Spoiler Alert! Unveiling the Plot in Thought Experiments and other Fictional Works

Daniele Molinari

University of Parma

Abstract

According to a recent philosophical claim, "works of fiction are thought experiments" (Elgin 2007: 47), though there are relevant differences, as the role of spoilers shows—they can ruin a novel but improve the understanding we can gain through a thought experiment. In the present article I will analyze the role of spoilers and argue for a more differentiated perspective on the relation between literature and thought experiments. I will start with a short discussion of different perspectives on thought experiments and argue that the mental-model view and the conception of games of make-believe are most promising for developing the present analogy. Then I will assess the similarities and differences between thought experiments and other works of fiction. I will focus on the role of spoilers and, more generally, on the foretaste context, of which they are part. This context guides readers of literary works of art to draw their attention to the literary and aesthetic quality of the text. In the case of thought experiments, on the other hand, it (i) prompts them to accept the presence of fictional elements in worldly-cognitive works and (ii) draws their attention towards cognitively relevant elements of the story. A discussion of Borges' Pierre Menard in the last part will show that literary works of art become thought experiments if they are embedded in an appropriate foretaste context. Spoilers, thus, unveil that even works which—due to their length or plenty of detail—usually are not considered thought experiments, can perform similar cognitive functions.

Keywords: Thought experiments, Works of fiction, Spoiler, Imagination, Foretaste context.

1. Introduction

A brand-new mystery story of your favorite author has just been published. As often, several plot anticipations begin to appear online. You cannot keep curiosity at bay and instantly try to find out who the new murderer is—a weak moment that just ruins your enjoyment of the long-awaited work. This love-hate relationship results from two conflicting tendencies: (i) the desire to immediately release

Argumenta 6, 1 (2020): 81-97
ISSN 2465-2334
© 2020 University of Sassari
DOI 10.14275/2465-2334/202011.mol

the tension and find out how the work ends and (ii) the desire to enjoy a good, unspoiled read, to work your way through the text step by step and slowly unveil the solution. This is particularly evident in the case of criminal novels or movies, where suspension builds up, holding the reader in tension, and twist turn their expectations upside down. It is the main purpose of works of this kind to keep the reader in suspense—and we admire them for that. Does this point apply to all works of fiction? If we look at thought experiments as fictional narratives,[1] spoilers seem to lose all their destructive aura and turn into a most helpful tool.

According to a recent claim, at least some works of literature are thought experiments (Carroll 2002; Davies 2007, 2018; Elgin 1996, 2007, 2014, 2017, 2019). This claim sheds light on a much-debated question: can fiction provide knowledge? It may be obvious that Nobel-laureates such as Wisława Szymborska and Gabriel Garcia Márquez wrote literary masterpieces of outstanding cultural and cognitive value. But it is less obvious how exactly these or other works of fiction impart relevant knowledge about the real world to the reader. Catherine Elgin provides a straightforward explanation: "works of fiction are thought experiments" (2007: 47). Thus, if a work of fiction can widen our cognitive horizons, it will do so in the same way a thought experiment does. This suggestion is, however, only the tip of the iceberg that hides, on a more profound level, a series of problems. Its plausibility depends essentially on the conception of thought experiments that one endorses, i.e. which position one opts for in the lively debate on the nature of thought experiments that is immersed in the icy waters, as it were: how can thought experiments—which are never actually carried out and often involve a fictional narrative—add to our empirical knowledge or contribute to our scientific understanding?

I will develop my argument in two steps. First I will briefly recall the main positions in the current debate on thought experiments and suggest that Elgin's claim is best suited to the view that thought experiments are mental model reasoning and works of literature are games of make-believe. Then, I will present some analogies between works of literature and thought experiments and show how the role of spoilers can help us to shed a new light on the differences between them. I will argue that spoilers can be useful in thought experiments and scientific papers, but counter-productive in many genres of fictional works as they ruin the reader's experience. It will be helpful to focus on the context in which spoilers can be embedded. There are works of fiction that can be used in two different ways, as literary works or as thought experiments, depending on context; in the first case, the author of the work will avoid spoilers so as not to diminish the aesthetic quality of the work, while in the second case spoilers can be helpful as they guide the reader's attention and so enhance the cognitive efficiency of the work.

2. Thought Experiments and Games of Make-Believe

2.1 How Do Thought Experiments Work?

Let's start with thought experiments. They give rise to an epistemological puzzle (Kuhn 1977; Davies 2007) that can easily seem paradoxical: they can and often do enhance our understanding of reality even though they do not provide new empirical data. In recent debate at least three main accounts can be distinguished:

[1] Why should not we? Thought experiments are devices with narrative structure in which fictional events dynamically occur (Egan 2016, Nersessian 1991, Willée 2019).

the *Argument View*, the *Platonic Perception View*, and the *Mental-Model Reasoning View*. Let's have a quick look at the three positions.

John Norton claims that, behind their aesthetic and narrative features, thought experiments always ground themselves in deductive or inductive argumentations (1996, 2004). Each thought experiment can, therefore, be reduced to an argument with no epistemic loss: aesthetic and narrative features are only used for illustrative purposes. In an antithetical way, James Brown argues for a Platonic solution, taking thought experiments as "telescopes" directed to the realm of abstract entities (1991, 2004). The way in which we gain knowledge concerning this kind of artifacts has nothing to do with empirical experience. Rather, it is grounded on an *intellectual perception*—a special activity that allows us to grasp independent and outside-from-space-and-time laws of nature.

The attraction of the *Platonic view* lies in the fact that it recognizes thought experiments as a peculiar source of knowledge, different from empirical data-gathering or logical reasoning. Norton's position, on the other hand, has the virtue of parsimony, while Brown can do justice to the point that thought experiments are both essential and peculiar.

While for Norton, thought experiments are only "arguments in disguise" and can easily be replaced by them, the *Platonic Perception View* attributes a central relevance to them, but does invite for serious concerns: first, it does not provide a reliable account of how abstract entities can be grasped in intuition; second, the unexplained use of the metaphor of "intellectual perception" raises suspicion as it comes along with inappropriate empirical commitments. Norton's *Argument View*, on the other hand, fails to account for imagination-based thought experiments, such as cases where you have to imagine a certain shade of blue or those in which you put yourself in someone's else position to assess moral judgement.

Supporters of the *Mental-Model Reasoning view* can combine the strengths of the other two positions and avoid their problems by arguing that thought experiments provide cognitive advancement grounded on experience in virtue of an essential use of imagination rather than collecting new empirical data. In virtue of imagination, simulation, and memory, thought experiments can reconfigure previously obtained empirical data in new ways, prompting new experience-related knowledge (Gendler 2010; Miščević 1992, 2007; Nersessian 1991, 1992, 1999, 2018). This solution takes up central insights from Ernst Mach and Philip Johnson-Laird. The former argues that thought experiments use an instinctive kind of experience stored in memory but not yet propositionally articulated (Mach 1976). According to the latter we use mental models in the understanding of narrative texts, in which we experience some sort of contemplation of a fictional situation (Johnson-Laird 1983).

In conducting a thought experiment, we deal with a hypothetical state of affairs in a "What would happen if…" style. We reflect the consequences a counterfactual, fictional state of affairs would have. In doing so, the reader employs many non-linguistic cognitive resources, such as her familiarity with her own body, spatial intuitions, and forms of tacit knowledges (*Know-how*). In thought experiments, we imagine *seriously*: it is thanks to the meticulousness with which these cognitive resources are used, with respect to the constraints designed by the author, that a thought experiment can play a role that is analogous to that of a real experiment. Successful thought experiments invite the reader to imagine scenarios that allow her to draw true, or at least plausible, conclusions regarding the real world.

Tamar Gendler, who also draws on Mach's account, further develops this position, using the terms "mental representation" (2010: 47). She argues that thought experiments are moments in which one is "contemplating the imaginary case in question" using "a store of unarticulated knowledge of the world which is not organized under any theoretical framework" (Gendler 2010: 39).

It is worth noting that the authors here considered explicitly state that their theories regard scientific thought experiments only. According to Rachel Cooper, this is a "strategy of caution" (2005: 329): there are different types of thought experiments, and scientific ones seem *prima facie* easier to classify. Gendler does state, however, that the only difference between scientific and non-scientific thought experiments lies in the fact that the first concern "features of the physical world" (2010: 45). For these reasons, I believe that a unifying account of thought experiments is to be preferred. In this article I will consider thought experiments in general.

2.2 Games of Make-Believe and Pierre Menard

The role of imagination and its limits have been of interest in the entire history of philosophy. In what follows—to spoil the choice right away—I will rely on Kendall Walton's account, which provides an interesting explanation of the analogy between works of literary fiction and thought experiments. In particular, Walton's notion of *fictional world*—which can be compared to "mental model", a "space" within which the development of a fictional state of affairs is imagined, will be of interest.

For our purposes, one highlight of *Mimesis as Make-Believe* (Walton 1990) is the concept of *principle of generation*—a more or less explicit rule that gives structures to a fictional world. For example, a bunch of bored children decide to play together, and imagine that the apples on the kitchen-table are hand bombs. From the beginning of the game, and until its conclusion, the principle of generation "apples are dangerous hand bombs" is pretended to be true and the children begin to behave accordingly. All children who accept this principle of generation become players attuned to the same fictional world—a micro-world within which apples are fictitious hand-grenades. In light of that, a player who eats an apple (i) either pretends to commit suicide, (ii) proposes a further principle of generation and pretends to defuse or hide the bomb, (iii) does not play correctly, (iv) or is just hungry and decides to "detune" from the fictional world and "re-attune" to the real world.

According to Walton, principles of generation are central also in other contexts: figurative sculptures, paintings, and works of fiction prompt the viewers'/readers' imagination. Principles of generation are, thus, best understood as rules that prescribe what to imagine and which every player has to respect in order to take part in the game. According to this conception, imagination is not an unregulated, creative faculty but "it is a realm in which the play of ideas is bound by constraints the imaginer sets" (Elgin 2014: 227).

However, the mere desire to comply with a rule is not enough to act in accordance with it (Wittgenstein 2009: §202). An external criterion is needed. In Walton's perspective, *props* as warrants of the correctness of a game perform this function. A *prop* is an object that makes it possible to retrieve principles of generation and give coherence to a fictional world. The children involved in the *apple-*

bomb world rely on real apples to generate the fictional truth that there are hand-grenades on the kitchen table. The apples, thus, serve as external criterion in this.

Similarly, also fictional texts could be understood as props in games of make-believe: Jorge Luis Borges' short story (1999) generates the fictional truth that Pierre Menard planned to re-write Cervantes' *Don Quixote*, just like John Searle's article (1980) made it fictionally true that the person in the room can answer questions formulated in Chinese without mastering the language—and so made it possible that all readers take part in the same game of make-believe, the *Chinese-room* thought experiment. All props work independently from particular acts of imagination, but not in isolation: a (more or less) explicit agreement, in the stipulation of the principles of generation that guide the game, is necessary to start it and to make sure that all can enjoy the same fiction work or thought experiment—that might deviate from one another only in minor details (Meynell 2018: 503). This shows that Walton's theory has a well-marked normative and social dimension.

According to Walton, thus, the great works of literature are fine props that prompt the readers' imagination in different ways: by surprising them, by making them reflect on determinate matters or on the psychology of a character, or by making them identify with strange beings, etc. Sometimes, a work of literature can be used to show, to test, or to argue for a certain theory or hypothesis. Thought experiments understood as mental model manipulation seem to work precisely in that way: thanks to a set of prescriptions the author of a thought experiment invites the readers to imagine a certain fictitious state of affairs and so designs a model that shows something meaningful.

A potential critique of the position that I have discussed so far could emerge from a recent suggestion, formulated by Fiora Salis and Roman Frigg (2020), according to which only propositional imagination is necessary for the performance of thought experiments and games of make-believe. The contents of propositional imagination are propositions. It has three main features: (i) one can freely imagine any proposition one desires, (ii) an imagined proposition imposes inferential commitments similar to those imposed by a proposition that is believed, and (iii) to imagine a proposition does not require one to believe it. The authors argue that games of make-believe are cases of propositional imagination which, in addition of the above features, are (iv) social activities structured through (v) normative aspects. Such a classification is inspired by Gregory Currie (1990), who argues that activities of make-believe are propositional attitudes similar to belief and desire. Salis and Frigg, following Currie, suggest that to play a game of make-believe might arouse mental images, but this is not a necessary feature, and therefore it is not a relevant element of make-believe.

According to Salis and Frigg, the five features that identify make-believe listed above are also shared by thought experiments: to imagine a set of propositions that describe a hypothetical scenario, to imagine their possible inferences, and to imagine the conclusion that is to be drawn from such a set of imagined propositions do not require the contemplation of a quasi-visual situation. The conclusions of thought experiments, therefore, presuppose the possession of linguistic competences, rather than pre-theoretical information stored in memory or any kind of "phenomenological" imagination.

Walton is less reductive on this point; he distinguishes different types of imagination: "imagining a proposition, imagining a *thing*, imagining *doing* something" (1990: 13) and states:

> Props prescribe nonpropositional imaginings as well as propositional ones. They do not thereby generate fictional truths, but the mandated nonpropositional imaginings are a distinctive and important part of our games of make-believe (Walton 1990: 43).

Although the term 'make-believe' suggests some resemblance to beliefs, and therefore to propositional attitudes, Walton does not privilege propositional imagination over other forms of imagination: depending on the fictional world that is presented in them, games of make-believe can involve different forms of imagination.

Mental model reasoning seems to provide further evidence for Walton's liberal position, as it underlines the fact that we can gain meaningful insights from different imaginative resources, such as imagine-that-color ("imagining a *thing*") and imagine-that-feeling ("imagining *doing* something", or "imagining *being affected* by something"). Hume's notorious thought experiment of the *Missing Shade of Blue* (1999: 9f.) and Thomson's *Dying Violinist* (1971: 48f.), essentially involve non-propositional imagination: the visual imagination of a particular nuance of blue in the first case and the imagination of feeling empathy for a character, in the latter. Since both types of imagination cannot be reduced to propositional imagination, the attempt to reduce the imaginative activities required by thought experiments to propositional imagination must fail.

We are now in a position to come back to Elgin's claim concerning an analogy between literature and thought experiments. Borges' *Pierre Menard, Author of the Quixote* seems to substantiate this view: it invites the reader to consider a counterfactual scenario, adopting the sober and direct style of a literary review. As a work of literature, the text exemplifies Borgesian humor and an excellent mix of essayistic and narrative genre. As a thought experiment, it develops an argument in favor of the thesis that a work of literature cannot easily be separated from the historical context and that extrinsic features of the work (partially) determine its identity-criteria (Bailey 1990; Danto 1981; Goodman and Elgin 1988; Lamarque 2009).

We should be cautious, however. The example discussed, which is right in the *Goldilocks zone* between works of literature and thought experiments, is likely to be more an exception than a prototypical example. It is true that works of literature and thought experiments have important analogies, but we have to take their differences into account to avoid a *petitio principii*. I will return to *Pierre Menard* later in this paper, when I will discuss the role of spoilers in works of fiction.

3. Exploring the Analogy

There are apparent similarities between literary works and thought experiments. In what follows, I will discuss four points that illustrate the strength of Elgin's analogy: (i) both develop their plot in a narrative, (ii) in indeterminate or incomplete contexts, (iii) are subject to reality constraints, (iv) and both can provide an advancement of the reader's understanding. Elgin's analogy has its limits, however, as I will try to show in section 3.2, where I will come back to the phenomenon of spoilers.

3.1 Similarities

(i) First of all, both works of literature and thought experiments develop a narrative. Both describe series of events that are causally related or groups of

similar but causally unrelated events.[2] It is worth noticing that thought experiments are usually presented *in medias res*, which likens them to standard experiments, that are also conducted *in medias res* (Elgin 2014: 225), while other fictional works may use different narrative devices depending on what type of game of make-believe they are meant to induce.

(ii) Second, narrated events always occur in partially indeterminate or incomplete contexts. Unlike possible worlds, at least in David Lewis' conception,[3] works of literature and thought experiments leave many aspects open. It is neither true nor false that Pierre Menard has a brother who is a musician, or that the *Society of Music Lovers* in Judith Thomson's thought experiment publishes a journal that is dedicated entirely to Chopin's style. These details are not mentioned in the text and, therefore, play no role in the games of make-believe that they prompt. Ignoring details of a fictional world that are not present in the plot helps the author to control the scenario and to highlight the ones that she wants to draw attention to. We find a similar strategy in real experiments conducted in laboratories: also here it is important to suspend all irrelevant elements that could distort results and lead the reader astray (Elgin, 2014: 222)—and in thought experimentation this selective ability is at its best use.

(iii) Moreover, elements of the real world are typically carried over to the fictional worlds that are described in works of fiction or in thought experiments. This is what Stacie Friend calls the *reality assumption* (2017: 31), i.e. the assumption that readers of fictional works usually import aspects of the plot from the real world. It is true in the story that Pierre Menard has got a brain, or that the *Society of Music Lovers* is not composed by a bunch of domesticated monkeys, even if Borges or Thomson did not explicitly mention any of these facts. This point does not conflict with (ii). Although it suggests that we add details not present in the plot, these are details that we take for granted in our world-view and that are not specific of a given state of affairs. Moreover, even though some aspects are imported from the real world, other are still left indeterminate.

(iv) One last analogy is about the advancement of understanding: both thought experiments and works of literature can show, defend or confute a hypothesis. Powerful works of literature can have a more lasting impact and make the reader to reflect on their themes in later moments in time. Engaged novels like Orwell's *Animal Farm*, dystopian ones like Huxley's *Brave New World*, and existential novels such as Camus' *The Stranger* are particularly apt to prompt the reader to think about the meaning of life, moral choices, the responsibility of technology etc. In the case of thought experiments, it is more obvious that they can serve cognitive goals—after all, this is the main purpose for which they have been devised.

Notwithstanding these analogies, however, we can note a relevant difference between literature and thought experiments, which becomes particularly evident

[2] In *Magnolia*, directed by Paul Thomas Anderson in 1999, the plot is developed in an alternating intertwining that made school in contemporary cinematography. The stories narrated are almost totally isolated from each other but connected by various themes, including that of cancer.

[3] Lewis argues that every assertion regarding to a possible world has truth-value (1986).

when we look at the way in which the narrative is presented; i.e. at aspects of aesthetic appreciation and of literary style. I will focus on one of these aspects in the next section.

3.2 A Significant Difference: Spoilers

Let me illustrate how the role of spoilers differs between thought experiments and other works of fiction by discussing two emblematic examples, the movie *The Sixth Sense* (directed by M. Night Shyamalan in 1999) and Judith Thomson's aforementioned thought experiment of the *Dying Violinist*. Attention: spoiler alert! In the following discussion, the plot of *The Sixth Sense* will be revealed. If you do not want to ruin your experience of the movie, you better skip the next two paragraphs.

In *The Sixth Sense*, the child psychologist Malcolm Crowe deals with an apparently common case: a 9-year-old boy called Cole feels strongly anxious in every life context. Dr. Crowe takes the child to heart but Cole confesses that his problem is not psychological: he claims to possess the extraordinary capacity to see dead people. Crowe works hard to give the boy a life-purpose and put him in a condition to accept his special capacity. The film ends with a masterful plot-twist: we realize that Dr. Crowe had already died and was dead throughout the movie. His interaction with Cole were possible only due to the latter's paranormal gift. Throughout the movie, the spectator was longing to find out whether or how Dr. Crowe was able to heal Cole from his anxieties—just to find out, at the end, that Dr. Crowe was dead and it was Cole who comforted the wandering dead all the way long.

Shyamalan's movie illustrates well the destructive effect a spoiler can have on the spectator's experience: one of the central pleasures of the film lies in the fact that the final twist turns upside down the storyline of the entire film and forces the spectator to interpret several scenes of the film in a completely new light. One might even want to see the film a second time, just to see whether the entire plot was consistent with the new interpretation and whether there were hidden clues that could have given the surprising final twist away. This second viewing would afford, if any, a very different kind of pleasure. If a potential viewer has never watched that movie but already knows that Crowe is a wandering dead, both kind of pleasure would be ruined. The first because the surprise element, which makes the movie so interesting, would be lost; the second because who knows about the final twist throughout the movie will interpret all relevant scenes "correctly" and is deprived of the enjoyment of performing a "check-reading", i.e. to go back in memory to the relevant moments and reflect whether they are coherent with the new interpretation.

It might be argued that a viewer who already knows about the final twist could still enjoy *The Sixth Sense*, appreciating different aspects of the movie, such as the photography, the way it is directed, or the performance of the actors. Yet, the enjoyment the movie was intended to arouse would be lost, the viewer would not deal it as a story, because "when you have negotiated the intricacies of the plot—when you have experienced the surprises, made the discoveries, had your expectations verified—you have realized the intentions of the novelist [or the screenwriter] *qua* story-teller" (Kivy 2011: 7f.). Peter Kivy suggests that works of narrative fiction can be enjoyed only once if read by the reader as a story to be

told. Further readings could not provide the same kind of pleasure.[4] Accordingly, even though *The Sixth Sense* can be viewed focusing on scenography or photography, or viewed a second time for a "check-reading", it seems quite obvious that the main quality of the film is related to the enjoyment that results from the final plot-twist. Shyamalan's movie is a particularly effective example for analyzing spoilers in fictions since it masterfully shows the destructive effect it can have on the experience of a work of fiction.

Let me illustrate the role of spoilers in thought experiments with a short discussion of Thomson's *Dying Violinist*. In this famous thought experiment the philosopher asks you to imagine yourself waking up in bed next to a famous violinist who, as you learn right after your awakening, suffers from a kidney disease and risks dying. The *Society of Music Lovers* has kidnapped you because you have the same rare blood-type as the violinist and could, with your circulatory system pumping blood also through the violinist's body, save the life of the violinist. The hospital director concisely states: to save his life, you have to stay connected to his body for nine months. At this point, Thomson asks the reader: "is it morally incumbent on you to accede to this situation?" (1971: 49). What should you do if the situation will last not for months but for years, or for the rest of your life? It would certainly be a kind action to save the violinist, but no one seems to be morally constrained to stay in bed and ruin her existence for saving another person's life. This thought experiment invites the reader to imagine a fictional world designed to conceive, by analogy, the possible relationship between a mother and her fetus, and to understand some moral implications of abortion that could easily be underestimated or neglected.

In this paper, I am not interested in discussing this experiment's moral or political implications, but in another question: if a person, who reads through the text and in doing so conducts the thought experiment in her mind, knows already about the final twist beforehand, would that ruin her experience or the effectiveness of the thought experiment? I do not think so. Moreover, it seems that also Thomson would agree, since the paper in which the Violinist's case is presented is called *A Defense of Abortion*. It just seems that with this title Thomson wants to get the reader "straight to the point", without trying to hide the cognitive instances with which she elaborates her fictional story. Rather, she uses the "spoiler" to attract the readers' attention and arouse their curiosity.

This observation can be explained by the fact that every thought experiment is part of a broader argumentation or theoretical context—which we can call *foretaste context*—in which spoilers about the cognitive instances of the fictional work are presented and which, therefore, makes it easier for the reader to accept that a fictional narrative is embedded in a theoretical, scientific discussion.[5]

It is possible to introduce in this context instructions that guide the interpretation of the thought experiment and to communicate to the reader which aspect

[4] An exception to this point could be due to the viewer's bad memory. If over time she has forgotten the plot, a new experience of the same movie would realize again its story-telling intentions (Kivy 2011: 8).

[5] Note, however, that spoilers are not a necessary element of a foretaste context. It can be generalized as the ever-present contextualization that embeds all texts. It may be more or less complex, contain spoilers or not, be marginalized or significant in order to serve the purposes of the author. Thus, the aim of this paper is to analyze the role that spoilers can play within any foretaste context.

of the narrative is salient for the argument. Even when the thought experiment leads to an unwelcome conclusion,[6] knowing about it in advance does not ruin its effectiveness; on the contrary, spoilers can stimulate and guide a thought experiment's reading[7]—which marks an essential difference to other works of fiction.

The role of spoilers, considered here as a narrative and epistemic device, is useful to unveil a significant difference between the narrative style of works of literature or movies on the one hand, and thought experiments, on the other. Whether or not spoilers can be an effective device depends on how the content of the narrative is presented: showing spoilers concerning the plot—or its extreme consequences—can guide the viewer's attention, or block her emotional involvement and enjoyment, depending on the purposes for which a work is used.[8]

In this paper I suggest that it is in virtue of an effective foretaste context, in which we can read spoilers about claims and conclusions of a fictional state of affairs, that the author of a thought experiment succeeds in two aims: (i) justifying the presence of a fictional element within an essay with worldly-cognitive purposes, and (ii) guiding the reader's attention in order to put her in the condition to better understand which elements of that fiction are salient and which are not. The first claim becomes evident if we imagine the bewilderment of a reader who, while she's struggling with an essay about artificial intelligence, suddenly comes across a paragraph regarding Chinese idioms and people locked up in isolated rooms.[9] An out-of-the-blue employment of fictional stories is not so common in scientific essays, therefore an effective foretaste context is needed to make the reader accept the use of fiction in this kind of context.

The second point highlights the role that the instructions provided by the author play in the reader's interpretation of a thought experiment. Fictional texts often present multiple interpretative layers that result from different aspects of the work; clear-cut instructions can, thus, be helpful to put aside irrelevant interpretations—at least, the ones that are irrelevant for the author's cognitive purposes. If this strategy is applied, even fictional texts that were not designed as thought experiments can be illuminating. This requires us to put the aesthetic qualities aside and focus on those aspects that deepen understanding.

A good example of these two points is, again, Thomson's paper, in which the reader accepts the use of a short fictional story within a theoretical context because it is presented as a counterfactual situation about the main topic, thanks to the title and the style-formula "It sounds plausible [that a fetus could be a person from the moment of conception.] But now let me ask you to imagine this" (1971: 48). The reader, thus, is guided to weigh the elements presented in the fictional narrative: the fictitious fact of having been kidnapped by the *Society of Music Lovers* is hardly relevant, while the fictitious fact of staying in bed for exactly nine

[6] For example, a reader may be skeptical about Thomson's argument but surprisedly conclude, after carrying out her thought experiment, that the author has got a point.

[7] Just briefly think about how much interest the question "do you want to know about that thought experiment which shows how computers cannot think?" can prompt.

[8] An "above board" presentation of what cognitive instances are actually in play could enhance the strength and understanding of an argument better than a plot-twist. This does not mean we always have to ignore emotive responses when our claims are epistemic, nor that emotions play no role in the constitution of beliefs or in the advancement of understanding (see Elgin 2002, 2008).

[9] Fortunately, Searle's *Chinese Room* is well contextualized and does not confuse the reader—at least not for its being fictional.

months immediately stand out as salient. The spoiler in the title suggests that the thought experiment is presented with the aim to advocate women's right to choose whether to carry on a pregnancy or not—this allows the reader to read the story through the interpretative lens intended by Thomson and to neglect, or at least bracket, alternative interpretations.

Complementary to Elgin, who states that "[true descriptions can be] embedded in a work of fiction, a context in which an author is free to take liberties with truth in order to serve his aesthetic ends" (2007: 43), I think it is important to note that the opposite situation can also occur. Just as it is possible that fictional works may contain true statements within them—so it is possible that worldly-cognitive works may contain fictional elements, if fictionality helps the author to pursue her cognitive ends.

In addition, I suggest that, with an effective foretaste context, an author may be able to make a fictional story work like a thought experiment. This does not hold for any story, though: there must be a cognitively relevant and not overly ambiguous content emerging from the narration. In literary works of art of this kind, spoilers guide the reader's attention, and focus it on internal argumentation present in a literary work, on plausible causal chains of fictional events, on a thesis shown in the work, rather than focusing on the aesthetic pleasure that is aroused by it.

Note that it is still possible to perform a thought experiment without any previous spoiler about the plot and the cognitive instances involved. In these cases, if we cannot properly speak of "spoilers", we can still find interpretative instructions that guide the readers' attention. When the focus of the reader's attention is guided "at a later stage", she could—and probably will—do a "check-reading" of the thought experiment, just like in *The Sixth Sense* mentioned above, and check whether the interpretation is consistent with the fictitious state of affairs.

Thus, the use of spoilers is not a necessary condition for carrying out a thought experiment, even though it is an interesting device. Could spoilers nevertheless be a sufficient condition for a text to be a thought experiment? It depends on what they are used for. If their purpose is primarily epistemic, the work is used as a thought experiment. Spoilers, however, can serve different functions: some deep spoilers of *Pierre Menard* can be helpful in a creative writing course as a virtuous example. Thus, spoilers alone are not sufficient; they can fulfill different functions in the foretaste context of a work.

The point here is that a useful condition for using a fictional story with cognitive purposes is a theoretical context that puts aspects like entertainment and suspense aside. The author's instructions for the interpretation can be placed before or after the presentation of the fictional part, but, in order to render it more likely that the reader will accept the presence of a fictional story and avoid bewilderment, some spoilers can be helpful. Granted this, in the next section I will discuss in more detail an example of a short story that can work as thought experiment by virtue of an effective foretaste context.

3.3 Spoilers and Pierre Menard

In section 2.2 I mentioned Jorge Luis Borges' short story *Pierre Menard* as a good example of both a thought experiment and a work of literature. In the present section I will take a closer look at the work, analyzing how it presents itself and how it works in its original context. Finally, I will discuss the question of whether

an efficient foretaste context can "transform" *Pierre Menard*'s short story into a thought experiment.

The cognitive force of *Pierre Menard* lies in its illustration of how two texts that are identical word for word can be considered two distinct works of art, if we consider elements such as the social, historical and cultural contexts, as well as the intentions of the authors. It will be useful to have a look at the way the story is presented and, with all due respect to Derrida's claim,[10] the author's intentions: the text has the form of a short story and it has been published in a collection of short stories entitled *Ficciones*.

The foretaste context in which the work is presented contains no spoiler. It seems that the author did not want to force a single interpretation onto the reader. In fact, we only need to know a little bit about Borges to understand the plausibility of this point: throughout his *oeuvre*, Borges play on ambiguity so effectively that this clearly contributes to its outstanding literary and aesthetic value.

Nonetheless, the meta-narrative level in which *Pierre Menard* is developed makes it particularly versatile. If we focus on certain passages of the work rather than others,[11] we can easily find an ontological theory that is presented in an aesthetically and cognitively successful way. These considerations have led some philosophers to take up *Pierre Menard* as a thought experiment, quoting it with the appropriate foretaste context. Let me give you some examples that illustrate this point as well as the potential of the appropriate foretaste context.

Discussing the identity criteria of texts, Arthur Danto argues for the possibility that two works that are indiscernible are not identical with one another. With reference to Borges' work, he writes:

> The possibility was first recognized, I believe, in connection with literary works, by Borges, who has the glory of having discovered it in his masterpiece, *Pierre Menard, Symbolist Poet*. There he describes two fragments of works, one of which is part of *Don Quixote* by Cervantes, and the other, like it in every graphic respect—like it, indeed, as much as two copies of the fragment by Cervantes could be—which happens to be by Pierre Menard and not by Cervantes. […] the books are written at different times by different authors of different nationalities and literary intentions: these facts are not external one; they serve to characterize the work(s) and of course to individuate them for all their graphic indiscernibility. […] Borges' contribution to the ontology of art is stupendous: you cannot isolate these factors from the work since they penetrate, so to speak, the essence of the work (Danto 1981: 33-36).

Although Danto recalls the work with a different title, in this section we can see how he uses *Pierre Menard* as a thought experiment. He describes the text—without any attempt to avoid spoilers—and uses it as an integral part of his own argument. If someone who has never read Borges' work would come across this passage, probably all these spoilers and the argumentative context, in which it was

[10] The reference here is to the lucky slogan "there is no outside the text" (Derrida 1976: 158).

[11] In the first part of the work, there is a long and ironic list of the visible work left by Pierre Menard, to be contrasted with his invisible fragment, the re-writing of parts of the *Don Quixote*. An effective foretaste context will take these elements to the background and guide the reader's attention to another part of the story, where comparative analyses between Menard's and Cervantes' texts clearly illustrate the thesis at stake.

embedded, would tempt her to read it. It is worth noting that, in this case, the spoilers would not ruin the experience of the short story. On the contrary, they would add to its enjoyment: the ontological thesis shown in the story is the main point of its plot—at least, in this theoretical context—and there is no need for plot twists or unexpected surprises.

Even more explicitly, Peter Lamarque uses *Pierre Menard* as thought experiment in *The Philosophy of Literature*:

> Jorge Luis Borges's witty short story "Pierre Menard, Author of the *Quixote*" has come to epitomize, for philosophers, thought-experiments about works and texts, supposedly offering a powerful fictional exemplification of the view that distinct works can have identical texts. In the story, Menard, a fictional early-twentieth-century Symbolist poet, has the ambition to write *Don Quixote*, not by merely copying the original, but by a fully inspired act of literary creation. Here is a key, and often quoted, passage from the story:
>
> > It is a revelation to compare Menard's *Don Quixote* with Cervantes's. The latter, for example, wrote (part one, chapter nine):
> >
> > > …truth, whose mother is history, rival of time, depository of deeds, witness of the past, exemplar and adviser to the present, and the future's counsellor.
> >
> > Written in the seventeenth century, written by the "lay genius" Cervantes, this enumeration is a mere rhetorical praise of history. Menard, on the other hand, writes:
> >
> > > …truth, whose mother is history, rival of time, depository of deeds, witness of the past, exemplar and adviser to the present, and the future's counsellor.
> >
> > History, the *mother* of truth: the idea is astounding. Menard, a contemporary of William James, does not define history as an inquiry into reality but as its origin. Historical truth, for him, is not what has happened: it is what we judge to have happened. The final phrases—*exemplar and adviser to the present, and the future's counsellor*—are brazenly pragmatic.
> >
> > The contrast in style is also vivid. The archaic style of Menard—quite foreign after all—suffers from a certain affectation. Not so that of his forerunner, who handles with ease the current Spanish of his time.
>
> […] Whether or not Borges's story in itself provides adequate grounds for distinguishing work from text, it shows in effect the way that distinction could be maintained. A single text could be shared by two distinct works if certain conditions are in place: at the least, the works must have different properties and the texts must be produced by independent creative acts (Lamarque 2009: 74ff.).

Unlike Danto, Lamarque adds two elements: he not only discusses the cognitive value of the work, quoting its most significant part; he explicitly refers to Borges' text using the term 'thought experiment'. Lamarque's term is uncontroversial and does not hurt the reader's common sense: the short story fits well within a theoretical and argumentative horizon that justifies its presence.

Not all philosophers agree with the ontological thesis presented by Borges. Nelson Goodman and Catherine Elgin, for example, consider it with critical intent, when they discuss the criteria of identity of a text and its relationship with different interpretations (1988: 60ff.). This clearly indicates that Goodman and Elgin tacitly treat Borges' work as a thought experiment. After all, it is good prac-

tice that a philosopher, who does not agree with the conclusion of a thought experiment, challenges the fictional scenario or its interpretation.[12] For Goodman and Elgin, therefore, *Pierre Menard* is a thought experiment that can raise a serious objection to their own claim and, thus, calls for discussion.

Finally, George Bailey dedicates an entire paper to the ontological debate around the *Pierre Menard*'s case. He recognizes in Goodman and Danto the main argumentative poles, arguing that Borges' story should be considered as a valid contribution to the ontology of artworks (Bailey 1990: 340). This shows the importance of Borges' short story, which has come to stand for a specific philosophical position within a prolific debate in contemporary aesthetics.

It is important to note, however, that Borges does not use a spoiler in the story's foretaste context to guide the reader towards a specific ontological conclusion. It rather seems plausible that he has played with the absence of any interpretative line, with the aim to prompt astonishment in the reader who, reading *Pierre Menard*, might wonder whether what she has just read is nothing but brilliant nonsense or whether, perhaps, Borges has got a point. The philosophers considered above certainly accept the second possibility and use *Pierre Menard* in a context which, at the cost of unveiling the short story's plot, gives more prominence to the thesis it can be taken to prove.

This shows a central point. If we put *Pierre Menard, Author of the Quixote* in a theoretical context which explicitly states the main point of its plot, the short story does not appear flawed. We would consider the argument to be properly developed—and might be surprised by how well it was written, especially when compared to other thought experiments.

Thus, Borges' work can be considered both a literary work of art and a thought experiment, depending on the context in which it is presented and used. If we grant the lesson that Borges seems to illustrate in his short story, perhaps we could argue that there are two distinct works—a literary work of art and a thought experiment—that are identical word for word, but differ in context, purpose, style and have to be interpreted in different ways. The whole thing about anticipation and explanation of a plot's main points therefore seems to be a constancy, and also an important literary device, in guiding the modes of access, interpretation and identification with which a reader approaches a fictional text in cognitive contexts. Although it has previously been argued that *Pierre Menard* is both an outstanding work of literature and an illuminating thought experiment, the "transformation" is not immediately established and we have to add some spoilers, as well as to change the context of presentation of the short story, i.e. we have to change what I have called *foretaste context*, to make it really work as a thought experiment.

It should be noted, however, that Borges' short story is a most effective example: it is particularly apt to become a thought experiment, since it is short enough and the main point of its plot can be easily recognized as a substantial thesis concerning the ontology of artworks. It is more common to find larger and more heterogeneous works of literature that cannot, if considered in their whole, act as thought experiments even though some of them might contain suitable passages that do. Even in these cases, the foretaste context has the task of guiding the

[12] In this context it can be interesting to recall that John Searle presents his thought experiment of the *Chinese Room Argument* in connection with a discussion of several replies that have been raised against it (1980: 419-24).

reader to accept the author's choice of a certain part of the text, and to focus her attention in order to bring out the cognitive instances that are so fundamental in thought experimentation, but that, without explicit instructions, can also serve the function to add to the pleasure of reading.[13]

References

Bailey, G. 1990, "Pierre Menard's Don Quixote", *The Jerusalem Philosophical Quarterly*, 39, 339-57.

Borges, J. 1999, *Collected Fictions*, London: Penguin Books.

Brown, J.R. 1991, "Thought Experiments: A Platonic Account", in Massery, G.J. and Horowitz, T. (eds.), *Thought Experiments in Science and Philosophy*, Maryland: Rowman and Littlefield, 119-28.

Brown, J.R. 2004, "Peeking into Plato's Heaven", *Philosophy of Science*, 71, 1126-38.

Carroll, N. 2002, "The Wheel of Virtue: Art, Literature and Moral Knowledge", *Journal of Aesthetic and Art Criticism*, 60, 3-26.

Cooper, R. 2005, "Thought Experiments", in *Metaphilosophy*, 36, 328-47.

Currie, G. 1990, *The Nature of Fiction*. Cambridge: Cambridge University Press.

Danto, A. 1981, *The Transfiguration of the Commonplace. A philosophy of Art*, Cambridge, MA: Harvard University Press.

Davies, D. 2007, "Thought Experiments and Fictional Narratives", *Croatian Journal of Philosophy*, 7, 29-45.

Davies, D. 2018, "Art and Thought Experiments", in Stuart M.T., Fehige, Y. and Brown, J.R. (eds.), *The Routledge Companion to Thought Experiments*, London and New York: Routledge, 512-25.

Derrida, J. 1976, *On Grammatology*, Baltimore: John Hopkins University Press.

Egan, D. 2016, "Literature and Thought Experiments", *The Journal of Aesthetics and Art Criticism*, 74, 139-50.

Elgin, C. 1996, *Considered Judgement*, Princeton: Princeton University Press.

Elgin, C. 2002, "Art in the Advancement of Understanding", *American Philosophical Quarterly*, 39, 1-12.

Elgin, C. 2007, "The Laboratory of the Mind", in Huemer, W., Gibson, J. and Pocci, L. (eds.), *A Sense of the World. Essays on Fiction, Narrative, and Knowledge*, New York: Routledge, 43-54.

Elgin, C. 2008, "Emotion and Understanding", in Brun, G., Dogouglu, U. and Kunzle, D. (eds.), *Epistemology and Emotions*, London: Ashgate, 33-49.

Elgin, C. 2014, "Fiction as Thought Experiment", *Perspectives on Science*, 22, 221-41.

Elgin, C. 2017, *True Enough*, Cambridge: MIT Press.

Elgin, C. 2019, "Imaginative Investigations: Thought Experiments in Science, Philosophy and Literature", in Bornmüller, F., Franzen, J. and Lessau, M. (eds.), *Literature as Thought Experiment? Perspectives from Philosophy and Literary Studies*, Paderborn: Fink, 1-16.

[13] I am deeply grateful to Wolfgang Huemer for his helpful suggestions on earlier versions of this paper.

Friend, S. 2017, "The Real Foundation of Fictional Worlds", *Australasian Journal of Philosophy*, 95, 29-42.

Gendler, T.S. 2010, *Intuition, Imagination and Philosophical Methodology*, Oxford: Oxford University Press.

Goodman, N. and Elgin, C. 1988, *Reconceptions in Philosophy and other Arts and Sciences*, Cambridge: Hackett.

Hume, D. 1999, *An Enquiry Concerning Human Understanding*, Oxford: Oxford University Press.

Johnson-Laird, P. 1983, *Mental Models*, Cambridge: Cambridge University Press.

Kivy, P. 2011, *Once-Told Tales. An Essay in Literary Aesthetics*, Hoboken: Wiley-Blackwell.

Kuhn, T. 1977, "A Function for Thought Experiments", in *The Essential Tension*, Chicago: University of Chicago Press, 240-65.

Lamarque, P. 2009, *The Philosophy of Literature*, Oxford: Blackwell.

Lewis, D. 1986, *On the Plurality of Worlds*, Oxford: Blackwell.

Mach, E. 1976, "On Thought Experiments", in McGuinness, B. (ed.), *Knowledge and Error*, Dordrecht-Boston: Reidel, 134-47.

Meynell, L. 2018, "Images and Imagination in Thought Experiments", in Stuart, M.T., Fehige, Y. and Brown, J.R. (eds.), *The Routledge Companion to Thought Experiments*, London and New York: Routledge, 498-511.

Miščević, N. 1992, "Mental Models and Thought Experiments", *International Studies in the Philosophy of Science*, 6, 215-26.

Miščević, N. 2007, "Modelling Intuitions and Thought Experiments", *Croatian Journal of Philosophy* , 7, 181-214.

Nersessian, N.J. 1991, "Why Do Thought Experiments work?", *Proceedings of the Cognitive Science Society* ,13, 480-86.

Nersessian, N.J. 1992, "In the Theoretician's Laboratory: Thought Experimenting as Mental Modeling", *Proceedings of the Biennial Meeting of the Philosophy of Science Association*, 2, 291-301.

Nersessian, N.J. 1999, "Model-Based Reasoning in Conceptual Change", in Magnani, L., Nersessian, N.J. and Thagard P. (eds.), *Model-Based Reasoning in Scientific Discovery*, New York: Kluwer Academic and Plenum, 5-22.

Nersessian, N.J. 2018, "Cognitive Science, Mental Modeling, and Thought Experiments", in Stuart, M.T., Fehige, Y. and Brown, J.R. (eds.), *The Routledge Companion to Thought Experiments*, London and New York: Routledge, 309-26.

Norton, J.D. 1996, "Are Thought Experiments Just What You Thought?", *Canadian Journal of Philosophy*, 26, 333-66.

Norton, J.D. 2004, "Why Thought Experiments Do Not Transcend Empiricism", in Hitchcock, C. (ed.), *Contemporary Debates in the Philosophy of Science*, Oxford: Blackwell, 44-66.

Salis, F. and Frigg, R. 2020, "Capturing the Scientific Imagination", in Levy, A. and Godfrey-Smith, P. (eds.), *The Scientific Imagination*, Oxford: Oxford University Press, 17-50.

Searle, J. 1980, "Minds, Brains and Programs", *Behavioral and Brain Sciences*, 3, 417-57.

Thomson, J. 1971, "A Defense of Abortion", *Philosophy & Public Affairs*, 1, 47-66.

Walton, K. 1990, *Mimesis as Make-Believe: On the Foundations of the Representational Arts*, Cambridge, MA: Harvard University Press.

Willée, A. 2019, "Thought Experiments as a Narrative Genre", in Bornmüller, F., Franzen, J. and Lessau, M. (eds.), *Literature as Thought Experiment? Perspectives from Philosophy and Literary Studies*, Paderborn: Fink, 83-96.

Wittgenstein, L. 2009, *Philosophical Investigations*. Oxford: Blackwell.