to reveal a paradox in the theory of intelligence presented by Jefferson, which tied logical kind to physical kind. This satisfies Kuhn's characterization of the function of thought experiments.

Further, this article has reconstructed Turing's heuristic use of his test (Sect. 6), showing that the test illustrates the emotional nature of the perception of intelligence. This explains why the practical value of the test necessarily depends on the judgment of (average) human interrogators. However, Turing also used his test to suggest the hypothesis that a learning machine may be created simple and educated

naturally, without reboots or special coaching, to play the imitation game well. This explains why running practical Turing tests on machines that have been specially coached to pass it is misguided. The focus of Turing's proposal was to provide both an empirical criterion to justify an existence proof of machine intelligence and a research strategy for fulfilling that criterion. The reconstruction of Turing's heuristic use of his test puts an end to the second horn of the dilemma.

Mach (1976 [1897]) observed that thought experiments based on continuous variation 'undoubtedly have led to enormous changes in our thinking and to an opening up of most important new paths of enquiry' (p. 138). This is the case with the Turing test.

Acknowledgements I thank Fabio Cozman (Universidade de São Paulo) and three anonymous reviewers for their valuable comments on previous versions of the manuscript. I owe Pio Garcia (Universidad Nacional de Córdoba) the early intuition that Ernst Mach's conception of thought experiments could shed light on the Turing test.

Funding This research has been supported by The Sao Paulo Research Foundation (FAPESP) under Grant Number #19/21489-4 'The Future of Artifcial Intelligence: The Logical Structure of Alan Turing's Argument'.

Declarations

Conflict of interest Author declares that he has no conflict of interest or financial ties to disclose.

References

Blum, P. R. (2010). Michael Polanyi: Can the mind be represented by a machine? Documents of the discussion in 1949. *Polanyiana, 19*(1–2), 35–60. Retrieved July 3, 2022, from https://philpapers.org/rec/BLUMPC-2

Brewster, E. T. (1912). *Natural wonders every child should know*. New York: Doubleday, Doran & Co., Inc.

Bringsjord, S., Bello, P., & Ferrucci, D. (2001). Creativity, the Turing Test, and the (better) Lovelace Test. *Minds and Machines, 11*, 3–27. https://doi.org/10.1023/A:1011206622741

Copeland, B. J. (2000). The Turing test. *Minds and Machines, 10*(4), 519–539. https://doi.org/10.1023/A:1011285919106

Copeland, B. J. (2004). *The essential turing: The ideas that gave birth to the computer age*. Oxford: University Press.

Dennett, D. (2006 [1984]). Can machines think? In C. Teuscher (Ed.), *Alan turing: Life and legacy of a great thinker* (pp. 295–316). Dordrecht: Springer. Reprinted from M. G. Shafto (Ed.), *How we know* (pp. 121–145), 1984 (Harper & Row), plus postscripts "Eyes, ears, hands and history" (1985) and (1997, no title).

Dennett, D. (2006 [1997]). Postscript (1997, no title) to "Can machines think?" (1984). In C. Teuscher (Ed.), *Alan turing: Life and legacy of a great thinker* (pp. 314–316). Dordrecht: Springer.

Epstein, R. (1992). The quest for the thinking computer. *AI Magazine, 13*(2), 81–95. https://doi.org/10.1609/aimag.v13i2.993

Floyd, J. (2017). Turing on "common sense": Cambridge resonances. In J. Floyd & A. Bokulich (Eds.), *Philosophical explorations of the legacy of alan turing* (Vol. 324, pp. 103–149). Dordrecht: Springer. https://doi.org/10.1007/978-3-319-53280-6_5

Fuseya, Y., Katsuno, H., Behnia, K., et al. (2021). Nanoscale Turing patterns in a bismuth monolayer. *Nature Physics, 17*, 1031–1036. https://doi.org/10.1038/s41567-021-01288-y

Genova, J. (1994). Turing's sexual guessing game. *Social Epistemology, 8*(4), 13–26. https://doi.org/10.1080/02691729408578758

Gonçalves, B. (2022). Can machines think? The controversy that led to the Turing test. *AI & Society*. https://doi.org/10.1007/s00146-021-01318-6

Hartree, D. (1947). *Calculating machines: Recent and prospective developments and their impact on mathematical physics*. Cambridge: University Press. (Inaugural lecture in October 1946).

Hartree, D. R. (1949). *Calculating instruments and machines*. University of Illinois Press.

Hayes, P., & Ford, K. (1995). Turing test considered harmful. In *Proceedings of the 14th international joint conference on artificial intelligence (IJCAI'95)*, 1995 (pp. 972–977).

Hodges, A. (2012) [1983]. *Alan turing: The enigma*. The Centenary Edition. Princeton: University Press.

Hovy, E., Navigli, R., & Ponzetto, S. P. (2013). Collaboratively built semi-structured content and Artificial Intelligence: The story so far. *Artificial Intelligence, 194*, 2–27. https://doi.org/10.1016/j.artint.2012.10.002

Jefferson, G. (1949). The mind of mechanical man. *British Medical Journal, 1*(4616), 1105–1110. https://doi.org/10.1136/bmj.1.4616.1105

Jones, A. (2004). Five 1951 BBC broadcasts on automatic calculating machines. *IEEE Annals of the History of Computing, 26*(2), 3–15. https://doi.org/10.1109/MAHC.2004.1299654

Koyré, A. (1953). An experiment in measurement. *Proceedings of the American Philosophical Society, 97*(2), 222–237. JSTOR. http://www.jstor.org/stable/3143896

Kuhn, T. (1977 [1964]). A function for thought experiments. In T. Kuhn (Ed.), *The essential tension: selected studies in scientific tradition and change* (pp. 240–265). Chicago: University of Press. Reprinted from A. Koyré (Ed.), *Mélanges Alexandre Koyré, publiés à l'occasion de son soixante-dixième anniversaire: L'aventure de la science* (Hermann, 1964, 2:307–334)

Lavington, S. (2012). *Alan turing and his contemporaries: Building the world's first computers*. British Conservation Society: The Chartered Institute for IT.

Mach, E. (1976 [1897]). On thought experiments. In E. N. Hiebert (Ed.), *Knowledge and error: Sketches on the psychology of enquiry* (pp. 134–147). D. Reidel. Translation of the 5th edition of *Erkenntnis und Irrtum: Skizzen zur Psychologie der Forschung* (Johann Ambrosius Barth, 1905), which included "Über Gedankenexperimente", in: *Zeitschrift für den physikalischen und chemischen Unterricht*, 10: 1–5, 1897. The German version is permanently available at: https://archive.org/details/erkenntnisundirr00machuoft. The English translation has https://doi.org/10.1007/978-94-010-1428-1

Marcus, G., Rossi, F., & Veloso, M. (2016). Beyond the Turing test. *AI Magazine,37*(1), 3–4. Special issue editorial. https://doi.org/10.1609/aimag.v37i1.2650

Mays, W. (1952). Can machines think? *Philosophy, 27*(101), 148–162. https://doi.org/10.1017/S0031819100022660X

Mays, W. (2000). Turing and Polanyi on minds and machines. *Appraisal, 3*(2), 55–62.

Mays, W. (2001). My reply to Turing: fiftieth anniversary. *Journal of the British Society for Phenomenology, 32*(1), 4–23. https://doi.org/10.1080/00071773.2001.11007314

McDermott, D. (2014). What was Alan Turing's imitation game? *The Critique.* (Part of an issue on Turing's imitation game.) Retrieved January 20, 2021, from http://www.thecritique.com/articles/what-was-alan-turings-imitation-game/

Michie, D. (2002). *Transcript of interview. Recollections of early AI in Britain: 1942–1965* (video for the BCS Computer Conservation Society's October 2002 conference on the history of AI in Britain). BCS Computer Conservation Society. Retrieved May 31, 2022, from http://www.aiai.ed.ac.uk/events/ccs2002/CCS-early-british-ai-dmichie.pdf

Moor, J. H. (1976). An analysis of the Turing test. *Philosophical Studies, 30*(4), 249–257. https://doi.org/10.1007/BF00372497

Moor, J. H. (2001). The status and future of the Turing test. *Minds and Machines, 11*(1), 77–93. https://doi.org/10.1023/A:1011218925467

Naylor, R. H. (1974). Galileo and the problem of free fall. *British Journal for the History of Science, 7*(2), 105–134. https://doi.org/10.1017/S0007087400013108
Newman, M. H. A. (1955). Alan Mathison Turing, 1912–1954. *Biographical Memoirs of Fellows of the Royal Society, 1*(November), 252–263. https://doi.org/10.1098/rsbm.1955.0019
Norton, J. (1996). Are thought experiments just what you thought? *Canadian Journal of Philosophy, 26*(3), 333–366. https://doi.org/10.1080/00455091.1996.10717457
Pettit, P. (1991). Realism and response-dependence. *Mind, 100*(4), 587–626. JSTOR. https://www.jstor.org/stable/2255012
Piccinini, G. (2000). Turing's rules for the imitation game. *Minds and Machines, 10*(4), 573–582. https://doi.org/10.1023/A:1011246220923
Polanyi, M. (1974 [1958]). *Personal knowledge: Towards a post-critical philosophy* (2nd ed.). Chicago: University Press.
Popper, K. (2002 [1959]). *The Logic of Scientific Discovery*. London: Routledge. English edition translated and extended from the 1935 German edition *Logik der Forschung: zur Erkenntnistheorie der Modernen Naturwissenschaft*. Dordrecht: Springer.
Proudfoot, D. (2011). Anthropomorphism and AI: Turing's much misunderstood imitation game. *Artificial Intelligence, 175*(5–6), 950–957. https://doi.org/10.1016/j.artint.2011.01.006
Proudfoot, D. (2013). Rethinking Turing's test. *The Journal of Philosophy, 110*(7), 391–411. JSTOR. https://www.jstor.org/stable/43820781
Proudfoot, D. (2017). Turing's concept of intelligence. In B. J. Copeland et al. (Eds.), *The turing guide* (pp. 301–307). Oxford: University Press. https://doi.org/10.1093/oso/9780198747826.003.0038
Ryle, G. (2000 [1949]). *The concept of mind*. Chicago: University Press.
Shah, H., & Warwick, K. (2015). Human or machine? *Communications of the ACM, 58*(4), 8. https://doi.org/10.1145/2740243
Shieber, S. M. (1994). Lessons from a restricted Turing test. *Communications of the ACM, 37*(6), 70–78. https://doi.org/10.1145/175208.175217
Shieber, S. M. (2007). The Turing test as interactive proof. *Nôus, XLI*(4), 686–713. https://doi.org/10.1111/j.1468-0068.2007.00636.x
Shieber, S. M. (2016). Principles for designing an AI competition, or why the Turing test fails as an inducement prize. *AI Magazine, 37*(1), 91–96. https://doi.org/10.1609/aimag.v37i1.2646
Sorensen, R. A. (1992). *Thought experiments*. Oxford: University Press.
Sterrett, S. G. (2000). Turing's two tests for intelligence. *Minds and Machines, 10*, 541–559. https://doi.org/10.1023/A:1011242120015
Sterrett, S. G. (2012). Bringing up Turing's child-machine. In S. B. Cooper, et al. (Eds.), *How the world computes: Proceedings of the turing centenary conference*, Cambridge, UK (pp. 703–713). Dordrecht: Springer. https://doi.org/10.1007/978-3-642-30870-3_71
Sterrett, S. G. (2020). The genius of the 'original imitation game' test. *Minds and Machines, 30*, 469–486. https://doi.org/10.1007/s11023-020-09543-6
Stuart, M. T. (2018). How thought experiments increase understanding. In M. T. Stuart, Y. Fehige, & J. R. Brown (Eds.), *The routledge companion to thought experiments*. London: Routledge. https://doi.org/10.4324/9781315175027
Stuart, M. T., Fehige, Y., & Brown, J. R. (Eds.). (2018). *The routledge companion to thought experiments*. London: Routledge. https://doi.org/10.4324/9781315175027
Traiger, S. (2000). Making the right identification in the Turing test. *Minds and Machines, 10*(4), 561–572. https://doi.org/10.1023/A:1011254505902
Turing, A. M. (1936). On computable numbers, with an application to the Entscheidungs problem. *Proceedings of the London Mathematical Society, s2–42*(1), 230–265. https://doi.org/10.1112/plms/s2-42.1.230
Turing, A. M. (1950). Computing machinery and intelligence. *Mind, LIX*(236), 433–460. https://doi.org/10.1093/mind/LIX.236.433
Turing, A. M. (1952). The chemical basis of morphogenesis. *Philosophical Transactions of the Royal Society B: Biological Sciences, 237*(641), 37–72. https://doi.org/10.1098/rstb.1952.0012
Turing, A. M. (2004 [1947]). Lecture on the automatic computing engine. In B. J. Copeland (Ed.), *The essential turing: the ideas that gave birth to the computer age* (pp. 378–394). Oxford: University Press.
Turing, A. M. (2004 [1948]). Intelligent machinery. In B. J. Copeland (Ed.), *The essential turing: The ideas that gave birth to the computer age* (pp. 410–432). Oxford: University Press.

Turing, A. M. (2004 [1951a]). Can digital computers think? In B. J. Copeland (Ed.), *The essential turing: The ideas that gave birth to the computer age* (pp. 482–486). Oxford: University Press.

Turing, A. M. (2004 [c. 1951b]). Intelligent machinery, a heretical theory. In B. J. Copeland (Ed.), *The essential turing: The ideas that gave birth to the computer age* (pp. 472–475). Oxford: University Press.

Turing, A. M., Braithwaite, R., Jefferson, G., & Newman, M. (2004 [1952]). Can automatic calculating machines be said to think? In B. J. Copeland (Ed.), *The Essential Turing: The Ideas that Gave Birth to the Computer Age* (pp. 494–506). Oxford: University Press.

Vardi, M. Y. (2014). Would Turing have passed the Turing Test? *Communications of the ACM, 57*(9), 5. https://doi.org/10.1145/2643596

Warwick, K., & Shah, H. (2015). Can machines think? A report on Turing test experiments at the Royal Society. *Journal of Experimental and Theoretical Artificial Intelligence*. https://doi.org/10.1080/0952813X.2021.1964003

Warwick, K., & Shah, H. (2016). *Turing's Imitation Game: Conversations with the Unknown*. Cambridge: University Press.

Weizenbaum, J. (1966). ELIZA: A computer program for the study of natural language communication between man and machine. *Communications of the ACM, 9*(1), 36–45. https://doi.org/10.1145/365153.365168