This is the penultimate draft. The final version of the paper is forthcoming in
HOPOS: The Journal of the International Society for the History of Philosophy of Science (2025).

The Normativity of Imagination and The Evolution of Thought Experiments

Daniele Molinari (University of Parma)

According to Bokulich and Frappier, understanding thought experiments as Waltonian props for the imagination cannot explain their evolution, since their content is fixed by prescriptions to imagine. That is, fictional truths constrain researchers' imagination not to imagine otherwise. I suggest that the normative dimension of imagination is more flexible than Walton claims, especially in the context of TEs. Feyerabend's philosophy shows this by highlighting the fruitful role of violating prescriptions to imagine. I focus on the power of subjective imaginings to develop TEs-based debates. Readers of TEs are receptive researchers who rely on an initial fictional narrative to challenge the epistemic stance for which a TE was designed. This dialectical process can be framed through Waltonian notions of "work world" and "game worlds": the intersubjective set of fictional truths provided by a prop and the subjective imaginings of participants in the game of make-believe. According to Walton, proper players create their game worlds by building on the work world and trying to minimize mismatches. However, Feyerabend's philosophy emphasizes the importance of divergent game worlds. By violating fiction-based constraints, researchers engaged in a TE give their creative game worlds such a prominent role that they eventually evolve the work world itself.

1 Why we should care about the evolution of TEs[1]

Thought experiments are fictional narratives (Elgin 2014; Willée 2019 among others) written for epistemic purposes – such as supporting, criticizing or illustrating a thesis – in which readers are invited to imagine what would happen in a hypothetical situation. Far from being unusual tools, TEs have often been used in philosophy and the sciences. Some even play an important role in scientific conceptual reforming, as Kuhn noted (1964/1977, 263). Galilei's *Falling Bodies* (1638/1974, 66ff), for example, highlights a contradiction in Aristotle's physics, thus prompting a groundbreaking reconceptualization of the natural philosophy of his time.

Moreover, it is easy to find scientific debates that develop by amending fictional scenarios. For example, during a discussion of Heisenberg's Uncertainty Principle, Einstein invites Bohr to imagine a situation in which it seems to be possible to measure both the time and the energy quantities of a particle by weighing a box embedded in a purpose-built

[1] I am grateful to Wolfgang Huemer, Michael Stuart, Rawad El Skaf, Alessandro Torza, Andrea Bianchi, Francesco Cagnin, the participants of *The Legacy of Feyerabend* conference (University of Milan, 3–5 July 2024), the participants of *Paul Feyerabend: A Century of Challenging Orthodoxies* conference (University of Maribor, 3–4 December 2024), as well as the anonymous referees and editor, for their helpful suggestions on earlier versions of this paper. This research was funded by the project *RICERCA DUSIC 2023—AZIONE 1: The Epistemic Value of Imagination* (D.R. rep n. 1956/2023, PROT. 273172).

mechanism. Bohr, not satisfied with the scenario proposed by his opponent, switches from a generic "weigh-the-box" to weighing an imaginary box that "is suspended in a spring-balance and is furnished with a pointer to read its position on a scale fixed to the balance support" (Bohr 1949/1970, 226) – thus claiming that Heisenberg's principle holds even in this scenario. Similar examples can be found in philosophy as well: in a discussion based on Thomson's *Dying Violinist* (1971, 48f) one could vary the scenario from people whose bodies are connected to a violinist after being kidnapped by members of the *Society of Music Lovers* to those who are in the same situation after their names are drawn in a lottery (Warren 1973, 51). These are just two examples of how TEs that have been presented in one canonical form by their author can be varied by other scholars who discuss them. As we will see, these changes in the narrative are not misunderstandings: they are key stages in which our imagination-based debates unfold.

The widespread practice of modifying narratives in TEs has led some philosophers to address the question of what counts as an instance of the same TE, and whether it is correct to talk about an "evolution" of TEs. In this regard, Hacking takes a skeptical stance and argues that TEs are "rather fixed, largely immutable" as "they have only one tension to expose" (1992, 307), even though their presentation may be improved with time. According to the philosopher, a TE is better conceived as a picture to be accepted or rejected as a whole, as a "once-told tale" that cannot be retooled in the imagination.

More recently, some authors advocated a more flexible nature of TEs, arguing that their evolution is a common practice and not an anomaly (Bokulich and Frappier 2018, 549), that they have "a life on their own" (Bokulich 2001, 287) and a crucial diachronic aspect (Miščević 2022, 76), and that, in order to replicate a TE as well as a *real* experiment, "it is not necessary (or even possible) to duplicate the original in all its details" (Bishop 1999, 540). From this perspective, a TE can be reworked, its details altered, and it can remain the same TE.

In order to understand the point of this debate, it seems appropriate to ask why we should investigate the identity of TEs. What is philosophically interesting about them being fixed or evolving objects? I suggest that focusing on their evolving dimension is interesting for at least two reasons. First, it sheds light on their dialectical role. The practice of TEs flourishes by amending and reframing fictional scenarios. If we fully capture the flexibility of norms to imagine which constitute TEs, we can better understand how the scientific and philosophical debates based on them unfold. Second, it also impacts their pragmatic dimension, which consists of our willingness to participate – and accept the participation

of others – in this kind of imaginative activity. When scholars engage with the same hypothetical scenarios, their imaginings may yield different results. This shows that the contrasting inputs of individual researchers are phases of a negotiation process. The more aware we are of these dynamics, the more likely we are to accept an author's invitation to imagine in the first place: it would be more difficult to accept it if we were not allowed to revise and explore the proposed scenario.

In a recent paper, Bokulich and Frappier explore this subject by framing it through the most popular theories concerning TEs. According to them, an epistemically controversial theory such as Brown's Platonism[2] is better suited to explain the evolution of TEs, since it "seems to identify a continuity in the central narrative as all that is required to replicate a thought experiment" (Bokulich and Frappier 2018, 550), even though some further questions arise, such as what identifies the "core" of the narrative. A more parsimonious theory such as Norton's Argument view[3], on the other hand, struggles to achieve the same explanatory power, insofar as a single change in the premises would lead to a different argument and, therefore, a different TE. Moreover, the fictionalist theory which understands TEs as Waltonian props for the imagination (Meynell 2014, 2018) suffers from the same weakness, as the content of a TE would be fixed once and for all by its prescriptions to imagine.

In what follows, I will assume that TEs do evolve and show how this assumption is compatible with a fictionalist account of TEs – though diverging slightly from the established Waltonianism. To do so, I will appeal to a more comprehensive understanding of the normative dimension of the imagination, which considers the dialectical aspect of TEs. I will balance Walton's emphasis on following prescriptions to imagine with Feyerabend's focus on the productive role of "errors" – which, in this paper, take the form of violation of fiction-based constraints. But before that, let us consider the strengths and limits of the Waltonian account of TEs.

2 Strengths and limits of the Waltonian account of TEs

[2] According to Brown, TEs allow us to peek into the abstract realm of natural laws (2004, 1131). That is, they prompt an intellectual perception of mind-independent entities outside space and time.
[3] Norton argues that TEs can be identified as embellished arguments and that such embellishments play no role in justifying their conclusions. Therefore, there would be no epistemic loss in reducing the narrative of a TE into deductive or inductive arguments (1996, 2004).

According to Walton, our imaginative activities are regulated by intersubjectively valid rules ("principles of generation") that are determined by objects ("props") such as dolls, paintings and texts. These rules define what is true in fiction ("fictional truths") and, thus, prescribe what has to be imagined. The floor in the "the-floor-is-lava" game, for example, is a prop that invites the participants to imagine that a given floor is lava. If someone accepts the invitation, she begins to participate in a game of make-believe in which it is the case that the floor is made of lava. Similarly, a copy of Williams' *Stoner* is a prop that prescribes the reader to imagine that "William Stoner entered the University of Missouri as a freshman in the year 1910" (1965/2012, 1). These principles of generation prescribe the participants what to imagine. If they refused to do so, they would not play the game in an "authorized" way.

According to Walton, thus, games of make-believe introduce intersubjective criteria of correctness about what to imagine: "True, the agreements are made, the rules established voluntarily, and their prescriptions are relative to one's role as a participant in the imaginative activity in question. But they do prescribe. Anyone who refuses to imagine what was agreed on refuses to "play the game" or plays it improperly. He breaks a rule." (Walton 1990, 39). Games of make-believe impose constraints on the participants' imagination that authorize some imaginings while restricting others. These constraints – which we may refer to as "fiction-based constraints" – play a crucial role in shaping the game's outcome. If the rules are violated, a mismatch will arise between what the game prescribes to imagine and the participants' actual imaginings.

Walton explains these dynamics by employing the notions of "work world" and "game world" (1990, 59). Each participant creates her personal fictional world by relying on that of the work[4]. The set of fictional truths that constitute a work world is an intersubjective criterion with which participants should tune their imaginings. A work world deploys a normative force that all players have to take into account to play the same, authorized game of make-believe and create their subjective game worlds. In a nutshell, the work world sets the fiction-based constraints on imagination[5]. In section 3.2, I will further

[4] Minor mismatches can occur even with careful imaginers. An immersed imaginer could imagine herself as an observer within the fictional world, even without this prescription to imagine. If adding minor details does not conflict with the work world, such an imagining is authorized.

[5] This does not entail that what is fictional in a work world is always in plain sight. Fictional worlds are often open-ended, and fictional truths might be implied or left undetermined. Walton argues that readers import beliefs and expectations from the real world into fictional ones, applying the *Reality* and *Mutual Belief* Principles (1990, 144ff). Moreover, other elements such as "conventions specific to the genre" (1990, 165) further shape fiction-based constraints.

explore the interplay between the work world and game worlds. For now, it is important to keep in mind that, according to Walton, fictional truths in the game world of an "infallible" imaginer (who complies with all the prescriptions to imagine) and those that constitute the work world would overlap[6].

Although Walton focuses on representational arts, his notion of "games of make-believe" is also well-suited to advance our understanding of the nature of TEs, as the contributions of (Salis and Frigg 2020) and (Meynell 2014, 2018) have shown. From this perspective, TEs are props that prescribe imaginings for epistemic purposes. A section of Galilei's *Two New Sciences*, for example, prescribes the reader to imagine two falling bodies of different weights in order to reject parts of Aristotelian physics. If the participants can imagine the consequences and implications derived from the relevant principles of generation, they could broaden their cognitive horizons and observe a phenomenon from a new perspective.

Walton's theory emphasizes the social dimension of TEs and the normative role that imagination plays in them. It can do so by focusing on the intersubjectively accessible objects involved in games of make-believe and by showing how props can generate the constraints that constitute a TE. Moreover, the notions of "work world" and "game worlds" can account for the identity of TEs: "We perform the same TE when we imagine two game worlds which share the same fictional truths as the work world of a TE. The fictional truths of the work world are what confer identity, which allows that rather different descriptions of a TE […] are still importantly the same TE." (Meynell 2014, 4166). According to Meynell, we participate in the same TE if there is an overlap of fictional truths between our game worlds and the work world. Therefore, disagreements between scholars engaged in a TE are cases where "two different people construct two different game worlds on the basis of the same work world" (2014, 4166). Differences between game worlds might be due to simple errors. In the cases of deep disagreement, however, they might result from a divergent application of the principles of generation. According to Meynell, also these latter cases count as errors: "in such cases we have reason to suspect that the TE may be poorly constructed. Perhaps it is unclear which are the appropriate principles of generation, or perhaps there is some background belief […] or convention that is assumed by the TE's

[6] Again, putting aside irrelevant but widespread discrepancies, such as projecting oneself as a witness of the imaginary situation. However, this mismatch might become relevant if the prop prescribes that there is no witness, as in the story of a tree falling in the forest and no one hears the sound.

author and a number of members of the community that is in fact controversial." (Meynell 2014, 4166). Thus, when game worlds diverge significantly from that of the work, the TE is presumably flawed. The fictional truths in the work world hold a normative power, and their modification would lead to an unauthorized TE. If a TE constantly prompts divergent game worlds, Meynell argues, there must be something wrong with its narrative.

According to Bokulich and Frappier, this Waltonian approach fails to explain the evolution of TEs (2018, 554). The fictional truths constituting a TE's work world are fixed: when promulgated, they become written in stone. Accordingly, violations are errors that require correction. This point is consistent with Hacking's: when scholars discuss TEs, what they imagine "is what was once thought" and performing a TE is akin to "acting a part in a play" (1992, 307). Plays are often understood as artworks to be fully accepted or rejected by the audience – it would be bizarre to edit the plot of *Othello* by adding scenes. Such "reverence" for the original fictional world echoes the Waltonian focus on the normativity of the imagination: once we choose to participate in a game of make-believe, we ought to imagine its fictional truths. Therefore, if a TE is a Waltonian prop – and performing a TE is like acting in a play – we must accept or reject it as a whole.

To summarize: diverging from the fictional truths of a work world creates an unauthorized game world, while altering the text results in a different TE. In both cases, we violate the relevant fiction-based constraints, requiring a revision of our game world. I argue that the main limit of this Waltonian approach to TEs lies in overly loyal obedience to the prescriptions to imagine. Such a reverence might be appropriate in our engagement with artworks in museums. However, things are different with props in our epistemic endeavors, which develop through dialectical processes. We can enhance a fictionalist approach to TEs by incorporating Feyerabend's insight that violating constraints is more common and epistemically fruitful in the sciences than is often recognized. As we will see, TEs dialectically evolve through violations to imagine.

3 A Feyerabendian approach to Fictionalism and the evolution of TEs

In this section, I build on a suggestion by Stuart (2020, 2021) and previous stages of my research. First, I propose that a Feyerabendian view of TEs is well-suited to a fictionalist interpretation that emphasizes the role of imagination and artistic means in epistemic

endeavours – while also exhibiting a lively irreverence[7]. Second, I highlight the connection between the evolution of TEs and the flexibility of prescriptions to imagine.

3.1 Fictionalism and Feyerabend's view on TEs

Stuart argues that Feyerabend has interesting insights on TEs and invites us to imagine what would happen if the contemporary debate on them had started with Feyerabend instead of Kuhn[8] (2021, 278). I suggest that a consequence of this *what-if* scenario might be the emergence of a fictionalist account of TEs in the early stages of the current debate. Unlike the Waltonian approach, this account would emphasize the epistemic role of breaking the constraints of the imagination. This would highlight the significant ways in which rules are violated – rather than followed. One of its outcomes would be an explanation of the evolution of TEs that focuses on the abundance of creative game worlds rather than conformity to a single work world.

I ground these suggestions on three intertwined points of Feyerabend's philosophy: his methodological dadaism, his widespread appeal to the epistemic role of the imagination and his focus on theatre and storytelling. His motto "anything goes" (1975/1993, 19) invites us to be epistemically both cautious and courageous. It is a call to unleash our most creative thoughts and, at the same time, carefully consider all possible developments of theories that seem outdated and inaccurate. In other words, it is an invitation to grasp the value of the plurality of practices and methods that enable us to better understand a subject. It often takes the form of a call to imagine hypotheses at odds with currently accepted scientific theories – what Feyerabend calls "counterinduction" (1975/1993, 22f) – and a broad appeal to not limit one's own imagination, as "a very silly idea can lead to a very solid result" (2011, 130f). Even what would be marked as scientifically unacceptable theories – such as those postulating the existence of angels – may turn out, as time and traditions go by, to be successful. For example, medieval theological theories on the movement of angels unintentionally contributed to the reformulation of the physical notion of "space" which brought it closer to Newton's. Aristotle's qualitative conception of "place", according to which every type of substance belongs to its natural place, could not explain the movement

[7] The term "irreverence" was suggested by an anonymous referee, whose contribution I greatly appreciate.

[8] The contemporary debate on TEs is grounded in Kuhn's influential work and his focus on TEs as justification devices in the sciences (Stuart et al. 2018, 9). Accordingly, we should explain how, without relying on new evidence, a TE can "lead to new knowledge or to new understanding of nature" (1964/1977, 241).

of spiritual substances such as angels. Medieval philosophers solved this problem by elaborating a more quantitative, that is, more modern, notion of "space" (see Suarez-Nani 2018 for further discussion). These authors carried out what we might call TEs in order to understand how angels could move, while also bracketing the impact of the dominant Aristotelian tradition of their time. In this way, the unrestricted use of imagination enables us to achieve scientific progress by creating counterfactual situations that test established theories and suggest new directions for future research.

Feyerabend's emphasis on the free use of imagination aligns with his focus on art and theatre as epistemic tools. For instance, he argues that dramatic plays and scientific hypotheses share important similarities, as both can offer alternative scenarios or objects in order to challenge the dominant views (1967a, 410f), whether about physical laws or social stereotypes. Additionally, he suggests that drama – along with other art forms – can advance our understanding of "the *psychological conditions* under which criticism can be expected to become effective" (1967b, 298) and reveal the underlying ideologies involved in choosing which scientific hypotheses to develop further (1967b, 302). This emphasis on the cognitive dimension of theatre resonates with what Stuart calls Feyerabend's "epistemology of drama", which he extends to the broader practice of storytelling (2021, 266).

By combining Feyerabend's methodological Dadaism with his focus on the unrestricted use of imagination and the epistemology of drama, I argue that a Feyerabendian approach to TEs can be understood as an evolved form of Fictionalism – one that shares similarities with Waltonianism but diverges in important ways. In this view, conducting a TE might be akin to playing a role in a theatrical performance, as suggested by Hacking's metaphor. However, while Hacking considers the script of a TE as a prescriptive source to be followed (thus fitting a Waltonian perspective) a Feyerabendian approach would adopt a more irreverent stance, suggesting that TEs are not fixed by their original scripts. This interpretation of Feyerabend's philosophy is further supported by his references to Ionesco's theory of drama, where he claims that "the 'mechanics of the drama' assume a 'life [of their own] on the stage'" (1967b, 302) and that there are "'images, objects, events and characters' that have 'escaped' the author" (1967b, 303). This idea highlights the evolving dynamics of a performance that, once started, develops beyond the author's control and the script's prescriptions to imagine. With regard to TEs, this suggests that their fictional scenarios may evolve unpredictably as they engage with new audiences and contexts. If we understand TEs as Feyerabendian plays, these *escaping-the-author*

dynamics extend beyond the audience's detached judgement of the performance. Rather, the researchers employing a TE become deeply engaged co-creators of the work, actively reshaping the original script itself – a process we will explore in more detail in section 3.2.

Stuart suggests that TEs, in continuity with dramas, poems and short stories among others[9], are narratives that function as imaginative epistemic tools by presenting new *natural interpretations* (2021, 270). This notion plays a crucial role in Feyerabend's philosophy, where he argues that Galilei was able to provoke his scientific reform by altering the dominant natural interpretations of his time (1975/1993, 65). According to Feyerabend, our observations of phenomena are not pure sensory impressions independent of our prior knowledge. Instead, they are always an interweaving of sensory impressions and linguistic interpretations that emerge as a unified mental operation. In many familiar observations, such as those learned in educational settings during childhood, we often overlook this twofold aspect, leading to the conviction that phenomena "speak for themselves without outside help or extraneous knowledge" (1975/1993, 57).

Although intertwined, we can distinguish between the "sensory" and "linguistic" components of our experiences. The latter, which Feyerabend refers to as "natural interpretation", is the cognitive process that closely accompanies a sensation and frames it within a coherent interpretation. Examples of natural interpretations include descriptions such as "this figure represents a duck" when viewing the duck-rabbit ambiguous image. If we firmly believe that the duck-rabbit image represents a rabbit, this natural interpretation will sound incorrect or nonsensical. By devising compelling TEs, Galilei achieved a conceptual reform, challenging the established natural interpretations of his time,

[9] Stuart also considers forms of non-fictional storytelling, including scientific prose and political speeches. Note that, in Feyerabend's later work, the fiction/non-fiction distinction appears increasingly blurred. In *Conquest of Abundance*, as Ambrosio points out, Feyerabend argues that both scientific and artistic means are "mode[s] of reconfiguring a pliable reality while imitating it" (2021, 38). Feyerabend's analysis of Renaissance linear perspective suggests that the painting, the depicted objects, and the geometric procedures form a stage created by the painter (1999, 100). Accordingly, it is not possible to compare the painting and what it represents as independent objects, since "whatever 'reality' is being taken into consideration is manufactured by the stage, it changes with the setting" (1999, 111). This blurring, however, does not deny the crucial role of explicitly fictional stories in Feyerabend's epistemology. Rather, it suggests that fiction is so central in shaping our understanding of reality that it becomes fully integrated into our epistemic efforts – as Feyerabend notes: "*we need a dream-world in order to discover the features of the real world*" (1975/1993, 22). In this paper, I focus on TEs as explicitly fictional storytelling. Building on Stuart's interpretation of Feyerabend, my original contribution is twofold: (i) I frame Feyerabend's insights within a fictionalist view of TEs, challenging the dominant Waltonian emphasis on imagining in compliance with fictional truths, and (ii) I highlight the evolution of TEs by focusing on the normative dimension of the imagination.

introducing a new observational language, and undermining Aristotelian physics and Ptolemaic cosmology (1975/1993, 65ff). His TEs are examples of how scientific progress can be realized by employing fictional narratives that dispute natural interpretations once believed to be fixed and derived from things themselves.

However, these products of the imagination are stories that can switch from fertile possibilities to fixed dogmas. The latter kind of narrative – also dubbed as "myth" – is "no longer "alive" and changeable" (Stuart 2021, 268), which reminds of Hacking's view when he argues that "once the thought experiment is written out in perfection it is an icon. Icons, to reiterate, do not have a life on their own" (1992, 307). Dogmas, myths and icons – such as the story about the superiority of Western culture or the view of the mind as a computer – are natural interpretations that can be epistemically harmful insofar as they are taken for granted by people, thus discouraging them from imagining otherwise. From this perspective, TEs can be propaganda tools that strengthen myths by employing fictional situations aligned with those myths, such as cases in which a Western man decides whether to kill Native American people (Williams 1973) or cases in which human brains are connected to computers that simulate conscious experience (Putnam 1982). Such compliance with pervasive myths can lead to "a deterioration of intellectual capabilities, of the power of the imagination" (1975/1993, 96f).

TEs, however, can also break the power of myths (Stuart 2021, 270ff; Feyerabend 1975/1993, 118). They can provide competing natural interpretations, which allows for a proliferation of possible objects, situations, and styles – limited only by the imaginative skills of their authors and critics. To avoid turning a TE into a myth, it is important to produce new fictional details or whole new stories. This exploration of imaginative possibilities is what tests the "adequacy of popular standards" and constantly "violates them" (1975/1993, 234). Thus, the ongoing proliferation of TEs helps to keep them discussed and modified, that is, "alive".

Stuart presents two cases where successful TEs violate logic-based[10] and accuracy-based[11] constraints (2020, 970). Proponents of these kinds of constraints conceive them as rules to comply with in order to design an epistemically fruitful fictional scenario – along with the natural interpretation it conveys – in the first place. In this paper, however, I am interested in the kind of constraints whose violation leads to the evolution of a TE, that is, cases in which a scientific achievement is obtained by building on and modifying a previously presented imaginary scenario. Therefore, I will focus on fiction-based constraints, according to which researchers have to imagine what is fictionally true.

3.2 Violating fiction-based constraints and the evolution of TEs

While logic-based and accuracy-based kinds of constraints are meant to help an author in designing a successful TE, fiction-based constraints are needed to present a TE to other researchers. They thus become significant when we focus on the social context in which TEs are received and discussed, that is, research communities[12]. Even though the creation of a TE can occur in various circumstances, its "experimental" dimension requires a form of "replicability", that is, it involves the participation of other researchers in the same imaginative project. Although the replicability of a TE by a research community contributes to its reliability (El Skaf 2021), it does not need to be faithful down to the smallest detail

[10] Logic-based constraints require scientists to employ true premises and valid inferences in their imaginative exercises. Variations of this account appear in (Norton 2004; Williamson 2016) among others. Stuart discusses Galilei's *Falling Bodies* TE as an example where violating such constraints led to scientific progress. According to the logic-based constraints of his time, Galilei's argument was not conclusive, as Aristotelians argued that additional premises were needed, such as those concerning the difference between "united" and "unified" bodies (Gendler 1998/2010, 30). Thus, full compliance with logic-based constraints would not have allowed Galilei to obtain his scientific achievement.

[11] Accuracy-based constraints require imagining situations that closely resemble reality. Such constraints are explored in mental model theories of TEs, which hold that performing a TE involves representing a plausible situation shaped by mental simulations and previous experiences (Gendler 1998/2010; Miščević 1992; Nersessian 1992). Stuart presents Einstein's *Chasing the Light* TE as an instance of cognitive advancement provided by violating accuracy-based constraints. This TE invites you to imagine yourself riding a beam of light and wonder what you would see in this situation (cf. Norton 2013, 123). If we follow accuracy-based constraints in this TE, we would imagine a catastrophic explosion (Stuart 2020, 974). By violating this accuracy-based constraint, Einstein pursued his research on special relativity.

[12] TEs are devices employed in our collective epistemic activities (Molinari 2022; Sorensen 1992, 162; Elgin 2017, 230; Salis and Frigg 2020, 44; Murphy 2024). Note that research is just one context in which TEs are commonly used. In educational settings, for instance, TEs promote student participation and can make scientific theories more applicable in everyday life (Stuart 2018). Depending on the purpose, there are various types of TEs and different methods for performing the same TE (Frappier 2023). Focusing on TEs in research communities highlights an important point: TEs should be included in the dialectical processes familiar to philosophy and science.

(Bishop 1999, 540). Far from it, the social proliferation of TEs unfolds through rejections and variations of the fictional scenario.

In previous papers, I distinguished three ways in which TEs are commonly criticized, suggesting that such cases of "imaginative disharmony" are at the core of TEs as social practice (Molinari 2022, 240ff; Binini et al. 2024): (i) rejecting the proposed scenario and presenting a new one, (ii) reorganizing fictional truths by filling in the blanks and (iii) slightly adjusting the principles of generation. These methods can be employed to violate the fiction-based constraints on the imagination, and at least two can also be helpful for our current purposes.

Strategy (i) consists of rejecting the proposed scenario and replacing it with a new one, which echoes Hacking's argument and can be understood as the replacement of "one picture by another" (1992, 307). Even though it may be an appropriate dialectical move in the debate, it does not prompt any evolution of a TE. It rather suggests that a TE is so unsuccessful that it needs to be rejected entirely.

On the other hand, strategies (ii) and (iii) prompt the evolution of TEs by partially rewriting the script or uncovering previously undetermined/neglected fictional truths. In the Waltonian framework, these techniques – such as Warren's variation of details in Thomson's *Dying Violinist* or Bohr's implementation of a new measuring apparatus in Einstein's *Clock in a Box* – are unauthorized. Participants in a game of make-believe using such methods are breaking some fiction-based constraints. A Waltonian author would consider this result an error or, at best, the presentation of another game of make-believe. Feyerabend's insights, on the other hand, allow us to incorporate these techniques into scientific practices. Far from errors that need correction, they are fruitful steps that keep TEs "alive" and powerful.

But how can we explain the evolution of TEs as imaginative props more aligned with Feyerabendian plays? To address the issue, I employ the notions of work world/game worlds – this time from a Feyerabendian perspective. The key idea is that by putting their divergent game worlds[13] into play, researchers engaged in a TE can make it evolve by

[13] Oliveira argues that the most compelling instances of theoretical proliferation in Feyerabend's work involve cosmological theses (2021, 424), that is, whole divergent worlds. In this paper, however, I employ the notion of "world" as Waltonian "fictional world". Specifically, fictional worlds are sets of fictional truths and are "sometimes impossible and usually incomplete" (Walton 1990, 64). Therefore, a "divergent [game] world" does not concern a cosmologically divergent worldview but, rather, involves the change of fictional truths within a game of make-believe. My strategy is to employ a language familiar to proponents of established theories as a *Trojan horse*, as discussed by Feyerabend (1975/1993, 60). This approach facilitates the introduction of new ideas

revising its work world. Although this move is discouraged in a Waltonian framework, it is promoted as an example of healthy science from a Feyerabendian perspective. As a result, we shift the focus from the constraining power of a work world to the creative abundance of game worlds. The point can be illustrated by examining two telling examples: the evolutions of the *Clock in a Box* and *Dying Violinist* TEs.

Bohr's disagreement with the fiction-based constraints in Einstein's *Clock in a Box* work world can be understood as a modification of the initial narrative by presenting his divergent game world, inviting others to follow his new prescription to imagine. Bohr employed the second technique by which TEs are often criticized: (ii) reorganizing fictional truths by filling in the blanks. This technique takes advantage of the widespread narrative gaps that can be found in TEs and other forms of narratives (Sorensen 2019, 792). Even though some processes and objects can be described in more detail than others, no author can make every detail of their story explicit. These gaps, whether obvious (such as Sherlock Holmes' two nostrils) or undetermined (such as the contents of the briefcase in *Pulp Fiction*), provide fertile ground for researchers to enrich the original work world by creating more detailed game worlds. In the first stage of the *Clock in a Box* TE, Einstein did not specify how to weigh the box – a detail Bohr considered crucial to the TE's outcome (see El Skaf 2017 for further discussion). Therefore, by expanding on the *Clock in a Box*'s work world, Bohr imagines a game world that diverges to the extent that it clearly defines the method of weighing the box. This technique not only exploits a narrative gap left by Einstein but also prompts the evolution of the TE itself. Once Bohr's new detail is accepted by Einstein and the other participants in the game, it becomes an integral part of the set of prescriptions that constitute the *Clock in a Box*'s intersubjective work world – or, in other words, its script. As a result, it will constrain how subsequent researchers create their game worlds based on this TE.

The Feyerabendian interpretation of the evolutive dynamics between work world and game worlds also applies to the third technique of criticizing a TE, namely (iii) slightly adjusting the principles of generation. Warren's disagreement with the fiction-based constraints set by Thomson in her *Dying Violinist* work world exemplifies such a technique. Since we don't need to (and often don't) remain overly loyal to the work world of a TE, we're ready to introduce our divergent game worlds – as long as they contribute to

into the debate, prompting skeptics to accept them more readily. Likewise, I draw upon Waltonian notions in order to bring Feyerabendian insights into the current Fictionalist debate on TEs.

advancing the debate – even if they explicitly alter some principles of generation that constitute the work world. Warren argues that full compliance with the *Dying Violinist*'s work world would limit its scope to abuse-related unwanted pregnancy, pointing to the brutal kidnapping as the key detail that causes such limitation. Therefore, she complies with almost all the prescriptions provided by the *Dying Violinist*'s work world but diverges by replacing the kidnapping with a voluntary connection after her name is drawn in a lottery. According to Warren, this evolution broadens the scope of the *Dying Violinist*, allowing it to include less dramatic situations.

These dynamics emphasize the importance of divergent game worlds in the constitution of effective TEs. Feyerabend teaches us that the constraints set by a work world are not insurmountable boundaries. Rather, they invite participants in a TE to employ the fictional story to challenge the author's epistemic stance and to express their own voice. Researchers recognize themselves as engaged in the same TE, prompting them to collaborate and negotiate. At the same time, they are receptive to the contributions of others that often violate the original prescriptions to imagine. These negotiations produce "spates of articles" that contain evolving stages of the same TEs, creating what Miščević calls a growing "trail", or a "tradition" (2022, 84) – which can serve as an external criterion for their identification over time.

Miščević suggests that this evolving process aims toward a reflective equilibrium (2022, 85), where each new contribution refines the original scenario, eventually incorporating divergent viewpoints into a convergent conclusion. In contrast, Feyerabend's philosophy suggests that these spates are an unstoppable proliferation of alternative interpretations and modifications of fictional stories. Accordingly, presenting a TE does not aim for a peaceful equilibrium but establishes an ongoing starting point: participants are invited to contribute to a growing "ocean of anomalies" (1975/1993, 39), and each new reworking enriches the story and keeps it alive.

Therefore, a TE is a collective enterprise that evolves through the dialectical interaction of a starting work world and the divergent game worlds proposed by researchers – an enterprise identifiable through peer recognition and the "spates of articles" it generates. Accepting this idea does not exclude the possibility that some TEs could have a core set of prescriptions to imagine that is hard to disregard. Rather, it suggests that, if a TE does possess such a core set, it will result from negotiations among the researchers who employ that TE, instead of being fixed by the original script, the author's intentions, or any underlying metaphysical commitments.

This Feyerabendian interpretation of the interplay between work world and game worlds integrates the evolution of TEs with the fictionalist view of TEs as props for imagination. Every theory regarding the nature of TEs should be able to account for their evolution since it is closely related to their dialectical role in the sciences. Therefore, overly loyal obedience to the prescriptions to imagine risks collapsing the idea of TEs as imaginative tools. As we have seen, in contrast to what Bokulich and Frappier argue, TEs are props for the imagination that can evolve – if we recognize the flexibility of fiction-based constraints in epistemic contexts.

Even though the practice of TEs is not as tightly constrained as it is often suggested, the emphasis on the violations of fiction-based constraints does not entail a denial of the normative dimension of imagination. There are two sides to this dimension, though: on the one hand, Walton emphasizes compliance with the prescriptions to imagine. On the other, Feyerabend focuses on the epistemic fecundity of violating the rules of imagination. We need to know how to comply with rules to break them. Therefore, both accounts are complementary and highlight the normativity of the imagination, even though from opposite perspectives. The evolution of TEs is based on our abilities to follow and break rules. If we always comply with prescriptions to imagine, we will be held captive by fixed myths. Conversely, if we always violate the rules, we will not take any step forward. Both extremes lead to fossilized and unproductive TEs. Therefore, by following the rules and at times breaking them, we keep TEs alive.

4 Conclusion

This paper aims to reconcile a Fictionalist view of TEs with the phenomenon of their evolution. To achieve this, I propose a Feyerabendian interpretation of the Waltonian notions of "work world" and "game worlds" in order to highlight the connection between the evolution of TEs and the normativity of imagination. By adopting a Feyerabendian perspective, we recognize that TEs are not fixed once and for all by their original scripts. On the contrary, they are plays open to rejection, reinterpretation, and modification by the research community. The integration of Feyerabend's and Walton's philosophies sets the stage for a promising Fictionalist approach – one that encourages further exploration of the power of imagination in prompting the evolution of TEs and highlights the essential role of each imaginer in expressing the plurality of voices within research communities.

References

Ambrosio, Chiara. 2021. “Feyerabend on Art and Science.” In *Interpreting Feyerabend*, ed. Karim Bschir and Jamie Shaw, 11–39. Cambridge: Cambridge University Press.

Binini, Irene, Huemer, Wolfgang A., and Molinari, Daniele. 2024. “A Game of Perspectives. On the Role of Imagination in Thought Experiments.” *Erkenntnis* https://doi.org/10.1007/s10670- 023-00.

Bishop, Michael A. 1999. “Why Thought Experiments are not Arguments.” *Philosophy of Science* 66(1): 534–541.

Bohr, Niels. 1949/1970. “Discussion with Einstein on Epistemological Problems in Atomic Physics.” In *Albert Einstein: Philosopher-scientist*, ed. Paul A. Schilpp, 201–241. Cambridge: Cambridge University Press.

Bokulich, Alisa. 2001. “Rethinking Thought Experiments.” *Perspectives on Science* 9(1): 285–307.

Bokulich, Alisa, and Frappier, Mélanie. 2018. “On the Identity of Thought Experiments. Thought Experiments Rethought.” In *The Routledge Companion to Thought Experiments*, ed. Michael T. Stuart, Yiftach J. Fehige, and James R. Brown, 545–557. London: Routledge.

Brown, James R. 2004. “Peeking into Plato’s Heaven.” *Philosophy of Science* 71(5): 1126–1138.

El Skaf, Rawad. 2017. “What Notion of Possibility should we use in Assessing Scientific Thought Experiments?” *LatoSensu* 4(1): 19–30.

———. 2021. “Probing Theoretical Statements with Thought Experiments.” *Synthese* 199: 6119–6147.

Elgin, Catherine Z. 2014. “Fiction as Thought Experiment.” *Perspectives on Science* 22(2): 221–241.

———. 2017. *True Enough*. Cambridge: MIT Press.

Feyerabend, Paul K. 1967a. “On the Improvement of the Sciences and the Arts, and the Possible Identity of the Two.” In *Boston Studies in the Philosophy of Science, vol. III*, ed. Robert S. Cohen, and Marx W. Wartofsky, 387–415. Dordrecht: Reidel.

———. 1967b. “The Theatre as an Instrument of the Criticism of Ideologies. Notes on Ionesco.” *Inquiry* 10: 298–312.

———. 1975/1993. *Against Method*, 3rd ed., London: Verso.

———. 1999. *Conquest of Abundance. A Tale of Abstraction Versus the Richness of Being*, ed. Bert Terpstra. Chicago: University of Chicago Press.

———. 2011. *The Tyranny of Science*, ed. Eric Oberheim. Cambridge: Polity.

Frappier, Mélanie. 2023. “Towards a Pluralistic Account of Thought Experiments.” *Philosophia* 52(3): 595–603.

Galilei, Galileo. 1638/1974. *Two New Sciences: Including Centers of Gravity and Force of Percussion*. Trans. and ed. Stillman Drake. Madison: University of Wisconsin Press.

Gendler, Tamar S. 1998/2010. “Galileo and the Indispensability of Scientific Thought Experiments.” Repr. in *Intuition, Imagination and Philosophical Methodology*, 21–41. Oxford: Oxford University Press.

Hacking, Ian 1992. “Do Thought Experiments have a Life on their Own? Comments on James Brown, Nancy Nersessian, and David Gooding.” In *Philosophy of Science Association 1992 vol. 2*, ed. David. L. Hull, Micky Forbes, and Kathleen Okruhlik, 302–308. East Lansing: Philosophy of Science Association.

Kuhn, Thomas S. 1964/1977. “A Function for Thought Experiments.” Repr. in *The Essential Tension. Selected Studies in Scientific Tradition and Change*, 240–265. Chicago: University of Chicago Press.

Meynell, Letitia. 2014. “Imagination and Insight: A New Account of the Content of Thought Experiments.” *Synthese* 191(17): 4149–4168.

———. 2018. “Images and Imagination in Thought Experiments.” In *The Routledge Companion to Thought Experiments*, ed. Michael T. Stuart, Yiftach J. Fehige, and James R. Brown, 498–511. London: Routledge.

Miščević, Nenad. 1992. “Mental Models and Thought Experiments.” *International Studies in the Philosophy of Science* 6(3): 215–226.

———. 2022. *Thought Experiments*. Cham: Springer.

Molinari, Daniele. 2022. “Thought Experiments as Social Practice and the Clash of Imaginers.” *Croatian Journal of Philosophy* 22(2): 229–247.

Murphy, Alice. 2024. “Imagination and Creativity in the Scientific Realm.” In *Oxford Handbook of Philosophy of Imagination and Creativity*, ed. Amy Kind, and Julia Langkau. Oxford: Oxford University Press.

Nersessian, Nancy J. 1992. “In the Theoretician’s Laboratory: Thought Experimenting as Mental Modeling.” In *Philosophy of Science Association 1992 vol. 2*, ed. David L. Hull, Micky Forbes, and Kathleen Okruhlik, 291–301. East Lansing: Philosophy of Science Association.

Norton, John D. 1996. “Are Thought Experiments Just What You Thought?” *Canadian Journal of Philosophy* 26(3): 333–366.

———. 2004. “On Thought Experiments: is There More to the Argument?” *Philosophy of Science* 71(5): 1139–1151.

———. 2013. “Chasing the Light. Einstein’s Most Famous Thought Experiment.” In *Thought Experiments in Philosophy, Sciences and the Arts*, ed. Mélanie Frappier, Letitia Meynell, and James R. Brown, 123–140. London: Routledge.

Oliveira, Deivide G. S. 2021. “The Cosmological Divergent Proliferation in Feyerabend’s Pluralism.” *Principia* 25(3): 421–454.

Putnam, Hilary. 1982. *Reason, Truth and History*. Cambridge: Cambridge University Press.

Salis, Fiora, and Frigg, Roman. 2020. “Capturing the Scientific Imagination.” In *The Scientific Imagination*, ed. Arnon Levy, and Peter Godfrey-Smith, 17–50. Oxford: Oxford University Press.

Sorensen, Roy A. 1992. *Thought Experiments*. Oxford: Oxford University Press.

———. 2019. “Smartfounding: Four Grades of Resistance to Thought Experiments.” *Topoi* 38(1): 791–800.

Stuart, Michael T. 2018. “How Thought Experiments Increase Understanding.” In *The Routledge Companion to Thought Experiments*, ed. Michael T. Stuart, Yiftach J. Fehige, and James R. Brown, 526–544. London: Routledge.

———. 2020. “The Productive Anarchy of Scientific Imagination.” *Philosophy of Science* 87(1): 968–978.

———. 2021. “Telling Stories in Science: Feyerabend and Thought Experiments.” *HOPOS* 11(1): 262–281.

Suarez-Nani, Tiziana. 2018. “Space and Movement in Medieval Thought: The Angelological Shift.” In *Space, Imagination and the Cosmos from Antiquity to the Early Modern Period*, ed. Frederik A. Bakker, Carla Rita Palmerino, and Delphine Bellis, 69–89. Cham: Springer.

Thomson, Judith J. 1971. “A Defense of Abortion.” *Philosophy and Public Affairs* 1(1): 47–66.

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

Warren, Mary A. 1973. “On the Moral and Legal Status of Abortion.” *The Monist* 57(4): 43–61.

Willée, Arne. 2019. “Thought Experiments as a Narrative Genre.” In *Literature as Thought Experiment? Perspectives from Philosophy and Literary Studies*, ed. Falk Bornmüller, Johannes Franzen, and Mathis Lessau, 83–96. Paderborn: Wilhelm Fink Verlag.

Williams, Bernard. 1973. “A Critique of Utilitarianism.” In *Utilitarianism: For and Against*, ed. John. J. C. Smart, and Bernard Williams, 77–150. Cambridge: Cambridge University Press.

Williams, John. 1965/2012. *Stoner*. Repr. Vintage, New York: Vintage Publishing.

Williamson, Timothy. 2016. “Knowing by Imagining.” In *Knowledge Through Imagination*, ed. Amy Kind, and Peter Kung, 113–123. Oxford: Oxford University Press.