Imagining at the Threshold: Thought Experiments and Constraints in Quantum Gravity

Enrico Maresca[1,2]

[1]Department of Civilisations and Forms of Knowledge, University of Pisa, via Pasquale Paoli, Pisa, 56126,Italy.
[2]Department of Humanities and Philosophy, University of Florence, via della Pergola 60, Florence, 50121, Italy.

Contributing authors: enrico.maresca@phd.unipi.it;

Abstract

Thought experiments (TEs) are indispensable conceptual tools in scientific research, particularly in the study of quantum gravity. Many scholars argue that the epistemic significance of TEs hinges on the proper and ineliminable use of imagination. However, there is disagreement regarding the specific nature of the imagination involved. A valuable perspective on this debate is provided by a TE proposed by Matvei Bronstein in 1936 to support a quantum theory of gravity. His contribution serves as a notable example of destructive TE, aiming to highlight the internal inconsistency within a unified theory of both quantum mechanics and general relativity. In this paper, I reconstruct Bronstein's TE in the context of recent discussions on the relationship between TEs and imagination. I argue that this case study challenges existing epistemological frameworks for understanding TEs. I contend that Bronstein's TE introduces a new form of imagination, termed *operational imagination*, as indispensable for reaching its intended conclusion. I conclude that operational imagination can be integrated into simulative model-based accounts of TEs.

Keywords: Thought experiments, Quantum gravity, Imagination, Constraints, Coherence

1 Introduction

Imagination is an essential aspect of scientific research which contributes to scientific practice in different ways and at different stages. Nevertheless, the exact contribution it offers to the sciences is the object of long-standing contention. Traditionally, imagination has been relegated to the context of the discovery of theories, due to a tendency to examine theories only in terms of their conditions of acceptance and confirmation. While imagination was acknowledged a primary function in suggesting new hypotheses for phenomena as yet unknown, theory acceptance could not rely on such imaginings. Instead, acceptance was to be assessed "by careful observation or experiment" (Hempel, 1966, 16). Similarly, Popper (1934/2002) emphasised how imagination, although attested in scientific practice, is ultimately irrelevant to the assessment of a theory, which instead should be based entirely on available empirical evidence.

Post-Kuhnian philosophy of science questioned the demarcation between contexts of discovery and justification. Consequently, the segregation of imagining to the former has shifted.[1] This transformation has prompted philosophers of science to re-examine imagination (Murphy, 2022, 2024), not only as an object of inquiry but also as a pivotal element in related debates, such as the study of scientific models.[2] Notably, the centrality of imagination has been extensively discussed in the context of thought experiments (TEs), both within the sciences and in general philosophical practice.

This paper explores the complex uses of imagination in TEs from a privileged viewpoint. It is primarily concerned with the epistemic contribution that imagination can provide to the execution of TEs in the context of frontier physical research. This application is motivated by the peculiar inaccessibility of the objects of inquiry of certain new physical theories to empirical investigation and experimentation. As a result, this limitation necessitates the development of new research strategies that allow physicists to evaluate their theories. In addition, philosophers are required to assess the epistemological support provided by these new methods to their conclusions.

In this context, Bronstein's 2012/1936 work offers a noteworthy case study. In this seminal paper, the author proposes a TE to illustrate the reciprocal incompatibility between two successful theories of modern physics, namely general relativity (GR) and quantum mechanics (QM). Its final goal was to emphasise the necessity for a novel theory of quantum spacetime for high energy regimes.

I argue that Bronstein's TE can reach its epistemic goal only if it involves the use of a specific type of non-propositional imagination, termed "operational imagination." This is defined as the capacity to design imaginary settings and perform operations on them under suitable sets of constraints. The manipulative nature of operational imagination makes it an adequate conceptual tool in certain destructive TEs, such as Bronstein's. The experiments yield an epistemologically interesting scenario wherein

[1] See (Levy & Godfrey-Smith, 2020, 1–2) and (Stokes, 2016, 252–256) for an examination of the segregation of imagination to the context of discovery and the shift of perspective in recent times. Moreover, see Schickore (2022) for a broader discussion on the development of the notion of scientific discovery that puts creativity and imagination to the forefront.

[2] See, e.g., Frigg (2010), *contra* Levy (2012); Toon (2012, 2016). In particular, Toon (2016) provides a background for the discussion by tracing a precise distinction between direct and indirect fictional theories.

the investigator's objective is to highlight an inconsistency within the set of constraints. In particular, operational imagination is contrasted with two other prominent accounts within the literature on TEs, namely Salis and Frigg (2020)'s Waltonian account and Nersessian (1992)'s simulative model-based account.

The argument is articulated as follows. Section 2 offers an overview of the epistemology of TEs, with particular emphasis on the contribution of imagination in reaching the intended epistemic goals. Section 3 reconstructs Bronstein's TE and emphasises the function of imagination for its execution, as contrasted to the use of mere supposition. Section 4 then explains the notion of operational imagination, exploring its theoretical underpinnings and applications. In particular, it compares this notion with other types of imagination as delineated in the literature. Consequently, it illustrates how the internal inconsistency within a set of constraints provides epistemic support for the conclusion of destructive TEs. Finally, Section 5 discusses three potential criticisms of the contribution of operational imagination, both within the context of Bronstein's TE and in general.

2 Imagination and TEs

Imagination often plays a pivotal role in the construction and narration of TEs. This is particularly evident in the field of physics. To illustrate this point, one may consider Galileo's suggestion that two bodies be thrown from the Tower of Pisa. By doing so, the audience is invited to reproduce, or simulate, the TE in their minds to reach the same conclusion: the two bodies fall with the same acceleration. This *invitation to imagine* is a recurring element in many TEs, given that they can be carried out, by definition, without any material support. Consequently, a salient question can be posited regarding the sufficiency of this invitation to imagine and the cognitive processes involved in the execution of the TE to ensure that the conclusion is epistemically supported by the imaginary setting (*question of internal validity*).

Furthermore, it can be questioned whether the TE, although internally valid, can reach further epistemic goals when confronted with empirical scenarios (*question of external validity*). For example, a TE could be used as a substitute for a material experiment due to technical challenges in conducting the latter. Furthermore, one may require that the TE be explanatorily relevant to a specific phenomenon. Finally (but one may go further), the TE may be intended to increase understanding of a phenomenon under study, for instance by highlighting its mechanism of production, although this mechanism may be directly unobservable.

These questions give rise to the issue of providing a precise definition of the epistemic role of TEs in scientific practice. The extension of the philosophical literature evidences the complexity of this task. In light of this, Section 2.1 provides an overview of the two main "camps" in the epistemology of TEs. Section 2.2 then assesses the possibility of employing imagination in TEs. Furthermore, it discusses some relevant attempts to identify appropriate principles to constrain its uses within scientific contexts.

2.1 What are scientific TEs?

The status of TEs has been the subject of considerable debate and controversy within the recent philosophical literature. A significant reason for the lack of consensus has been the vast number of exemplary cases, discussed both in the sciences and in general philosophy. Accordingly, numerous attempts at the identification of clear and shared patterns of construction and execution, as well as the assessment of their epistemological function, have produced a somewhat fragmented landscape. A notable epistemological problem concerning TEs pertains to whether (and, if so, how) they provide new knowledge about the empirical world.

Some philosophers (termed the "no-camp" by Sartori (2023)) take a negative stance on this prospect, asserting that TEs, while applicable in scientific practice, contribute limited epistemic value. To illustrate, TEs may serve as "intuition pumps," thereby prompting our intuitions and increasing understanding of a certain phenomenon (Dennett, 1996). Alternatively, they may contribute to the teaching of science, for instance by emphasising certain interpretations of the physical laws by means of pictorial representations (Brown & Fehige, 2023). However, TEs are insufficient to generate new knowledge of the external world, nor can they provide sufficient epistemic justification for their own conclusions.

In contrast, other philosophers (the "yes-camp") put forth a more optimistic perspective, thus answering affirmatively to the aforementioned question. For instance, TEs have been proposed as a means to have Platonic intuition of the general laws behind phenomena (Brown, 2011). Alternatively, they may serve as conceptual tools for visualising and manipulating model systems of a physical target, thereby providing new knowledge about how the latter should behave in reality (Gendler, 2004; Miščević, 1992). This camp contends that imagination does indeed play an active epistemic role in reaching the conclusion of a TE.[3] However, internal differences in how the conclusion is obtained by means of imaginings are crucial.

The literature distinguishes different kinds of imagination, depending on the goal and structure of a TE. A prime distinction can be made between two categories. *Propositional imagination* is defined as the capacity to imagine *that* something is the case. To illustrate, one may imagine that a wooden chair is soft without necessarily visualising it. This kind of imagination dispenses of mental images as irrelevant for the imaginative process to be epistemically significant.

In contrast, *non-propositional imagination* assigns epistemic functions to non-propositional constructions and pertains to *how* something is the case. It is applicable to those imaginings that involve the production of mental images (*objectual or imagistic imagination*) and concerns how objects are portrayed in mental images. Similarly, non-propositional imagination may apply to imagining other phenomenal activities, such as seeing, hearing, and so on (*phenomenal imagination*).

Certain authors (notably, Norton (1991, 1996, 2004a, 2004b)) have contended that imaginative elements may occur in a TE, yet they ultimately have no active epistemic role and thus can be eliminated. According to this position, TEs possess

[3] By "active epistemic role" I intend to indicate the fact that imagination fulfils a significant function within the TE, one that allows the initial mental setting to reach a final state or conclusion. In absence of imaginative factors (whose nature should be further specified), the TE could not be executed from its initial setting to the intended conclusion.

epistemic significance merely due to their intrinsic argumentative structure, whereas additional factors are considered superfluous from an epistemic standpoint. In other words, Norton claims that TEs are essentially ordinary arguments disguised in pictorial form. Whenever a TE exhibits imaginative elements in its narrative, it can be reconstructed as a logical argument (*reconstruction thesis*) and entirely reduced to its argumentative structure (*elimination thesis*). Norton further contends that TEs are epistemically reliable due to their fundamental nature as arguments (*reliability thesis*). These arguments are constructed on empirical premises, facilitating their connection to reality.

Norton's account faces substantial criticism.[4] Firstly, the introduction of the reconstruction and elimination theses leads Norton to conflate their epistemic reach. As for the former, reconstruction does not provide epistemic support to TEs. The epistemic value of reconstructions is a matter of considerable debate, yet Norton does not take any position on that, nor does he specify epistemological implications of a reconstruction with respect to the original pictorial form.

Similarly, the elimination of pictorial elements may facilitate the elucidation of the inferential structure of TEs. However, this thesis entails a reduction in the scope of the experiment's broader epistemic functions. To illustrate, the elimination thesis arguably renders the TE unsuitable for understanding a phenomenon, provided that understanding is not limited to argumentative perspicuity.

Finally, both theses are presumed to work in principle, yet neither is guaranteed to succeed unless proven otherwise. In this sense, Norton omits any demonstration of the possibility of reconstructing a TE and eliminating its imaginative aspects; rather, he either assumes that everything works as intended, or uses specific examples of difficult generalisation.

Secondly, it is my contention that Norton's reliability thesis fails to take into account the existence of alternative epistemologically reliable means within science. Indeed, logical reliability is very restrictive as a condition of epistemic salience for scientific investigations. While I agree that logical reasoning is deductively reliable, it is not guaranteed to yield empirically adequate conclusions, despite the conditions that Norton imposes on the premises of a TE. It can be argued that empirical adequacy is contingent upon appropriate choices of language and logical system for investigating or reconstructing a phenomenon of interest. Furthermore, in the majority of cases, scientists apply more encompassing reasoning schemes, such as abductive reasoning, rather than strict deductive, logical reasoning. It could thus be argued that a logical truth is reliable, yet uninformative due to its deductive (thus non-ampliative) nature.

In conclusion, Norton ultimately makes the epistemic content of TEs collapse onto that of logical arguments, but then conflates the latter without adequately addressing several significant concerns.

In contrast, non-propositional accounts of TEs assign active epistemic functions to imaginative elements. To illustrate the point, Nersessian (1992, 2017) characterises TEs as instances of *simulative model-based reasoning*, which involves the manipulation of the mental models of a situation depicted in the narrative of the experiment. A *mental model* can be defined as "a structural analog of a real-world or imaginary

[4] Additional criticisms have been raised, e.g., by Salis and Frigg (2020).

situation, event, or process that the mind constructs to reason with." Consequently, "it embodies a representation of the spatial and temporal relations among and the causal structure connecting the events and entities depicted" (Nersessian, 1992, 293). In Nersessian's view, mental models exhibit non-propositional characteristics, yet do not necessitate any introspective process which would generate an image of it. Put differently, mental models are more malleable than iconic representations and deeply embedded in the activity of manipulating features within the experimental context.[5]

TEs involve the construction of a mental model and the act of drawing inferences from it. The narrative serves to describe the setting and sequence of steps in the execution of the TE, as well as being important for its communication. Consequently, narrative elements play a crucial yet auxiliary function, providing guidance to the reader in the construction of the model and the execution of the relevant inferences. Contrary to Norton, however, narrative elements are indispensable. This is due to the fact that the modelling function cannot be supplemented by reducing the TE to an argument. Any gaps in the narrative can also be integrated by pre-existing and real-world knowledge, while situational constraints built into the model are effective for directing the possible manipulations. These conditions ensure that the TE proceeds "*as one would in the real world*" (p. 295; italics in the original).

A middle ground between argumentative and non-propositional accounts is proposed by Salis and Frigg (2020). In their view, TEs are instances of Waltonian make-believe games characterised by the exclusive use of propositional imagination. This conception relies on the capacity to imagine propositions (*freedom*) and to derive inferential commitments from them (*mirroring*). Although the inferential procedure resembles belief, the authors specify that imagining is additionally characterised by *quarantining*: imagining that p does not entail believing that p. In contrast, non-propositional imagination is not a prerequisite for attaining the epistemic goal of a TE.

In this framework, thought-experimentation is defined as the construction of make-believe games where derived truths can be obtained from primary truths via principles of generation. The statement "it is fictional that p" can therefore be understood to mean "it is to be imagined that p." This prescription to imagine entails that the TE works as an exploration of a fictional scenario, guided by constraints and social aspects. The execution of a TE is thus motivated by the objective of determining what is true within it.[6]

[5]This point is worth stressing against an objection that has been raised in the literature. (Salis & Frigg, 2020, 38–40) conceive mental models as iconic representations described in "mentalese," and thus criticise Nersessian's account on two levels. Firstly, they claim that imagistic reasoning may be insufficient in a number of interesting cases. Second, they argue that it can even be unnecessary for a TE, as one needs to grasp the relations between elements, rather than the mental image they delineate. Given the above reconstruction of Nersessian's account, however, I consider these two objections convincing, but somehow misdirected.

[6]It is noteworthy that Salis and Frigg's reliance on propositional imagination has raised some concerns. Murphy (2020) expresses the worry that their position may be, after all, too close to Norton's account, or even more restrictive than the latter. Moreover, she argues, Salis and Frigg characterise *both* TEs and models as instances of make-believe games, and so they blur the differences between them.

2.2 Constrained imagination

A significant concern regarding the function of imagination in TEs is raised by the unreliability of its everyday uses.[7] It is, of course, possible to imagine a tomato that tastes of tabacco, although, to the best of my knowledge, no such product is currently available on the market or outside the tv series *The Simpsons*. In a TE focused on the dynamics of said tomato, taste would undoubtedly be an irrelevant detail, and so the freedom to imagine further properties would not be detrimental to the execution of the experiment. However, an alternative scenario could be imagined in which, rather than descending towards the centre of the Earth, tomatoes (and *only tomatoes*) are propelled towards the centre of Jupiter for reasons that may appear arbitrary. The resulting dynamics for tomatoes, derived from this TE, would be inapplicable to real-world scenarios and disconfirmed by a material experiment.

Such instances of free imagining substantiate the proposition of the sceptic who wishes to dissociate the use of imagination from the practice of thought-experimentation. In other words, the voluntary nature of imagination restricts its epistemic power, confining it to the realm of modality, as exemplified by the construction of counterfactuals. Consequently, no additional epistemic function can be attributed to imagination beyond this domain (*charge of epistemic irrelevance*; see Kind (2018)).

In contrast, a more nuanced perspective posits that imagination is not always entirely unconstrained. To illustrate, consider the case of an elementary school teacher who invites students to imagine what the interior of a pyramid would look like. Setting aside the implausible scenarios of Indiana Jones, this TE may have pedagogical purposes, as evidenced by the case of considering where the safest place might be to locate the treasure of the pharaoh. Such a process would be based on two main factors: common sense and previous knowledge. For instance, the location of the treasure should not be easily accessible, and the student should have prior experience of images depicting narrow corridors in pyramids or analogous structures. Imagination thus functions on a continuum of applications, ranging from its unrestricted, transcendent use to a more constrained, instructive one (Kind & Kung, 2016, 5). Each application is then displayed in a suitable manner in the context of different epistemic activities.

It is crucial to note that scientists do engage in imaginative activities. Scientific practice indicates that imagining can be an epistemologically *significant* activity even outside the scope of modal reasoning, provided that certain conditions are met. It thus becomes necessary to determine the extent to which imagination must be constrained in order for it to be considered reliable. This conception of imagination within scientific practice is an instance of what the literature calls *constrained imagination*, or "imagination under constraints" (Kind, 2018).

To illustrate, (Weisberg, 2020, 255) argues that "reality-proneness," or the tendency to constrain imaginings to what is perceived as "realistic" or "close to actuality," represents a default assumption for children when engaging in imaginative activities. The imposition of real-world structures and causal connections serves to constrain the imagined world, both in everyday situations and investigative contexts, with the potential for specific biases to influence the result. This connection to reality implies

[7] See McAllister (2013) for a comprehensive discussion of the different approaches to this problem.

that, should imagination prove unreliable, the resulting impact on our picture of reality would not be as significant as one might initially assume. Conversely, Weisberg contends that this significantly limits our capacity to envisage alternative scenarios in counterfactual reasoning. Consequently, the identification of an appropriate set of constraints becomes a pressing issue for the appropriate use of imaginings in any TE that purports to maintain some connection with reality.

The question of which constraints should be selected remains open. A general classification proposed by Murphy (2024) differentiates between *logic-based* and *model-based constraints*. The former ensure that imaginings align with the principles of sound reasoning, thus ensuring that all steps in the execution of a TE are constrained to be logically admissible, despite being performed in the realm of imagination. In contrast, the latter ensure that the imaginings accurately represent the target of the experiment and that all its manipulations track its realistic evolution over time. In particular, model-based constraints address the necessity of maintaining a degree of realism in the imaginings in order for the TE to offer insights into the actual world.

Nevertheless, this is not always the case. In addressing the conditions under which a TE may be considered externally valid, Sartori (2023) highlights that a TE may intentionally deviate from realism, potentially failing to align with a concrete target outside the realm of imagination or even misrepresenting it (for instance, through intense idealisation). Consequently, in Sartori's view, Murphy's model-based constraints should be refined to align with a more comprehensive framework of the representation of the target within the TE.[8]

An additional set of constraints is proposed by Walton (1990) and accepted by the so-called Waltonian accounts of imagination. In these approaches, the elements of the TE are conceptualised as *props* in a *make-believe game*. These props are defined as any object that can affect our senses and serve as prescription to imagine. Therefore, they direct the course of imaginings and fix the content of a fictional world, namely, the scenario described by the TE. In the TE's domain, *primary truths* are explicit conditions to be met and stated in the description of the experiment, whereas *derived truths* follow from the former by *principles of generation*. The execution of the TE (and similarly, the process of modelling) consists in the identification of the derived truths that concern the modelled target.

In his analysis, Walton identifies two principles of generation that warrant particular attention. Firstly, it is crucial to maintain a close proximity between the fictional and the real world (*reality principle*: (Walton, 1990, 145 ff.)). In the context of scientific TEs, this ensures that the performance and conclusion of the TE remains sufficiently aligned with real-world scenarios to teach us something about the latter. Secondly, the mutual beliefs held by a community can imply new truths within the narrative (or, in the case of a TE, within its execution) that they generate (*mutual belief principle*: (Walton, 1990, 150 ff.)). In other words, the beliefs held by a community engaged in a make-believe game can provide supplementary information that was not initially

[8]Sartori identifies this viewpoint in the so-called *DEKI* account (see Frigg and Nguyen (2020), especially ch. 8). Here, TEs are taken to describe fictional scenarios that denote a target and exemplify its relevant properties in a chosen context. These properties are keyed-up, i.e. mapped onto those of the material target via a "key." The key allows re-adapting the conclusions of the TE to the real target by removing the approximations involved in its execution. Finally, the keyed-up properties can be meaningfully imputed to the material target.

specified. This highly constrained structure, as outlined by Salis (2020), renders the Waltonian account of TEs extremely successful in science.

However, as Todd (2020) points out, the Waltonian account is lacking in details. Nevertheless, he concurs that when dealing with models and imaginings, one is ultimately engaging in a type of surrogative reasoning with the aim of gaining insight into the target through the application of principles of generation. These principles function as constraints on the model or imagining. However, the provenance and justification behind these principles remain unclear. For Todd, the constraining principles must precede the imagining process; otherwise, they would be products of that same process. Furthermore, these principles are not discovered, in the sense that discovery might entail the very imagining that they intend to constrain. Rather, their function is to restrict the potential direction of investigation.

In contrast, Stuart (2020) advocates the concept of epistemological anarchy in the utilisation of constraints on imagination. Upon examination of the current epistemology of imagination, he notes that these constraints predominantly pertain to either rules of good reasoning or the necessity for an accurate representation of the target system. Nevertheless, constraint-based perspectives ultimately offer principles that are either excessively permissive or unduly restrictive. The danger is that a set of norms will be imposed on scientific practice without allowing for free ("anarchic") investigation. The limitation arises from the aspiration for exhaustive coverage, which substantiates attempts to categorise all forms of scientific imagination as constrained by these two families of principles. In contrast, Stuart highlights instances where scientists have challenged these constraints to facilitate advancements. While this violation of constraints may introduce potential for error, he asserts that such missteps can also foster epistemic growth.

3 Bronstein's thought experiment: A case study

Stuart identifies a pivotal issue for present research in modern physics. To illustrate, consider the following quote by (El Skaf, 2021, 6132):

> Recently, many TEs have been used in black hole physics (since the 1970s). In these TEs, statements from QM, GR and thermodynamics, our best current theories, are grouped together. Their aim is precisely to reveal an inconsistency between these well-established theoretical statements; the "information paradox" [...]. These TEs are an important tool in contemporary foundational (philosophy of) physics, seeing that the object of inquiry is *inaccessible* – which renders any direct investigation impossible, at least for the time being. In addition, these theoretical statements come from theories with different objects and scales of applicability and thus are not easily grouped together.

El Skaf then proceeds to recognise that "we now need an account of TEs which appraises both their theoretical and experimental character" (*ivi*). This is evidenced by many research programmes in frontier physics and, in particular, by the case of quantum gravity (QG).

Historically, research in QG has not been driven by the presence of inconsistencies within the available data, in contrast to other theoretical frameworks. Instead, it has emerged as a consequence of a *theoretical* inconsistency between two of the primary theories of modern physics: QM and GR. The need to reconcile these theories has been substantiated by numerous TEs published since the 1930s. The objective of these efforts was to emphasise the necessity of formulating a novel quantum theory of gravity, with the aim of studying regimes of significant spacetime curvature. As will be discussed in the subsequent sections, these TEs operated by indicating an inconsistency between two sets of principles, one for each theory. They concluded that the resolution of the inconsistency necessitates the elimination or reformulation of some principle. However, given the absence of a theory of QG, or even a conceptual framework for what such a theory should entail, numerous alternative principles emerged, which may or may not interact with each other. Furthermore, these same principles may assume different roles in each account to QG, thereby acquiring a different status in the processes of building or assessing a theory (see e.g. Crowther (2021)).

Research in QG heavily relies on TEs, leading to the exploration of these strategies for theory construction by philosophers of science. For instance, Shumelda (2013) examines a renowned TE by Eppley and Hannah (1977), which has motivated the pursuit of a novel theory of quantum spacetime. In this regard, Shumelda contends that Eppley and Hannah's TE is intended to restrict the possible future theories of QG, even in the absence of empirical evidence.[9]

Among these TEs, Bronstein (2012/1936) offers a noteworthy study case. In his seminal paper, *Quantum theory of weak gravitational fields*, Bronstein argues in favour of the construction of a *quantum* theory of gravitation for high-energy regimes, in contrast to the success of GR at low energies. Despite the limited influence and circulation of the paper at the time, it was among the first to contemplate the possibility of a quantum spacetime structure. Bronstein's most significant contribution to the subject of this work is a TE that demonstrates, under specific operational assumptions, that the accuracy of measurement procedures in spacetime is bounded at high-energy scales. The limitations imposed by the operational definition of spatiotemporal objects, in accordance with the measurement-oriented approach of the time, intersect with the construction of the spacetime structure at those regimes.

Bronstein's TE poses an interesting challenge to current philosophical debates on the epistemology of TEs. Despite that, recent theoretical developments in physics have had little impact on the existing philosophical literature, which continues to prioritise older TEs and speculative inquiries. In contrast, there is a growing necessity to examine the role of TEs in shaping the contemporary research landscape of physics, with QG serving as a privileged example. To this end, I present a detailed account of Bronstein's TE and its implications. This preliminary study paves the way for a precise investigation into the function of imagination in deriving its conclusion.

3.1 The thought experiment

(Bronstein, 2012/1936, 274) proposes his TE "[i]n order to understand somewhat better the physical content of the quantum theory of the gravitational field" he intends to promote. To this end, he seeks to establish the experimental scenario, imposing two metatheoretical constraints:

[9] However, it is important to note that the results of this TE have been criticised by Mattingly (2006).

(B1) GR provides an accurate description of spacetime at all scales.
(B2) It is always possible to find a probe that is adequate for measuring the desired quantity of a system.

(B1) validates the extension of GR's domain of applicability beyond its intended macroscopic regime. This allows its application to microscopic regimes where quantum effects are anticipated to become relevant for our description of spacetime geometry. Consequently, all the laws of GR are incorporated into the experimental scenario as governing the possible behaviour of the systems involved.

Conversely, **(B2)** postulates the satisfaction of an operationalist methodology, which Bronstein shares with numerous other physicists working on the problem of localitation at high energy at that time.[10] In this context, the term "adequate" is used to describe a probing procedure that yields a precise measurement of the quantity under scrutiny. Consequently, **(B2)** introduces an additional constraint on the acceptability of possible conclusions. Should a conclusion violate operationalism, it would be rejected in order to maintain the consistency of the scenario of the TE.

The TE invites the reader to imagine performing a measurement of the Christoffel symbols of the gravitational field. The Christoffel symbols encode information regarding the curvature of spacetime, thereby providing insight into the gravitational effects within a general relativistic framework. All additional details are irrelevant; the TE is conducted in a vacuum, wherein the sole objects under consideration are the classical gravitational field (the target of the imagined measurement, described by GR according to **(B1)**) and a quantum probe. Such a probe is a particle that is sent into a region of interest and subsequently detected at the end of the experiment. This process enables the indirect acquisition of information regarding the properties of spacetime within that region.

(B2) necessitates the properties of the probe to enable an extremely rapid measurement of its momentum. This requires the minimisation of two uncertainties affecting the particle. The first term stems from Heisenberg's uncertainty principle, which arises from the measurement of the probe's momentum. The second uncertainty is attributed to the recoil of the probe subsequent to its interaction with the gravitational field. In order to minimise these two uncertainties, it is necessary to increase the mass of the probe. This is permissible within the context of a weak gravitational field, wherein spacetime is governed by special relativity. In a strong gravitational regime, however, the dominant theory is GR. In particular, GR imposes a bound on the mass density (ρ) of the probe based on its volume (V):

$$\rho \lesssim \frac{c^2}{G_N V^{2/3}}, \tag{1}$$

where c is the speed of light and G_N is Newton's gravitational constant.

It can be shown that the increase of mass required to eliminate the uncertainties exceeds the upper bound. As a result, GR predicts that spacetime will react by forming a black hole in the region of interest, thereby preventing information from escaping the event horizon. This outcome invalidates the operationalist motivation of

[10] For an extensive investigation into this methodology, see Hagar (2014).

(B2), as precise data regarding the value of the Christoffel symbols in that region cannot be transmitted to the observer. Conversely, if the aforementioned bound is not exceeded, this result will inevitably be uncertain, thereby violating the primary motivation behind the TE.

In conclusion, the TE gives rise to a paradox due to the conjunction of three constraints: operationalism (as required by **(B2)**); strong gravity (as required by a choice of setting); and the arbitrary extension of GR (as required by **(B1)**). This results in a failure of operationalism and an incompatibility between the quantum nature of the probe and the classical description of spacetime.[11] In this sense, Bronstein's TE is *destructive*, as it illustrates the inconsistency of special combinations of constraints.

3.2 Assessing the conclusion

Bronstein's TE demonstrates that a fundamental quantity bounds the regime of applicability of GR. This bound plays a pivotal role in the TE, although it is never calculated explicitly. Instead, it can be evaluated through the comparison with other analogous TEs produced between the 1930s and the early 1990s and identified with the Planck length. This fundamental scale is demonstrated to coincide with the Schwarzschild radius of the black hole, that is to say, the size of the expanding probe beyond which the mass collapses and forms a singularity.

It is noteworthy that the TE bears resemblance to Heisenberg's microscope experiment (see Heisenberg 1927). This parallel is substantiated by the recent literature.[12] Both TEs rely on operationalist assumptions with analogous settings and both address the uncertainty of the probes. However, the conclusions that are reached are different. Bronstein emphasises a limitation of the classical theory of gravitation, whereas Heisenberg is concerned with the limitations of the classical theory of particles. Furthermore, the uncertainties involved in Bronstein's setting are not limited to the quantum nature of the probe, as in Heisenberg's, but have also gravitational origin.

Stuart (2016) offers a commentary on Heisenger's TE, arguing that its objective is to underscore a limitation of the classical theory in defining localisation within quantum regimes. The TE, therefore, has the potential to facilitate a more profound understanding of the behaviour of systems at these scales. The extent to which our understanding of this uncertain behaviour is enhanced by this approach is unclear.[13] Nevertheless, it is indisputable that *some* understanding is gained by reproducing this TE, even in an imaginary setting. A similar conclusion may be reached in relation to Bronstein's TE, which may facilitate a comparable deeper understanding of the spacetime structure at high-energy regimes.

It is also noteworthy that Bronstein's result extends beyond the scope of a single theory. The closest analogy to Heisenberg's case is to argue that Bronstein's TE helps

[11] A similar argument has been made in the opposite scenario, namely when a classical probe interacts with a quantum spacetime. In this case, the main problem addressed by the TE is the possibility of having sharp localisation of systems within regions of arbitrarily small size. The paradox suggests that classical probes are inadequate to localise systems in quantum spacetime, hence they should be replaced by *quantum* probes. See Doplicher, Fredenhagen, and Roberts (1995).

[12] See e.g. Lizzi (2019); Maresca (2015), and the similarities of Bronstein's TE with Mead (1964) and Doplicher et al. (1995).

[13] See e.g. Hilgevoord and Uffink (2016) on the interpretations of the Heisenberg uncertainty principle.

to facilitate a more nuanced understanding of GR.[14] This is evidenced by premise **(B1)**. By extending operationalism as a metatheoretical constraint (via **(B2)**) and extending the domain of applicability of GR to new scales, the TE demonstrates that the new systems cannot be meaningfully described by GR. This establishes a lower bound to the domain of applicability of the former theory. However, this interpretation of the conclusion of the TE provides no details regarding the structure of spacetime in *high*-energy regimes, which after all was the primary goal of Bronstein. In light of the original objective of the experiment, it is unclear how further understanding could be gained without extending the conclusion.[15]

While Bronstein may be delineating precise thresholds within the domain of GR, claiming that this was the precise goal of his TE would be an overstatement of the conclusion. The TE does not provide a precise value for the fundamental bound, and there are several alternative methods for circumventing the paradoxical scenario, each stemming from a slightly different interpretation of the conclusion. Of greater significance is the absence of epistemic evidence that would allow for the preferential selection of one option over another, given the current lack of empirical support for or against any alternative theory of QG.[16]

It is crucial to emphasise that Bronstein's TE does not provide guidance on which assumption should be dismissed to circumvent the conclusion. Three main options can be identified.

Firstly, one may reject **(B1)** and instead limit the domain of GR to low energies. This option necessitates the development of a novel theory of spacetime at high energies that accounts for quantum effects. Put differently, it brings forth the construction of a theory of QG.

Secondly, **(B2)** may be rejected. This would render operationalism unsuitable as a theoretical framework at high-energy scales. Consequently, this option would entail the identification of alternative criteria of physicality and the investigation of the persistence of the paradox in a different metatheoretical framework.

Thirdly, operationalism may be restricted to low energies, where the paradox does not arise. This would be the case, for example, in the context of special relativity. Conversely, operationalism would be untenable in the vicinity of the Planck length, as this defined the domain of classicality.

It is important to note that these options are not mutually exclusive, but the first can be considered in conjunction with either the second or the third. Furthermore,

[14] The extension of this result to an interpretation of GR is an open question: see e.g. Curiel (2009).

[15] An interesting parallel here is provided by the Einstein-Bohr debate concerning the double-slit TE. (Lupacchini, 1995, 158–168) highlights that this debate reveals two different attitudes towards the use of TEs in physics. On the one hand, Einstein assumes a set of principles and builds the TE in order to display their contradictions. Consequently, the TE is destructive and intended to motivate research for an alternative solution. On the other, Bohr builds on a preliminary phenomenological analysis of concepts and designs his TEs to illustrate experimental settings which may be in accordance or not with the initial set of principles. Consequently, his TEs are constructive, insofar as the experimental setting substantiates the new principle. If this conclusion does not follow, then the experimental setup is deemed inadequate with respect to the theoretical framework. In this context, Bronstein's TE seems closer in spirit to Einstein, whereas, as will be discussed, any attempt at converting his conclusion to a constructive TE may align with Bohr's attitude.

[16] Although several approaches have tried to tie the theory back to experimental situations by pointing out how they would be explained, were their theory of QG to be true. This is especially the case in the field of QG-phenomenology.

additional options may be identified. Historically, different approaches to QG have responded to these kinds of arguments by advocating different clusters of them.

4 Imagination at work

As outlined in Section 2.1, the epistemological status of several forms of imagination remains an open issue within the epistemology of science. Nevertheless, TEs in frontier physics solicit the identification of appropriate types of imagination for the derivation of their conclusions. This is a primary concern to address, yet each proposal is already open to questions. In particular, an answer must align with the TE's goal and narrative. Furthermore, it must be compared with the relevant literature to assess its epistemic gains and weaknesses.

Bronstein's TE constitutes an interesting case study in this regard. Section 4.1 delineates the contribution of imagination in the experiment, as opposed to the role of mere supposition. Section 4.2 then proposes to analyse Bronstein's TE as an example of operational imagination, emphasising its operative nature and bringing the TE closer to simulative accounts of scientific imagination. Finally, Section 4.3 reconciles operational imagination with the imposition of constraints on its use in the TE.

4.1 Imagination *vs* supposition

A fundamental concern with Bronstein's TE, as with many others, is whether it truly relies on the use of imagination. In other words, a philosophical investigation raises the question of whether imagination is epistemically significant for reaching the conclusion of the TE. The rejection of one of the constraints at the conclusion of the experiment, unless one is to fall into a paradox, may indeed prompt a reinterpretation of the entire setting in accordance with the notion of *supposition*, as opposed to imagination. In this section, I address the issue and argue that interpreting Bronstein's TE in terms of supposition is inadequate.

The so-called *common nature thesis* (CNT) posits that imagination and supposition represent distinct instances of the same cognitive capacity. However, while these faculties may share a common origin, this does not inherently entail a similar use. In fact, Jackson (2016) has contended that imagination and supposition fulfil distinct epistemic roles. He posits that the CNT is invalid in regard to imagination and supposition, as the latter does not necessitate the simulation of belief, in contrast to the former. In other words, supposition permits a sentence p to be accepted as true temporarily for a specific purpose. This enables the individual to engage in an inference-making task, which ultimately justifies the belief in a final conclusion that has already been discovered but is not yet justified.[17] This is achieved *without explicitly committing to believing p's truth.*

In contrast, imagination is characterised by a certain simulative quality. When an individual engages in imaginative activities, she interacts with the image in question as if it were true, thereby simulating the corresponding belief. This allows her to control and manipulate the content of her imagination, facilitating the execution of the TE.

[17] Paradigmatic examples of supposition are proofs by contradiction. In this case, one wants to prove that a sentence p holds. To this end, she *supposes* that $\neg p$ and derives, by contradiction, p.

The conclusion of the TE is thus reached and justified on the basis of the constraints and setting, rather than being known prior to their investigation.

In Bronstein's TE, the conclusion is a *negative* statement: there is an incompatibility between the constraints, namely **(B1-2)**, and strong gravity. In other words, the TE yields an *impossibility result*. Should a *positive* conclusion be sought, a number of potential investigative paths can be pursued, yet, as previously outlined, the experiment points to none in particular. In other words, the extension of Bronstein's TE to a *constructive* one is not unique.

Consequently, it appears that positing the premises as suppositions does not align with Bronstein's original intent. The negative conclusion has not been derived from other arguments; thus, it is not possible to suppose that one of the three premises is true, contrary to previous knowledge, in order to arrive at its paradoxical consequences. Instead, we imagine a scenario with constraints that are arguably plausible and demonstrate how these constraints lead to the negative conclusion. The conclusion is justified by the derivation itself, in contrast to the case of supposition.

Nonetheless, one may still attempt to reverse the TE. It is not possible for the three constraints to be compatible; however, *suppose* that they could. This possible inquiry appears promising for further investigation. However, a potential TE of this nature would diverge from Bronstein's original one and from his intention, as previously outlined. It is important to recall that Bronstein proposed his TE in an attempt to make his mathematical insight physically perspicuous. He highlighted the necessity for a theory of *quantum* spacetime by developing a fictional scenario (i) that exemplifies the mathematical structures underlying spacetime physics and (ii) where the incompatibility of constraints can be elucidated through an imaginative, thought-experimental process.

Supposition rather pertains to the possible extensions of GR to high-energy theories of QG. In that case, it is unclear what form this theory would take. In this context, supposition and imagination cooperate in order to investigate an unknown regime, where certain constraints of our currently successful physical theories may be violated, as anticipated by some physicists.[18]

It is noteworthy that Salis and Frigg (2020) also distinguish supposition from a specific instance of imagination, namely Waltonian make-believe. In their view, TEs fall under the latter category, but share with supposition their origin as instances of imagination. The authors trace the distinction between supposition and make-believe back to the use of props. This alternative may help to reconcile the aforementioned divide in Bronstein's case. Nevertheless, it also requires the advocate of Salis and Frigg's proposal to provide a more detailed account of the type of imagination involved in the TE. In other words, does the TE constitute an instance of supposition or one of make-believe? Put in yet another way, this alternative still permits a hypothetical objector to reiterate her argument that imagination may *not* be implicated.

[18] For example, Lorentz-invariance has been questioned based on its incompatibility with the introduction of a fundamental length scale. See Maresca (2015) for a reconstruction of the problem.

4.2 Operational imagination and mental models

At first glance, Bronstein's TE can be considered as a particular instance of propositional imagination. In essence, the experiment is rehearsed mathematically, whereby an appropriate description of the scenario is defined and the relevant equations are followed. The execution of the experiment is not informed by any mental image of the actual performance, nor is it influenced by any reference to perceptual experiences, such as hearing, touching, etc.

While a propositional reading of the TE may appear plausible, at first, I contend that it is ultimately insufficient and overly broad in scope. This approach is inadequate for explaining the fundamental constraint of the TE, namely the operationalist requirement. Indeed, confining the epistemic content of the experiment to its propositional aspect at best makes the operationalist requirement a constraint on a par with the others. However, Bronstein clearly places a higher value on the operationalist requirement, regarding it as some kind of "higher principle" or "meta-principle." The propositional reading is unable to adequately account for the narrative's emphasis on the experimental performance and the selection of an appropriate setup. The crucial aspect is not *that* the probe has this and that property, or *that* its position is uncertain; rather, it is *how* the probe interacts with the gravitational field and *how* the uncertainty can be dealt with by increasing the mass. In summary, the propositional reading is too broad to make any specific claim about the functioning of this argument. If there are propositional elements in the TE, which could be easily accepted, they cannot do all the work.

Conversely, a phenomenal, non-propositional account of imagination also fails to address the core issue. It is possible to imagine that the probing particle is blue or red; however, these perceptual features are irrelevant for reaching the conclusion. Furthermore, the experiment does not make any mention of a human or human-like observer. As beings capable of perception, we do not engage with the execution of the experiment within the fictional world, unless one wishes to revisit long-gone debates about observers and consciousness. The sole reference to an observer is directly linked to the operationalist assumption. Nevertheless, this is a remarkably broad conception of observer, as it can readily accommodate a secondary measurement apparatus capable of detecting the probing particle post-experiment and displaying the momentum value of the particle on a screen. This ability to detect the uncertainty carried by the probe after its interaction with the spacetime region is all that is required by the TE to reach the conclusion.

Finally, there is a possibility that the TE engages in objectual imagination, which, I maintain, is an overly restrictive interpretation. Naturally, visualising the rehearsal of the TE has a certain degree of cognitive value, as it helps the individual to gain a deeper understanding of the process. The creation of a mental image of the trajectory of the probe, represented for example by a series of coloured lines on a screen, facilitates a more accurate visualisation of the TE. However, it is not the *visualisation* component of this mental image that facilitates understanding of the conclusion; rather, it is the fact that, in principle, we are unable to form such a visual representation of the trajectory, unless we contradict the conclusion of the TE. The position uncertainty precludes us from doing so.

In light of the aforementioned considerations, I suggest to regard Bronstein's TE as an instance of a novel form of non-propositional imagination, which may be termed *operational imagination.* This is set forth in two stages. Firstly, it involves the cognitive ability to imagine and design a mental setup for an experiment (*design function*). This setup is distinguished from the set of principles and constraints in that, whereas the former involves the disposition and establishment of the properties of the pieces of the experiment, the latter defines what is nomologically possible in that scenario and what is not. Furthermore, the design of the setup may exclude certain details and specifics that are deemed irrelevant for the TE. In other words, it can "quarantine" specific aspects of the real world without concern for the reality principle.

Secondly, operational imagination involves the capacity to imagine the progression and performance of the experiment's operations, step by step, ultimately resulting in the desired outcome (*performance function*). This also permits intervention on the imagined parameters of the experiment through the implementation of arbitrary modifications. To illustrate, the momentum of a particle may be modified to accommodate adjustments to the experiment, or the curvature of a specific region of spacetime may be increased to prompt the particle to generate *ad hoc* singularities within the scenario of the TE.

Operational imagination differs from *manipulative imagination*, in that the latter is concerned solely with the performance of operations in the experiment, whereas the former also encompasses the preparation of the setup. Moreover, operational imagination differs from *interventionist imagination*, in that it is not necessarily linked to counterfactual scenarios. Additionally, operational imagination is distinguished from *abductive imagination*, in that it is not exploratory; it does not imagine a solution to an established problem (a backward-looking process), but rather is forward-looking. Finally, operational imagination is distinct from propositional accounts, in that it involves the active imagining of the preparation and performance procedures of the experiment, whereas propositional imagination is better conceived of as tracking its different steps.

Moreover, operational imagination is compatible with TEs that purport to demonstrate the ultimate impossibility of certain scenarios in reality. This is exemplified by Bronstein's TE. The experiment may be interpreted as ultimately suggesting that operationalism in high-curvature regimes is meaningless and does not properly apply. Therefore, any suggestion to operationalise within that regime should be rejected outright. Nonetheless, in order to arrive at its paradoxical conclusion, the TE does invite us to imagine the possibility of such operations being performed. The absence of material constraints allows for the step-by-step execution of imaginary operations and ultimately demonstrates their inconsistency with the constraints. This can be employed as a means of establishing the impossibility results, despite the potential self-defeating nature of operationalism.

At this juncture, it is necessary to compare the use of operational imagination with the alternative accounts of TEs presented in Section 2.1. The primary objective here is to ascertain whether these accounts can accommodate operational imagination as an epistemically significant concept for Bronstein's TE.

It can be argued that a Waltonian account can be developed in this direction. In its *design function*, operational imagination is concerned with the selection of appropriate mental constructions, which may be conceived of as props. To illustrate, the probe may be conceived of as a prop endowed with a specific set of properties, such as being a quantum object or being able to attain progressively greater energy without divergence in the underlying equations. Spacetime will be another prop with classical features described by GR (due to constrain **(B1)**). The experiment will be conducted in a fictional scenario, wherein the probe will be isolated in a vacuum, interacting solely with the specified spacetime region.

In this setting, the constraints and laws of the theories involved (QM for the probe, GR for spacetime) serve as principles of generation. In accordance with these principles, the TE can be performed in its entirety. The experimental steps may involve increasing the mass of the probe or detecting the final uncertainty in order to make sense of the operationalist constraint; any such step falls under the *performance function* of operational imagination. In other words, this type of imagination enables the implementation of abstract principles of generation into mental manipulations of the props. The TE demonstrates that, in accordance with the established laws and constraints, the act of imagining specific operations and properties of the props in question contravenes the aforementioned constraints. The epistemic reach of the TE lies in the actualisation of these operations within the experimenter's cognitive domain, up to the requisite level of details.

It is important to note that this Waltonian account is subject to two significant limitations. Firstly, the association of the props with material systems is weak. It would be exceedingly difficult, if not impossible, to isolate a probe with the requisite properties. Furthermore, it would be unfeasible to increase its mass up to the desired value. Secondly, the Waltonian account is typically articulated in terms of propositional imagination, which, as previously argued, is too general to account for the specifics of Bronstein's TE.

I contend that the simulative model-based account put forth by Nersessian is more promising to study Bronstein's experiment. This account is sufficiently comprehensive to encompass the design function of operational imagination, as well as to facilitate the execution of all operations falling within the performance function. Indeed, once the mental model has been constructed, it can be manipulated so that it aligns with the conclusion of the TE, with the narrative providing the necessary constraints.

Notably, these constraints originate from already accepted theories, namely GR and QM, as evidenced in Bronstein's original paper. The discrepancy between the mental model and the actual physical system is permitted by the simulative nature of the former: the probe and the spacetime region can be isolated from further interactions with the rest of the fictional environment. Consequently, additional negligible information is not required for the TE. Furthermore, the non-iconic account of non-propositional imagination accounts for the operational imagination involved in Bronstein's TE, while avoiding the aforementioned criticisms.

4.3 Internal validity and coherence

Nersessian's simulative model-based account facilitates the focus on the "operational" aspect of the TE. Following the construction of the mental model of the setting, the TE progresses by executing operations and transformations on the setting and deriving a conclusion from its final configuration. At each operational stage, the setting undergoes a transition from one configuration to another. To elucidate the stages of Bronstein's TE, I suggest the following classification.

Firstly, it is possible to isolate a set of *principles of the setting*. These are conditions pertaining to the possible configurations of the setting. A subset of these principles includes the *constraints* on the possible configurations. These constraints are designed to restrict the range of operations that can be mentally performed on the setting at each stage. The setting must transition from one permitted configuration to another, and thus must always satisfy the constraints. In Bronstein's case, the setting must always satisfy **(B1-2)**.

An additional subset of these principles is predetermined by the initial configuration of the system. To illustrate, Bronstein's TE necessitates the existence of probes and spacetime regions, as well as the location of the TE in a regime of strong gravity.

Secondly, it is possible to identify a set of *rules of experimenting*. These principles resemble those of the Waltonian accounts. The rules of experimenting delineate the methodology for performing the experimental steps within the TE. In particular, they illustrate how the setting can evolve from one possible configuration to another, contingent on the satisfaction of the constraints. It is important to note that these rules of experimentation must explicitly incorporate the act of performing an operation; otherwise, they are inconsistent with the underlying operational imagination that is being employed.

To illustrate, in Bronstein's TE, a rule of experimenting states that the mass of the probe can be increased arbitrarily. Consequently, it permits the transition from a low-energy to a high-energy configuration of the setting. Similarly, another rule posits that the probe can be detected and its momentum measured, although no specification regarding the sharpness of the outcome is necessary (otherwise the experiment is automatically invalid).

Finally, the TE requires the introduction of a set of *laws* derived from the theories under examination. In Bronstein's case, these are the laws of GR and QM, which regulate the evolution of the systems under examination at each configuration of the setting.

It is noteworthy that, in contrast to the rules of experimenting, the laws do not require operational imagination to be applicable; rather, they function as instances of propositional imagination. Following the application of the rules of experimenting, the system is allowed to evolve according to its laws before applying the subsequent rule. To illustrate, the laws of Bronstein's TE may be defined as follows: an increase in the energy of the probe is equal to an increase in its mass; additionally, a mini-black hole is formed beyond the Schwarzschild radius.

This distinction provides greater insight into the relationship between constraints, rules of experimenting and laws. Constraint **(B1)** imposes the laws of GR on the setting at each stage, whereas the laws of QM originate from the initial configuration of the setting, which includes a quantum probe. Furthermore, the setting must permit a rule of experimentation according to which the system can evolve from a low-energy to a high-energy configuration as the mass increases. This is a crucial aspect of the TE. The absence of this rule renders constraint **(B1)** trivial in a low-energy, special relativistic setting, thereby preventing the formation of a black hole (or the minimisation of the gravitational uncertainty).

Conversely, constraint **(B2)** restricts the possible designs of the setting to those that satisfy the operationalist requirement. In light of this, it can be posited that the setting must include a probe. Therefore, the presence of a probe in the initial setting necessitates the establishment of its own set of rules of experimenting that govern functioning.

In its final stage, Bronstein's TE considers a *fixed* configuration of the setting, specified by its high-energy regime. In this configuration, an incompatibility of laws is established. This incompatibility of laws in turn translates into an incompatibility of the original constraints, since the latter allow for inconsistent settings. Specifically, the latter are such that either (i) the uncertainty is completely eliminated by arbitrarily increasing the mass of the probe, thus satisfying the operationalist requirement but violating GR, or (ii) a black hole is formed, violating the operationalist requirement while GR is satisfied. In this sense, the TE is destructive: not of the laws, but of the *conjunction of constraints*. By following the evolution of the setting from its initial allowed configuration to the final one, it is diagnosed that the original set of constraints is *inconsistent*, i.e., the TE exhibits a case of *internal inconsistency*. This diagnosis provides a rationale for the range of possible responses to the conclusion, as outlined in Section 3.2: all proposals reject or reformulate either **(B1)** or **(B2)**.

Two specifications are necessary. First, I speak of internal consistency (or lack thereof) as opposed to internal *validity*. As Sartori (2023) correctly points out, there is no agreed notion of internal validity for a TE. This is a significant point of contention between different positions, whether they seek to reduce TEs to arguments or not. Sartori proposes a Waltonian characterisation of internal validity, whereby the steps of the TE must proceed from primitive to derived truths, or from derived to derived truths, by means of principles of generation. In contrast to the situation previously outlined, this condition is imposed on the *steps* of the TE, rather than on its initial design. In the classification established above, it is a condition imposed on the rules of experimenting and the laws of the TE, rather than one imposed on the constraints. Consequently, the issue of internal validity does not arise in Bronstein's TE; the problem is more profound.

Second, the paradoxical conclusion may not be due to an inconsistent set of constraints, but rather due to an issue with the rules of experimenting. Admittedly, at least one of these rules is unrealistic, namely, allowing an arbitrary increase in the mass of the probe. This observation alone may suffice to cast reasonable doubt on the adequacy of the rules of experimenting with respect to their ultimate goal.

While I agree that arbitrarily increasing the mass of a probe is unrealistic, I do not believe that this constitutes an issue for the TE. Put in another way, the proximity of the rules of experimenting to real-world scenarios does not impinge upon the internal consistency or validity of the TE. Conversely, it must be carefully considered for the replicability of the TE as a material experiment, i.e., as an issue of *external* validity.

External validity is frequently discussed as a necessary condition for TEs to acquire physical significance, a point particularly emphasised by the Neokantian account within the epistemology of TEs. This account claims that TEs work as anticipations of material experiments. The fundamental difference between this account and its alternatives lies concerns the objectives: Neokantians focus on the possibility of conducting material experiments, while the others are frequently concerned with the issue of the actual realisation of the TEs. In this sense, Neokantians regard TEs as paradigmatic cases of counterfactual reasoning, and external validity imposes the resuirement that the results of a TE coincide with those obtained by the execution of a material experiments. This relation is summarised, for example, by (Buzzoni, 2010, 15; 18):

> *A thought experiment is both the project in thought of a real experiment that is in principle realisable, and the linguistic-discursive anticipation of nature's specific answer to the question implicit in that real experiment.* [... A]*ll thought experiments must be thought of as translatable into real ones, and all real experiments as realisations of thought ones. Thought experiments are conceivable as preparing and anticipating real ones: they anticipate a connection between objects which, when thought of as realised, makes the concepts of experiment and thought experiment coincide completely.* (italics in the original)

Whilst concurring that external validity is a crucial aspect for many TEs that have been proposed over time, I contend that its characterisation as necessary for the epistemic significance of a TE is overly restrictive. This necessity can be questioned in two contexts.

Firstly, external validity may be impossible to assess. This is the case of TEs that involve increasing degrees of idealisations. The consequence of this idealisation is the absence of substantial guidance in the realisation of a material implementation of the TE. In such instances, the challenge in assessing external validity between the TE and its material implementation has no bearing on the epistemic support provided by the TE's setting for its intended conclusion.

Secondly, external validity is inapplicable in cases where there is no empirically accessible counterpart to the domain of application of the TE. If the intended regime is inaccessible by material experimental means, then it is not possible to assess whether the conclusions of material and TE coincide, since we were never capable of performing the material experiment in the first place. Instead, the TE is intended to *replace* a possible material experiment in response to the impossibility of the latter. While the legitimacy of this research task may be contested, it is evident that numerous investigations in QG employ TEs as epistemic resources to a significant extent, given the absence (by definition) of material counterparts. Consequently, confining TEs exclusively to anticipations of material experiments, and thereby imposing external validity as a prerequisite for their epistemic significance within scientific research, is a restrictive perspective.

To further illustrate this point, it is noteworthy that Bronstein does *not* address external validity. Consequently, once it is agreed that external validity should be investigated, one could attempt to replicate the TE as a material experiment and look for discrepancies. It is only in the material experiment that we will have a constraint on the arbitrary increase in the mass of the probe, and hence a limitation on the rules of experimenting, if the TE is to be externally valid. However, this is left as future work, and the very possibility of finding such material experiment can be questioned on physical grounds.

5 Three possible objections

In the preceding sections, I proposed an interpretation of Bronstein's TE based on the notion of operational imagination. In particular, I emphasised similarities and differences with both the Waltonian account and Nersessian's simulative model-based account. In this concluding section, I will address three potential objections to my analysis.

Firstly, the definition of operational imagination may be criticised due to its generality. It may be contended that operational imagination is epistemically significant within a TE only if it is associated with an appropriately detailed specification of the setup. In this sense, the performance function is, in itself, insufficient for supporting the conclusion of the TE under examination. This specification is required to account for the methodology of the experiment, the details of the setup, and so forth. In Bronstein's case, for instance, one might inquire about the type of particle employed as a probe. The fewer specifications in the setup of the TE, the more distant it becomes from material experiments, thereby diminishing the epistemic significance of the conclusion.

In response to this objection, it can be contended that such a requirement of specificity is unduly stringent and rarely applicable to well-accepted TEs. It is not uncommon for TEs to rely on approximations and idealisations included in the preparation of the setup, given that some information may be irrelevant or counter-effective with respect to the fictional scenario. This is also evidenced, for example, by Salis and Frigg's quarantining or by specific conditions that Nersessian imposes on the appropriate narratives.

Moreover, in the case of Bronstein's TE, the information regarding the type of particle employed is redundant. Indeed, the specifics contribute nothing to the impossibility result that the TE aims to conclude; they are irrelevant in terms of achieving the internal consistency of the constraints. Conversely, the reach of the TE is guaranteed by its generality, as it attempts to tackle a universal problem concerning the compatibility between GR and QM at high energies under operationalist assumptions. It is evident that the specification of the type of particle employed as a probe would serve to restrict the scope of the TE, in contrast to its original objective.

A second potential objection may be raised by the advocates of Norton's account. It is indeed possible to question whether internal consistency is not merely an euphemism for the logical consistency of the premises of an argument. Accordingly, the manipulations of the mental setup in Bronstein's experiment would be reconstructed as logical steps in a stream of reasoning.

However, it must be noted that a purely argumentative reconstruction would fail to achieve the objective of the TE. Norton's position effectively reduces the epistemic use of the TE to that of an argument. This is incompatible with Bronstein's original goal, which, as previously stated, is to elucidate the physical picture underlying the mathematical construction of a quantum theory of spacetime. It is open to question whether arguments contribute to the understanding of a physical theory. While one may make a case for their contribution, it is important to note that understanding may demand a more nuanced approach than a straightforward delineation of the logical progression of a physical theory. Furthermore, as previously highlighted, the application of rules of experimenting bears only a vague resemblance to the use of admissible inference rules in an argument. Such a reduction would, in fact, eliminate any reference to operational imagination, thereby rendering the main epistemic endeavour behind Bronstein's manipulation of the setup meaningless.

Thirdly, it may be possible to transform Bronstein's TE (a destructive experiment) into a constructive experiment, concluding with a conjecture to be further explored. While this conversion is undoubtely feasible, the identification of a positive conclusion is a contentious matter. As discussed at the end of Section 3.2, there is no consensus among physicists on how to develop the conclusion into a positive conjecture for investigation. It is the burden of the objector to formulate this positive conclusion. Furthermore, such a positive conjecture would not have the same epistemic reach. It is unclear how this constructive conclusion would contribute to the original goal of the TE in question, especially in light of the controversies between approaches to QG.

6 Conclusion

The existing literature on TEs and the epistemic function of imagination is extensive, yet the absence of consensus on a common ground remains evident. A substantial portion of the literature has sought to generalise specific findings in order to characterise the entire class of TEs. However, this approach has frequently been shown inadequate. The diversity of TEs proposed throughout the history of science appears to contradict any attempt to find a common ground. Indeed, it is even open to question whether the cognitive and epistemic processes that underpin their conclusions are identical in all cases.

Conversely, it would appear that different kinds of TE require the use of imagination as an epistemic means of reaching a conclusion in different ways. This finding aligns with the recent epistemological shift in philosophy of science towards *pluralism*. As Murphy (2020) has argued, TEs appeal to a variety of imaginative capacities and invite different types of cognitive activity. The plurality of functions associated with a given TE evokes a spectrum of imaginative engagements and diverse imaginative capacities. A pluralistic approach is the most appropriate to account for this. I propose that those TEs, which purport to contravene a set of constraints by demonstrating reciprocal inconsistency, are distinct from other forms of constructive TEs and point towards pluralism.

Bronstein's TE is an example of such destructive TEs. In this case, I argued that the execution of the experiment involves a novel form of non-propositional imagination, termed operational imagination, which aligns well with simulation model-based accounts. In contrast, the epistemic content of the TE cannot be reduced to arguments or to propositional imagination. The experiment consequently arrives at a paradoxical conclusion, whereby the laws of GR and QM appear to contradict each other. This TE therefore demonstrates the inherent inconsistency of the set of constraints from which these laws were derived. It is reasonable to hypothesise that many other similar examples will be found over time.

Acknowledgements. I am grateful to the participants of the conference *Metamorphosis of Imagination* for their feedback on an earlier draft of this paper. I am especially grateful to Lorenzo Biagini, Fedele Lizzi, and Rossella Lupacchini for comments on the paper and discussions on the topic.

References

Bronstein, M. (2012/1936). Republication of: Quantum theory of weak gravitational field. *General Relativity and Gravitation*, *44*, 367–383, https://doi.org/10.1007/s10714-011-1285-4

Brown, J.R. (2011). *Laboratory of the mind: Thought experiments in the natural sciences (second edition)*. London: Routledge.

Brown, J.R., & Fehige, Y. (2023). Thought experiments. E.N. Zalta (Ed.), *The stanford encyclopedia of philosophy (sep)*. Retrieved from https://plato.stanford.edu/entries/thought-experiment/

Buzzoni, M. (2010). Empirical thought experiments: A transcendental-operational view. *Epistemologia*, *33*, 5–26,

Crowther, K. (2021). Defining a crisis: the role of principles in the search for a theory of quantum gravity. *Synthese*, *198*(Suppl 14), S3489–S3516, https://doi.org/10.1007/s11229-018-01970-4

Curiel, E. (2009). General relativity needs no interpretation. *Philosophy of Science*, *76*, 44–72, https://doi.org/10.1086/599277

Dennett, D. (1996). Intuition pumps. J. Brockman (Ed.), *Third culture: Beyond the scientific revolution* (pp. 181–197). New York: Simon and Schuster.

Doplicher, S., Fredenhagen, K., Roberts, J.E. (1995). The quantum structure of spacetime at the Planck scale and quantum fields. *Communications in Mathematical Physics*, *172*, 187–220, https://doi.org/10.1007/BF02104515

El Skaf, R. (2021). Probing theoretical statements with thought experiments. *Synthese*, *199*, 6119–6147, https://doi.org/10.1007/s11229-021-03063-1

Eppley, K., & Hannah, E. (1977). The necessity of quantizing the gravitational field. *Foundations of Physics*, *7*(1/2), 51–68, https://doi.org/10.1007/BF00715241

Frigg, R. (2010). Fiction and scientific representation. R. Frigg & M. Hunter (Eds.), *Beyond mimesis and convention: Representation in art and science* (pp. 97–138). Dordrecht: Springer.

Frigg, R., & Nguyen, J. (2020). *Modelling nature: An opinionated introduction to scientific representation.* Cham: Springer.

Gendler, T.S. (2004). Thought experiments rethought–and reperceived. *Philosophy of Science*, *71*, 1152–1164, https://doi.org/10.1086/425239

Hagar, A. (2014). *Discrete or continuous? the quest for fundamental length in modern physics.* Cambridge: Cambridge University Press.

Hempel, C.G. (1966). *Philosophy of natural science.* Upper Saddle River: Prentice Hall.

Hilgevoord, J., & Uffink, J. (2016). The uncertainty principle. E.N. Zalta (Ed.), *The stanford encyclopedia of philosophy (sep).* Retrieved from https://plato.stanford.edu/entries/qt-uncertainty/

Jackson, M.B. (2016). On the epistemic value of imagining, supposing, and conceiving. A. Kind & P. Kung (Eds.), *Knowledge through imagination* (pp. 41–60). New York: Oxford University Press.

Kind, A. (2018). How imagination gives rise to knowledge. F. Macpherson & F. Dorsch (Eds.), *Perceptual imagination and perceptual memory* (pp. 227–246). New York: Oxford University Press.

Kind, A., & Kung, P. (2016). Introduction: The puzzle of imaginative use. A. Kind & P. Kung (Eds.), *Knowledge through imagination* (pp. 1–37). New York: Oxford University Press.

Levy, A. (2012). Models, fictions, and realism: Two packages. *Philosophy of Science*, *79*(5), 738–748, https://doi.org/10.1086/667992

Levy, A., & Godfrey-Smith, P. (2020). Introduction. A. Levy & P. Godfrey-Smith (Eds.), *The scientific imagination: Philosophical and psychological perspectives* (pp. 1–16). New York: Oxford University Press.

Lizzi, F. (2019). *Points. Lack thereof.* Retrieved from https://arxiv.org/abs/1905.01653

Lupacchini, R. (1995). *Esperimenti mentali e realtà fisica.* Bologna: CLUEB.

Maresca, E. (2015). Noncommutative geometry and spacetime: A historical reconstruction. *Journal of Physics: Conference Series*, *2948*, 012011, https://doi.org/10.1088/1742-6596/2948/1/012011

Mattingly, J. (2006). Why Eppley and Hannah's thought experiment fails. *Physical Review D*, *73*, 064025, https://doi.org/10.1103/PhysRevD.73.064025

McAllister, J. (2013). Thought experiment and the exercise of imaginatiion in science. M. Frappier, L. Meynell, & J.R. Brown (Eds.), *Thought experiments in philosophy, science, and the arts* (pp. 11–29). London and New York: Routledge.

Mead, A.C. (1964). Possible connection between gravitation and fundamental length. *Physical Review*, *135*(3B), B849–B862, https://doi.org/10.1103/PhysRev.135.B849

Miščević, N. (1992). Mental models and thought experiments. *International Studies in the Philosophy of Science*, *6*, 215–226, https://doi.org/10.1080/02698599208573432

Murphy, A. (2020). Towards a pluralist account of the imagination in science. *Philosophy of Science*, *87*(5), 957–967, https://doi.org/10.1086/710620

Murphy, A. (2022). Imagination in science. *Philosophy Compass*, *17*, e12836, https://doi.org/10.1111/phc3.12836

Murphy, A. (2024). Imagination and creativity in the scientific realm. A. Kind & J. Langkau (Eds.), *Oxford handbook of philosophy of imagination and creativity.* New York: Oxford University Press.

Nersessian, N.J. (1992). In the theoretician's laboratory: Thought experimenting as mental modeling. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, *2*, 291–301,

Nersessian, N.J. (2017). Cognitive science, mental modeling, and thought experiments. M.T. Stuart, Y. Fehige, & J.R. Brown (Eds.), *The routledge companion to thought experiments* (pp. 309–326). London: Routledge.

Norton, J.D. (1991). Thought experiments in Einstein's work. T. Horowitz & G. Massey (Eds.), *Thought experiments in science and philosophy* (pp. 129–148). Lanham: Rowman & Littlefield.

Norton, J.D. (1996). Are thought experiments just what you thought? *Canadian Journal of Philosophy*, *29*, 333–366,

Norton, J.D. (2004a). On thought experiments: Is there more to the argument? *Philosophy of Science*, *71*(5), 1139–1151, https://doi.org/10.1086/425238

Norton, J.D. (2004b). Why thought experiments do not transcend empiricism. C. Hitchcock (Ed.), *Contemporary debates in the philosophy of science* (pp. 44–66). Oxford: Blackwell.

Popper, K. (1934/2002). *The logic of scientific discovery.* London and New York: Routledge.

Salis, F. (2020). Learning through the scientific imagination. *Argumenta*, *6*(1), 65–80,

Salis, F., & Frigg, R. (2020). Capturing the scientific imagination. A. Levy & P. Godfrey-Smith (Eds.), *The scientific imagination: Philosophical and psychological perspectives* (pp. 17–50). New York: Oxford University Press.

Sartori, L. (2023). Putting 'experiment' back into the 'thought experiment'. *Synthese*, *201*, 34, https://doi.org/10.1007/s11229-022-04011-3

Schickore, J. (2022). Scientific discovery. E.N. Zalta (Ed.), *The stanford encyclopedia of philosophy (sep).* Retrieved from https://plato.stanford.edu/entries/scientific-discovery/

Shumelda, M. (2013). At the limits of possibility: Thought experiments in quantum gravity. M. Frappier, L. Meynell, & J.R. Brown (Eds.), *Thought experiments in philosophy, science, and the arts* (pp. 141–159). London and New York: Routledge.

Stokes, D. (2016). Imagination and creativity. A. Kind (Ed.), *The routledge handbook of philosophy of imagination* (pp. 247–261). London and New York: Routledge.

Stuart, M.T. (2016). Taming theory with thought experiments: Understanding and scientific progress. *Studies in History and Philosophy of Science*, *58*, 24–33, https://doi.org/10.1016/j.shpsa.2016.04.002

Stuart, M.T. (2020). The productive anarchy of scientific imagination. *Philosophy of Science*, *87*, 968–978, https://doi.org/10.1086/710629

Todd, C. (2020). Imagination, aesthetic feelings, and scientific reasoning. M. Ivanova & S. French (Eds.), *Aesthetics and science: Beauty, imagination and understanding* (pp. 63–85). London and New York: Routledge.

Toon, A. (2012). *Models as make-believe.* Basingstoke: Palgrave Macmillan UK.

Toon, A. (2016). Imagination in scientific modeling. A. Kind (Ed.), *The routledge handbook of philosophy of imagination* (pp. 451–462). London and New York: Routledge.

Walton, K.L. (1990). *Mimesis as make-believe. on the foundations of the representational arts.* Cambridge, Mass and London: Harvard University Press.

Weisberg, D.S. (2020). Is imagination constrained enough for science? A. Levy & P. Godfrey-Smith (Eds.), *The scientific imagination: Philosophical and psychological perspectives* (pp. 250–261). New York: Oxford University Press.