Theoria, 2024, 39(1), 43-65
https://doi.org/10.1387/theoria.24674

Ὁ θεὸς ἀριθμετίζει
THEORIA

ISSN 0495-4548 – eISSN 2171-679X

Thought experiments in the Jefferson-Turing controversy: A Kuhnian perspective

(Experimentos de pensamiento en la controversia Jefferson-Turing: una perspectiva kuhniana)

Pío García*
Universidad Nacional de Córdoba

ABSTRACT: In this article we propose an analysis of the controversy between Geoffrey Jefferson and Alan Turing in terms of a Kuhnian account of thought experiments. In this account, the main task is not to evaluate intuitions or (only) to rearrange concepts. Instead, we propose that the main task is to construct scenarios by proposing relevant experiences in which shared assumptions and conflicting lines of inquiry can be made explicit. From this perspective, we can understand the arguments and assumptions in the Jefferson-Turing thinking machine controversy.

KEYWORDS: thought experiment, Turing; Jefferson, thinking machines.

RESUMEN: *En este artículo proponemos un análisis de la controversia entre Geoffrey Jefferson y Alan Turing en términos de una versión kuhniana de experimentos de pensamiento. En esta versión, la tarea principal no es evaluar intuiciones o (sólo) reordenar conceptos. En su lugar, proponemos que la tarea principal consiste en construir escenarios sugiriendo experiencias relevantes en las que puedan explicitarse supuestos compartidos y líneas de investigación conflictivas. Desde esta perspectiva, podemos entender los argumentos y supuestos de la controversia Jefferson-Turing sobre la máquina pensante.*

PALABRAS CLAVE: experimento de pensamiento, Turing; Jefferson, máquinas que piensan.

* ***Correspondence to:*** Pío García. Universidad Nacional de Córdoba, Centro de Investigaciones María Saleme de Burnichon (CIFFyH). Pabellón Agustín Tosco, 1.^{er} piso, Ciudad Universitaria (5000 Córdoba, Argentina) – piogarcia@ffyh.unc.edu.ar – https://orcid.org/0000-0001-7450-6539

How to cite: García, Pío. (2024). «Thought experiments in the Jefferson-Turing controversy: A Kuhnian perspective»; *Theoria. An International Journal for Theory, History and Foundations of Science,* 39(1), 43-65. (https://doi.org/10.1387/theoria.24674).

Received: 22 March, 2023; Final version: 02 February, 2024.

ISSN 0495-4548 - eISSN 2171-679X / © UPV/EHU Press

1. Introduction

In 1950, Alan Turing published a famous paper exploring the possibility of mechanical thought (Turing, 1950). Many of Turing's interesting and original arguments can be seen as a response to a critical paper published a year earlier by Geoffrey Jefferson, a prominent neurosurgeon (Jefferson, 1949b). Until recently, however, most attention has focused on the imitation game that Turing presented as a way of evaluating (and reframing) the question of thinking machines. If a digital computer can deceive a judge by answering questions as a human would, then the machine can pass the test. For Turing, the imitation game can be a surrogate for the problematic question of thinking machines and provides a suitable means for determining whether or not a machine is able to emulate human intellectual behaviour (Copeland, 2000, p. 530).[1]

The imitation game has been interpreted frequently as a thought experiment. Some philosophers assume, without any discussion, that Turing proposes a thought experiment. For example, Stuart Shieber, in his excellent compilation, commenting on the difficulties for a unified account of Turing's test, said that "[n]o two respondents to Turing's proposals share the same interpretation of the Turing Test. In part, this is what makes the Test such a fascinating thought experiment" (Shieber, 2004, p. 139). Richard Purtill discusses the Turing test with alternative thought experiments, assuming that he must use the same methodology as Turing's (Purtill, 1971). Daniel Dennett also presents the Turing test in terms of thought experiments (Dennett, 2004, 2013). And, of course, we can refer to the philosophical literature around Searle's Chinese Room argument (Searle, 1980, 1984). Recently, Bernardo Gonçalves have argued that the Turing test can be understood as a specific kind of thought experiment (Gonçalves, 2021, 2023b).

In this article we will follow the tradition of interpreting Turing's work in terms of thought experiments. However, instead of focusing on the imitation game, we will revisit the debate between Jefferson and Turing in terms of thought experiments. Taking into account a proposal by Thomas Kuhn, we will offer an alternative understanding of thought experiments in Turing's work. The main implication of this approach is that when we look for instances of thought experiments, the imitation game is not the main target. We do not claim that this was Turing's original intention. But we argue that this approach allows us to identify relevant aspects of Jefferson and Turing controversy over thinking machines. Thus, we will suggest that a Kuhnian account of thought experiments is better suited to the Jefferson-Turing controversy than other proposals. In our account, this controversy can be reconstructed in terms of a contrast between the relevant situations or experiences that each participant considers, and how these situations or experiences can be used to support and argue for alternative positions. Typically, accounts of thought experiments are based on an evaluation of intuitions, concepts or ideas in terms of exceptional cases or exemplars. Although these aspects play a role in the Jefferson-Turing controversy, the contrast between relevant experiences marks the pace of the exchange between the contenders. Our aim is to understand the assumptions, strategies and relevant situations proposed by both sides of the controversy. As we shall see, Jefferson and Turing usually present these aspects in the form of situations and experiences that each side considers relevant. And the task of

[1] Copeland refers to this as the Turing principle (Copeland, 2000).

selecting and making plausible relevant experiences for testing concepts, intuitions or ideas is highlighted in a Kuhnian account of thought experiments.

We will divide our article into four main sections. After a brief general introduction in the first section, we present our account of thought experiments in the second section, focusing on an interpretation of Kuhn's proposal. In the third section, using the conceptual tools presented in the previous section, we will revisit the Jefferson-Turing controversy. Finally, in the fourth section, we present our conclusions.

2. *Thought experiments and Kuhn's proposal*

2.1. Thought experiments accounts

There are several philosophical accounts of thought experiments[2]. From some perspectives, thought experiments have been seen as intuition pumps (Dennett, 2013), tools of the imagination (Stuart, 2022), sorts of paradoxes (Sorensen, 1992) arguments (Norton, 2004a, 2004b), mental models (Miščević, 2007, Nersessian, 2007), scientific models (Mettini, 2020) or closely related to physical experiments (Mach, 1905/1976). However, despite the differences between philosophical approaches, there are epistemic challenges and general features that most of these perspectives share.[3]

The question of whether thought experiments generate new knowledge or merely present or make explicit what is already implicit in the problem statement is usually dubbed as their main epistemic challenge.[4] Among the general features, it is typically suggested that a thought experiment presents an (imagined) scenario (Gendler, 2000, 2004) in which intuitions (Dennett, 2013, Brown, 2004), ideas or concepts (Kuhn, 1977) are assessed. An important aspect of thought experiment, tied with the construction of an imagined scenario, is their speculative character. In any case, a core issue for thought experiments is how these intuitions, ideas, or concepts are evaluated. We can identify this topic as the most relevant for specifying the epistemic challenge that thought experiments face, and where some major differences emerge. For some philosophers thought experiments magnify implicit intuitions (Dennett), test modal consequences (Sorensen), generate (a sort of) *a priori* knowledge (Brown) or can be reduced to a vivid presentation of an argument (Norton). There are other viewpoints that emphasize the epistemic role of thought experiments in terms of how our mental models are constructed, developed and manipulated. (Nersessian, 2007). And, again, the answer that each account offers to the question of how the business of evaluation is conducted is usually linked to this issue of the kind of knowledge that these resources enable or generate. The dispute between Brown and Norton about the *a priori*, empirical or rhetorical nature of the results of thought experiments can be seen in this light.

[2] For a general overview of the problems and bibliography related to thought experiments see Brown & Fehige (2019).

[3] Gendler (2000) suggested that thought experiment perspectives can be divided into those based on examples or prototypes and those with some distinguishing features (privileged feature theories). But even a privileged feature account relies on exemplary cases. We will suggest that this is the case.

[4] Gendler called this issue the central puzzle surrounding scientific thought experiments. That is, how an imaginary scenario can lead to new knowledge (Gendler, 2004, p. 1152).

There is also a tendency to see thought experiments as an 'exceptional case' from which we can learn something. This tendency is most evident in the exemplar cases discussed in the literature. Gendler (2000), for example, structures her book around three different thought experiments. The first one is the famous Galileo's discussion with Aristotelian about falling bodies and weight: "Imagine that a heavy and light body are strapped together and dropped from a significant height. What would the Aristotelian expect to be the natural speed of their combination?" (Gendler, 2000, p. xii).

The second exemplar case is the *Ship of Theseus* and the third one is a thought experiment about personal identity and anti-reductionism. The rationale behind exemplary cases lies in the exceptional nature of thought experiments and, probably, in the assumption that what is mainly evaluated are (unusual) intuitions, ideas or concepts. It is undeniable that most historical cases of thought experiments can be regarded in this way. But in cases where a controversy develops in terms of multiple scenarios in which the 'world' is experienced in alternative ways, a different pattern is needed. We defend that a thought experiment account based on a Kuhnian perspective is well suited for some philosophical controversies, like the one that is illustrated by Jefferson and Turing.

2.2. Following a Machian tradition

The question "What can we learn from a thought experiment?" can be answered in several ways. Two main answers are that we can learn something about the "world", or we can learn something about our concepts or ideas. This is the starting point of Kuhn's analysis of thought experiments. Kuhn considers thought experiments as a potent tool for increasing man's understanding of nature (Kuhn, 1977, p. 240). With this statement, Kuhn is, in some sense, following a Machian tradition[5]. Let us give a brief account of Ernst Mach's proposal for thought experiments. This will help to highlight some of the similarities and differences with Kuhn's account. Most contemporaries' approach to thought experiments are indebted to Mach's proposal (Sorensen, 1992). As is well documented, the term thought-experiment (*Gedankenexperiment*) was coined by Hans Christian Ørsterd (Ierodiakonou & Roux, 2011, p. 4). But it is Mach who made the first conceptual contribution to this notion.

Mach's first presentation of thought experiments was in 1896-97 (Buzzoni, 2018; Mach, 1905/1976, p. XXVllI) and then as a chapter in *Knowledge and Error* (Mach, 1905/1976).[6] For Mach there is a strong link between thought experiment and regular physical experiments. It is a truism to point out that human beings —and animals— learn from experience. And we can sometimes learn by observing a situation. But we can learn even more by finding a way to change a situation, for example by moving our bodies (changing our perspective) or by some kind of intervention. The instinctive propensity of learning by changing an experimented situation is a basic skill that we share with other animals (Mach, 1905/1976 p. 134-135). Mach also supposes that "[w]e can hardly doubt that there is no sharp dividing line between instinctive and thought-guided experiments"

[5] Gonçalves suggests an alternative account of Mach's thought experiments (2023b).

[6] In the *Science of Mechanics*, Mach discusses thought experiments like the one proposed by Stevinus (Mach, 1893/2013, p. 24ss) in terms of instinctive knowledge.

(Mach, 1905/1976 p. 134). There is a strong link between basic cognitive skills, traditional experimentation and thought experiments. So, Mach presents experiments in science in terms of physical ones and those made "on a higher intellectual level" (thought experiments). Thought experiments are used not only by philosophers: the planner, the builder of castles in the air, the novelist, the author of social and technological utopias is experimenting with thoughts (Mach, 1905/1976, p. 136) And also the enquirer and the serious inventor. All of them would use a common scheme of inquiry: "[They] imagine conditions, and connect with them their expectations and surmise of certain consequences: they gain a thought experience" (Mach, 1905/1976, p. 136).

However, for Mach, there is a difference between the "castle builder" and the "serious inventor and researcher":

> [W]hile the former combine in fantasy certain conditions that never occur together in reality, or imagine these conditions accompanied by consequences that are not connected with them, the latter, whose ideas are good representations of the facts, will keep fairly close to reality in their thinking. (Mach, 1905/1976, p. 136)

From this perspective, thought experiments can be described in terms of conditions that are arranged in a way that are followed by consequences. And because the conditions are representations that are linked with facts, the thought experiment is not a fairy tale. At this point appears one of Mach's most controversial hypotheses: "Indeed, it is the more or less non-arbitrary representation of facts in our ideas that makes thought experiments possible. For we can find in memory details that we failed to notice when directly observing the facts" (Mach, 1905/1976, p. 136).

This last part of Mach's defense of thought experiments is related to his Humean empiricist view. As a general account, however, the Machian thought experiment can be seen as a good starting point for our discussion. Taking in consideration the general dimensions presented in the previous section, we can see how a Machian approach addresses them. First, there is a concern with how experimentation (a basic cognitive skill) is related to the variation of situations. We can identify this basic skill as the way in which Mach faces the epistemic challenge of thought experiments: we can learn (to a different degree) when we can make a change or a variation in situations or experiences. We also have general features of a thought experiment in terms of a scheme: imagine conditions associated with expectation and some subsequent consequences. As a thought experiment is generally understood, those imagined conditions represent some sort of speculative conditions. Maybe the core issue related with a Machian account is how to evaluate, in general terms, a good scenario: we have to focus on the conditions for "good representation". And despite Mach's commitment to a Humean empiricism, there is a genuine interest in non-arbitrary representations or conditions.[7]

A Machian account is not only historically interesting. Its insistence on the relevance of (varied) experience can be seen as a precursor of a Kuhnian account. But in order to justify this reconstruction, we must first discuss the reception of the Kuhnian perspective on thought experiments. As we will see, the literature on thought experiments usually sees a

[7] Recently Brecevic (2021) has defended Mach's position on thought experiments against an extreme phenomenological reductionist version.

Kuhnian account as defending a conceptual perspective rather than relying on (varied) experiences.

2.3. A Kuhnian Account

Kuhn's account of thought experiments has been interpreted as a perspective that emphasizes concepts over experiences[8]. For example, Layman sees Kuhn as defending a specific account of models "Kuhn 1964 is best understood as being about the semantic role that thought experiments can play" (Horowitz, 1991, p. 189). Brown interprets Kuhn's account in terms of the conceptual framework of the *Structure of Scientific Revolutions* (Kuhn, 1962) Then according to this view, thought experiments "help us to see the old data in a new way reconceptualized" (Brown, 2011, p. 111). And when it comes to what we can expect from thought experiments, Brown thinks that the main goal of Kuhn's proposal is to learn from our conceptual apparatus: "The big difference between us is this: Kuhn and Gendler think we learn about our conceptual scheme, and only derivatively about the world, while I think we learn about the world, and only secondarily about our conceptual scheme" (Brown, 2011, p. 113).

In other places, Brown emphasizes this "conceptual" perspective in interpreting Kuhn: "Conceptual constructivism was first proposed by Thomas Kuhn (1964) [...] Thought experiments can teach us something new about the world, even though we have no new empirical data, by helping us to re-conceptualize the world in a new way" (Brown & Fehige, 2019). Kuhn's 1964 paper is understood, in general, as defending that the central role of thought experiments is to learn something about scientists' conceptual apparatus.[9] Let us present the reasons why we think Kuhn's proposal can be read in another way.

As noted above, Kuhn's initial strategy is to contrast two main possible functions for thought experiments. It seems that when we engage in a thought experiment, we learn or understand something new. But are we learning or understanding something about the conceptual apparatus of the scientist or about 'nature'? In the first case, the function of a thought experiment is to "assist in the elimination of prior confusion" (Kuhn, 1977, p. 242). The second case is not easy to defend because the knowledge supposed in a thought experiment is apparently not relying, at least in a direct sense, on empirical data. If the main function of thought experiments is to collaborate in the elimination of prior confusion, then this resource allows us to understand the scientist's conceptual apparatus. We do not need additional empirical data. For example, it seems natural to describe the result of Galileo's thought experiment, presented above, in terms of a conceptual contradiction: "The result [of Galileo's thought experiment], of course, is paradox, and that is the way, or one of them, in which Galileo prepared his contemporaries for a change in the concepts employed when discussing, analyzing, or experimenting upon motion" (Kuhn, 1977, p. 251). Even more, some prior information about the world is embodied in a (good)

[8] In presenting Kuhn's proposal, it should be noted that his study focuses on thought experiments in physics. However, he explicitly recognizes that the category of thought experiment could be applied to several fields. We will argue that important aspects of Kuhn's analysis can be applied to philosophical controversies.

[9] There is a short communication by Ana Butkovic that challenges this general interpretation (2007). There is also a more sophisticated Kuhnian account in Moue, Masavetas, & Karayianni (2006).

thought experiment. And that information is not what is under analysis but what is well-known and generally accepted.[10]

In this interpretation, the scientist only recognizes a contradiction. Eliminating confusion and self-contradictory aspects from an account seems to be the main epistemic task of a thought experiment. In both cases, we have to specify verisimilitude conditions. These conditions have a similar function to Mach's "non-arbitrary" representations. Thus, according to Kuhn, the first verisimilitude condition for understanding thought experiments is that: "The imagined situation must be one in which the scientist can apply his concepts in the way he has normally employed them before" (Kuhn, 1977, p. 242). But Kuhn defended that there is more than a logical or a conceptual issue involved in learning from thought experiments. When we analyze historical examples, other features of thought experiments appear: "that description suggests that the effects of thought experimentation, even though it presents no new data, are much closer to those of actual experimentation than has usually been supposed" (Kuhn,1977, p. 242). Again, the Machian heritage is evident in this quote: If a thought experiment can be considered closer to a traditional experiment, then there must be more than a conceptual contradiction involved.[11] It is important to point out that Kuhn said that one result of thought experiment is a reformulation or readjustment of existing concepts. However, it is crucial to address how this readjustment is done. If the problem posed by a thought experiment is only a logical one, or a question of consistency among the scientist's concepts, then even if the result is a "reformulation of an existing concept," the problem is primarily a conceptual one.

Kuhn analyzes similarities between psychological experiments (from Piaget) and examples in the history of physics (Galileo). In both cases, there is not only a conflict between concepts (a sort of contradiction). There is also a dispute among different criteria for applying concepts. For learning something from a thought experiment, identifying a scenario of controversy is the first (fundamental) step.

In this sense, Kuhn defends that the supposed paradox that underlines Galileo's thought experiment is not mainly related with logical aspects of concepts (contradiction, for example) but with their application criteria. This difference points to another aspect that we want to underline in Kuhn's analysis: the relevant experiences where some concepts are successfully applied (a familiar context) and some other experiences where our concepts fail to apply. Or, in other words, scenarios where we can test the application criteria of our concepts. These aspects are important because an interesting (effective) thought experiment must first provide a common ground scenario and then a contrasting scenario for the concepts or perspectives being compared (familiar experiences and new ones). In Kuhn's words:

> If this sort of thought experiment is to be effective, it must allow those who perform or study it to employ concepts in the same ways they have been employed before. Only if that condition is met can the thought experiment confront its audience with unanticipated consequences of their

[10] There is a parallel here with Mach's idea of "instinctive knowledge". It can be defended that even though they are different (Kuhn does not have an empiricist commitment), they serve a similar function.

[11] Gonçalves (2023b) has an interesting analysis of thought experiments in terms of variability. These restrictions can be seen as nonarbitrary representations in Mach's language.

> normal conceptual operations... Nothing about the imagined situation may be entirely unfamiliar or strange. (Kuhn, 1977, p. 252).

As we will see in the next section, the construction or establishment of common ground is crucial to the application of a thought experiment in a philosophical controversy. The issue, then, is how to select the experiences that are relevant. This problem brings us back to the verisimilitude constraint. This aspect is important for two reasons.[12] First, because not all cases require the same kind of verisimilitude restrictions. Kuhn said that Galileo's thought experiment had verisimilitude restrictions (physical restrictions) that Piaget's experiment does not have (Kuhn, 1977, p. 246). Second, if a thought experiment is intended to do more than make some assumption explicit, then there must be a shared scenario of controversy in which experiences are judged plausible (verisimilar) by both sides, even if they disagree about their particular relevance. In other words, there must be a scenario in which selected experiences have the appropriate verisimilitude. For example, if a thought experiment is one where there is a discussion about the concept of "faster" and "speed" in a scientific community, then a physical verisimilitude is usually required. With those tools, Kuhn wants to argue that:

> from thought experiments most people learn about their concepts and the world together. In learning about the concept of speed Galileo's readers also learn something about how bodies move. What happens to them is very similar to what happens to a man, like Lavoisier, who must assimilate the result of a new unexpected experimental discovery. (Kuhn, 1977, p. 253)

Then, the result of a thought experiment is more than resolving a confusion or a contradiction, because self-contradictory exemplars cannot be exemplified by any possible world. Or, in the vocabulary that we are discussing, we cannot find, in those exemplars, experiences where our familiar concepts can be applied. If there is a sense of confusion that can be used in empirical thought experiments, it must be one where there are some experiences that are familiar. Experiences that naturally fit with our previous concepts. And there must be some experiences that challenge those concepts. Pointing out, in abstract, that there are some contexts where our empirical concepts do not apply seems to be trivial. If we are working with empirical statements, and not tautologies, then there must be conditions where those concepts should fail to be applied. The crucial question is how to challenge a previous concept by confronting it not only with a new concept, but also with a novel scenario in which there are experiences that are judged to be relevant. These experiences are in conformity with a 'world' in which our ideas or concepts are successfully in use: "we cannot, I think, find any intrinsic defect in the concept by itself. Its defects lay not in its logical consistency but in its failure to fit the full fine structure of the world to which it was expected to apply" (Kuhn, 1977, p. 258). The reference to "the full fine structure of the world" is a way of expressing the dual nature of a scenario in which there are some shared experiences and some experiences that resist the conceptual application of one of the contenders. Moving to the context of history of science, Kuhn underlines an issue that has been discussed extensively by the

[12] Although Kuhn presents this problem, he does not discuss it. But the issue of how to choose verisimilitude constraints is central because our intention is to outline a general account of thought experiments.

philosophy of models: “It follows that those concepts were not intended for application to any possible world, but only to the world as the scientist saw it” (Kuhn, 1977, p. 260).

The confrontation that characterizes a thought experiment is in part motivated by a conceptual change, but also is triggered by nature or the world: “[n]ature rather than logic alone was responsible for the apparent confusion” (Kuhn, 1977, p. 261). Kuhn concludes the article by suggesting a problem-solving approach to theory change. And how a problem-solving activity narrows down the relevant issues and puts anomalies on the periphery. The image of what is in focus and what is in the periphery is the best way to underline why the activity of selecting relevant experiences constitutes a creative intellectual task. It also sheds light on how the relevant concepts are applied. Both tasks, selecting relevant experiences and the application of concepts, explain the conceptual and empirical side of a thought experiment. But, to understand some of their functions, we must not separate the conceptual and the empirical aspects.

At this point, we have some tools to analyze the Jefferson-Turing controversy. To understand the assumptions, strategy and relevant situations proposed from both sides of the controversy, we will use some key aspects of Kuhn’s thought experiment account. Going back to the general dimensions that we present above for thought experiments we can see general features from a Kuhnian account. We need, first, a common ground. This aspect is the background of the scenario of controversy, which is composed of concepts but mainly of experiences that Jefferson and Turing evaluate as relevant. Taking into account the epistemic challenge, a Kuhnian account is committed to a kind of empirical side of thought experiments because of the task of selecting relevant experiences in contrasting scenarios. The concepts and experiences of a scenario could be in conflict, but are considered admissible by both sides. As we will observe, Jefferson and Turing share, for example, a common ground about how to study, in very general terms, the problem of thinking machines. However, there are significant differences in how to study the problem in particular. Again, these differences are broadly seen as appropriate by both sides, but the judgments of their relevance are contrasting.

In our analysis, how to choose relevant experiences is the key aspect of a Kuhnian thought experiment. Assumptions or restrictions from verisimilitude evaluations could explain the selection of each contender. When we examine the controversy from this point of view, several controversial scenarios emerge. Some of them are general, and others are more specific. However, each scenario presents significant experiences that create worlds in which problems and concepts can show their value and impact. In other words, as we understand the controversy between Jefferson and Turing, it is more about the task of selecting and making plausible some piece of the “world they both share” (relevant experiences) than it is about the task of introducing or making plausible a concept. Of course, the result, as Kuhn observed, is ultimately how a concept might or might not apply to a scenario. But this result is a consequence of providing and defending appropriate experiences.

3. *Controversy scenarios about thinking machines: Jefferson and Turing*

“Computing Machinery and Intelligence” by Alan Turing was published in 1950 in the *Mind* journal. It is considered a starting point in the contemporary discussion about thinking machines (Epstein, Roberts, & Beber, 2009). It is also regarded as a milestone of Turing's previous work (Copeland, 2004, p. 353ss) or as a crystallization of a controversy with several contemporaries (Gonçalves, 2023a). In this article, however, we want to focus on a more specific issue: the responses that Turing presents to "The Mind of a Mechanical Man", a conference that Jefferson gave to the Royal College of Surgeons of England in June 1949. There are several reasons why Jefferson's paper is important for understanding some of Turing's ideas. First, through Jefferson's paper, Turing's attention is drawn to Descartes's ideas about thinking machines (Abramson, 2011). And Jefferson not only quotes Descartes' work but also gives an interpretation of key Cartesian concepts like the notion of "soul". Second, as we will see, there are several arguments in Turing's paper that are illustrated or motivated by Jefferson's objections.[13]

The interest of Jefferson in Descartes' work is far from being anecdotal. In 1949, Jefferson published another paper, this time in the *Irish Journal of Medical Science*, entitled "Rene Descartes on the Localization of the Soul". Even though this paper was published in September 1949, it was delivered as a Lecture in May of the same year, a month before the paper that was read later by Turing.[14] This timeline is important because in the paper about Descartes, Jefferson analyzes the concept of soul and he concludes that even when the echoes of the term are religious or metaphysical, there is an "entirely psychological sense, as the psyche without any religious implications. This usage has survived to our own day" (Jefferson, 1949a, p. 697). And this psychological use plays a major role in Jefferson's later work. We will return to this issue.

Turing's paper contains nine objections, and corresponding answers, to the possibility of thinking machines. What are the objections that can be traced back to Jefferson?[15] In answering this question, we will begin with the tools we discussed in the previous section on a Kuhnian account of thought experiments. First, we need to characterize the scenario of the controversy that allows us to identify the relevant experiences with restrictions of verisimilitude that ultimately structure thought experiments. If we seek the explicit Jefferson's quotes in Turing's paper, we will find in the fourth objection the main reference: the argument from Consciousness:

> This argument is very well expressed in Professor Jefferson's Lister Oration for 1949, from which I quote. Not until a machine can write a sonnet or compose a concerto because of thoughts and emotions felt, and not by the chance fall of symbols, could we agree that machine equals brain, that is, not only write it but know that it had written it. No mechanism could feel (and not merely artificially signal, an easy contrivance) pleasure at its successes, grief when its valves fuse, be warmed by flattery, be made miserable by its mistakes, be charmed by sex, be angry or depressed when it cannot get what it wants. (Turing, 1950, p. 445-446)

And later, in the fifth objection, Turing refers to Jefferson paper again (p. 450), but only to recall the fourth objection. It seems that Turing appeals to Jefferson only to discuss the

[13] In this article we will not consider other contemporaries who may have influenced Turing's work. A broader context for Turing's work is discussed in Copeland & Proudfoot (2009) and Gonçalves (2023a).

[14] Jefferson's paper was published on June 25, and the conference was delivered on June 9.

[15] There is at least one other important source, a discussion between Turing, Newman, R. B. Braithwaite & G. Jefferson, recorded by the BBC on January 10, 1952 (Copeland, 2004, p. 487ss).

Consciousness objection, as is explicit in the final statements of the fifth objection (Arguments from Various Disabilities). According to this reading, if we want to compare Turing responses to Jefferson's position, we have to restrict ourselves to the fourth objection.

Even when consciousness is a key point in the comparison between Turing and Jefferson, we defend that there are other places to look for. As we said, we are going to develop this idea with the structure of a thought experiment that we presented in the previous section. Following the Kuhnian structure, we need first a common general ground where the controversy takes place. This is what we will call the first scenario of the controversy. Then, once a common ground is established, contestants propose more experiences that are considered relevant, and new scenarios emerge. With this approach, we hope to make explicit the problems and assumptions of each context, and we also expect to evaluate their particular verisimilitude. We will start by addressing the issue of the scope of thinking machines and which machines are considered meaningful. The first issue can be divided into two parts: the epistemological status of the inquiry and the appropriate science for the research. To explore this, we will first examine a common scenario and then a contrasting one, which includes relevant conflict experience. Furthermore, the speculative nature of the initial (common) scenario provides a compelling reason to approach this controversy through thought experiments. Then we will develop additional scenarios that present contrasting experiences.

3.1. The scope of the question about thinking machines

The first problem, the epistemological status of the inquiry, is presented by Jefferson quoting another scientist, Hughlings Jackson, who defends the invention of the hypothesis as part of a usual research task. This is a way to escape from the restricted method of following only "facts" or, as is described by Jackson, from "Baconian inductions" (Jefferson, 1949b, p. 1105). Jefferson saw this perspective as "to proceed in the hope that, although we shall not arrive at certainty, we may discover some illumination on the way". Jefferson expresses a similar opinion in 1948:

> Whatever means science theoretically should use, the scientist is a man more imaginative than Bacon would allow... He is a good deal more rational, more emotional, in a word more human, than argument can hold him to be. Hence his scepticism must be wilful. (As cited in Jefferson, 1984, p. 3)

Turing certainly read Jefferson's ideas as a defense of the speculative stance where he discusses the hypothesis of thinking machines. As is stated by Turing, a world where computers could play the imitation game well was (at least) 50 years away.[16] The point is why ask the question now when the answer belongs to the future: "Conjectures are of great importance since they suggest useful lines of research." (Turing, 1950, p. 442). Thus, for the two scientists, a mathematician and a neurosurgeon, who in principle felt compelled to limit themselves to facts, this epistemological reflection on hypothesis and conjectures legitimizes a common speculative arena, a scenario where the controversy can be settled, to dis-

[16] A few years later, in response to a question from Newman, Turing said that a computer would beat the imitation game in 100 years (Copeland, 2004, p. 495).

cuss the possibility of mechanical thinking. Again, the conjectural space in which most of Turing's ideas are discussed favors an interpretation of the 1950 paper, not only of the imitation game, in the sense of a thought experiment, as a setting to test possible answers and to try out new ways of investigation. This is a decisive step to understand how much of Turing's perspective is indebted to Jefferson paper. And it is also important because this context constitutes a first scenario of controversy. A scenario in which a speculative pursuit is acceptable. From this general point of departure, Jefferson and Turing suggest other scenarios where more specific compromises and commitments are presented.

Related to the second problem that we mention above, the science appropriate for the research, in the first line of his paper Jefferson starts, following the title of the conference, by pointing out the apparent difference between the brain and the mind. The contrast between properties of the brain (finite) and properties of the mind (amorphous and elusive) seems to be more an approximation to our knowledge than a proper description of the subject. But it is stressed here that something needs to be investigated: the issue about how to study the problem of the relationship between the brain and the mind: "It is a subject which at present awakes a renewed interest, because we are invaded by the physicists and mathematicians —an invasion by no means unwelcome, bringing as it does new suggestions for analogy and comparison" (Jefferson, 1949b, p. 1105). Besides the connotation of the "invasion" metaphor, Jefferson suggests that the study of brain-mind phenomena can be approached by different tools and scientific resources. Is it mainly a *physical*, a *mathematical*, or a *biological* problem? Jefferson believed that life sciences, especially medicine, have the main responsibility for answering the problem of the relationship between the brain and the mind.

Not only is medicine cited by Jefferson as a specific area where to discuss the mind-brain issue, but he also warns us about the analogies of pure science that go beyond the limits allowed by practical research. This is why medicine (and not just biology or the life sciences in general) is so important to Jefferson. Physics and mathematics can bring useful analogies, but we have to be careful not to go beyond their limits. They can only provide some clues in the epistemological realm or as a heuristic tool. The history of medicine, said Jefferson, is plenty with examples of analogies that went too far. In a nutshell, we can get hints from mathematics or physics, but a proper answer must be given by medicine or applied biology.

When Turing proposed digital computers as a solution to the thinking machine problem, he implicitly challenged the claim that the proper answer must be found in medicine. So, having agreed on how to discuss the problem of thinking machines, we now have a contrasting scenario when trying to answer the crucial question of which discipline is more appropriate. On a Kuhnian account, the important question here is how each contender makes his contrasting experiences convincing.

In order to make this proposal plausible, Turing combined two of the problems that we regard as problems of scope. Namely, the scope of the question of thinking machines, or the issue of what science is to be considered and what kind of machine is relevant. With this move, Turing is not only settling the question about "where" we have to discuss the question (scenario of controversy) but also start to challenge some of the main suppositions defended by Jefferson (restrictions of verisimilitude). And by exploring the differences between both scientists we can discover the assumptions that allow a development of the discussion about mechanical thinking, as we suggest at the end of the previous section.

In summary, until now, the justification of a space of conjecture is the first common commitment that allows a shared arena of discussion. And the discussion about how to investigate the problem settles the contrast between, in principle, a plausible path of research (computer science and medicine) that is evaluated very differently by both scientists. If we consider the first scenario where both contenders can propose conjectures, this is a second scenario where the differences started to emerge. In this second space of controversy, there is a quarrel about which experiences are more relevant to understand a human mind: the domain of computers or the field of organisms.

3.2. Kinds of machines

As we noted above, the main claim of Turing's 1950s paper is that computers could successfully play the imitation game (a replacement for the question about thinking machines). Jefferson is clearly against the possibility of thinking machines. But what kind of machines computers are is the important question for Turing. He was fully aware, at that time, that the plasticity of computers and the properties of some of these machines, like their universality property, challenge the traditional limits of mechanism. Classic automata and modern computers are very different kinds of machines. So a question for those assumptions (restrictions in verisimilitude) is in order. Jefferson knows also that automata and computers are different. However, the issue is that they do not agree about where to find the difference.

Jefferson's assessment regarding which machines we have to consider is not direct. It combines, at least, an argument from a Cartesian tradition and an argument from life sciences in a historical context. The last argument occurs in the classical contrast between mechanistic and other kinds of explanations (mainly chemical and biological). In the final sentences of his paper, Jefferson is clear about his general position about mechanism: "I end by ranging myself with the humanist Shakespeare rather than the mechanists" (Jefferson, 1949b, p. 1110).

Jefferson compares the ancient automata that imitate life forms to the modern one. If the ancient automata were constructed mainly for entertainment but nevertheless were impressive, more can be expected by modern automata that were made by serious scientists that want to make a "cunning replica of a living thing" (Jefferson, 1949b, p. 1106). Jefferson also concentrates on Descartes' account of automata as an imitation of life forms. Jefferson summarizes two main arguments from part V of the *Discourse on the Method*: the limitation that a machine could have in managing a language and the link between a particular mechanism and particular actions.[17] Jefferson finally quotes Descartes and concludes that a machine cannot imitate a human being because "it has no mind". So, there has to be, following Descartes' arguments, a link between the diversity of behaviors and the mind. At the same time, Jefferson wants to challenge some other Cartesian suppositions. Besides some superficial analogy, the body is not a "sum of mechanism". In a living organism there are hidden "all kinds of biochemical ingenuities" (Jefferson, 1949b, p. 1106).[18] And the same can be said for the brain and the mind. Even if we can say that there

[17] Here Jefferson refers to Capek's robots to draw an analogy between old automata and modern devices.

[18] Jefferson points out the historical importance of mechanism in life sciences: it contributes to dismissing some mystery. But analogy is not identity.

are “some nervous mechanism in isolation”, they are also “so complicated ... by endocrines, so coloured is thought by emotion” (Jefferson, 1949b, p. 1106) that they cannot be considered just a mechanism. Chemical and biological “ingenuities” cannot be reduced to a mechanical arrangement.[19]

Then, against Cartesian suppositions, Jefferson defends that an organism is not a machine. But Jefferson wants, at the same time, to defend that the behaviour and even what can be called some sort of conscious mental process of animals have a “variety of behaviour that confuses us”. The reason why Jefferson is so interested in arguing for an “animal mind” is that he agrees with Descartes that the variability of behavior is the key against the mechanization of the mind. But he also disagrees with Descartes about the scope (life is not mechanical) and the plausible cause of this plastic behavior. “It seems to me likely that the number of synapses in a nervous system is the key to the possible variations in its behaviour” (Jefferson, 1949b, p. 1107).

This second scenario, in which some empirical findings about computers and organisms compete to develop some features of a mind, is linked with a classical Cartesian scenario about thinking machines. Following the analysis of Abramson (2011) and Laudan (1981), we can see that there are other aspects of the Cartesian scenario that are important for understanding the Jefferson-Turing controversy.

3.3. A cartesian scenario: Causes, behaviour and inferences

Jefferson and Turing are engaged in a quest for an appropriate cause of a mind. From this point of view, is the same scenario as Descartes’s discussion about minds and machines. However, Descartes presents at least two very different scenarios of controversy regarding the possibility of mechanical minds. The link that Descartes build between a mechanism and their resulting behaviour [20]constitutes the first Cartesian scenario. In part V of the *Dis¬course on the Method* (1637/2006), Descartes tells us that a mechanical behavior —such as that generated by an animal— can be fully explained by the arrangement of its organs. Here arrangement of the organs refers to some mechanism or particular organization of parts that could explain the generation of movement. Thus, Descartes, after explaining the mechanism of the heart and blood circulation, says:

> Finally, so that those who do not know the force of mathematical proof and are not used to distinguish true reasoning from plausible reasoning, should not venture to deny all this without examining it, I would like to point out to them that the movement I have just explained follows necessarily from the mere disposition of organs that one can see with the naked eye in the heart, from the heat which one can feel there with one’s fingers, and from the nature of blood which one can know from observation, in the same way as the movement of a clock follows from the force, position, and shape of its counterweights and wheels. (Descartes, 1637/2006, p. 41-42)

[19] Jefferson is not alone in this. There is an important anti-mechanistic and anti-reductionist tradition in the life sciences in England (see Allen, 2005, Lenoir 1989). For a critical account of the methodological side of this research program, see Roll-Hansen (1984).

[20] We are not assuming a behaviorist interpretation with this Cartesian scenario. It is just a place where arguments are articulated. In this sense, our understanding of this scenario is compatible with the interpretation of the Turing test presented by Proudfoot (2013).

With knowledge of the cause or the structure of a mechanism, we can deduce their behaviour. As Descartes said, it is like a mathematical proof. This first Cartesian scenario in which we can deduce behaviours from causes is not the context of controversy where Jefferson and Turing discuss their ideas. But following the analysis of Abramson (2011) and Laudan (1981), we can say that Descartes also presents a second Cartesian scenario. When Descartes suggests his arguments from variability and language, it is in the situation where from a behaviour we have to infer the appropriate cause. And this other scenario is one where a plausible inference is made from behaviour to causes. Here, a clarification must be made. Against his own advice, Descartes has to allow plausible reasoning in the context where an unknown cause has to be inferred from a behaviour. In the *Principles of Philosophy*, Descartes faces the problem of how to infer a mechanical cause (imperceptible corpuscles) from the perceived world. And at this point, Descartes is fully aware that only a conjecture can be made:

> For just as the same artisan can make two clocks which indicate the hours equally well and are exactly similar externally, but are internally composed of an entirely dissimilar combination of small wheels: so there is no doubt that the greatest Artificer of things could have made all those things which we see in many diverse ways. (Descartes, 1644/1983, p. 286)

This context constitutes another scenario inspired by the *Discourse on the Method's* arguments. It is no more a scenario where we have to deduce a behavior from a known cause (mechanism). Now, in this new scenario, we have two Cartesian machines, a human body and an automaton, and the task is to infer the plausible cause from an interesting (variable) behaviour. There is not, at least in principle, a dualistic commitment. There is a quest where plausible reasoning is the only game in town. And this is the scenario of controversy that Jefferson cites in their 1949 paper.

This is a third scenario where the investigation of the possibility of thinking machines is pursued in terms of a non-dualistic commitment where plausible reasoning is the tool to be used. And in this scenario, Jefferson could advocate the study of organisms (not computers) from the point of view of applied biology (not mathematics).

3.4. A materialistic (or naturalistic) realm

We can identify another complementary scenario from materialistic assumptions in Jefferson and Turing. However, considering the historical context, Jefferson's materialistic commitments can be in question. On 25 June 1949, the editor of the *British Medical Journal* made a few comments on Jefferson's paper. For issues like the body-mind problem, we need "scientists and mathematicians to become their own philosophers" (BMJ, 1949, p. 1129). In particular, scientists-philosophers are needed to evaluate affirmations about a "mechanical brain". A Turing's interview with the Times in 1949 is cited as an example of a mathematician that works in an electrical brain. But, the main problem for the Journal's editor is materialism: "There is an undeniable danger in the facile acceptance of materialism, for the materialist finds values and ethics an insoluble problem" (BMJ, 1949, p. 1130). Then, it seems that Jefferson's engagements are non-materialistic. As we will see in this section, there is evidence against this simplistic interpretation. The editor cites at the end of their comment Whitehead's *Science and Modern World* to defend an anti-rationalistic version of science[21]. Jefferson seems to be more committed to an anti-reductionistic account than to an anti-materialistic position. And an anti-reductionistic account is more directly relevant to the ethical problem motivated by the Journal's editor. The main target in this critique is an interpretation of mechanism in terms of a rationalistic and reductionist account of science. But Jefferson and Turing were both committed to finding an answer to the problem of thinking machines in the sciences. The disagreement begins, as we have seen, when they specify which science bears the burden of the answer.

For Jefferson, the corresponding increase of complexity and variability in behaviour, as opposed to mechanical determinism, can explain several (human) phenomena, including what we call free will. Even when the argument is Cartesian, the causes involved are not more non-physical substances (*res cogitans*). In this sense, the common arena of controversy with Turing is in a materialistic realm (or at least a naturalistic one). Jefferson has an argument from the complexity of the nervous system and the convergence of non-mechanical causes.[22] Here it is important to recall Jefferson's interpretation of the Cartesian soul. He can use Cartesian arguments with a materialistic (but not mechanical) commitment because he is discussing a (Cartesian) psychological soul. As we have seen, Jefferson can justify this conceptual move.

Jefferson not only considers automata that were paradigmatic for Descartes. For Jefferson, the imitation of life and mind is related to "modern automata" and "calculating machines" respectively. And besides the analogy that emerges from the nervous impulse and the electric machines, biological and chemical considerations indicate to us that there is more than an electrical circuit involved in the nervous system. Part of the argument rests on a relative anti-reductionism about scientific disciplines and kinds of explanation.

Jefferson defends anti-reductionism in several steps. First, it is suggested that the plasticity of human behavior (mind) may be the result not of some mechanical fragments but of a "whole integrated nervous system of man" (Jefferson, 1949b, p. 1108)[23] Second, there is a gap in our understanding of how high-level intellectual activities are generated. We do not yet know whether the way in which high intellectual activity is generated differs from the way in which it is generated at lower levels. We are unaware of "the final process of brain activity that results in what we call, for convenience, mind" (Jefferson, 1949b, p. 1108). This argument from what is unknown is considered for Jefferson against a mechanical mind. But to make it plausible, it must be linked with suppositions from the first argument: only fragments of the nervous system can be described in terms of a mechanical

[21] At the time, it was very common among biologists to appeal to Whitehead as a champion against reductionism. For example, in a book that was very influential for the anti-reductionist philosophy of biology in the early 20th century, Woodger's *Biological Principles* (1929), Whitehead was by far the most cited philosopher. The problematic account for these biologists was a strong and narrow mechanistic perspective on the organism. The appeal to Whitehead was then a common anti-reductionist move among biologists such as the editor of the *British Medical Journal.*

[22] Jefferson reminds us that this problem can even be presented in physics. Niels Bohr and the dual nature of the electron is the example to illustrate this idea.

[23] The most common anti-reductionist argument in the life sciences in the first half of the twentieth century in England was constructed mainly from the whole-part distinction. For example, Frederick Hopkins, a major figure in the establishment of biochemistry as a science, uses the idea of a dynamic "whole" in several places (Weatherall & Kamminga, 1996).

cause. So if a mechanical description is inadequate at the lower level, we cannot expect it to be transformed into an adequate one at the higher level. Third, to undermine the analogy between the nervous system and electronic computers, Jefferson refers to the case of injury to the human brain:

> Damage to large parts of the human brain, entailing vast cell losses, can occur without serious loss of memory, and that is not true of calculating machines so far, though so large a one might be imagined that parts of it might be rendered inoperative without total loss of function. (Jefferson, 1949b, p. 1109)

We now have a fourth scenario in which a new shared ground is established: a materialistic (or at least naturalistic) realm. The speculative inquiry defended by both scientists is a common commitment to a non-dualistic explanation. The nature of the experiences to which Turing and Jefferson refer (characteristics of computers and organisms) is another clue to the plausibility of this fourth scenario. A fifth scenario is suggested when a special feature is proposed that organisms have but computers do not: variability.

3.5. A special feature: variability

The plasticity of behavior can be attributed to the plasticity of the cause.[24] Then the first challenge for Turing is how to explain variability from a (simple) mechanical cause. Most likely, Turing was a materialist about the mind (Hodges, 2019). In a common ground of a materialistic realm (fourth scenario), the discussion focuses on the appropriate cause that brings about the kind of variability that is present in life and especially in the mind. This could be the main reason why the first objection considered by Turing, the one based on a non-material soul, is interpreted in a strongly ironic way. A non-material cause is excluded and not (seriously) discussed. However, the link that Jefferson made between variability of behavior and an anti-mechanistic (but material) cause has to be addressed by Turing. The answer that Turing proposed involved an important change in the level of abstraction at which the problem must be presented. As we have seen, Jefferson was convinced that in order to study the mind, we must study its material cause (some kind of living thing). And Turing thought that the problem of thinking machines could be solved at the level of computers. This level of abstraction involved a shift in the problem: instead of focusing on studying the material cause (nervous system), we can study the behavior generated for computers. We can see that some important aspects of this controversy can be understood in terms of what experiences or contexts are relevant. Here is one of the places where the previous scenarios are important. Jefferson built a scenario where he links anti-mechanical causes and variability. And this scenario prevents the alternative of understanding the disagreement as two complementary ways of investigating the possibility of thinking machines. There are contrasting scenarios where the appropriate cause of a plastic behaviour is different. We have seen how Jefferson presents the relevant experiences for his scenario. Now we must review the particular experiences that Turing empathizes, in order to build his scenario.

[24] Later, Jefferson refers to Wiener's suggestion of "machine disease". But Jefferson thinks that the deterministic, localized, and simple nature of the machine works against the usefulness of the analogy.

But first, Turing has to evaluate whether the limits of automata are the same as those of computers. This is how Turing could make a contrast with Jefferson's scenario. The property of universality of some computers is an essential part of Turing's answer. Also important is the plasticity of computers, linked to the property of programmability. And, the final answer is learning machines. It is important to notice that even when Turing supposes an important degree of abstraction when he suggests using computers to study mechanical thinking, at some point there is a significant relevance of practical problems, for example, when we are engaged in programming a computer to imitate a learning child.

For Turing, there is also the problem of how Jefferson connects computers to human brains. At this level of abstraction, the analogy to the nervous system associated with electricity is irrelevant:

> The fact that Babbage's Analytical Engine was to be entirely mechanical will help us to rid ourselves of a superstition. Importance is often attached to the fact that modem digital computers are electrical, and that the nervous system also is electrical. Since Babbage's machine was not electrical, and since all digital computers are in a sense equivalent, we see that this use of electricity cannot be of theoretical importance...The feature of using electricity is thus seen to be only a very superficial similarity. (Turing, 1950, p. 439)

There are also two other differences that Jefferson suggests between minds and machines. The machine can only answer questions "prearranged by its operator". Here Jefferson considers the objection that we do not have a blueprint for a human being. As we said above, Jefferson's discussion is presented in a materialistic arena, so his answer cannot rest in a dualistic position. The proposal is that, unlike electric machines, human beings "build" their minds by education and experience data.

Finally, Jefferson thinks that high intellectual process is related to language. Language is not static, as history of science has shown, and depends on conceptual thinking: "It is not enough, therefore, to build a machine that could use words (if that were possible), it would have to be able to create concepts and to find for itself suitable words in which to express additions to knowledge that it brought about" (Jefferson, 1949b, p. 1110).

The Cartesian aspect of the argument is undeniable. But, to make clear where the link between language and conceptual thinking came from, Jefferson remarks on the main idea of the paper. Computers can be fast, but:

> The great difference in favor of the calculating machine as compared with the crane, and I willingly allow it, is that the means employed are basically so similar to some single nervous layouts. As I have said, the schism arises over the use of words and lies above all in the machines' lack of opinions, of creative thinking in verbal concepts. (Jefferson, 1949b, p. 1110)

And, returning to the discussion in the previous sections, here is where a discipline is needed that does not omit complexity and details through idealization: "I am quite sure that the extreme variety, flexibility, and complexity of nervous mechanisms are greatly underestimated by the physicists, who naturally omit everything unfavorable to a point of view" (Jefferson, 1949b, p. 1110).

For Turing, the argument of variability and the challenge to computers from some particular mechanistic assumptions are important. Turing starts to answer those challenges in the Lady Lovelace objection section. First, as we mentioned, there is an in principle reply: universal machines. Then the question is about how to define novelty in psychological terms. And finally, the main answer is learning machines. Part of Jefferson's objection is related to variability. And here a mechanistic cause interpreted as deterministic is a major conceptual problem. Turing saw the difficulty, and the possibility of computer mistakes and the elaboration of two kinds of errors is crucial: "Errors of functioning are due to some mechanical or electrical fault which causes the machine to behave otherwise than it was designed to do [...] Errors of conclusion can only arise when some meaning is attached to the output signals from the machine" (Turing, 1950, p. 454).

Of course, a computer can be programmed to make mistakes that are trivial (a perverse example). The interest case is where a computer models an inductive inference. "To take a less perverse example, it might have some method for drawing conclusions by scientific induction. We must expect such a method to lead occasionally to erroneous results" (Turing, 1950, p. 454).

In this way, an anti-mechanical objection built into a naive deterministic assumption is undermined. At some level of analysis, a computer is deterministic, but the main issue is how to interpret a computer's output and what is being modeled. The importance of this problem for Turing cannot be overstated. A general deterministic argument is not enough to exclude the variability of a computer. And for Turing, intelligence was associated with error or some non-deterministic factor. In his 1950s paper Turing remarks: "Intelligent behaviour presumably consists in a departure from the completely disciplined behaviour involved in computation" (Turing,1950, p. 459). And in Intelligent "Machinery, A Heretical Theory", Turing said:

> I believe that ... [the] danger of the mathematician making mistakes is an unavoidable corollary of his power of sometimes hitting upon an entirely new method. This seems to be conformed by the well known fact that the most reliable people will not usually hit upon really new methods... My contention is that machines can be constructed which will simulate the behaviour of the human mind very closely. They will make mistakes at times, and at times they may make new and very interesting statements. (Copeland, 2004, p. 472)

Copeland even suggests a link between some types of error and a heuristic search for problem solving (Copeland, 2004, pp. 469-470). This is another argumentative move in which the assessment of verisimilitude (plausible cause of variability), which depends on some insightful understanding of technical resources (modeling inductive inference), allows the development of an alternative scenario.

4. *Final words*

Our main goal in this article is to understand the assumptions and identify the relevant aspects in the Jefferson-Turing controversy about thinking machines. Following a traditional analysis of Turing's 1950 paper, we propose to use an account of thought experiments to accomplish this task. However, instead of focusing on the imitation game, we posit a broader context in which Turing's 1950 paper and 1949 work by Jefferson can be considered. In order to address different aspects of the controversy, a particular account of the thought experiment is required. Most accounts assume that thought experiments are exceptional cases, special exemplars. But it seems that the way Jefferson and Turing formulated their positions has more to do with situations that are deemed plausible and relevant. We argue that a Kuhnian account of thought experiments is well suited to cases like the Jefferson-Turing controversy.

In this sense we propose an alternative interpretation of Kuhn's thought experiments, because his account is usually understood in the philosophical literature in terms of a conceptual rearrangement. In our interpretation of Kuhn's thought experiments, the main task is not primarily the evaluation of intuitions or (only) conceptual rearrangement. Instead, we contend that the main task is to construct scenarios by providing relevant experiences in which shared assumptions and conflicting lines of inquiry can be made explicit. We defend this interpretation as closer to Kuhn's original proposal. In this way, the main task in constructing thought experiments is to identify the relevant experiences that bring common and conflicting scenarios to life.

We have presented several scenarios in which the Jefferson-Turing controversy can be assessed. A first (common) scenario is one in which conjectures are permissible. It cannot be stressed enough how important this scenario is, because it makes the other scenarios feasible, it provides a common arena for discussion, and it sets the philosophical tone for the rest of the controversy. Jefferson and Turing believe that it is worth pursuing the problem in a speculative space, even if they cannot resolve the question in (strictly) empirical terms. In a second scenario, important differences emerge. The question of which scientific discipline is more adequate for investigating the issue of thinking machines, and which "kinds of machines" are involved, is decisive. Applied biology or what we now call computer science represented two divergent intellectual bets in which the nervous system or computing machines became the plausible causes. And, in time, this scenario raises the question of the appropriate analogy that allows us to use computers to study the mind. A third scenario is advanced by Jefferson, where the investigation of the possibility of thinking machines is pursued in terms of a Cartesian view. However, there are at least two very different Cartesian paths of this subject. The traditional one is committed to a dualistic and deductive research program, where the main question is how to derive an interesting behavior from an appropriate cause. There is another account in which the main task is to infer the appropriate cause from a relevant behavior. This second Cartesian account involves a non-dualistic commitment in which some kind of plausible inference must be used. When Jefferson quotes Descartes' account of thinking machines, it is this last alternative path that is defended. And Turing shared this way of presenting the problem. But again, in this new scenario, the contrast between what each contender considers relevant about organisms and machines is recreated, except that now the main issue is plasticity or variability. In direct relation to this scenario, a fourth can be presented in which another shared ground is established. There is strong evidence for Turing's materialist commitment. It is debatable whether Jefferson can be considered a materialist. The last words of his 1949 paper seem to oppose this assumption. However, from the antimechanistic arguments raised by Jefferson and their explicit anti-reductionist perspective, it could be claimed in favor of this (shared) scenario. In particular, it could be argued that the kind of explanation Jefferson is looking for is in terms of the complexity of organisms and an anti-mechanistic account of them. So, at least in terms of the kind of explanation that Jefferson used, it can be defended that there is a materialist (or naturalistic) commitment in both contenders. A fifth and final scenario, in which more detailed experiences are contrasted, is used to specify the nature of plastic behavior. The importance of error is emphasized by Jefferson and Turing. For Jefferson, following common arguments in anti-reductionist biology of the first half of the twentieth century, the specific kind of plasticity that organisms show is evident in how they deal with damage, or in his machinist version, error. Turing saw in a particular kind of error (error of inference) the actual possibility of intelligence. Again, while they share a basic appreciation of the importance of this issue, they disagree on the particular experiences that support alternative worlds. It is worth noting that even when this fifth scenario is similar to the third one, more specific experiences (about kinds of error or damage) are called upon.

Sometimes thought experiments are seen primarily as a way to contrast intuitions or concepts. However, an investigation using a Kuhnian account of thought experiments could encourage an understanding of controversies in cases where alternative scenarios are built by contrasting relevant experiences. In this sense, it can be defended that we can learn, understand and, more importantly, open up new lines of research, contrasting and evaluating alternative scenarios, where the choice of relevant experiences and the assessment of assumptions are clarified.

Acknowledgments

An earlier version of this paper was presented in Montevideo in 2022. I am grateful to Marisa Velasco, Andrés Ilcic, and Bernardo Gonçalves for insightful comments and corrections on previous drafts of this paper. I also wish to thank the *Modelar* Research Group of the University of Córdoba and two anonymous referees for providing several suggestions and criticisms.

REFERENCES

Abramson, D. (2011). Descartes' influence on Turing. *Studies in History and Philosophy of Science-Part A*, *42*(4), 544-551.

Allen, G. E. (2005). Mechanism, vitalism and organicism in late nineteenth and twentieth-century biology: The importance of historical context. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, *36*(2), 261-283.

BMJ (1949). Mind, machine, and man. *The British Medical Journal*, *1*(4616), 1129-1130.

Brown, J. R. (2004). Why thought experiments transcend empiricism. In Hitchcock, C. (Ed.) *Contemporary debates in the philosophy of science* (pp. 23-43). Oxford: Blackwell.

Brown, J. R. (1991). *The laboratory of the mind: thought experiments in the natural sciences*. London/New York: Routledge.

Brown, J. R., & Fehige, Y. (2019). Thought experiments. In E. N. Zalta (Ed.),*The Stanford encyclopedia of philosophy* (Winter 2019). Retrieved from https://plato.stanford.edu/archives/win2019/entries/thought-experiment/

Butkovic, A. (2007). What is the function of thought experiments: Kuhn vs. Brown. *Croatian Journal of Philosophy*, *7*(19), 63-67.

Buzzoni, M. (2018). Pierre Duhem and Ernst Mach on thought experiments. *Hopos: the Journal of the International Society for the History of Philosophy of Science, 8*(1), 1-27. doi:10.1086/695720.

Brecevic, C. (2021). The role of imagination in Ernst Mach's philosophy of science: a biologic-economical View. *Hopos: The Journal of the International Society for the History of Philosophy of Science*, *11*(1), 241-261.

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

Copeland, B.J. (Ed.). (2004). *The essential Turing: seminal writings in computing, logic, philosophy, artificial intelligence, and artificial life, plus the secrets of Enigma*. Oxford: Clarendon Press.

Copeland, B.J. & Proudfoot, D. (2009). Turing's test. In Epstein, R., Roberts, G., Beber, G. (Eds.), *Parsing the Turing test* (pp. 119-138). Dordrecht: Springer.

Dennett, D. C. (2004). Can machines think? In Teuscher, C. (Eds.), *Alan Turing: life and legacy of a great thinker* (pp. 295-316). Heidelberg: Springer.

Dennett, D. C. (2013). *Intuition pumps and other tools for thinking*. Oxford: Norton.

Descartes, R. (1637/2006). *A discourse on the method*. Oxford: Oxford University Press.

Descartes, R. (1644/1983). *Principles of philosophy* (VR Miller and RP Miller, Trans.). Dordrecht: Reidel Publishing Company (original work published 1664).

Epstein, R., Roberts, G. & Beber, G. (Eds.) (2009). *Parsing the Turing test*. Dordrecht: Springer.

Gendler, T. S. (2000) *Thought experiment: on the powers and limits of imaginary cases*. New York: Garland Press (now Routledge).

Gendler, T. S. (2004). Thought experiments: rethought—and reperceived. *Philosophy of Science*, *71*(5), 1152-1163. https://doi.org/10.1086/425239

Gonçalves, B. (2021). *Machines will think: structure and interpretation of Alan Turing's imitation game*. (Doctoral dissertation). Retrieved from *The Digital Library of Theses and Dissertations of the University of São Paulo*.

Gonçalves, B. (2023a). Can machines think? The controversy that led to the Turing test. *AI & SOCIETY*, *38*(6), 2499-2509.

Gonçalves, B. (2023b). The Turing test is a thought experiment. *Minds and Machines*, *33*(1), 1-31. doi:10.1007/s11023-022-09616-8

Hodges, A., Alan Turing. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2019 Edition). <https://plato.stanford.edu/archives/win2019/entries/turing/>.

Horowitz, T. & Massey, G. J. (eds). (1991). *Thought experiments in science and philosophy*. Savage, Maryland: Rowman and Littlefield.

Ierodiakonou, K., & Roux, S. (Eds.) (2011). *Thought experiments in methodological and historical contexts*. Boston: Brill.

Jefferson, A. (1984). Geoffrey Jefferson 1886-1961. *Surgical Neurology*, *22*(1), 1-4.

Jefferson, G. (1949a). René Descartes on the localisation of the soul. *Irish Journal of Medical Science* (1926-1967), *24*(9), 691-706.

Jefferson, G. (1949b). The mind of a mechanical man. *British Medical Journal*, *1*(4616), 1105.

Kuhn, T. S. (1962). *The structure of scientific revolutions*. Chicago: University of Chicago Press.

Kuhn, T. S. (1977). *The essential tension: selected studies in scientific tradition and change*. Chicago: University of Chicago Press.

Laudan, L. (1981). *Science and hypothesis: historical essays on scientific methodology*. Dordrecht: Springer.

Lenoir, T. (1989). *The strategy of life: teleology and mechanism in nineteenth-century biology*. Chicago/London: University of Chicago Press.

Mach, E. (1893/2013). *The science of mechanics: a critical and historical exposition of its principles*. (T. McCormack, Trans.). Cambridge: Cambridge University Press (original work published 1893).

Mach, E. (1905/1976). *Knowledge and error*. Dordrecht: Reidel. (Original work published 1905)

Mettini, G. (2020). Los experimentos mentales como modelos científicos. *Revista Colombiana de Filosofía de la Ciencia*, *20*(40), 199-223. https://doi.org/10.18270/rcfc.v20i40.3237

Miščević, N. (2007). Modelling intuitions and thought experiments, *Croatian Journal of Philosophy*, *7*(20), 181-214.

Moue, A. S., Masavetas, K. A., & Karayianni, H. (2006). Tracing the development of thought experiments in the philosophy of natural sciences. *Journal for General Philosophy of Science*, *37*(1), 61-75. http://www.jstor.org/stable/25171335

Nersessian, N. J. (2007). Thought experimenting as mental modeling. *Croatian Journal of Philosophy*, *7* (2), 125-161.

Norton, J. D. (2004a). Why thought experiments do not transcend empiricism. In Hitchcock, C. (Ed.) *Contemporary debates in the philosophy of science* (pp. 23-43). Oxford: Blackwell.

Norton, J. D. (2004b). On thought experiments: is there more to the argument?.*Philosophy of Science*, *71*(5), 1139-1151.

Proudfoot, D. (2013). Rethinking Turing's test.*Journal of Philosophy*, *110*(7), 391-411.

Purtill, R. L. (1971). Beating the imitation game. *Mind*, *80*(318), 290-294.

Roll-Hansen, N. (1984). E. S. Russell and J. H. Woodger: The failure of two twentieth-century opponents of mechanistic biology. *Journal of the History of Biology*, *17*(3), 399-428. https://doi.org/10.1007/BF00126370

Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, *3*(3), 417-424.

Searle, J. R. (1984). *Minds, brains and science*. Cambridge, MA: Harvard University Press.

Shieber, S. M. (Ed.) (2004). *The Turing test: verbal behavior as the hallmark of intelligence*. Cambridge, MA: MIT Press.

Sorensen, R. A., (1992). *Thought experiments*. Oxford: Oxford University Press.

Stuart, M. T. (2022). Sharpening the tools of imagination. *Synthese*, *200*(6), 451. https://doi.org/10.1007/s11229-022-03939-w

Turing, A. (1950). Computing machinery and intelligence. *Mind*, 59(236), 433-460.

Weatherall, M. W., & Kamminga, H. (1996). The making of a biochemist ii: The construction of Frederick Gowland Hopkins' reputation. *Medical History*, *40*(4), 415-436.

Woodger, J. H. (1929). *Biological principles: a critical study*. London: Routledge.

Pío García is a professor in the Department of Philosophy at the National University of Cordoba (UNC), Argentina. He holds a doctorate in philosophy from UNC. His research interests are in the philosophy of science and the philosophy of computers.

Address: Universidad Nacional de Córdoba, Centro de Investigaciones María Saleme de Burnichon (CIFFyH). Pabellón Agustín Tosco, 1.er piso, Ciudad Universitaria, Córdoba, C.P. 5000, Argentina. E-mail: piogarcia@ffyh.unc.edu.ar - ORCID: https://orcid.org/0000-0001-7450-6539