Thought Experiments and the Scientific Imagination

Alice Murphy

Submitted in accordance with the requirements for the degree of
Doctor of Philosophy

The University of Leeds
School of Philosophy, Religion, and History of Science

September 2020

The candidate confirms that the work submitted is her own and that appropriate credit has been given where reference has been made to the work of others.

Acknowledgements

Firstly, a huge thank you to my supervisors, Steven French and Aaron Meskin. It’s hard to do justice to the support and insights that you have each provided over the past four years. Steven has been endlessly generous, and has an open minded approach to research that is a constant source of inspiration. Aaron’s meticulous attention to detail has made me a more careful writer and thinker. You’ve both enabled me to form sketchy ideas into arguments, spurred me on when I’ve been lacking in confidence, and have made the whole process energising and, dare I say it, fun.

At the risk of causing too high of expectations of the thesis that follows, I also want to say thank you to Milena Ivanova, Andrea Blomqvist, Mike Stuart, Max Jones, Fiora Salis, Karim Thébault and Adrian Currie for incredibly valuable conversations and feedback. I am grateful to David Davies for hosting me during my visit to McGill, and Elay Shech, James Shelley and Keren Gorodeisky for hosting me during my time at Auburn.

I have benefitted enormously from the mentorship of Víctor Durà-Vilà over the past couple of years, and I also want to thank Nick Jones, Callum Duguid, and Alison Toop for helping me develop as a teacher. Thank you to Debbie Foy, Paul McShane and the rest of the PRHS administrators for their immense patience. I am very grateful to WRoCAH and the Royal Institute of Philosophy for funding this project. Thank you to Roman Frigg and Juha Saasti for a stimulating viva, and to Adrian Wilson for chairing.

The School of PRHS is a brilliant and welcoming community. A big thanks to my fellow MAP committee members for providing opportunities to work towards making philosophy more inclusive (and a space to get angry about things in the meantime) and to all the members of the Philosophy of Science reading group for broadening my research horizons. Thank you to Emily Herring, Laura Sellars, Coreen McGuire, Jade Fletcher, Christina Nick, Hollie Gowan, Alex Aylward, Tadhg O’Laoghaire, Joe Saunders, Mikey Cannon, Adina Covaci, Will Gamester, Miriam Bowen, David Heering, Simon Newey, Arthur Carlyle and Mathieu Rees for your friendship and guidance. And to Vishnu Radhakrishnan for being my best bud from the moment I moved to Leeds. A massive thank you, and an even bigger apology, to Caitlin Molloy for all the weekend visits and for listening to me whinge about obscure “problems”.

Finishing a PhD during a global pandemic wasn’t exactly easy, but it would have been a whole lot harder without the support of Nathan Wood. Thank you for making me laugh even in my moodiest of moody moments and for being the highlight of my day every day. Finally, I want to say thank you to all my family, especially my g-dad Morty, my brother Tom (who also proofread many of these chapters), and most of all to my mum and dad for 27 years of boundless love and encouragement.

Abstract

Thought experiments (TEs) are important tools in science, used to both undermine and support theories, and communicate and explain complex phenomena. Their interest within philosophy of science has been dominated by a narrow question: How do TEs increase knowledge? My aim is to push beyond this to consider their broader value in scientific practice. I do this through an investigation into the scientific imagination. Part one explores questions regarding TEs as "experiments in the imagination" via a debate concerning the epistemic status of computer simulations in science. I outline how, against Hacking, TEs also have "a life of their own" and I argue against accounts that privilege experiments over simulations (and by extension TEs) in light of their capacity to surprise in a productive way. Part two develops a pluralist account of the nature of the imagination in science. At its core, my view is that when we attend to a number of examples of TEs and consider the context in which they are used, we see that TEs engage a variety of our imaginative capacities. Existing monistic views fail to recognise the richness of the imagination and its potential in science. Part three looks to the "beauty" of TEs which is currently overlooked in the aesthetics of science literature. I put forward a new account that demonstrates the epistemic value of aesthetic features in science by showing how an appropriate fit between form and content enhances the usability of a TE, and its effectiveness as a prompt for our imagination. This also enables a more nuanced take on the proposed similarities between TEs and literary fictions. In the concluding chapter, I outline ways in which the core features of my account can be extended beyond TEs to illuminate the significance of the imagination and aesthetic values in other areas of science.

Contents

List of Figures

Introduction

Thought Experiments in Science

Philosophical discussions of the imagination have been primarily located in the philosophy of mind and aesthetics and until recently, rarely discussed in the case of science. My project sets out to show that thought experiments are an excellent case study for thinking about the value of the imagination across the sciences. This is because thought experiments directly appeal to the imagination and have led to significant scientific advances.

Famous examples of thought experiments include Galileo's falling bodies, Einstein's elevator, Darwin's "imaginary illustrations", Schrödinger's cat, Newton's bucket and Maxwell's demon, to name just a few. Thought experiments are not limited to the sciences. Their use in philosophy is well known, and they can also be found in history, mathematics, political theory and theology. Additionally (as I discuss in chapter 5) some artworks have also been described as thought experiments, which, it is argued, explains how many of us come away from great works of literature or films with a sense that we have gained something of epistemic value.

What is a thought experiment? There is no accepted definition, but as Brown states, we 'recognize them when we see them' (2011, 1). They often take a narrative form and begin with a description of a hypothetical scenario. We imagine the state of affairs in order to "see" or determine what would happen if such a scenario were to occur, and the conclusions that we draw from them are taken to tell us something beyond the particulars of the case. Many (but not all) thought experiments would be impossible to conduct as a physical experiment. Perhaps a precise definition can be given, and throughout the thesis, I compare thought experiments with other features of scientific practice, including experiments, computer simulations and models more generally, in order to highlight thought experiments' distinctive qualities. However, like Brown, I think it is best to proceed with examples from different areas of science rather than to put strict boundaries on what constitutes a thought experiment.

While their use can be seen throughout history, and they have been the subject of philosophical interest in the works of Kant and Lichtenberg amongst others (Fehige and Stuart 2014), the term "thought experiment" was not coined until the eighteenth century. Ørsted is known to have come up with the term, but it is Mach (1896) who popularised it in his paper “Über Gedankenexperimente” (“On Thought Experiments”) (ibid.). Furthermore, the discussion of thought experiments in the philosophy of science did not really take off until the late 1980s with a debate between Brown and Norton which continues to shape the literature to this day. Brown and Norton are each concerned with what is often known as Kuhn’s (1977) “paradox” or “puzzle” of thought experiments (Stuart et. al. 2018, 9). This is the question of how thought experiments can provide new information about the world. Kuhn’s original question was: ‘How, then, relying exclusively upon familiar data, can a thought experiment lead to new knowledge or to a new understanding of nature?’ (1977, 241).

With Kuhn, Brown and Norton agree that thought experiments can provide knowledge of the world. Further, they each take it that their positions represent the only possible ways of responding to Kuhn’s question. But that is about as far as their agreement goes.

According to Brown (1986, 2004a, 2004b, 2011), there is a class of thought experiments that allow us to ‘intuit the laws of nature’ which exist as platonic objects in an abstract realm. In particular, Brown holds that natural laws are necessary relations between universals, as proposed by the Dretske-Tooley-Armstrong (DTA) theory and rejects the Humean line that the laws of nature are merely regularities. As Tooley states ‘the fact that universals stand in certain relationships may logically necessitate some corresponding generalization about particulars, and that when this is the case, the generalization in question expresses a law’ (1977, 672). Brown therefore commits to a governing conception of laws; they ‘necessitate the regularities that we experience in the empirical world’ (2004b, 1131).

Brown’s view is a rationalist approach to thought experiments; they enable us to learn new things about the world independently of empirical evidence, that is, simply by thinking about a scenario. He states that intuitions ‘are nonsensory perceptions of abstract entities. Because they do not involve the senses, they transcend experience, and give us a priori knowledge of the laws of nature’ (2004a, 34). He compares his view to Gödel’s mathematical platonism. For Gödel, mathematical objects such as numbers and sets are “real” entities that exist independently of human minds. Both Brown and Gödel draw analogies with ordinary perception; although not a sensory experience, we perceive some mathematical propositions as truths, or perceive the laws of nature through our faculty of intuition. This is thus a seeing with the “mind’s eye”. In chapter 2 I consider a further connection between these two views via the way in which thought experiments, like physical objects, can surprise us.

Brown states that contra Norton’s view, discussed below, the process of a thought experiment cannot be reduced to an argument. To support the view that sometimes we can “see” the truth, Brown provides examples used in mathematics such as the “picture proof” for the theorem ‘$n^2/2 + 1/2n = 1 + 2 + 3 + \ldots$n’ and argues that in such cases, we are not required to work through a mathematical equation. Rather, the evidence in this case is visual.

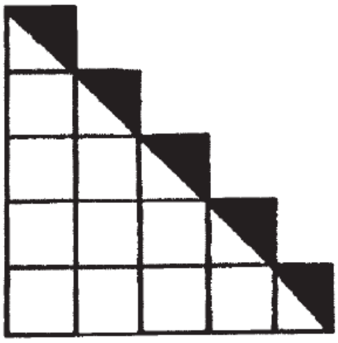

Figure 1: A picture proof (from Brown 2004a)

Analogously, when we conduct a thought experiment we see why the results must follow, rather than inferring this via argumentation (Brown 2004a, 32). Brown presents Galileo’s famous thought experiment against Aristotle as an illustrative example. The thought experiment undermines Aristotle’s theory that heavier bodies fall faster than lighter ones by a *reductio ad absurdum*, and also generates a new theory, making it, what Brown labels, both a destructive and a constructive thought experiment. Galileo asks us to imagine attaching two balls together, a heavy cannonball and a lighter musket ball, and dropping them from the leaning tower of Pisa. What does Aristotle’s theory predict? Both that the combined balls will fall faster than the heavier one on its own, as the combined object is heavier, and that the combined object will fall slower, as the lighter ball is inclined to fall slower and so, will drag the heavier body back. From this, Galileo proposes a new theory; all objects fall at the same speed.

Brown emphasises that not only does this thought experiment undermine Aristotle’s theory by a *reductio ad absurdum,* it also plays a constructive role. That is, it proposes the basis of a new theory of motion—that all objects fall at the same rate. And he claims that no new empirical data is being used when we move from Aristotle to Galileo’s theory, nor does the thought experiment illustrate a logical truth. Therefore, the argument goes, with platonist thought experiments, that is, ones which have the capacity to refute one theory (by exposing its inconsistency) whilst proposing another, we are given a priori knowledge of nature. I come back to this example and Brown's analysis of it in chapter 2.

Norton has dismissed Brown's approach as 'epistemic magic' and aims to provide an empiricist alternative (2004, 44). As mentioned, Norton's key focus is to answer Kuhn's "puzzle" of thought experiments. Norton phrases the issue as the 'epistemological problem' of thought experiments which he outlines as the following question: 'Thought experiments are supposed to give us knowledge of the natural world. From where does this knowledge come?' (ibid.). In response, Norton has defended a view in which thought experiments are arguments, although often disguised in a more "picturesque" form (1991, 1996, 2004).

More precisely, Norton states that thought experiments are arguments which '(i) posit hypothetical or counterfactual states of affairs, and (ii) involve particulars irrelevant to the generality of the conclusion' (1991, 129). Thought experiments transform what we already know (whether explicitly or tacitly) through deduction and/or induction. The outcome of a thought experiment is reliable only insofar as the reorganisation of existing knowledge preserves the truth of the premises, or in inductive cases, there is a strong chance of its preservation. To support his view, Norton has reconstructed many thought experiments into arguments, where the result is stated as a conclusion, and he claims that this process can extend to any example. Norton outlines two ways in which thought experiments should be thought of as arguments:

> '(Ia) (*Context of justification)* All thought experiments can be reconstructed as arguments based on tacit or explicit assumptions. Belief in the outcome-conclusion of the thought experiment is justified only insofar as the reconstructed argument can justify the conclusion.
>
> (1b) *(Context of discovery)* The actual conduct of a thought experiment consists of the execution of an argument, although this may not be obvious, since the argument may appear only in abbreviated form and with suppressed premises' (2004, 50).

Norton's (1a), can be further outlined as comprising two claims. Firstly, the "reconstruction thesis" holds that all thought experiments can be reconstructed as arguments by filling in their assumptions. Secondly, the "elimination thesis" entails that the particularities of a thought experiment have no epistemic force. Norton does not mention any role for the imagination, mental imagery, intuition, nor for their narrative structure, arguing that there is no important factor that thought experiments have, but their argument forms lack. These factors are merely picturesque qualities and can be eliminated without losing anything epistemically relevant.

Although they have shaped the discussion, neither Brown nor Norton's view have gained much support. Brown's view is rejected in light of its controversial platonism regarding the metaphysics and epistemology of the laws of nature. Further, his account applies to only a few cases. Although I will come back to Norton's account at various points throughout the thesis, it is worth mentioning some of the worries positioned against his view.

Some have accepted some version of (1a), arguing that insofar as a thought experiment can play a justificatory role, they must provide an argument, or at least, they agree that representing thought experiments as arguments can be an illuminating way of checking the quality of the inferences involved in their conduct. But this does not commit us to Norton's (1b), that is, the claim that the performance of a thought experiment just is that of an argument. Häggqvist highlights that argumentation is also important to concrete experiments, the results of which can also be set out in argument form, yet we do not class experiments as arguments (2009, 61). Further, many have highlighted, with Brown, that the phenomenology of conducting a thought experiment is very different than a reconstructed argument. And this can be maintained whilst rejecting the claim that thought experiments are "windows to Plato's heaven" as on Brown's view. One way to articulate this is to focus on the imagination and/or "mental models" (Nersessian 1992, 2007). It has also been pointed out how a reconstructed argument version of a thought experiment lacks features that form part of the epistemic value of the example when conducted in its original form. We shall see this in chapter 2 via Gendler's (1998) criticism of Norton.

A further source of worry is that Norton's view has changed over the years and is a 'moving target' (Brendel 2018, 283; Stuart 2016b). Norton's initial formulation of the view was restricted to deductive and inductive reasoning, but in later papers, in order to reconstruct many thought experiments, Norton has expanded this to include inferences to the best explanation, informal logics, diagrams, and arguments from analogy (1996, 2004). A case of the latter is Freiling's "refutation" of the continuum hypothesis which crucially involves 'some kind of visual or figurative evidence', namely an analogy to throwing darts at a dartboard (Brendel 2018, 286; see also Norton 2004). The problem is that Norton's view has been modified so much that any thought experiment can be reconstructed into an argument if we take "argument" to be such a broad camp. As Stuart states, this makes the claim that thought experiments can also be reconstructed as arguments ‘a good deal less impressive’ (2016b, 453).

And so, there are difficulties with the two major approaches to thought experiments. Further, as previously mentioned, Norton and Brown are interested in a narrow question regarding thought experiments: How do they give us knowledge? But as I have already indicated, thought experiments play a variety of roles in science, and I set out to illuminate those that go beyond increasing knowledge. Now that we have an understanding of the contours of the central debate on thought experiments, I’ll outline the thesis that follows.

Thesis Outline

The first part of my thesis explores the sense in which thought experiments can be considered “experiments” in the imagination. Since Mach, there has been a set of views that treat thought experiments as genuinely experimental, pointing out that thought experiments also work by what is called “the method of variation”. I outline how a parallel debate occurs in the literature on computer simulation methods and that drawing on this debate can allow us to make progress in our comparisons between thought experiments and ordinary, physical experiments.

In chapter 1, I consider the claim that experiments have a “life of their own” independent of theory, which motivates an epistemology of experiments. I outline the central features that give experiments a “life” through a prominent account of experiments provided by Hacking (1983). In a short paper on thought experiments, Hacking (1992) denies that thought experiments have a life of their own. Rather, he claims, they are “icons”. Unlike experiments, the argument goes, thought experiments do not evolve and their role is limited to the tension that they were designed to expose. I present three ways in which Hacking claims that experiments have a life independent of theory, and I demonstrate that thought experiments also have a life according to these criteria.

In chapter 2, I consider two arguments for the privileging of experiments over computer simulations: The materiality thesis and the argument from surprise. A major proponent of both of these arguments is Mary Morgan (2003, 2005). The materiality thesis claims that there is an essential difference concerning the relation between the object of an investigation and the ultimate target of an investigation in an experiment compared with that of a simulation. This is said to have an important consequence; we are more justified in our claim that we have learned something about the world in an experiment than in a computer simulation. The second argument that Morgan proposes is the argument from surprise which she links with Hacking’s views on experiments. The core claim is that given that a computer simulation is programmed by scientists, and does not interact with a part of nature, they cannot surprise in the way experiments can. Morgan labels the valuable type of surprise “confoundment” and highlights its productivity; it can force scientists to revise their theories in order to accommodate the results. I show how, if these arguments are true, they each have implications for the status of thought experiments.

However, neither succeed in giving automatic epistemic privilege to experiments. I utilise discussions on computer simulations in order to put pressure on the materiality thesis before turning to my main concern in the chapter; the argument from surprise. I begin by considering surprise from the perspective of Brown and Norton’s accounts before discussing the differences between thought experiments on one hand, and computer simulations and arguments on the other. Gendler (1998) presents an argument against Norton’s view through the example of Galileo’s falling bodies. She argues that the thought experiment cannot be reconstructed into an argument form without losing part of its epistemic value. Through considering this objection, I argue that we can see that the surprise that this example brings about cannot be adequately captured on the argument view. Rather, the use of imagination in thought experiments means that we can be surprised in a distinctive way. The “freedom” of imagination allows for reasoning that cannot be captured within the framework of more formal reasoning.

I end the first part of the thesis by coming back to Hacking’s claim that thought experiments are “icons”. Interestingly, Hacking also includes plays such as *Othello* within this category. Hacking himself does not add much detail as to what he means by this statement and so, I attempt to flesh out his claim. While Hacking uses this comparison to deny the life of thought experiments, I argue that his comparison actually opens up some interesting questions which are at the heart of the rest of the thesis: What is the significance of the imagination in thought experiments? And to what extent do thought experiments share features with works of art, in particular, literary fictions?

In the second part of the thesis, I explore the role of the imagination in science and propose a pluralist view of the nature of imagination in thought experiments. In chapter 3, I begin with an overview of the changing attitudes surrounding the imagination in science. I then point out that while the imagination is starting to gain attention in the philosophy of science, most notably through a collection of views that characterise models as fictions, there is a central question that remains under explored: What is meant by the imagination?

There has been a general assumption in philosophy of science that the imagination consists in mental images. In particular, views that have included the imagination in their analysis of thought experiments have taken it to mean that we form visual “pictures” in our minds of the described scenario. Salis and Frigg (2020) challenge this assumption in a recent paper. As on my account, they draw on discussions of imagination in other areas of philosophy to increase our understanding of thought experiments and models. Salis and Frigg argue that mental imagery is never required for the conduct of a thought experiment (or a scientific model). Rather, they claim that the only type of imagination that is relevant in the scientific domain is of a propositional kind; we imagine *that* something is the case. They take their position to be extendable to all scientific thought experiments (and models).

Against these existing monistic views, I develop a pluralist account of the nature of the imagination in science. In chapter 4, I begin with the ways in which thought experiments and scientific models differ. These differences impact the nature and role of the imagination in their use. In particular, I show that an underlying motivation for the propositional view—the complexity of models and hence the limitations of mental images—does not carry over to the use of thought experiments. The upshot of this is that while there are clearly connections to be made in terms of the imagination in thought experiments and other areas of science, we should not assume that details regarding the imagination in modelling can be automatically generalised to thought experiments.

I then turn to my main argument in the chapter. I emphasise that we need to pay close attention to particular examples of thought experiments and their use in scientific practice in order to determine the type of imagination that the example invites. This includes the different functions of thought experiments, as well as a consideration of the context in which they are formulated, and who they are formulated for. Through outlining a number of examples, I demonstrate how some thought experiments clearly invite merely a propositional form of imagination, whereas others invite an imaginative activity that goes beyond this. In cases of the latter, we are also invited to form objectual imaginings, in particular, to visualise a state of affairs. The outcome of this is that we can attempt to reconstruct thought experiments into a propositional or argument form, but this will ignore significant ways in which they engage the imagination. My pluralist alternative accommodates how thought experiments call on the imagination in an effective way to communicate and explain.

The final part of the thesis continues with this focus on the imagination in scientific practice to consider the role of aesthetic values in thought experiments, and the ways in which thought experiments can be said to share qualities with literary works. The aesthetics of science literature primarily focuses on the "beauty" of theories and mathematical proofs as well as whether such judgements can indicate the truth of a theory, or can play some other epistemic role. I outline how thought experiments should also be part of the discussion; they are often described using aesthetic language (whether positive or negative). Furthermore, connections have been drawn between thought experiments and artworks. In particular, there is a prominent set of views in philosophy of art that defend the cognitive value of narrative art, usually literary fiction, by arguing that artworks can function as thought experiments and hence, can provide us with insights about the world and our place within in (Elgin 2014).

However, chapter 5 also includes the ways in which thought experiments and literary fictions differ which, it is argued, undermines the ways in which we can say they provide insights in a similar way. Further to this, I present a dilemma for existing aesthetics of science projects as introduced by Todd (2008). It appears that either aesthetic language should be understood as a proxy for epistemic features, or aesthetic language is literal, but we are left with a need to explain why such values are significant in science and consequently why they should be of any interest to philosophers of science.

Responding to this dilemma is my central aim in chapter 6. I outline a new approach to an aspect of aesthetics of science that offers a way of defending the epistemic importance of aesthetic values, without reducing them to epistemic features. I turn to accounts in aesthetics that discuss the interrelation between form and content as a source of aesthetic value. On these views, the aesthetic pleasure of artworks is rooted in our appreciation of the ways in which the form of a work is well suited to an artwork's overarching content. I then discuss this in the context of thought experiments through focusing on two examples from Darwin. I argue that the usefulness of a thought experiment in scientific practice is impacted by the way in which its content is conveyed through the concrete particulars given in the narrative, as well as through the use of diagrams and other visual aids. Well formulated thought experiments are better prompts for our imagination, they aid our understanding, and can contribute to thought experiments' demonstrative force.

Once my proposal for a promising and novel view of aesthetics in science is in place, I come back to the proposed similarities between thought experiments and works of art. I claim that we should not be too quick to defend the cognitive value of all literary fictions via an emphasis on their similarities with thought experiments. Instead, I argue, via a consideration of genre conventions in science, that we should be more selective in our comparisons. In particular, I outline how thought experiments are significantly similar to works of speculative fiction as well as to fables and parables.

The thesis concludes by summing up my key findings in the previous chapters which together show how an exploration of the varieties of the imagination in thought experiments sheds new light on their value in scientific practice. This includes how thought experiments can surprise in a fruitful way and can aid our understanding of theories and phenomena, as well as demonstrating the salience of aesthetics in science for forming useful imaginings. I also suggest the ways in which my core contributions can be extended to other areas of scientific practice, and I highlight the future research topics suggested by my thesis.

1. Thought Experiments and Experiments

1.1 Introduction

One way in which the thought experiment literature is divided is through the relations between thought experiments and ordinary experiments. Some argue that thought experiments are experiments or share significant features with the latter. Whereas others argue that thought experiments lack key features of genuine experiments, such as the fact that they do not intervene on the world and therefore do not belong in the class of experiments. However, much of the answers to these questions appear down to a matter of emphasis; views either focus on certain qualities of thought experiments that make them appear experimental (and thereby include them in the class of experiments) or alternatively, they focus on the ways in which they differ and so, claim they are something else (for example, arguments).

In this chapter and the one that follows, I move away from identity questions (that is, are thought experiments *experiments?)* and instead provide a closer look at the ways in which thought experiments compare to physical experiments. I do this via attending to issues in philosophy of experiment as well as by drawing on debates surrounding the epistemic status of computer simulations compared to ordinary, physical experiments which is a central topic in the philosophy of computer simulations in science.

This chapter begins with an overview of some of the existing accounts of thought experiments and physical experiments, and highlights the commonalities between these discussions and those that surround computer simulations. I then turn to Hacking's view that experiments have a "life of their own", independent from theory, which he denies for thought experiments. But I set out to show how thought experiments also have a life of their own. In the next chapter, I continue with these comparisons. I outline two arguments for privileging experiments—the materiality thesis and my main focus, the surprise argument—which are central in the literature on computer simulations. I demonstrate how, if true, these arguments would also have implications for the epistemic status of thought experiments. However, I argue that despite there being important differences between experiments and these other practices, neither of the arguments are successful in establishing the superiority of experiments in general. Rather, different practices are useful in different circumstances.

Thought experiments, I argue, have a sui generis status from experiments and computer simulations despite similarities in their methodologies. Discussions that identify thought experiments with other practices try to fit thought experiments into one box and in doing so, they lose part of thought experiments' value. These two chapters set up the rest of the thesis as I then go on to explore features which might not be unique to thought experiments but play a central role in their use. The second part of the thesis looks at the nature and role of the imagination in thought experiments and the final part looks at their aesthetic features and how they compare with literary works.

1.2 Thought Experiments, Experiments and Computer Simulations

What is the relationship between thought experiments and physical experiments? Should we understand the use of "experiment" in thought experiments as a mere metaphor? This issue has divided the literature on scientific thought experiments. On one side, there are those who take seriously the "experimental" aspect of thought experiments. Proponents claim, for example, that thought experiments are experiments in their own right or are on a continuum with experiments. On the other side are those who draw a sharp line between thought experiments and experiments. This is a key difference between Norton and Brown's approaches. As seen (in the introductory chapter), for Norton, thought experiments are arguments. As they work by inferences and do not involve interacting with, manipulating nor observing the natural world, any similarities with experiments are superficial. Whereas for Brown, that thought experiments take place in, as he puts in, the "laboratory of the mind" does not entail that they are not experimental. Thought experiments involve quasi-observation; a system is represented and then observed by the mind's eye in a way that is analogous to physical experiments.

Highlighting the similarities between thought and physical experiments can be used as part of a defence of the epistemic value of thought experiments given the acceptance of experiments as a crucial part of acquiring scientific knowledge. An early reflection on the methodology of thought experiments did just this. According to Mach, thought experiments are experiments that operate at a 'higher intellectual level' than physical experiments (1896/1973, 452).[1] He

[1] Although he focuses on cases in physics and mathematics, Mach also looks beyond science and draws connections with the arts when discussing experiments and thought experiments: 'The dreamer, the builder of castles in the air, the poet of social or technological utopias all experiment in thought' (1896,

argues that ‘it can be seen that the basic method of thought experiment is just like that of physical experiment, namely, the method of variation. By varying the circumstances (continuously, if possible) the range of validity of an idea (expectation) related to these circumstances is increased’ (ibid., 453). In the design of an experiment, certain factors are isolated, variables are controlled, and irrelevant aspects are idealised away. The experimenter then observes what follows and interprets the results. Thought experiments, too, involve this method of variation. But in their case, the manipulation it not of the world itself. Rather, thought experiments are conducted in the imagination. For Mach, then, there are similarities in the methodologies of thought experiments and physical experiments. Further to this, Mach argues that carrying out experiments in our minds is advantageous, as our ‘own ideas are more easily and readily at our disposal than physical facts. We experiment with thought, so to say, at little expense’ (ibid., 452). And this explains why many thought experiments are precursors to physical experiments; ‘it shouldn’t surprise us that, oftentimes, the thought experiment precedes the physical experiment and prepares the way for it…Every inventor and every experimenter must have in his mind the detailed order before he actualizes it’ (ibid.). It remains unclear whether Mach regards thought experiments as mere instances of imagining about concrete experiments, such as in the latter’s design, since he does not distinguish between this kind of imagined experiment and thought experiments.

While thought experiments can precede experiments in this sense, Mach highlights how the former are ultimately dependent on the latter. The success of a thought experiment relies on knowledge that has been gained by previous experience with the world. That is, they work by drawing on these past observations, allowing us to see features that had gone unnoticed. Additionally, for Mach, if there is any doubt regarding the result of a thought experiment, then a physical experiment needs to be carried out. Mach acknowledges that the result of some thought experiments ‘can be so definite and decisive that any further test by means of a physical experiment, whether rightly or wrongly, may seem unnecessary to the author’ (ibid.). But he was wary of those who take the outcome of a thought experiment as decisive. He argued that the ‘more uncertain and more indefinite the results are, however, the more the thought experiment necessitates the physical experiment as its natural continuation which must now delimit and determine the experiment’ (ibid.). As a consequence, some have argued

451). I come back to connections between thought experiments and the arts in chapters 5 and 6 where I discuss accounts that understand literary narratives as thought experiments.

that for Mach, ‘physical experiment *always* prevails over thought experiment’ (Sorensen 1992a, 74; see also Buzzoni 2018).

Other accounts of thought experiments can be said to be part of the Machian tradition. For example, Sorensen has argued that thought experiments are a limiting case of experiment (1992a, 1992b). Sorensen builds on Mach’s analysis and also highlights the advantages of conducting thought experiments over physical experiments in that they are easier to conduct and replicate (see also Gooding 1992). But Sorensen emphasises thought experiments’ independent value from physical experiments by stating that not all thought experiments can be realised as a physical experiment and further, this is not required for belief in their results. For Sorensen, thought experiments are a limiting case of experiment; a ‘thought experiment is an experiment that purports to answer (or raise) its question by mere contemplation of its design’ (1992b, 16). Similarly, Elgin regards thought experiments as ‘not actual, and often not even possible, experiments’ (2017, 229). Like Mach, Elgin emphasises the similarities in the methodology of thought experiments and ordinary, physical experiments. She points out how in laboratory experiments, ‘scientists simplify, streamline, manipulate, and omit, so that the effects of the potentially confounding factors are minimized, marginalized, or cancelled out’. Laboratory experiments thus involve “departures from nature” and Elgin argues that thought experiments share these features but involve even further distancing (ibid.).

We can see that a similar discussion occurs in the philosophy of computer simulations. Their use to study a range of complex phenomena is prevalent throughout the sciences. Winsberg presents a narrow and a broad definition of a computer simulation. In the narrow sense, a computer simulation is a computer program that uses ‘step-by-step methods to explore the approximate behaviour of a mathematical model’. This can be, but is not always, a model of a real-world system. A more broader use of computer simulation ‘includes choosing a model; finding a way of implementing that model in a form that can be run on a computer; calculating the output of the algorithm; and visualising and studying the resultant data’ (Winsberg 2019, §1).[2] A key issue is how they compare to ordinary, physical experiments, and they have been classed as, for example, experiments in silico, numerical experiments, virtual experiments or

[2] Similarly, Brown states that a narrow definition of an experiment or thought experiment is the setup and the behaviour we observe. A broad definition includes any initial theorizing, draw conclusions and so on: ‘The narrow sense of experiment (whether real experiment or thought experiment) is what we observe, the phenomenon, the middle of the schema. The broad sense includes the whole thing from theory and background assumptions to the final result’ (2007, 158).

experiments without materiality. And some have claimed that computer simulations just are experiments (see Barberousse et al (2009) for an overview). But their status as genuinely experimental is contested. This is because, like thought experiments, computer simulations do not intervene on the natural world and instead explore "hypothetical worlds" (Lenhard 2018).

As Arcangeli points out, it is fairly easy to see from Winsberg's description of a computer simulation that Mach's "method of variation" can also be applied to their case. Choosing the model (and finding a way to implement it) belongs with the setup of an experiment (or thought experiment) and the selection and isolation of certain factors. The running of the computer program belongs with the stage in which the variables interact. And the calculating of the output as well as studying/visualising the results accords with the experimenter's (or thought experimenter's) observations of what happens, and their interpretation of the results (2018, 9).[3] We also saw that Mach highlighted how conducting thought experiments can be easier than physical experiments. Computer simulations also come "at little expense" in the sense that it can be easier to conduct a computer simulation than perform a physical experiment to learn about particular systems. Finally, while we can maintain that thought experiments and computer simulations have a role beyond preparing physical experiments, Winsberg makes a point similar to Mach's when he argues that physical experiments have priority over computer simulations in that the latter rely on data and information that has been previously collected from experiment and observation (2009, 591).[4]

While there is a way, then, of highlighting similarities between the three practices with regards to their methodologies, there are also comparisons that have focused predominately on the similarities that thought experiments and computer simulations share, given that unlike experiments, neither involve intervening on the world. Some have commented on how computer simulations are thought experiments, albeit ones that use a computer program. For example, Dennett (1994) and Swan (2009) have suggested that the Tierra model of artificial

[3] Applying Mach's insights in this way is similar to El Skaf and Imbert's (2013) view that physical experiments, thought experiments and computer simulations each involve the construction of a scenario which is then "unfolded" and interpreted.

[4] I'll come back to issues regarding the privileging of experiments over simulations in the next chapter but it is important to note that this is not entailed by Winsberg's statement regarding the priority of experiments. He states: 'There may have been a time in the history of science, perhaps before Newton, perhaps even earlier, when we did not have sufficient systematic knowledge of nature—enough of a toolkit of trustworthy model building principles—for a simulation to ever be as reliable a source of knowledge as even the crudest experiments, but that time has long passed' (2009, 591).

life is a kind of thought experiment. As Swan states: ‘Alife provides a wide variety of means for rethinking our conceptions of life and our understanding of evolutionary processes by creating imaginative alternatives to what is—allowing us to entertain what might be, or what could have been, given different parameters—which is essentially how thought experiments work’ (ibid., 688). Others have extended existing analyses of thought experiments to computer simulations. For example, on Chandrasekharan et al’s (2012) view, computer simulations are an extension of the capacity for mental simulation. They compare this with how instruments such as telescopes are used in order to extend our visual abilities. Both thought experiments and computer simulations involve an exploration of counterfactual situations but because computer simulations can deal with more complex simulations, Chandrasekharan et al predict that they will come to replace thought experiments.

Beisbart and Norton (2012) utilise Norton’s work on thought experiments in order to draw a sharp contrast between physical experiments on one hand, and thought experiments and computer simulations on the other. Beisbart and Norton argue that like thought experiments, computer simulations are arguments. For the argument account, the important difference between these practices compared to experiments is that thought experiments and computer simulations involve inference, whereas experiments involve discovery. And so, any similarities raised above on the Machian analysis are irrelevant to thought experiments and computer simulations’ epistemology. I’ll come back to this account in the next chapter when considering the ways in which thought experiments can surprise. Similarly, Humphreys has separated thought experiments and computer simulations from real experiments, arguing that they involve theoretical as opposed to practical manipulation and thus ‘lie much closer to theory than to the world’ (1994, 218).

And so, there has been some developments regarding the similarities (and differences) between the three practices. However, I am inclined to agree with El Skaf and Imbert that the existing literature ‘offers more a battlefield than a steadily developing domain; in particular, it is not completely clear how much the described similarities between these activities are deep or shallow and merely rhetorical’ (2013, 3453; see also Cooper 2005). Whether or not thought experiments ought to be *identified* as experiments or not depends largely down to what features are emphasised, and what features one takes to be essential to an experiment.[5] But then the discussion gets somewhat “stuck” (see also Arcangeli 2018). And so, my aim in

[5] Winsberg (2019, §5) makes a similar point regarding debates on computer simulations.

the rest of this chapter and the next, is to offer a more productive discussion on the commonalities of these practices.

There are many ways in which the existing observations could be meaningfully developed in order to build upon the similarities and differences between thought experiments, physical experiments and computer simulations that are relevant to their epistemic significance and their fruitfulness in science. For the remainder of this chapter, I will turn to important work in the philosophy of experiment that has also been picked up on in the computer simulation literature (Winsberg 2010). This is Hacking's view that physical experiments have a "life of their own". In the chapter that follows, I focus on two arguments for the privileging of experiments over computer simulations and apply them to the case of thought experiments.

1.3 The Life of an Experiment

Hacking argues that a key difference between thought experiments and ordinary, physical experiments are that the latter, but not the former, have a "life of their own". He states:

> 'experiments are organic, develop, change, and yet retain a certain long-term development which makes us talk about repeating and replicating experiments…I think of experiments as having a life: maturing, evolving, adapting, being not only recycled but also, quite literally, being retooled. But thought experiments are rather fixed, largely immutable. That is yet another respect that they are like mathematical proofs, but good proofs have proof ideas that can be used over and over in new contexts—which is not, in general the case with thought experiments. They have just one tension to expose… Once the thought experiment is written out in perfection it is an icon. Icons, to reiterate, do not have a life of their own' (1992, 307).[6]

To understand Hacking's comments on thought experiments, we first need to look at what he means by the "life of an experiment". His paper on thought experiments does not go into detail on experiments and so, I will turn to his *Representing and Intervening* (1983) to expand the above claims.[7] In the introduction to his ground breaking text, Hacking states that for a long

[6] It is interesting that Hacking compares thought experiments to icons. Sadly, Hacking does not go on to give a detailed explanation for this comparison. I'll say some more about the connection with icons at the end of the next chapter.

[7] Bokulich (2001) and Bokulich and Frappier (2018) also explore Hacking's claims regarding the life of an experiment in the context of thought experiments which I return to below. But they focus only on

time, philosophers and historians of science have heavily focused on theories and representations, and experiments have been severely overlooked. As a response to this, he sets out to 'initiate a Back-to-Bacon movement, in which we attend more seriously to experimental science' (1983, 150). *Representing and Intervening* was one of the first philosophical works to attend to experiments in scientific practice, and the "life" of an experiment is a powerful metaphor which has been subsequently adopted by many. Hacking argues that the lack of philosophical and historical analysis of experiment results from seeing experimentation as an activity dictated by theory in which the value of experiments consists in their role in testing theory. In outlining experiments' independence from theory, Hacking carves out the space for an epistemology of experiment.[8]

He cites chemist Liebig and Popper as exemplars of the view that experimentation is a theory-driven practice. For Liebig, an 'experiment not preceded by theory, i.e. by an idea, bears the same relation to scientific research as a child's rattle to music' (1863, 49 in Hacking 1983, 153). Hacking notes that there is a weak version and a strong version of Liebig's view. The weak version claims that you must have 'some ideas about nature and your apparatus before you conduct an experiment' (1983, 153). A strong version maintains that testing a theory about the experimental object is the only way an experimental result can be significant. Take Popper's view that theoreticians propose well-formed questions to the experimenter whose role is to provide definite answers to those questions and thereby assess the validity of a theoretician's claim. For Popper, theory 'dominates the experimental world from its initial planning up to the finishing touches in the laboratory' (1934, 107 in Hacking 1983, 155). Hacking grants the weak claim—experiments should be conducted in light of some understanding or ideas about the experimental setup—but sets out to undermine the strong version. For Hacking, "theory" is thus more developed than mere ideas surrounding an experiment and its apparatus. Hacking aims to show how fundamental research can proceed without relevant theory in this latter sense and to defend the independent status of experiments.[9]

a couple of features, and do not draw on Hacking's work on experiment to flesh out his claims. Here, I follow Shinod (2017) and turn to Hacking's earlier work in order to get clearer on his comparison between experiments and thought experiments.

[8] Franklin (1986, 2013) and Franklin and Perovic (2019) also adopt the term "the life of an experiment".

[9] Hacking does, however, suggest that even the weak version could be challenged: 'The physicist George Darwin used to say that every once in a while one should do a completely crazy experiment, like blowing trumpets to the tulips every morning for a month. Probably nothing will happen, but if something did happen, that would be a stupendous discovery' (1983, 154).

Hacking's work on experiments set against these views expressed by Liebig and Popper is extensive and cannot be fully covered here. But I will highlight a few of the central features of the life of experiments, which is sufficient for the comparison with thought experiments.

1.3.1 The Evolution of Experiments

The first point that we learn from Hacking is that experiments can evolve. Hacking outlines how experiments can evolve when instruments develop and become more refined. For example, in the development of microscopes. Another sense in which they can evolve is when scientists become more adept at observing such as in the case of Caroline Hershel's tireless observations of the night's sky that resulted in her discovery of many comets, including a particularly productive period in which she discovered eight in a single year (1983, 180). An example that Hacking discusses in detail is the Michelson-Morley experiment. The aim of the experiment was to detect the velocity of the Earth with respect to the ether, a hypothesised medium in space for the propagation of light waves. Hacking outlines how a series of experiments, beginning with Michelson's experiment in 1881 and ending in the 1920s have been labelled "the Michelson-Morley experiment". The most famous instance took place in 1887 and lasted over a year: 'This included making and remaking the apparatus and getting it to work, and above all acquiring the curious knack of knowing when the apparatus is working' (1983, 174).

Hacking traces the experiment's history and its relation to various theories of the ether. Michelson's 1881 experiment consisted in splitting a ray of light into two beams by sending it through a half-silvered mirror so that the beams travelled at right angles to one another, one in the direction of the earth's motion. The beams were then reflected back into the middle and this is where any effect of the ether would be seen; the beam travelling in the direction of the earth's motion should take longer to be reflected back than the beam travelling at a right angle. But this effect was not observed. Michelson faced problems with his apparatus in his experiment, and he went on to work with Morley, a gifted chemist, to carry out improved versions of the experiment with more precise equipment in 1886 and 1887.

The experiment failed to show what Michelson and Morley intended, and the results of the experiment were initially taken to refute the ether hypotheses. Later, the experiment was regarded as support for special relativity (1983, 254).[10]

1.3.2 The Robustness of Experimental Results

The second feature of the "life" of experiments that we learn from Hacking is that experimental results are robust despite changes in the theory of the phenomena or of how the relevant apparatus works. Hacking draws on the example of microscopes. Theories of how microscopes work have varied extensively, and Hacking notes that the correct explanation did not emerge until 1873 with the work of Abbe (1983, 194). He argues that there has been continued belief in the visual representations that microscopes produce despite prominent change in the theoretical domain (ibid., 199). For Hacking, this challenges the view that experiments are always theory-laden, that is, our observations, what we see in an experiment, are determined by whatever theoretical presuppositions we have.[11] Such a commitment would predict that there would be changes in the results of experiments once new theories are introduced and theoretical presuppositions change.

One reason that Hacking provides for the robustness of experimental results is to do with the fact that scientists manipulate their objects of study. For instance, when observing cells under a microscope, a scientist might use a glass needle to inject a fluid into the wall of the cell. They will observe the needle going into the cell through the microscope, and have predictions about the effect this will have, that is, the cell's shape will change. When the scientist observes this, this strengthens their belief that the microscope is producing '"true" images which are, in some sense, "like" the specimen' (ibid., 190).[12]

Franklin notes that while Hacking's discussion of "robustness" applies well to this example in which we can manipulate the object of study in the above sense, as well as use different instruments but yield the same results, this is not so simple for all experiments. In light of this,

[10] Hacking (1983, 254-261) presents a more detailed account of the experiment and its relation to various theoretical work. See also Franklin and Laymon (2019).

[11] See Schindler (2013) for a detailed discussion of the ways in which observations can be understood as theory-laden, and for various responses from the "New Experimentalists", including Hacking.

[12] Hacking also offers the example of a particular constellation of dots, 'dense bodies' in cells, is seen with different types of microscopes to argue for the validity of such observations, that is, to show they are not an artefact of the apparatus (1983, 201-202).

Franklin (1986, 1989) and Franklin and Perovic (2019) outline other strategies that are used in order to validate experimental results. Confidence in results can increase when, for example: (1) the results are replicated in other experiments, (2) experimental bias is eliminated, (3) the system is well isolated and possible sources of error are eliminated and (4) the apparatus used is calibrated and is based on well-developed theory.

1.3.3 Experiments Can Precede Theory

Finally, Hacking sets out how experiments can precede relevant theory. He offers the examples of the early development of optics which 'depended on simply noticing some surprising phenomenon' such as in the case of Bartholin's discovery of the double refraction of light in Iceland spar in 1689. The phenomenon could not be integrated into the current theory of optics at the time. Hacking is clear that Bartholin, amongst the other scientists he refers to in this discussion, was of course 'curious, inquisitive and reflective' and was setting out to form theories, but these observations came before theory (1983, 155-156). In addition, Hacking offers examples from thermodynamics where the invention and improvement of steam engines came before a theoretical explanation of thermal engines; 'the very word 'thermodynamics' recalls that this science arose from a profound analysis of a notable sequence of inventions' (ibid., 164).

Now that we have a sense of what is meant by the different dimensions of the "life of an experiment", we can turn to Hacking's argument that thought experiments do not have a life of their own.

1.4 The Life of a Thought Experiment

As seen, Hacking argues experiments have a life independent of theory. He states: 'I think of experiments as having a life: maturing, evolving, adapting, being not only recycled but also, quite literally, being retooled' (1992, 307). But he denies this for thought experiments; 'thought experiments are rather fixed, largely immutable' (ibid.).

Why does Hacking regard thought experiments as "fixed"? Hacking argues that while experiments perform many roles in science, there is just one role for thought experiments. Following Kuhn, he argues that thought experiments are valuable in so far as they 'reveal tensions between one vision of the world and another. That is their job, their once and future job' (1992, 307). Furthermore, Hacking argues that a thought experiment has 'just one tension to expose' (ibid.). As mentioned above, he contrasts this with mathematical proofs as the latter can be used in different contexts. Hacking presents the case of Plato's account of Socrates and the slave boy in *Meno* who is asked to double the size of a square, used as support for the view that all learning is recollection. Although this example has been repeated throughout history, it does not change. In Hacking's terminology, we draw the same diagrams and reach the same conclusions every time. Hacking grants that thought experiments will go through changes in their initial development but the key point is that once they are written out in a finalised form, they cease to evolve. This highlights a key difference between experiments and thought experiments' relation with theory. For Hacking, thought experiments are fixed to their original context and therefore to the theoretical ideas that they were initially built upon. For this reason, they cannot be retooled in the way that experiments can. As Hacking puts it, 'what they think is what was once thought' (1992, 307). It is in light of these qualities that Hacking regards thought experiments as "icons".

In the previous section, I outlined three ways in which experiments have a life independent of theory. I will now show how thought experiments also have a life of their own in accordance with Hacking's criteria.

1.4.1 The Evolution of Thought Experiments

The first feature of the life of an experiment that I outlined was to do with how they can evolve. Can the same be said for thought experiments? Bokulich suggests that the reason why thought experiments may appear as static entities is due to the fact that they are often used pedagogically. This means that they are generally presented with a lack of attention to the context in which they were devised or to how they have changed over time, and therefore have become simplified (2001, 304-305). It is also clear that some scientific thought experiments have not undergone significant changes since their inception such as Galileo's falling bodies. Galileo may have modified the thought experiment in its initial development, but since publishing the example in *The Two Dialogues*, it has remained stable.[13] However, there are contrasting cases.

[13] See Palmieri (2005) for a historical overview of the development of Galileo's falling bodies.

One example of a thought experiment that can be compared to the way in which experiments evolve is the clock-in-the-box as presented by Einstein in 1930. Einstein used the clock-in-the-box in an attempt to undermine Heisenberg's uncertainty principle. The set up of the scenario involves imagining a box containing photons. On one of the sides of the box there is a shutter which is controlled by a clock. We imagine taking a measurement of the weight of the box. We further imagine the clock opening the shutter for the exact amount of time that it would take for just one photon to pass through. We then measure the weight of the box again. By working out the difference in weight of the box before and after the photon was emitted, we can also work out the mass of the photon. Further, by using $E=mc^2$, we can also work out the photon's energy. Einstein uses the thought experiment to argue that in principle, 'we can measure the photon's energy and its time of passage to any arbitrary degree of accuracy'. Thus, the argument goes, the thought experiment demonstrates that the uncertainty principle is incorrect (Bishop 1999, 536).

Bokulich and Frappier outline the evolution of this example at the hands of Einstein and Bohr, starting from its initial formulation in 1927. Bohr (1949) recounted how, before the thought experiment was described in the above way, Einstein described a 'particle passing through a narrow slit in a diaphragm that was placed some distance before a photographic plate'. One modification was to add in another diaphragm with two slits, between the first diaphragm and the photographic plate. Another modification, once the slit had been replaced by a shutter attached to a clock (as in the version above), was then presented with a '"pseudorealistic" diagram of Einstein's clock-shutter device, with the support bolted to the table and the box suspended from the support by a string allowing the box to move up and down' (Bokulich and Frappier 2018, 547). The thought experiment, then, has evolved in a similar way to how Hacking argues experiments can evolve. The clock-in-the-box went through a series of revisions in order to be a better, more refined thought experiment.

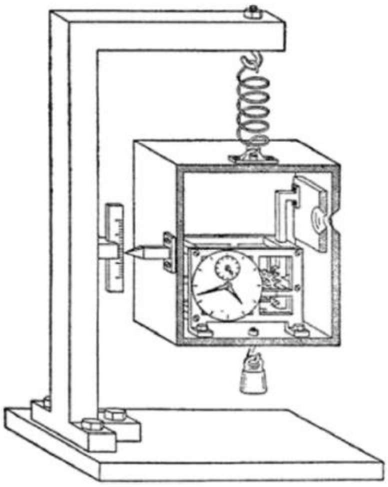

Figure 2: Bohr's "pseudorealistic drawing"of the clock in the box (from Bohr 1949)

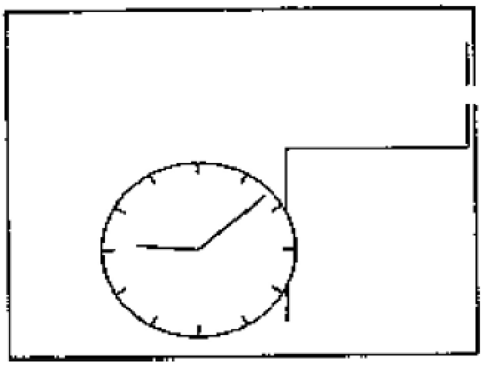

Figure 3: Bohr's depiction of Einstein's clock in the box (from Bohr 1949)

Another interesting feature of this case is that Bohr did not agree that the thought experiment successfully undermined the uncertainty principle. At first, the version of the thought experiment presented above baffled Bohr, but he was quickly able to respond to Einstein. By concentrating on the practicalities of measuring the photon's energy at a particular time, Bohr argued that in order to weigh the box, the apparatus would have to move in a gravitational field which would change the rate of the clock. This thus leads to the indeterminacy that Einstein was attempting to falsify. Einstein agreed that Bohr's reply had successfully shown that his thought experiment failed (Bishop 1999, 536). However, despite Einstein accepting the result, the thought experiment was not thereby considered a done deal. This example continues to evolve. It is still being revised, with different formulations resulting in different theoretical conclusions. The impact of the clock-in-the-box therefore remains a contentious issue (Stuart 2016a, 28).[14]

Another example of a thought experiment that evolves is Maxwell's demon, first devised by Maxwell in 1867 to show how the second law of thermodynamics could be violated. I say more about Maxwell's use of the thought experiment in my discussion of the nature of the imagination in science in chapter 4. The important point for now is that the example has been modified over time, with various critiques and analyses of the scenario presented, and what the

[14] More recent reformulations of the thought experiment have been given by Treder (1975), De la Torre et al. (2000) and Hilgevoord (1998)

thought experiment shows is still a matter of debate (see Norton (2013) and Earman and Norton (1998)).[15]

1.4.2 The Robustness of Thought Experiments

The second way in which experiments have a life was to do with the robustness of experimental results despite theoretical changes. As seen in the experiment case, Hacking linked the robustness of results with the fact that experiments intervene on nature. As thought experiments do not intervene on the world, this helps explain why on Hacking's view, the results of a thought experiment are tied to the theory that they were initially built upon. But there are also cases of thought experimental results that are robust in that there is agreement on what would happen in a thought experiment, but the results are explained by appealing to different theories. In Bokulich's (2001) terminology, how thought experiments can also be "rethought" from various theoretical standpoints. We can look to the rocket and thread thought experiment as an example. The setup is as follows:

> 'Imagine two identically constructed rockets, B and C, both initially at rest in an inertial reference frame, *S.* The two rockets are arranged one behind the other, 100 meters apart in *S* and they are connected by a thin piece of thread just long enough to connect the two rockets… Now imagine that both rockets fire up their engines simultaneously in this frame and gently accelerate to relativistic velocities. Once they reach four-fifths the speed of light relative to *S,* they simultaneously stop accelerating, and are now moving with a uniform velocity. According to an observer at rest in *S,* the two rockets have been moving in tandem and are still 100 meters apart. The question now is whether or not the thread will break' (Bokulich 2001, 290).

The thought experiment was initially devised by Dewan and Beran in 1959 who evaluated the scenario from the perspective of special relativity. For an observer at S, the thread will be Lorentz contracted and can no longer reach the two rockets and so, it will break. The distance between the rockets remains the same because they accelerated simultaneously and equally in S. Dewan and Beran also show that the thread will also break for an observer on rocket A at rest in the rockets' final inertial frame, S', but for a different reason. For an observer in S', the

[15] There is thus an interesting comparison to be made with questions regarding the "end" of an experiment, as discussed by Galison (1987).

distance between the rockets increases (the accelerations of the two rockets were not simultaneous) but the length of the thread stays the same and therefore breaks.

For Dewan and Beran, the aim of the thought experiment was to show that Lorentz contraction (the phenomenon in which the length of an object is shorter when it is moving than when it is at rest) can cause measurable stress on moving bodies. The thought experiment was then used again by Bell in 1976. Bell did not dispute the result of the thought experiment; he agreed that the thread would break. Rather, he demonstrated how this result could be reached but from a different theoretical standpoint. Bell showed how Lorentz's ether theory could also explain the breaking of the thread for an observer at S and S' (Bokulich 2001, 291-292). This example demonstrates that there can be agreement on what a thought experiment shows, but there can be different, opposing theoretical explanations of the result.[16]

In considering Hacking's example of the robustness of results in the case of microscopes, I also outlined further ways in which experimental results can be validated through drawing on Franklin's work. Given the use of computer simulations and thought experiments in science, we also need reason to believe in the reliability of the results they provide. One way in which we can do this, which would accord with Norton's view, as well as Norton and Beisbart (2012) (discussed in the next chapter) would be to analyse them as arguments; we assess them by checking for validity and soundness. However, another way in which they can be assessed is through drawing on the epistemology of experiment. I will not go into detail here, but Stuart (2016b) has offered a discussion of what this might look like in the case of thought experiments and Parker (2008) and Winsberg (2010) have applied these strategies to computer simulations.

Focusing on Stuart's analysis of thought experiments, we can consider how the criteria may apply: (1) we check that thought experiment results can be replicated by getting a wide range of people to conduct the same example, which is of course a large practice of thought

[16] Bokulich ties her discussion of this example to Duhem's scepticism of thought experiments, or what he called "fictitious experiments", which he rooted in his scepticism concerning crucial experiments, that is, those which decide in favour of one theory over another. For Duhem, experimental evidence is always underdetermined by theory choice such that an isolated hypothesis is never tested in an experiment. Rather, it is always connected with a set of background theories. However, Bokulich states that 'that thought experiments are no more bound to any one particular theory than ordinary physical experiments are, and second, they can underdetermine theory choice in the same way too' (2001, 293). That is to say, the interpretation of thought experiments also presupposes background theory and hence, thought experiments cannot be "crucial" experiments either. See Buzzoni (2018) for a recent paper that challenges the view that Duhem had a general wariness of thought experiments through a discussion of the commonalities between Duhem's position and Mach's as presented above.

experiments; they are written down, shared and analysed by the scientific community, this also helps to (2) eliminate bias in their results. (3) We isolate the system in our imagination through idealisation and abstraction in order to trust that possible sources of error are eliminated. Possible sources of error include ‘inaccurate representations and weak imaginations’ (Stuart 2016b, 461). Finally (4) the apparatus used is calibrated and is based on well-developed theory. As Stuart explains, a greater understanding of our imagination is gained through studies in philosophy of mind, cognitive science and psychology. Further, Stuart argues that there are good reasons to think that our imaginations are calibrated. This is because uses of the imagination are ubiquitous; most of us use our imaginations every day when considering hypotheticals and in problem solving, for example (ibid.).

There is clearly more to be said concerning the ways in which the epistemology of experiment can carry over to thought experiments. My aim here was just to indicate some promising work in that direction. I’ll now turn to the final feature of the “life” of an experiment.

1.4.3 Thought Experiments Can Precede Theory

Finally, we saw how experiments can precede theory and can be conducted without the aim of testing or supporting established theories. We saw this quality through the examples of observing surprising phenomenon, or in the case of theories being formulated in order to explain how existing inventions work. It is difficult to see what an analogous case of a thought experiment would be with regards to the latter. And we can see that many of the thought experiments that are frequently discussed follow some theory. Take Galileo’s thought experiments against Aristotle. They start by assuming a theory in Aristotelian physics then proceed by showing how it cannot be the case. Similarly, Einstein’s clock-in-the-box thought experiment sets out to show an inconsistency in Heisenberg’s quantum mechanics and Maxwell’s demon does not precede the second law of thermodynamics, but rather, is introduced in order to aid our understanding of it.

However, we can see that there are cases of thought experiments which have the exploratory quality that Hacking highlights for experiments.[17] To recall, Hacking does not intend to argue

[17] My use of “exploratory” here comes from making a connection to models, in particular, Gelfert’s discussion of exploratory models. Gelfert draws an analogy with exploratory experiments, that is, those which are conducted in the absence of well-developed theory. He argues: ‘a model always have to render an empirical phenomenon amenable to subsumption under a pre-existing theory… an analysis of its [models] exploratory uses is needed to account for situations where an underlying theory is

that important experimental work can be done without having some ideas in mind about the phenomenon under investigation, and he recognises that in the observations he refers to, there were some general aims towards developing theories. In light of this, we can consider examples of thought experiments that also meet this criterion. One case would be Einstein's chasing a light beam thought experiment. In his *Autobiographical Notes,* Einstein outlines a thought experiment that he performed when he was 16. In the thought experiment, Einstein imagines himself chasing after a beam of light at the speed of light and considers what he would observe if he caught up to it. This lead to an important realisation. According to Newtonian mechanics, if we travelled at the speed of light next to a beam of light, then the light would appear frozen or stationary. Just as say, a car would appear at rest to an observer travelling in another car at the same speed. But Maxwell's theory takes light to be an electromagnetic wave, composed of oscillating electromagnetic fields. On Maxwell's theory, the speed of light has a constant value. If we experienced a stationary light beam, then the speed of light would not be constant. It would instead be relative to inertial reference frames. And so, the thought experiment exposed the inconsistency in accepting both Newtonian mechanics and Maxwell's theory. Einstein states how this was an important part of the development of special relativity. He states that in the thought experiment, 'the germ of special relativity theory is already contained' (1979, 51).

Another case, also from Einstein, is the elevator thought experiment. Special relativity had only accounted for the relativity of uniform motion, and Einstein set out to show how the relativity of motion also extends to accelerated motion. The thought experiment allowed Einstein to 'proceed to important results precisely when there is no background theory' (Norton 1996, 346). It led to the equivalence principle which was a first step in his arrival at general relativity.

This third feature of experiments brings us to yet another feature discussed by Hacking that I am yet to outline; the role of experiments in unexpected discoveries that prompts the development of new theories. This is taken up by Morgan (2003, 2005) and leads to her distinction between mere surprise and a more significant form of surprise that she labels confoundment. This argument contributes to Morgan's view that experiments have epistemic superiority over models and simulations. I will come back to this in the next chapter where I

unavailable, or where—as James Clerk Maxwell put it—it is essential 'to avoid the dangers arising from a premature theory'' (2016. 75). This practice is not completely "theory free"; scientists utilise 'significant background knowledge, including background theories' (ibid., 78). See also Franklin (2005) for an account of exploratory experiments, that is, those which are not guided by theory.

demonstrate how the surprise/confoundment distinction applies to thought experiments, before defending a view of how thought experiments can productively surprise.

In this section, I have aimed to show that thought experiments, too, have a life of their own. Consequently, this feature fails as a way to demarcate experiments from thought experiments. Rather, responding to Hacking's comments offers a fruitful way of exploring the commonalities between experiments and thought experiments. Continuing the comparison with computer simulations, it is worth noting that Winsberg has argued that computer simulations also have a life in that they evolve and can be retooled to different ends. He uses the example of the 'piecewise parabolic method' (PPM) an algorithm that can be used to simulate fluid flow. Various versions have been developed to be applicable to a plurality of physical systems and it has undergone revisions in order to increase its accuracy; 'the simulation practices for calculating discontinuous compressible flows, in which the PPM has figured prominently, has had its own independent history… it has matured, evolved, been adapted, recycled, and retooled. It has had a life of its own' (2010, 48).

As Bokulich and Frappier (2018) highlight, reflecting on how thought experiments have a life of their own leads to issues regarding the identity of thought experiments. As they can be retooled in the above sense, this raises the question of when we have an instance of the same thought experiment but in a changed form, or with a different interpretation, or when we have a new thought experiment. Depending on what a thought experiment is, the view regarding their identity conditions will vary. Take Norton's stance for example. Given that he holds that thought experiments are arguments, any example that changes in such a way that the reconstructed argument form becomes different, that is, we get different premises and/or a different conclusion, would be an instance of a new thought experiment. And so, thought experiments are immutable. Of course, we could refer to different particulars or change some of the details but this would count as the same thought experiment only as long as the changes do not affect its argument form.[18]

Questions around identity will also arise for laboratory experiments, as well as for computer simulations and models more generally. It is not just a concern for thought experiments and does not provide a way of differentiating them from these other practices in science. Because of this, while the identity of thought experiments (and whether this is established in a way that

[18] See Bokulich and Frappier (2018) for a detailed overview of issues surrounding the identity of thought experiments, and for an outline of the variety of views available on this topic.

is similar to ordinary, physical experiments) is an interesting issue, I am not going to pursue it here. My own view is that thought experiments can undergo significant changes whilst still being the same thought experiment. I agree with Bokulich that we should take a liberal stance; we should establish whether a thought experiment counts as a version of the same thought experiment by attending to any ‘resemblance of the central narratives and a continuity through historical connection’ (2001, 286).

1.5 Conclusion

I began this chapter with a discussion of the ways in which thought experiments have been compared with ordinary, physical experiments. While some have taken the “experimental” components of thought experiments seriously and consequently, argued that thought experiments just are a species of experiment, others have focused on the ways in which they differ. I set out how there is a parallel debate concerning the relations between computer simulation methods and ordinary experiments which has gained a great amount of interest in philosophy of science. Like thought experiments, computer simulations can be said to work by the “method of variation” but unlike physical experiments, they do not intervene on the world. Bringing these two debates together can shed light on the connections between the three practices. I then turned to a prominent account of the “life” of experiments as put forward by Hacking. Hacking denies that thought experiments have a life, but I outlined how he is mistaken: Thought experiments can evolve in order to become more suited to their purposes, thought experimental results can be “robust” in the sense that they withstand changes in the theoretical realm and further, thought experiments can precede theory.

In the next chapter, I continue with my comparisons between thought experiments on one hand, and computer simulations and experiments on the other. I consider two arguments for the privileging of experiments over computer simulations, the materiality thesis and the argument from surprise, and consider them in the context of thought experiments.

2. Privileging Experiments

2.1 Introduction

In this chapter, I continue with the comparisons between thought experiments, computer simulations and experiments. I present two arguments for the privileging of experiments over computer simulations. This is the argument from materiality and the argument from surprise. A major proponent of these views is Mary Morgan (2003, 2005) and given her centrality in the debate, I predominantly focus on her presentation of the arguments. I demonstrate that Morgan's reasons for privileging experiments has implications for the epistemic status of thought experiments, given the similarities between the former and computer simulations as outlined in the previous chapter.

I'll begin with the materiality thesis which argues that the difference between the object-target relation in an experiment compared with that in a computer simulation gives reason to privilege experimental results; scientists are more justified in saying that they learn about their target in an experiment compared to in a simulation. I present worries with this argument as presented by those who defend the use of computer simulations against Morgan's claims by undermining the claim that materiality (always) matters. My main focus in this chapter is the second argument that Morgan presents; the surprise argument. This is my main focus as I argue that there is something distinctive about thought experiments regarding how they bring about surprise. For Morgan, there is a particular type of surprise, namely "confoundment", that only experiments can provide. This is a type of surprise that is productive; it is potentially disruptive in the sense that it can lead scientists to reconsider existing bodies of knowledge regarding their target system. I outline Morgan's arguments before demonstrating how computer simulations and thought experiments can also surprise in this sense, and end by demonstrating the ways in which thought experiments importantly differ from computer simulations which impacts the way in which they can productively surprise.

2.2 The Materiality Thesis

The first argument I'll consider in this chapter concerns the relation between what is studied in an experiment or a simulation and its relation to the ultimate target of the investigation.

The argument is often referred to as the materiality thesis and it has various forms. A strong version is put forward by Gilbert and Troitzsch who argue that a fundamental difference between an experiment and a simulation is that in an experiment, 'one is controlling the actual object of interest (for example, in a chemistry experiment, the chemicals under investigation)', whereas in a simulation, 'one is experimenting with a model rather than the phenomena itself' (1999, 13). But we can note that this difference is not as straightforward as they assume. There are of course natural experiments which involve interacting with a system in nature but often in an experiment, it is not simply the case that 'one is controlling the actual object of interest'. Rather, an experimental system under investigation is artificial in the sense that it differs from the natural systems that is the scientists' ultimate target.

Take Winsberg's example of two physicists interested in the interaction of a pair of fluids at supersonic speeds. One physicist investigates this via conducting an experiment in a laboratory. They have 'a tank of fluid containing simple spherical and cylindrical shapes, bubbles of gas, and a physical mechanism for causing a shock wave to propagate through the tank'. The second physicist uses models from fluid dynamic theory in order to build 'an algorithm suitable for simulating the relevant class of flow problems and transforms that algorithm into a computer program that runs on her computer' (2010, 49). In each of these cases, the scientist is studying their system—the laboratory setup or the computer program—to learn about something else. If we take it that the system that the two physicists are trying to learn about are the supersonic gas jets which are formed when gasses fall into the gravity well of a black hole, then it is clear that neither scientist is controlling the actual object of interest in Gilbert and Troitzsch's sense (ibid., 52).[19]

This means that for experiments conducted in a laboratory, there is also work to be done to demonstrate that the system informs us of the world outside the experimental setup. In the process of conducting experiments and analysing their results, scientists need to show why their experimental system is informative of their ultimate target, that is, they need to establish its external validity. And this is the subject of extensive discussion in the epistemology of experiment and science studies more generally. It is not a simple case of a scientist having immediate access to the target.

[19] Winsberg also highlights that it is unclear as to what it means to "experiment with a model", given that models are abstract entities. Instead, what is "manipulated" in a computer simulation is a computer, a physical entity. See Parker (2009) for an argument that computer simulations are material experiments as scientists observe the behaviour of a material system, that is, the computer.

And so, there is a need to acknowledge that both simulations and experiments involve studying an object that "stands in" for the system that the scientist is ultimately interested in. But for Morgan, the force of the materiality claim comes from recognising a core ontological difference: the object in an experiment replicates part of the world it stands for (albeit in a way that is domesticated, that is, simpler to manipulate), whereas the object of study in a simulation only represents the world outside the simulation. And so, there is no "shared ontology" between object and target in a simulation. Guala characterises this difference by stating that in an experiment, 'the correspondence holds at a 'deep 'material' level', and the same causal processes are present. On the other hand, in a simulation, 'the similarity is admittedly only 'abstract' and 'formal'' (2002, 66).

Morgan makes a persuasive case through discussing examples of experiments and models in economics. She highlights similarities between models and experiments and argues that models (and simulations) also have a "life of their own".[20] They are 'autonomous instruments of investigation' from theories and are said to 'mediate' between theory and the world (Morgan and Morrison 1999). However, Morgan argues that the difference between the relationship between object and target in experiments compared to simulations has consequences for the epistemic status of these practices; 'The fact that the same materials are in the experiment and the world makes inferences to the world possible if not easy…The shared ontology has epistemological implications. We are more justified in claiming to learn something about the world from experiment because the world and experiment share the same stuff' (2005, 323). In other words, 'on the grounds of inference, experiment remains the preferable mode of inquiry because ontological equivalence provides epistemological power' (ibid., 326). On Morgan's view, it is more difficult to establish a computer simulation's external validity, that is, how suited the object is as a stand in for the target and how valid scientists' inferences are from the experimental setup to the natural world, compared to the external validity of a physical experiment where a scientist has a version or sample of the target phenomenon.

Morgan focuses on economic experiments, but the claim is taken as carrying over to all sciences. Take cases of learning about how a new drug may affect the human body. It appears

[20] While Morgan speaks in terms of the comparison of models and experiments, I follow Parke (2014), Parker (2009), and Winsberg (2010) amongst others and take Morgan to be clearly speaking broadly of models thus including computer simulations, and consequently, that her arguments regarding the object-target relationship to carry over.

that in such a case, testing the drug on human volunteers, or at least other mammals such as mice, may give us a better indication of the drugs' effects than if we carried out a computer simulation. As we saw in Winsberg's example presented above, in both cases (the physicist conducting the experiment and the physicist using a computer simulation) the scientist is manipulating an object that is different from the target, that is, the intergalactic gas jet. Yet in the experiment, the physicist is manipulating something that composed of fluids which 'really have different densities, and they really flow past each other at supersonic speeds' (2010, 56). The thought is that this offers assurance that in at least some ways, the results of the experiment will be applicable to the target. The computer simulation, however, lacks these shared material features with the target. As Winsberg explains the view, computer simulations can only be reliably informative 'in virtue of being suitably programmed; the reliability of the results depends *entirely* on having chosen the right model and the right algorithm' (ibid.). It appears that the materiality claim will also have implications for the external validity of thought experiments. In Guala's terminology, thought experiments are also an instance of a formal, rather than a material, analogy.

Despite its intuitive plausibility, there are many problems with the materiality argument that have been highlighted in the computer simulation literature. For example, Parke (2014) has discussed issues establishing exactly what "sharing the same stuff" means. An uncharitable way of interpreting it would be the same stuff at a fundamental level. But, Parke explains if this was all that was meant then we could say that all biological experiments would have material continuity to whatever target system in the sense that 'their objects and targets are made of carbon, hydrogen, nitrogen, oxygen, phosphorus, and sulfur' (ibid., 523). Another way is to understand the claim as meaning that there are certain degrees of object-target correspondence. At the highest degree, the object in an experience would be identical to the target system. Laboratory experiments would not meet this, but natural experiments (if we take them to be experiments) might qualify as in such cases, there is no control of the system involved. Another level of material correspondence, Parke explains, would be an 'object that is a token of the same relevant ontological category as the target, at a sufficiently fine-grained level for the purpose at hand', for example chemists using study samples of an element to learn about its properties in general or one living organism of one species to learn about another (but both say, mammals). Further removed, would be instances of say using a plastic model of the target to learn about the system.

Somewhere along this ‘various degrees of material correspondence’, we might try to locate other types of simulations, including analogue simulations. This might be a difficult task given that analogue simulations are cases in which another physical system is used to simulate a target system. Often this is because the target system is often impossible or difficult to test. For example, Dardashti et al. (2017) discuss the difficulties of testing Hawking radiation, that is, the radiation predicted to be released by black holes. Given that scientists cannot construct black holes to test this nor can the radiation be detected through observation (as it is too weak), an analogue system has been developed. Unruh (1981) proposed the use of fluids to simulate black holes because there are relevant similarities between the two, labelling the set up a “dumb hole”. He highlighted that fluids could move so fast that it would mean that sound waves would have to move faster than the speed of sound in order to escape from them. This is analogous to how black holes have a horizon in which light cannot escape.[21] A key issue regarding analogue simulations is how they relate to both computer simulations in that that they can be said to involve a ‘programming’ of a system, albeit a physical one (Dardashti et al. 2017, 74). Another central question is how they relate to physical laboratory experiments (in the sense that it is the use of a physical system that is different from the ultimate target system).

And so, it appears that there might be degrees with regard to what material correspondence consists in and one way of interpreting the materiality thesis is to claim that we should be able to locate different research practices somewhere along this continuum. But, as Parke goes on to argue, establishing how well a research programme meets object-target correspondence is really difficult. Even if we can pin down how well material correspondence has been achieved between object and target, there are other difficulties with the materiality thesis. By drawing on the case of weather forecasting, Parker (2009) highlights the sheer difficulty of setting up and carrying out a laboratory experiment in order to make temperature predictions of various cities given the complexity of the temperature structure of the atmosphere. Given their historical reliability at providing at least approximately accurate temperature predictions for the target cities, Parker argues that scientists are more justified in drawing conclusions about their target when utilising a computer simulation. This is despite the material similarity that holds between the experimental setup and the target (2009, 492). So there will be technical reasons as to why

[21] See also Crowther et al. (2017) for a discussion of analogue simulations and their role in theory confirmation.

we cannot experiment on a system at all, or with the same level of experimental control that we can with a simulation. And this is the same for thought experiments; many thought experiments cannot be realised as a physical experiment.[22]

Parker has further argued that we ought to focus on relevant similarities, rather than on material similarities. What counts as a relevant similarity is dependent upon what scientists are trying to find out about their target system: 'the relevant similarities might be formal similarities, material similarities or some combination of the two, depending on the type of experiment and the target question a hand. But, crucially, having experimental and target systems made out of the same materials does not guarantee that all of the relevant similarities obtain' (2009, 493). With reference to the weather forecasting example above, Parker explains in the case of a laboratory experiment that is materially relevant to the target system, relevant similarity can fail to obtain. For example, one feature that has to be similar to the target in the experiment is the various dimensionless parameters of fluid flow. But as some of the values of the parameters 'depends on other things other than just which fluid is being used—e.g. on such things as the depth of the fluid and the size, shape, roughness and movement of any container holding it' then this means that even when the same materials are used, there will be reason to maintain that relevant similarity has not been met.[23] Similarly, Frigg and Reiss have discussed how in some cases, 'it is structural properties that enable reliable inferences and such properties can be shared between physical systems and mathematical models' (2009, 609) They offer the example of the stability of an ankle joint. This is not dependent upon the material components of the bones but rather on the shape of the joint which can be successfully captured in a model (ibid., 610).

The upshot of this is that there will be particular situations in which an experiment will take epistemic priority over a computer simulation or a thought experiment in virtue of its

[22] Parker acknowledges that Morgan's claim could be best understood as an "all other things being equal" claim: 'Ceteris paribus, inferences about target systems are more justified when experimental and target systems are made of the "same stuff" than when they are not' (2019, 492). And so, we should understand the claim to be that if an experiment could be performed, then scientists would be in a better position to make inferences about their target than in the case of a simulation. But Parker goes on to explain the difficulties in determining what the "all else" that must be equal is. Instead of trying to suss out what this could be, Parker moves on to her other worry with the materiality claim; it is relevant similarity that ought to be focused on.

[23] See also Roush (2018) for a recent defence of the epistemic priority of experiment that does not rely on the materiality thesis and depends instead on the property of natural kinds. Further, drawing on Parker, Currie and Levy argue that the experimental object must share focal properties with the target but these do not need to be materially similar ones. For instance, a focal property might be the property of being a predator or a being a parasite, rather than being made of the "same stuff" (2019, 1072).

material similarity with the target system. It might be that this is particularly apparent in cases in which scientists do not know much about their target system. However, the notion that the materiality argument applies in general is problematic. Therefore, against Morgan, ontological equivalence does not necessarily provide epistemological power.

2.3 The Value of Surprise

Now that I have discussed Morgan's materiality thesis, I'll turn to my main focus of the chapter which is her second argument for privileging experiments. This is the surprise argument which states that computer simulations do not have the capacity to surprise us in the same way that physical experiments can. This is said to have consequences for their epistemic status. Morgan's argument is yet to be applied to thought experiments, but I'll demonstrate how it carries over before saying more about the ways in which thought experiments can surprise in a productive way.

As previously mentioned, surprise is underexplored in the philosophy of science. The way in which surprised has been spoken about can be characterised as one of the following two forms 1) surprising phenomena or 2) surprising or novel predictions. An example of the first would be Becquerel's accidental discovery of radioactivity. As Bedessem and Ruphy discuss, Becquerel was studying a crystal that contained uranium in order to assess how uranium crystals are affected by light. But he noticed something surprising; the crystal had fogged a photo plate which happened to be left next to it. This led to the hypothesis that uranium is radioactive (2019, 3). An example of the second sense in which surprise has been considered in philosophy of science would be the case of Fresnel's theory of light that was used to make the novel prediction of the bright spot (Douglas and Magnus 2013). The sense of surprise that I discuss is related to (1), that is, surprising phenomena. This is orthogonal to discussions of novel and successful predictions and their place in debates regarding scientific realism and so I will not say anything about surprise in the sense of (2).

In presenting the distinction between surprise and confoundment, Morgan focuses on the comparison between modelling and experiment in economics. However, Boumans (2012), Parke (2014), Currie (2018) and Beisbart (2018) have extended the discussion to computer simulations and their use across science. Morgan distinguishes between 'mere surprise' and 'confoundment' (or 'productive surprise' (Currie 2018)) and argues that the latter carries more epistemic value than the former. Mere surprise consists in an unexpected result or behaviour but one that can ultimately be explained by existing theory. Whereas in cases of confoundment, a type of surprise emerges that can force the scientific community to re-evaluate their theories in a radical way. As Morgan states, confoundment consists in results ‘which are both surprising and unexplainable within the given realm of theory’ (2005, 324).

Morgan claims that while both simulations and experiments can achieve mere surprise, only the latter can achieve confoundment. This is explained by highlighting key differences between the objects of study in experiments compared to those in computer simulations. The surprise argument is presented alongside Morgan’s claims regarding the *materiality* of experiments. As we have seen, Morgan grants that both simulations and experiments involve studying a system that “stands in” for the system that the scientist is ultimately interested in. But she argues that there is a core ontological difference; the object in an experiment *replicates* part of the world it stands for (albeit in a way that is domesticated, that is, simpler to manipulate), whereas the object of study in a simulation only *represents* the world outside of the simulation.

This ontological difference then underpins that between confoundment and ‘mere’ surprise via the issue of *control*: As physical experiments are said to capture or reproduce the parts of the natural world the scientist is interested in, the object in an experiment is a version of the object or system in nature. This means scientists are not in complete control of the experiment’s results. In a computer simulation, on the other hand, scientists are studying something artificial that they programmed themselves and over which they, ultimately, retain control. This thought is echoed by Sniegowski, who states: ‘Although surprises do emerge in simulations, in general what goes into a simulation is well known and surprises are not anticipated. In contrast, surprises and exceptions to anticipated results are fairly common in experimental systems’ (quoted in Parke 2014, 527).

The relation between the materiality thesis (the view that experimental results are more reliable due to the object and target being made of the “same stuff”) as discussed in the previous section and the surprise argument is worth attending to. While Morgan presents the two in the same paper, and regards them as related, it does appear that they are independent arguments. Similarly, Parke argues against the materiality claim, but then presents another way in which experiments could be privileged, that is, due to their capacity to surprise in the confounding sense, which demonstrates that she also understands the arguments as distinct. Here, I will also treat them separately, whilst acknowledging that the grounds for the claim that experiments can confound has to do with materiality. I take it that what is relevant for Morgan's surprise claim is that in an experiment, scientists are studying a material system and therefore a system that they themselves have not programmed. Hence, there is a certain freedom that the system has when compared with simulations. This is different from the claim that confoundment is rooted in the material similarity between object and target. Morgan does not discuss analogue simulations, and analogue simulations have not been considered in the existing responses to Morgan's surprise argument. But it does appear that on Morgan's view, analogue simulations can also confound since they too are a material system (granted that for Morgan, it appears there will be more difficulties in establishing their external validity). Furthermore, treating the two arguments as separate allows for the surprise argument to be considered even though there are difficulties with the materiality thesis as seen above.[24]

To see the difference between surprise and confoundment, and why Morgan takes it that only experiments can achieve the latter, I will start with outlining how simulations can merely surprise on her view. Scientists are often ignorant about their simulations and even if they know everything about the starting assumption of their models and the rules for how the system will change over time, these can be very complex, and they will not know all the consequences of the conditions that they started with. And as Morgan highlights, finding out what follows from the initial conditions is the goal of running the simulation, and sometimes what follows can be unexpected. However, she states: 'The constraints on the model's behaviour are set, however opaque they may be, by the scientist who built the model so that however unexpected the model outcomes, they can be traced back to, and re-examined in terms of, the model' (2005, 325). That is to say, in computer simulations, the model used necessitates the result of the simulation. And so, a simulation's result can be fully explained by its design and implementation and hence is under the control of the scientists' involved. Consequently, simulations cannot confound.

As mentioned, Morgan argues that it is only physical experiments (and not simulations) that can confound. This is because the behaviour of the object that is being investigated in an

[24] Currie's discussion of surprise presents Morgan's two arguments are more closely intertwined. On his view, the surprise argument focuses on results that are not only surprising, but also tell us something about our target system. One way of making this connection to the target system would be through the materiality claim, but see Currie and Levy (2019) for how Currie's view differs from Morgan. In his paper on surprise, Currie shows how simulations can also surprise in a productive sense in that they produce surprising results that bear on the target of the investigation.

experiment is not completely controlled by the design of the experiment, and therefore, when we intervene on nature, genuinely new phenomena can emerge:

> ‘Such new behaviour patterns, ones that surprise and at first confound the profession, are only possible if experiments are set up with a certain degree of freedom… [so that its] behaviour is not totally determined by the theory involved, nor by the rules of the experiment’ (Morgan 2005, 324).

There is, then, this important condition of “no over-control” in the case of experiments that have the potential to confound rather than merely surprise. In conducting a physical experiment, a scientist sets out to discover how a system will respond to an intervention. But if the system is over-controlled, then the system will not be able to react in this way. Instead, its behaviour is dictated by the set up and, as Beisbart puts it, ‘nature doesn’t have anything to say’ (2018, 187). When experimental set-ups meet this condition of “no-over control”, then scientists can gain results which are not implied by the assumptions of the set up. We can consider again the example of Becquerel and the fogged photo plates—the world can surprise us, and physical experiments, despite being tamed in a sense, maintain this quality.

To summarise Morgan’s argument: In a computer simulation, surprising results only arise because we do not have epistemic access to all the consequences of our model before we run the simulation. But knowing the facts about the experimental design in an experiment is not always enough to explain the surprising results. Even within the setting of a laboratory there can be ‘potential for independent action’ (2005, 325). And when there is, we can be confronted with new phenomena that are both surprising and, importantly, ‘unexplainable within the given realm of theory’ (ibid, 324).

As I noted above, Morgan has limited her discussion of surprise and confoundment to economics. In particular, Morgan refers to Chamberlin’s classroom experiments on how markets work. In these experiments, students were told that they were either buyers or sellers, and they were presented with cards with numbers on. These numbers represented the maximum buying price or the minimum selling price, taken from a supply and demand model. As Morgan states, the ‘numbers generated by the model, and the ‘rules’ of the market, provided some constraint on the subject’s behaviour but also allowed them freedom to trade at some range of prices’. The students moved around the room, so that buyers and sellers could engage with each other and then trade at an agreed price. The results of the experiment were unexpected; ‘the average price was lower than predicted and sales higher than predicted'. Morgan explains how these experiments were not merely surprising, but confounding: 'To Chamberlin, the recalcitrance of his results—the phenomena of a pattern of behaviour—seemed not just surprising, but sufficiently un-expected and startling enough to make him think seriously about a fundamental assumption in microeconomics. On the basis of his results, he came to doubt that there was even a tendency towards equilibrium' (2005, 326).

An interesting question is whether the difference is particularly stark when applied to economics given the complexity of people and their behaviour in decision making. And so, an issue to consider is whether some of Morgan's claims ought to be taken as domain specific and do not carry over to all of the sciences, or have more weight in certain cases over others. Nevertheless, as mentioned, those who have discussed Morgan's account have not limited it to economics and the core thought remains. That is, we lose some general openness to nature when we conduct a simulation and this can affect the type of surprise that can arise.

To summarise, the epistemic value of confoundment lies in the fact that the relevant phenomena cry out for explanation. Confounding results are disruptive in a productive way: they have the potential to force us to think seriously about our existing theories and to motivate new research in order to find a way of accommodating the surprising results. Ritson, in her discussion of novelty at the Large Hadron Collider, has a similar view to Morgan. She states 'the kinds of novelty framed as most valuable are those that violate expectations and are difficult to incorporate into existing structures of knowledge' (2019, 1). For Morgan, it is only experiments that can achieve confoundment and this is one reason to privilege the experiments over computer simulations. I will now apply Morgan's view to thought experiments.

2.4 Surprise and Confoundment in Thought Experiments

What does Morgan's distinction between surprise and confoundment mean for thought experiments? On one hand, there are clear examples of thought experiments that have produced unexpected and significant outcomes. Take Einstein's chasing a beam of light example, as presented in the previous chapter. This thought experiment exposes the surprising tensions between Newtonian mechanics and Maxwell's equations. On the other hand, however, thought experiments, like computer simulations, do not involve interaction with the world. And so, we can ask: Should the surprise that arises from thought experiments be considered a less valuable kind of surprise as Morgan argues for computer simulations?

First, I shall show that depending on the account of what a thought experiment is, and how they work, there are alternative views as to how they can surprise us (and whether they can confound). I begin by outlining what the surprise/confoundment distinction might look like on Brown's platonist view of thought experiments, before turning to Norton's account which maintains that thought experiments are arguments. I then suggest an alternative position which attends to the role that the imagination plays in thought experiments that demonstrates how they can bring about confoundment in a distinctive way.

2.4.1 Brown's View: Thought Experiments and Platonism

As I discussed in the introduction, Brown argues that there is a particular group of thought experiments that provide knowledge of the world through "transcending empiricism"; they allow us access to the laws of nature that exist as platonic entities. Brown presents Galileo's famous thought experiment against Aristotle as an illustrative example. The thought experiment undermines Aristotle's theory that heavier bodies fall faster than lighter ones by a *reductio ad absurdum*, and also generates a new theory, making it, what Brown labels, both a destructive and a constructive thought experiment. Galileo asks us to imagine attaching two balls together, a heavy cannonball and a lighter musket ball, and dropping them from the leaning tower of Pisa. What does Aristotle's theory predict? Both that the combined balls will fall faster than the heavier one on its own, as the combined object is heavier, and that the combined object will fall slower, as the lighter ball is inclined to fall slower and so, will drag the heavier body back. From this, Galileo proposes a new theory; all objects fall at the same speed.

Brown states that in this example, 'we have a transition from one theory to another which is quite remarkable. There has been *no* new empirical evidence. The old theory was rationally believed before the thought experiment, but was shown to be absurd by it. The thought experiment established rational belief in a new theory' (1986, 10). For Brown, this is *a priori* knowledge; the belief in Galileo's theory is not based on new empirical data and importantly, against Norton's argument view (considered below), it is not logically derivable from old data.

As seen, Brown takes the analogy between thought experiments and physical experiments seriously. Thought experiments involve quasi-observation; a system is represented and then observed by the mind's eye in a way that is analogous to physical experiments. And if we think about Morgan's distinction between surprise and confoundment (when applied to thought experiments) on Brown's view, it appears that platonic thought experiments can also confound. This is because, despite relying on non-experiential perception, this class of thought experiments produce results that cannot be traced back to or explained in terms of the initial conditions of the thought experiment. And so, for Brown, the insights gained from platonic thought experiments are not simply a matter of 'seeing old empirical data in a new way' (ibid., 11) (by inferring the new theory from already known data) but rather, involve genuine *discovery.*

As mentioned, Brown's platonist view has not been popular in the literature, given its metaphysical commitments. And in what follows, I will go on to explore alternatives to his way of characterising thought experiments, including the type of surprise that they can bring about. Before I do, I want to outline a similar thought regarding the links between surprise and mind-independence that comes up in the philosophy of mathematics and the ontology of scientific theories, as discussed by Simons (unpublished) and Leng (2011), and French and Vickers (2011) and French (2020a) respectively.

We saw that Brown compares his view to mathematical platonism, in particular, he refers to Gödel's platonist position. As Leng (2011) and Simons (unpublished) discuss, the phenomenon of surprise can motivate a commitment to abstract objects in mathematics. The thought is that the phenomenology of mathematical theorising, explained as a feeling of discovery as opposed to creation, is best explained by platonism which maintains that mathematicians are indeed observing a mathematical reality that lies outside human minds. Similarly, in 1905, Einstein published a paper in which he derived his famous equation E=mc^2. Over seventy years later, Popper discussed this derivation, arguing that Einstein 'must have been surprised' to have obtained the equation from the core principles of special relativity. For Popper, this surprise offers support for his view that theories are 'real' in some sense. While they are not platonic objects (as they can have a causal effect on the world, are created and are subject to change), they do exist as abstract entities in 'world 3' along with artworks and social institutions. World 3 is 'the world of products of the human mind' (Popper 1972, 299).

Popper states that theories exist as hypotheses plus their logical consequences. They cannot be mental representations as it is not possible to have a 'full understanding of all the possibilities inherent in a theory' (ibid). Instead, they must be mind-independent: 'something objective and objectively existing—an object we study, something we try to grasp' (ibid.). And so, Popper took this aspect of surprise (as in the case of Einstein's derivation) to be evidence for the reality of theories. In the same sense that the physical world includes unexpected and surprising phenomena, theories too have hidden features which can be revealed to us. Given that the surprise that arises from physical objects depend on their independent existence (as in the Becquerel case), the mind-independence of theories explains how we can be surprised when we explore and discover more things about them (French and Vickers 2011, 797). I'll come back to this below. For now, we can consider an alternative to Brown's view.

2.4.2 Norton's View: Thought Experiments are Arguments

As already noted, Norton's advocates the view that thought experiments are arguments. In answering the question of how they can have novel empirical import, Norton claims that there is 'only one non-controversial source from which this information can come: it is elicited from information we already have by an identifiable argument… The alternative to this view is to suppose that thought experiments provide some new and even mysterious route to knowledge of the physical world' (1991, 129). Thought experiments give us knowledge about the world because we draw on previous knowledge. Starting from this, an argument leads us to new beliefs, and if sound, this provides knowledge.

Norton's view is often separated into two claims. The first is a reconstruction thesis (context of justification claim): The epistemic power of a thought experiment is that of its reconstructed argument form. Norton provides support for this by reconstructing many thought experiments into arguments, noting that we can do so without losing anything epistemically, and he argues that no case cannot be handled in such a way. The second claim is about the performance of a thought experiment (context of discovery claim): Norton takes it to be the case that the conduct of a thought experiment just is that of an argument. Revisiting Galileo's thought experiment, it can be reconstructed as an argument (uncovering an inconsistency in Aristotle's physics) as follows:

(i) Natural speed is directly proportional to weight

(ii) Weight is additive

(iii) Natural speed is mediative

From (ii) and (iii), we get the negation of (i)

This reconstruction is limited to the 'destructive' part of Galileo's thought experiment. But as I outlined when discussing Brown's view, the thought experiment is also a 'constructive' dimension. That is, it introduces a new theory that all bodies fall at the same speed. Norton grants that this part of the reconstruction is more difficult than the previous steps and concedes that this is where Brown's platonist view is at its most compelling. However, Norton points out that the move to the new theory depends on the addition of an assumption (which he takes Brown to hold that we make this tacitly when we conduct the thought experiment) that 'the speed of fall of bodies depends only on their weights' (Norton 1996, 342). He goes on to outline how this assumption is problematic. Salviati (Galileo's mouthpiece) was not able to make such an assumption at that time, as it would involve assuming away the medium in which the bodies fall. This is because the 'broader focus of discussion, the very point that raised the question of falling bodies, is the possibility of the existence of a vacuum. To assume the possibility of a vacuum at this point would be to beg the main question under discussion'. And Salviati goes on to allows that the same weight of the same material, for example gold, would fall differently depending on its shape (ibid., 344). Because of this, Norton concludes that 'this final step now looks more like a clumsy fudge or a stumble than a leap into the Platonic world of laws' (ibid., 345).

In a paper with Beisbart, Norton (2012) claims that computer simulations are also arguments (see also Beisbart 2012, 2018). The thought is that computer simulations raise a parallel issue to the above question: How do they provide knowledge about a real-world target without any observation of that target? Beisbart and Norton's answer is that thought experiments and computer simulations provide knowledge in the same way: we build what we know into their construction, that is, the description of the thought experiment or the assumptions of the computer simulation, and this knowledge is then transformed through a logical process. Like Norton, Beisbart reconstructs computer simulations into arguments, and argues that their epistemic force is not thereby lost. And further, that 'the reconstructing argument is executed when a computer simulation is carried out' (2012, 419-420). I will not consider any problems with such a view of computer simulations (or any view of how computer simulations provide knowledge) but I will come back to some of the worries of the argument view when applied to thought experiments.[25]

Returning to surprise, it appears that on this view, thought experiments can only achieve mere surprise in Morgan's sense; as all of the implications of the thought experimental setup are not known, these can be surprising. In Galileo's thought experiment, if we limit the example to its destructive component, we can say that the contradiction in Aristotelian physics was already, in some sense "there"; the thought experiment simply exposed it. But still, it was surprising. However thought experiments, like computer simulations, do not involve new observations of the world and instead they transform existing knowledge. Beisbart and Norton (2012) characterise this distinction as one between 'discovery' and 'inferring'; thought experiments and computer simulations allow us to explore what we already know. But because of their lack of contact with the world, we cannot discover anything that falls outside of our existing knowledge.[26]

As French and Vickers (2011) discuss in the context of theories, Wittgenstein dismissed the value of surprise in deductive contexts. In his *Remarks of the Foundations of Mathematics* (1978), Wittgenstein asserts that mathematical reasoning does not involve discovery as on the platonist view. He argues (in contrast with conducting an experiment or more generally, interacting with mind-independent objects) that we can explain the reason why people might be surprised in cases of drawing conclusions through logical reasoning by pointing to their limitations:

> "The demonstration has a surprising result!"--If you are surprised, then you have not understood it yet. For surprise is not legitimate here, as it is with the issue of an experiment. There--I should like to say--it is permissible to yield to its charm; but not when the surprise comes to you at the end of a chain of inference. For here it is only a sign that unclarity or some misunderstanding still reigns' (ibid., 111).

Here, Wittgenstein argues that surprise can be explained away by pointing out how mathematicians (and people more generally) have epistemic limitations: As Simons explains

[25] See Lusk (2016) and Boge (2019) for critical discussions of Beisbart and Norton's view of simulations.

[26] This is not to say that Beisbart and Norton take it that we cannot gain new knowledge from thought experiments and computer simulations. They emphasise that 'the results inferred were not known prior to investigations' (2012, 409). And in the context of thought experiments only, Norton states even if a hypothesis 'is in some sense implicit in the premises, it can still be novel in the sense that it was unanticipated' (1996, 346).

'a proof is too long to keep all its steps in mind, so something is lost from an individual's view' (unpublished, 7).[27] Whereas if a mathematician could hold the whole proof in their mind then they would see how each step would follow and hence, they would not be surprised. If we were to follow this line, along with Norton's presentation of thought experiments as arguments, then there might seem to be little of any interest to say about surprise in this context. French and Vickers use Wittgenstein's dismissal of surprise to undermine Popper's view, suggesting that the reason why surprise might arise from theoretical derivations lies in scientists' cognitive limitations. They argue that own our thoughts can surprise us 'if we accept that those thoughts (or their propositional representations) have consequences we haven't deduced yet…we have certain beliefs plus certain rules for generating new beliefs therefrom, not because the new beliefs actually already exist and we *discover* them as we make our inferences' (ibid., 797). This line of thought thus allows a resistance to the view that surprise motivates a realist picture of theories as world 3 entities (or a commitment to the laws of nature as platonic entities, as on Brown's view of thought experiments).

However, it is important to note that Norton's reconstructions are not limited to deductive arguments; they can also include inductive steps, as in the example of Einstein's elevator as mentioned in the previous chapter. The first two steps of Norton's reconstruction go as follows: '(1) In an opaque chest, an observer will see free bodies move identically in case the box is uniformly accelerated in gravitation free space and in case the box is at rest in a homogenous gravitational field. (2) Inductive step: (a) the case is typical and will hold for all observable phenomena and (b) the presence of the chest and observer are inessential to the equivalence' (1991, 137).[28]

And so, Norton's view of thought experiments allow for steps that are ampliative; they go beyond what is stated in the premises. The same holds for his and Beisbart's account of

[27] Leng also draws on Wittgenstein's notion of the 'hardness of the logical must' in order to argues that any feelings of discovery in mathematical theorising can be explained by seeing them as 'sudden realisations of what had to hold, given the constraints'. These constraints are not necessarily 'imposed by an independent realm of objects about which our theories must assert truths' but rather, arise from what questions we are setting out to answer, and what options count as appropriate given these questions (2011, 64).

[28] As mentioned in the introduction, Norton's notion of logical reasoning in thought experiments has expanded over the years to include steps beyond deduction and induction to informal inferences and reasoning from analogy. Here, I will not discuss this further given the criticism that it renders the argument account 'vacuously true' (Brendel 2018, 287; see also Stuart 2016). I come back to this in chapter 3 and 4 when discussing the nature of the imagination in science.

computer simulations; these can also transform the assumptions in the model in a way that preserves the probability of truth (2012, 411).

What is the epistemic value of the surprise we can gain from thought experiments on the argument view? In contrast to Brown, Norton's view is that the information we gain through deductive and inductive inferences do not constitute genuine discovery as in the case of experiments. And Beisbart has explicitly endorsed Morgan's view when discussing the epistemic status of simulations. Beisbart offers the example of the Michaelson-Morley experiment that undermined the view that the earth has a non-zero velocity with respect to the ether. As Beisbart argues, this experiment 'has a complicated set-up, and a number of assumptions are needed to interpret its data as having implications about the ether. But this does not imply what the result of the experiment is'. If in place of this experiment, a simulation was used, it would not have confounded as there would be an assumption regarding the earth's velocity with respect to the ether in the simulation's programming (Beisbart 2018, 12). Similarly, we could say that a thought experiment would not give us the confounding result.

And so, there are clearly cases in which an experiment only could confound. Further to this, there are cases of thought experiments that may bring about mere surprise, in the sense that they result in some unexpected behaviour, but do not confound. That is to say, the result can be explained by existing theory. One example, discussed in the previous chapter, is Dewan and Beran's rockets and thread. The thought experiment draws out a physical implication of special relativity. As Bokulich explains, 'one might understand the special theory of relativity, and know the laws that this theory postulates, but still not be aware that they imply the existence of relativistic stress effects' (2001, 301). What the rockets and thread thought experiment does is makes these consequences (which otherwise might have remained hidden) explicit.[29]

However, it can also be highlighted that thought experiments can surprise in a disruptive way. That is, in a way that can force us to re-evaluate our existing theories. In the case of the reconstructed argument form of Galileo's thought experiment, although there are no new empirical discoveries being made, the scenario we are asked to imagine exposes a

[29] Another example of a thought experiment that reveals a surprising consequence, but perhaps cannot be said to meet Morgan's sense of confoundment, is Schrödinger's cat. The thought experiment is introduced to demonstrate a bizarre feature of the Copenhagen Interpretation of quantum mechanics (namely that on such a view, the cat is both dead and alive).

contradiction in Aristotelian physics. We could just dismiss this along Wittgenstenian lines, but this would overlook how the result of the thought experiment cannot be explained within the domain of the old theory and prompts the development of a new theory. A similar claim has been made in the context of simulations. Currie and Parke give examples of confounding simulations that produce results that not only go against expectations, but also 'promote changes to, or re-examinations of, explanatory resources pertaining to the target' (Currie 2017, 654).[30] This therefore puts pressure on Morgan's argument that the different sources of surprise impact the epistemic status of the feature under consideration.

I agree that there is a difference between thought experiments and computer simulations on the one hand, and experiments on the other, in that the surprise arises in a different way. However, all can be disruptive. At this point, it is helpful to drawn on Currie's account of surprise in simulations. As we have seen, physical experiments involve bringing our theoretical knowledge into contact with the world (granted the differences between an experimental object and the target) and we gain new empirical results. When experiments meet the condition of "no over control", they have the potential to bring about results that can force us to revise our theoretical knowledge. That is, they can confound. Simulations importantly differ from experiments in certain respects. As Currie states, the design and running of a simulation is a way 'of filling out, making explicit, and probing our theoretical, conceptual and empirical ideas' (2018, 656). This is still a way of generating knowledge—and can bring about productive surprises—but unlike the experiment case, it does not involve this 'contact with new empirical results' (ibid). Like simulations, thought experiments also probe our theoretical, conceptual and empirical ideas rather than generating knowledge through bringing theoretical knowledge into contact with the world. And thus, they also differ from experiments in important respects. However, I will now look to important differences between thought experiments and computer simulations which illustrates how

[30] Parke outlines the ABM (agent-based model) Sugarscape. This model had "hidden features" that were previously unknown. In this example, the surprise (in the confounding sense) 'consists in the emergence of familiar macrostructures from the bottom up…from the simple local rules… The fact that these are sufficient to generate the phenomena is the surprise. The object of study that looks very simple has generative properties that one would have never known about until studying it' which in turn prompts questions regarding the features of the object of study (2014, 531-531). Similarly, Currie outlines a simulation that uses a computer model to work out the gait of a sauropod. The simulation productively surprises (that is, confounds): 'What is surprising is how it walked: no known animal combines an elephant-like stride with knuckle-walking. So, the results went against expectations insofar as there was no expectation for that gait to emerge. It wasn't, for instance, a pre-existing hypothesis to be tested' (2018, 654). This in turn prompts reflection on the explanatory resources on the target, that is, sauropods.

they probe this knowledge in different ways. And this has implications for how they bring about productive surprises.

2.5 Thought Experiments, the Imagination and Surprise

One view which highlights some of the differences between computer simulations and thought experiments is that computer simulations are simply more complex thought experiments. Di Paolo et al (2000) characterise simulations as 'opaque' thought experiments. By this, they mean that each explore hypothetical scenarios, but in simulations this is done via a computer and scientists do not follow each step. Similarly, Lenhard (2018) has argued that due to their complexity and opacity, simulations are more likely to surprise than thought experiments. And this forms part of Chandrasekheran et al's (2012) view that thought experiments will be replaced by computer simulations. Depending on what you take a thought experiment or a computer simulation to be, such a view could supplement Beisbart and Norton's argument view.

In one sense, it seems that simulations are more transparent in that we know that they work by many simple steps. As Beisbart and Norton highlight, the imagination does not enter into the picture, which for them, can cloud the fact that ultimately thought experiments work by 'prosaic inference', and are arguments (2012, 409). However, what Lenhard means by the transparency of thought experiments is that they 'have to meet high standards of intelligibility, because the whole process takes place in cognition. If it is ever unclear what happens next, that is, if one cannot comprehend why a certain outcome should happen, the thought experiment fails' (2018, 485). This contrasts with a computer simulation, which are not opaque 'because it would be unclear how one step follows from its predecessors' but rather, 'it is the multitude of interrelated steps that can render the overall process opaque' (ibid., 489). Lenhard references Morgan, but he does not discuss the distinction between surprise and confoundment. And so, it is difficult to see whether he takes the differences between thought experiments and simulations to impact whether they can merely surprise or confound. I take his view to be that i) surprise is valuable in science, including in simulations and thought experiments (even if in these cases, it is limited to mere surprise, that is, unexpected behaviours) and ii) we are more likely to get surprising results from computer simulations than from thought experiments, due to the former's opacity.

One of the examples Lenhard discusses is Schelling's (1978) model of social segregation which shows how individual's preferences regarding their neighbours lead to segregated neighbourhoods. The neighbourhood is represented by a grid, made up of individual cells. Each agent is a member of one of two groups (say, based on race) and is located within a cell on the grid. Agents are then able to move around the grid, that is, relocate, until their preference regarding their neighbours is met. The strength of the agent's preferences for neighbours within their own group can be altered. In cases where agents preferences are very strong, or even consist in wanting a neighbourhood entirely made up of others in their own group, then segregation will happen very quickly. But as Lenhard explains, what is surprising about Schelling's model is that even when the preference for neighbours of own group-type is low—when agents are happy with neighbourhood that is a mix of their own group and the other group (but not too dominated by the other group)—segregation still occurs. Lenhard explains:

> 'What does "too dominated" mean? Reasoning does not help here, one needs to try out and actually perform a great number of iterations. In the model, one cell is inspected after the other and it is determined whether inhabitants want to move. After all cells have been checked, the process starts over. After many iterations, an equilibrium will occur and then one can see whether segregation has happened or not. The intriguing question is how weak preferences have to be to prevent segregation. This question can be answered only by exploratory trials with varying parameter values' (2018, 491).

Here, Lenhard highlights how the important aspect of Schelling's model is not that segregation can occur, but rather, how weak the preferences have to be in order for a neighbourhood to become segregated. The simulation's explanatory power is rooted in its ability to determine such parameters. Lenhard highlights that we could not get this from a thought experiment; 'everything depends on the actual range of parameters that generate segregation. This range can be determined only by a great number of iterations' (ibid.). Similarly, Chandrasekharan et al have argued that unlike computer simulations, thought experiments 'do not *naturally* support the simulation of counterfactual scenarios beyond the one generated by the specific scenario and its elements, as the mental simulation process is driver by the behavior of concrete components' (2012, 259).

In discussing the Michelson-Morley example as presented by Beisbart, we saw that an experiment only (and not a simulation) could bring about the sense of surprise that the example brings about. Similarly, we can grant that in some cases, a thought experiment in place of a computer simulation would not produce the surprising (possibly confounding) outcome and we can explain this by referring to the differences in their complexity compared with simulations. But here I want to argue that we should not be too quick to regard the surprise in the context of thought experiments as less valuable than either experiments or computer simulations.

Firstly, characterising computer simulations as more complex or opaque thought experiments misses something important about the latter. Thought experiments can be surprising despite their simplicity. Galileo’s thought experiment had such significance in the history of science yet is a simple imagined scenario, involving the behaviour of bodies being dropped from a tower. Again, a comparison can be drawn here with surprise in the context of theoretical derivations. We cannot explain away the surprise that arises from Einstein’s derivation by appealing to its complexity; Einstein’s paper in which he derives $E=mc^2$ is only three pages long.

A second difference between simulations and thought experiments involves attending to the richness of the imagination with regards to the latter, which brings about surprise in a distinctive way. Before I say more about this, it is interesting to note that the Wittgensteinian dismissal of surprise in deductions as introduced can also be linked to a claim that he makes about the imagination. According to Wittgenstein, the imagination cannot provide us with new information, it ‘does not instruct us about the external world’ because it is subject to the will (1980, §80). Whereas physical objects in the world (and our perceptions of them) are of course not up to us in this sense, and thus can turn out differently to what we expect once we further study them and find out more about them. As in the case of mathematics, Wittgenstein states that in imagining we are creating, as opposed to observing or discovering. It is not just Wittgenstein who has made such a claim. For example, Sartre claims ‘it is impossible to find in the [mental] image anything more than what one puts into it; in other words, the [mental] image teaches nothing’ (2004, 103). Likewise, White insists that ‘one can’t be surprised by the features of what one imagines, since one put them there' (1990, 92). This mirrors the view of simulations presented earlier; our imagination gives us only what we put into it.[31]

It appears that most of this discussion has been directed towards the imagistic imagination, that is, the imagination in terms of mental images (whether visual, auditory and so on). However, in her discussion of these arguments, Stock (2006) has argued that such a position, if true, would also apply to the propositional imagination, that is, imagining *that* something is the case. I turn to the nature of the imagination in science in the following chapter and say more about this distinction there. But the thought is that if the contents of our imaginings are up to us, and further, are "transparently accessible" then it means that you cannot be mistaken and hence cannot be surprised by such imaginings, whether they are of a propositional or of an imagistic nature.

We have already seen how logical reasoning in the case of simulations and arguments can bring about surprise and thus we can put pressure on the thought that our mental contents and their consequences are fully accessible to us. Further, Stock (2006), Todd (2020) and Kind (2016) have offered responses to the claim when thinking about the role of imagination in learning more generally. And through focusing on scientific thought experiments we can see that it is just not obviously the case that we have clear access to our imaginings and the connections between them. Let's return to Galileo's thought experiment. So far, I've analysed the thought experiment through the lenses of both Brown's platonist and Norton's argument account. While I have granted that deductive reasoning can lead to productive surprises, I'll now show how Norton's reconstruction of the destructive component of Galileo's thought experiment is not as straightforward as he sets out.

Gendler has complicated the issue of whether the thought experiment shows Aristotelian physics to be inconsistent. This is because, she argues, it is unclear whether all the propositions in the reconstructed argument form ought to be considered part of Aristotle's theory. In particular, it has been asked why we should consider (iii) as part of the theory—that natural speed is mediative, or more specifically that 'Natural speed is a property such that if a body A has natural speed 1, and a body B has natural speed 2, the natural speed of the combined body A-B will fall between 1 and 2' (1998, p. 404). Without this assumption, the inconsistency claim is unfounded. As a result, there are various logically possible ways

[31] Here, I just offer a brief idea of what the claims surrounding surprise and imagination consist in. Stock (2006) offers a thorough investigation, breaking down the views into separate claims in order to show how they do not succeed.

out for the Aristotelian. For example, they can ask—are the bodies that are tied together one object or two? If one object, then it will fall at the speed that is proportional to the combined weight. Indeed, it has been argued that an Aristotelian could have chosen this option—as there is no commitment at the time on this issue (Vickers 2013, 196).

Gendler takes from this the idea that the thought experiment is indispensable; it cannot be reconstructed into an argument as on Norton's view without losing its demonstrative force.[32] This suggests that the imagination allows kinds of jumps that cannot be accommodated within the framework of more formal reasoning. A crucial issue in debates regarding imagination and learning revolves around whether and how the imagination can be appropriately constrained in order for us to gain insights about the world. The point I want to emphasise here is not that the imagination is totally unconstrained when it is fruitful in science but rather that the imagination can allow for reasoning that is less restrictive than that in arguments or computer simulations and this has the potential to bring about productive surprises.

In presenting Morgan and Beisbart's arguments, I outlined how experiments' capacity to confound is dependent upon them meeting the condition of no over-control. Whereas in the examples of confounding simulations presented by Currie and Parke, we see that simulations' capacity to surprise does not depend on such a condition. As Currie explains, simulations 'do not need freedom to produce [productive] surprise. Rather, careful control allows us to bring our ideas and hypotheses together, and it is in these new combinations that new knowledge arises'. While I argued that thought experiments also 'probe' existing knowledge, there is a sense in which attending to their use of imagination demonstrates how like experiments, thought experiments (at least in some cases) should not be too-over controlled. While of course thought experiments in order to be successful necessitate us to control our imaginings in the right kind of ways[33], Gendler's discussion of the Galileo example shows that if we restricted our reasoning in thought experiments to deductive or inductive steps from initial set up of the scenario, then we would not get a result that has

[32] For Gendler, the result of Galileo's thought experiment is justified because it taps into our previously unarticulated knowledge regarding the world. In this sense, her account denies the claim that mental images (or imaginings more generally) are constituted by the person who is imagining (which was key to Wittgenstein's scepticism) since the background beliefs that contribute to the formation of the imagining come from the world and the imaginer's experience of it, rather than solely from the imaginer themselves.

[33] In the next chapter, we will see how accounts that emphasise the potential epistemic value of imagination highlight how it has to be constrained in order for it to be of value.

been hugely productive in the history of science. As Stuart states in the context of constraints on imaginings in science (and their limits): ‘Either we reject Galileo’s use of imagination as part of sanctioned scientific progress, or we reject the idea that imagined scenarios should not always follow the rules of logic’ (forthcoming, 11).

Understanding thought experiments as merely involving argumentative reasoning fails to fully capture their potential to productively surprise, a feature that at least partly characterises their role in scientific practice. Importantly, attending to the role of the imagination gives us a way of capturing this sense of surprise without committing to Brown’s platonist view of thought experiments.

2.6 Conclusion

This chapter has presented two arguments for the privileging of experiments over computer simulations: The materiality thesis and the argument from surprise. I demonstrate how each can be applied to thought experiments, given their similarities with computer simulations. While the materiality thesis may be persuasive, I outlined how it ultimately fails. To begin with, it is unclear how we might go about assessing how materially similar an experimental object is to its target system. Further, there are clear cases in which the relevant similarity is not a material one. Hence, experiments cannot be automatically privileged for this reason. I then turned to my main focus; the argument from surprise. This is the view that experiments can “confound” that is, surprise in a productive way, whereas computer simulations can merely surprise that is, display unexpected behaviours. I outlined how this argument applies to different views on thought experiments, before demonstrating that we need to look to the ways in which the imagination plays a role in thought experiments in order to understand how thought experiments can be a distinctive source of productive surprise. While thought experiments can be said to “probe” our existing knowledge, the way in which they do so cannot always be characterised in terms of a process of deductive or inductive inferences.

To end, I want to return to a statement from Hacking as presented in the previous chapter. Hacking regards thought experiments as “icons”. He argued that while they might be repeated at different times throughout history, ‘what they think was once thought’. I have outlined how, against Hacking, thought experiments have a life of their own. They are not “fixed”. Rather, they can be developed and refined, and retooled for different purposes.

While I disagree with Hacking's claims, I do think it is interesting that he draws the connection with icons.[34]

Icons are of great significance within various religions and depict holy people, such as Jesus or Mary, or a scene as described within a religious text. Icons are importantly simple, and have been used throughout history to convey parts of religious orthodoxy to a broad audience. Furthermore, in depicting holy people or events, the thought is that icons also maintain a certain connection with them. As Kenna explains, an icon is a '"true likeness" in an analogical sense, and what it is true to is not the outward form but the "sanctity and glory of its prototype" in which it shares' (1985, 349). That is to say, icons are not just images of the divine, but rather are meant to give both access to, and knowledge of, the divine.

In chapter 5 and 6, I discuss the relations between thought experiments and literary fictions. As we shall see, a view put forward by Egan in the context of these relations argues that we read thought experiments in "allegorical" terms; each of the concrete elements of a thought experiment "stands in" for a more abstract concept. He states that the 'concrete story in the text is like a map whose purpose is exhausted by helping us see our way more clearly through the abstract territory it maps' (2016, 144). Icons also share this quality. For example, colour is often used in a symbolic way. For example, the use of gold in Byzantine works which 'imbues the figure of the saint or the surroundings with God's power and grace' (Kenna 1985, 352) or the use of colour of Mary's clothes; her blue cloak which symbolises purity and was also the colour of Byzantine royalty, which is worn over red which symbolises her devotion to motherhood. Hence, they rely on their audience knowing how to "decode" these various elements.

While Hacking describes thought experiments as "icons" to deny their life, I argue that Hacking's paper actually reveals some crucial aspects of their value which I will now go on to explore. The life of thought experiments and hence their epistemology is bound up their appeal to the imagination. I turn to the imagination in the next two chapters. Secondly, there

[34] And at the end of the paper, Hacking also mentions works of art. In considering how the same thought experiment can be replicated 'from generation to generation' Hacking states: 'But is this life? Is it not like acting a part in a play? Olivier brought new life to Othello, but no matter how profound the power of Othello or Lear, those characters do not have a life of their own' (1992, 307). It is interesting that Hacking uses a play to make this point. We might think that artworks such as paintings would be more useful to Hacking's aims given that Shakespeare's plays are not just repeated but undergo vastly different performances in order to emphasise various themes. In this sense, plays seem to very much have a "life" but I will leave this for another time.

are interesting connections to be made between thought experiments and artworks. In chapters 5 and 6, I discuss the comparisons between thought experiments and literary fictions, including allegories such as fables and parables.

3. The Scientific Imagination

3.1 Introduction

The previous chapters focused on the ways in which thought experiments compare with other practices such as experiments and computer simulations. In considering the value of surprise in science, I argued that attending to how thought experiments call upon our imaginative capacities shows how thought experiments can productively surprise in a way that differs from experiments and simulations. In this chapter and the one that follows, my aim is to build a new account of the imagination in the conduct of thought experiments. I begin with an overview of discussions of the imagination in science including reasons why we might be sceptical that the imagination has a part to play in the scientific realm. The most prominent accounts in philosophy of science that make reference to the imagination are those that understand modelling as a process of creating and analysing fictions, and some accounts extend that analysis to thought experiments. While there is a broad literature on fictions and make-believe in science, an important issue has, for the most part, been neglected. This is the question of what kind of imagination is involved in scientific modelling and thought experiments. Answering this is my central concern in the two chapters that follow.

To do this, I utilise discussions of the imagination in philosophy of mind and aesthetics in order to provide a starting point for thinking about the different forms the imagination can take. As mentioned, explicit discussions of the nature of imagination in thought experiments (and modelling) are limited, but I demonstrate that the default position takes the imagination as imagistic. That is to say, when engaging with thought experiments, we form pictures in our mind of the scenario described. I then discuss a recent paper that addresses this question head on, and provides an alternative to the assumed view. This is Salis and Frigg's propositional account which sets out to demonstrate that imagery plays no important role. This account is embedded within a "models as make-believe" view that offers a comparison between models and thought experiments on one hand, and works of literary fiction on the other, by drawing on Walton's framework of mimesis as make-believe.

The aim of this chapter is to offer an overview of current discussions of the imagination in science. In chapter 4, I then move beyond these to outline my own view of the nature of the imagination in science. While I agree with Salis and Frigg that we ought to include the propositional imagination in our analysis of the scientific imagination, I do not think this tells the whole story. Consequently, I present a different proposal. I argue that we ought to reject monistic approaches to the type of imagination involved in scientific thought experiments, and instead adopt a pluralist stance. On this view, we need to attend to the function of the thought experiment in order to address what kind of imagination the example invites. I outline a number of examples to demonstrate that my pluralist stance best captures the ways in which scientists use thought experiments for a range of purposes.

To begin, we can think more generally about connections between science and imagination.

3.2 Science and the Imagination

Discussions of the imagination have been limited in philosophy of science. Why might this be the case? One reason is that we typically associate the imagination with creativity and the arts, practices that we often regard as standing in sharp contrast with science. We can imagine things that we know are not the case; we can escape and explore other worlds, whether in the context of our engagement with fictions such as novels and films which invite us to imagine alternative worlds and characters and their lives or by playing games of make-believe as children. Furthermore, we celebrate artists and their abilities to richly conjure up new worlds and ideas. The imagination as characterised in this sense, that is, as aiding us to look beyond the world as it is, is part of why we value our imagination, and the imaginations of others, so highly.[35] We can note that 'creative' is often used interchangeably with 'imaginative'.[36] This connection can be traced back to Kant who took genius to consist in an imaginative activity that is not restricted by rules. Genius is 'a talent for producing that for which no determinate rule can be given, not a predisposition of skill for that which can be learned in accordance with some rule, consequently that originality must be its primary characteristic' (1781/2000, 186).

Kant argued that scientists, unlike artists, cannot fall under the category of genius; scientific discovery radically differs from the creative acts that he attributes to the arts. Kant claimed

[35] This kind of imaginative activity has been labelled the transcendent use of imagination and has been contrasted with so-called instructive uses, that is, those that enable us to learn something new about the world (Kind and Kung 2016, 1).

[36] See Gaut (2003) and Stokes (2016) for discussions of the links between imagination and creativity.

that Newton, for example, while clearly a ‘a great mind’ could ‘make all the steps that he had to take, from the first elements of geometry to his great and profound discoveries, entirely intuitive not only to himself but also to everyone else and thus set them out for posterity quite determinately’ (1781/2000, 187). Here, the thought is that the process of scientific discovery can be outlined in a systematic way, following step-by-step methods and rules and hence, is something that can be taught and passed on to others. Whereas ‘one cannot learn to write inspired poetry’ (ibid. 187-188). For Kant, the faculty for creativity in art is innate rather than something that can be learned. Further, unlike the process of scientific discovery, artistic creativity cannot be explained to an audience of artworks, nor is it intuitively accessible to the artist themselves.

The issue of what creativity consists in is widely disputed but many have accepted the notion that creativity and imagination are closely linked. While there has been important work on heuristics in philosophy of science (see Schickore and Steinle 2006), the Kantian view of how scientific discovery works has long been discarded. Further, contemporary accounts of creativity tend to broaden their scope beyond artistic geniuses and recognise that creative activity is also of significant import in science (Wenzel 2018). We clearly praise scientists for their innovative uses of imagination that lead to important and often surprising discoveries and insights. Take, for example, Friedrich Kekulé’s vision of ouroboros, the symbol of a snake holding its own tail in its mouth, which he claimed led him to the idea that the structure of a benzene molecule is ring-shaped. Or Nikola Tesla’s description of the creative processes that led to his inventions that involved visualising his constructions and augmenting them in his mind’s eye, without the aid of physical drawings of their design (Kind 2016, 154).

So there seems to be at least some role for imagination in science; the imagination can be crucial in the process of coming up with theories, can lead to interesting lines of research and can contribute to the formation of new technologies and inventions. But perhaps the reason why philosophers have been dismissive of the import of imagination in science is because it clashes with how we think scientists make epistemic progress, and what we take scientific methodology to consist in.

One way in which this can be articulated is by appealing to the renowned distinction between the context of discovery and the context of justification as introduced by Reichenbach (1938). This distinction separates questions about how new ideas are generated from questions about how the validity of those ideas are assessed. According to those who adopt the distinction, issues surrounding the context of discovery may be interesting for psychological or sociological studies of science, but not for philosophical ones. This is because such issues do not have any impact on the epistemic value of those ideas generated. We can see this expressed in Popper's work:

> 'The initial state, the act of conceiving or inventing a theory, seems to me neither to call for logical analysis not to be susceptible of it. The question how it happens that a new idea occurs to a man—whether it is a musical theme, dramatic conflict, or a scientific theory—may be of great interest to empirical psychology; but it is irrelevant to the logical analysis of scientific knowledge' (1934/2002, 7-8).

For Popper, justification ought to be bracketed from any references to the mental activities that produced the hypothesis in question as the latter are not susceptible to logical analysis. Discovery is thus subject matter for psychology, rather than philosophy. We saw that Kant overlooked the imagination (in the creative sense) in science as he took discovery as being a rational procedure that differed from creative genius in the arts. Those who adopt the distinction between the context of discovery and the context of justification, dismiss creativity for the opposite reason: It is because scientific discovery is not subject to rational analysis that it cannot be the proper subject matter for philosophy. Scientists may conjure up different theories and ideas about the world in their imaginations. However, this alone gives no reason to think that any of these are true. In the same sense that I can happily fantasise about being in the winning team of the World Cup but this does not give me any reason to believe that this has or will actually happen.

And so, even if the imagination sometimes plays an essential role in science, say in coming up with new scientific theories or raising possibilities, ultimately these must be tested and evaluated before we can claim that they that carry any evidential significance. That is, before we can say they are justified. In light of this, philosophers of science tended to focus on the features of scientific practice that follow the generation of new ideas (including the involvement of the imagination), emphasising that justification should be the central concern.

However, in the second half of the 20th century, the discovery-justification distinction came under scrutiny. There is a huge literature on the distinction and its problems.[37] For my purposes, it is enough to note that in light of historical and sociological studies of science, as

[37] See Schickore (2018) for an overview.

well as attending to scientific practice, philosophers have emphasised the difficulties in separating the features involved in discovery from those that are involved in justification, claiming that this shows how the distinction collapses (Kuhn 1962, Nickles 1980, Post 1993). Another way in which it has been challenged is to demonstrate that discovery is not a wholly unconstrained realm, either by attempting to set out a logic of discovery or by demonstrating that it is at least a process that can be subject to analysis. And this can be done whilst maintaining that the Kantian view is overly simplified and ought to be rejected. As a consequence, discovery becomes a legitimate topic for philosophers of science. Many of those who have provided accounts of discovery have utilised developments in cognitive science to do so (Nersessian 2009, Thagard 1984).

Furthermore, attitudes towards imagination in science are beginning to change. In recent years, we have seen books dedicated to exploring issues that connect art and science and often make references to the role of fiction and imagination in thought experiments and models (Frigg and Hunter 2010, Frappier et al. 2012, Bueno et al. 2018, Ivanova and French 2020). There has also been a volume dedicated purely to investigating the scientific imagination edited by Levy and Godfrey-Smith who, in the introduction to the collection, state how the imagination is important in 'conceiving new theoretical ideas, in exploring the explanatory resources of these ideas, and in working out how to bring theoretical ideas into contact with empirical constraints' (2020, 2). Additionally, there has been a boom in studies of the imagination's epistemic role outside of philosophy of science, including the ways in which it can aid decision-making and provide a way of acquiring modal knowledge (Kind and Kung 2016, Gendler and Hawthorne 2002).

There has therefore been an increase in philosophical attention towards the epistemic value of imagination. In the next section, I'll outline some of the discussions in which the imagination has received most attention in philosophy of science. This is via accounts that regard models as imaginary entities.

3.3 Models as Fictions

Modelling is a central part of scientific practice. Models can be material—as in cases of ball-and-stick models used in chemistry, or the hydraulic scale model of the San Francisco Bay (Weisberg 2013). But it is the use of theoretical models that have particularly sparked interest from philosophers of science.[38] In such cases, it is argued, the "face-value practice" of modelling is one of scientists describing and engaging with "missing systems" (Thomson-Jones 2010). For example, frictionless pendulums, models of the solar system in which the planets are perfectly spherical, predator-prey models where the system has just two species, and the assumptions of ideal, rational agents in economic markets. Scientists often talk about these idealised and simplified models as if they were real, concrete systems, yet they engage with them knowing full well that they do not describe anything that actually obtains in the world. So a question arises: How do we make sense of this aspect of scientific practice?

A set of views have attempted to answer these questions by comparing models with fictions in art. Traces of this view have been present for a long time in philosophy of science, such as in the work of Vaihinger's philosophy of the "as if" (1911/2004), and more recently in Cartwright's *How the Laws of Physics Lie* (1983) in which she offers an analysis of models (including Galilean thought experiments) as works of fictions, especially fables and parables.[39] A helpful summary of the approach is provided by Frigg, who claims:

> 'The core of the fiction view of model-systems is the claim that model-systems are akin to places and characters in literary fiction. When modelling the solar system as consisting of ten perfectly spherical spinning tops physicists describe (and *take themselves* to be describing) an imaginary physical system; when considering an ecosystem with only one species biologists describe an imaginary population; and when investigating an economy without money and transaction costs economists describe an imaginary economy. These imaginary scenarios are tellingly like the places and characters in works of fiction like Madame Bovary and Sherlock Holmes. These are scenarios we can talk about and make claims about, yet they don't exist' (Frigg 2010, 101).

There are various versions of this approach but the core agreement is that understanding models as fictions helps us accommodate the face-value features of modelling as described above. The views are set against alternative proposals such as those who take model systems to be abstract objects. These systems are defined by scientists' model descriptions, and the system is taken to represent a real system (Giere 1999, 2004). The abstract object view has been popular in philosophy of science, with different variations provided. For example, some

[38] Although see Toon (2012, 2016) for a discussion of the imagination in physical modelling.

[39] I come back to the comparison between thought experiments, and fables and parables in chapter 6.

have offered a more formal approach than Giere in which models are set theoretic structures (Suppes 1960) which has been adopted by the semantic approach (French and Ladyman 1999), or they are understood as mathematical structures more generally (Weisberg 2013). Such accounts, it is argued, fail to capture the ways in which scientists talk about models. In modelling, it does not seem that scientists are simply going through various mathematical calculations in their mind. Rather, they are engaging their imaginations in order to contemplate the concrete scenarios that the model describes (Toon 2016, 454).[40]

Those who advocate for the fiction view of models are not merely highlighting that the systems we imagine are idealised or depart from accurate representations of the world. Rather, the claim is that modelling involves engaging with fictions in a way that is analogous to our engagement with fictions in art, such as reading literature and watching films. A prominent version of the models as fiction view takes models (and/or thought experiments (Meynell 2014; Salis and Frigg 2020)) as examples of make-believe, and draw on Walton's (1990) framework developed in the context of representation in art. Make-believe is 'the use of (external) props in imaginative activities' (Walton 1990, 67). For example, the book Harry Potter functions as a prop that mandates the games of make-believe associated with the wizarding world and its characters. Applied to scientific models and thought experiments, the view is that models or thought experiment narratives function as "props" that guide imaginings.

The prescriptions to imagine licensed by a work of fiction fix the content of the fictional world. Some of what is "true in the fiction" will be explicitly stated by the text (these are the primary truths), whereas other truths are implied. For example, we take it as "true in the fiction" of Harry Potter that Hermione has blood in her veins, even though this is never made explicit in the book. When we are engaging with fictions, we "fill in" the gaps of the fictional world. We do this through principles of generation. For example, the Reality Principle authorises the use of real-world truths to fill in the background of a story.[41] This operates in the case of Hermione; we rely on real world truths about humans to fill in the background of her character. As it is never stated otherwise, we assume that she has blood in her veins. Or, in the cases where parts of the story are set in London, we take it to be true in the fiction that London is the capital of England even if it is not stated in the work. Another principle that

[40] Although see Weisberg (2013) and French (2020a, 2020b) for discussions of imagination in abstract model views.

[41] For a detailed discussion of this principle and its limits in the case of fiction see Friend (2017).

Walton highlights is the mutual belief principle according to which we are prescribed to imagine the content of any of the mutual beliefs of the author's society (unless explicitly excluded). For example, Salis and Frigg state that many of the implied truths of Dante's *Divine Comedy* 'are generated from the primary truths of the story and the medieval belief in the main tenants of the Ptolemaic geometric system' (2020, 35-36). It is important to highlight that Walton emphasises how these principles are not limited to these (1990, chapter 4). For instance, the genre of a work might affect what we take to be the implied truths; 'experienced readers of whodunnits immediately recognise that the obvious suspect is not guilty' (Friend 2020, 112).[42]

The Waltonian framework applied to science means that in the case of models and thought experiments, scientists fill in the gaps and there is general agreement regarding what these inferred propositions are. As Frigg states it is 'true that the Newtonian model-system representing the solar system is stable and that the model-earth is moves in an elliptic orbit; but none of this is part of the explicit content of the model-system's original specification' (2010, 102). Further to this, if we imagine something that departs from these (explicit or implied) prescriptions, then this constitutes a misuse of our imagination in this context. For example, while I am able to imagine that Hermione's skin turns blue whenever she is angry, this would not be an imagining that is authorised by the novel. Some imaginings are therefore authorised, while others are not (Walton 1990, 51). Similarly, while predator-prey models allow for some variation with regards to what particular species are imagined (for example, I could choose between foxes and rabbits, bears and fish, or lions and zebras), it would be a misuse of my imagination if I imagined, say, foxes as the prey and zebras as the predator (Friend 2020, 114). Make-believe is thus a constrained activity and proponents of this view highlight how this makes it appropriate for science. The practice of modelling (or the conduct of a thought experiment) is a matter of finding out the implied truths of the scenario described—that is, finding out what follows from the explicit prescriptions.

The Waltonian accounts of modelling are often divided between direct and indirect views.[43] According to the former, when scientists represent the world they do so in an indirect way,

[42] I revisit genre in chapter 6 where I consider the view that thought experiments are a genre of fiction.
[43] See Toon (2016), Frigg and Nguyen (2016), Friend (2020) and French (2020a) for recent criticisms and responses of the two approaches. Other views outside of the Waltonian framework can be considered direct or indirect. For example, Giere's abstract object view is an example of the latter. I do not take a position on this debate here, nor do I set out to defend the Waltonian view of modelling in general.

via a model system. For example, Frigg argues that in reading the description of the Newtonian model of the solar system, ‘we imagine an entity which has all the properties that the description specifies. The result of this process is the *model system*, the fictional scenario which is the vehicle of our reasoning: an imagined entity consisting of two spheres, etc.’ (2010, 133). Once the model system is imagined it is then connected with the target system. For example, the scientist might specify that ‘the sphere with mass m_e in the model system corresponds to the earth and the sphere with mass m_s to the sun’ (ibid., 134). Then the scientist can translate facts about the model system into claims about the target system (ibid., 135).

In contrast, on direct views such as those proposed by Toon (2012, 2016) and Levy (2012, 2015), models such as the Newtonian model of the solar system represent their targets directly rather than via a model system. That is to say, the model prescribes us to imagine things about the actual solar system such as imagining that the sun is perfectly spherical. For the direct view, model descriptions are thus ‘descriptions to imagine the world in a certain way’ (Toon 2016, 458). One way in which the difference between the two views can be seen is by highlighting that those who defend the indirect view often compare models to fictional characters such as Madame Bovary or Sherlock Holmes, as seen in Frigg’s outline of the position above. Whereas on the direct view, models are more like works of historical fiction in the sense that the latter represent real places and people. For example, Toon compares model descriptions to works such as Graves’ *I, Claudius* which prescribes us to imagining propositions (whether true or false) about the actual emperor Augustus (ibid.).

In this section, I have outlined views that highlight how models are fictional or imaginary entities, some of which have applied a prominent view of representation in the arts to shed light on the practice of modelling. While this view outlines how modelling involves props that prescribes imaginings, there is still a lot to be said about what these imaginings actually consist in. In the beginning of *Mimesis and Make-Believe,* Walton himself admits the sheer difficulty in pinning down what the imagination is. He asks: ‘What is it to imagine? We have examined a number of dimensions along which imaginings can vary; shouldn’t we now spell out what they have in common?—Yes, if we can. But I can’t’ (1990, 19). And so, the question I want to focus on for the remainder of these two chapters is the following: What is meant by “imagination” in the scientific context? This question, I argue, is central to discovering the ways in which the imagination functions in science. To make progress on this question, I will now turn to discussions on the varieties of imagination.

3.4 The Varieties of Imagination

It is widely recognised that the term ‘imagination’ picks out various acts and is too broad for a single characterisation. Given what Kind (2013) refers to as the ‘heterogeneity’ of the imagination, we can consider what different forms it can take. In philosophy of mind, there have been many different suggestions as to how the imagination can be categorised. To take just a few examples, some have distinguished between spontaneous imaginings, those that occur without conscious effort, and deliberate imaginings, those that require such effort (Walton 1990). Others have distinguished between dramatic imaginings, those that involve adopting a perspective (including emotional responses) and hypothetical imaginings, those that involve seeing what follows from a given proposition (Moran 1994). Yet another distinction, as mentioned above, is between transcendent imaginings, those that are involved when we engage in fantasy and escape from the world, and instructive imaginings, those that are constrained in such a way that enable us to learn about the world (Kind and Kung 2016).

It is not always clear whether and how the various distinctions between kinds of imaginings overlap with one another. This is further complicated by the fact that the applications of imagination in philosophy are vast. Imagination has been brought in to explain certain features in ethical reasoning, modal epistemology, mindreading (predicting and explaining others’ mental states), metaphors, fiction, and of course, thought experiments and models. To avoid confusion, I will limit myself to certain ways in which different forms of imagination have been distinguished which are most useful for thinking about the nature of the imagination in thought experiments and models.

One way to categorise the imagination, which follows the distinction Salis and Frigg present in their account of the scientific imagination, is to distinguish between 1) ‘propositional imagining’ and 2) ‘objectual imagining’ (Yablo 1993). Take the example of imagining a vase of flowers on a table. If we propositionally imagine this, then we imagine that there is a vase of flowers on a table. Propositional imagination consists in imagining *that* something is the case and does not require us picturing in our minds the flowers on the table. This kind of imagination, sometimes also referred to as the cognitive imagination, is similar to other propositional attitudes such as belief and desire, but it is the comparison with belief that is most relevant to the scientific cases.[44]

We can compare propositional imagination to imagination in the second sense, that is, objectual imagination. Here, we imagine a vase of flowers on a table by forming an image in our minds of the flowers on the table. This kind of imagination is often defined as being "perception-like" or having a "quasi-sensory" character. The thought here is that imagining in this sense has a phenomenal quality that is similar to ordinary perception. This kind of imagination is not limited to visual imagery; we can have mental imagery that correlates with other sense modalities as well. For example, I can imagine the smell of the flowers, or I can imagine a certain piece of music in this sensory way.

This should be distinguished from cases in which an imagining is about a perceptual experience, that is, imagining seeing-X, or imagining hearing-X, imagining feeling-X and so on (Arcangeli 2019, 4). For example, imagining *seeing a vase of flowers on a table*, or imagining *hearing a dog barking*. In these cases, imagery has perception as part of its content. But mental imagery does not require this. Rather, the key feature is that imagery has a character that is similar to perception, whether visual or auditory and so on (Currie and Ravenscroft 2002, 27). And there is a propositional analogue here too. We could imagine *that* we are seeing the vase of flowers. This would be an instance of the propositional imagination as perception is part of the content of the imagining, as opposed to the imagining being perception-like. Similarly, imagining in a way that is belief-like (in the case of the propositional imagination) can be distinguished from imagining believing-X which would be a case in which belief is part of the content of the imagining.

In propositionally imagining that there is a vase of flowers on the table, we imagine in a way that is similar to believing that there is a vase of flowers on the table. This is similar

[44] There is also a rich debate regarding the relation between propositional imagination and supposition given that the latter also involves considering a hypothetical. For some, such as Salis and Frigg (2020) and Arcangeli (2018), supposition is a type of imagining. Others argue for a discontinuity between supposition and imagining, including Balcerak Jackson (2016) whose work I look at in the next chapter. Whether or not supposition counts as a form of imagining, it has been differentiated from propositional imagining for various reasons. For example, some have claimed that supposition does not usually bring about affective responses whereas the imagination can (Weinberg and Meskin 2006). Others have argued that it is not subject to imaginative resistance. For example, while I can suppose that something morally reprehensible is morally right, I cannot imagine it to be (Gendler 2000). Others have claimed that supposition is less constrained; we can suppose, but cannot imagine, contradictions. See also French (2020b) for a discussion of supposition and imagination in science.

phenomenologically speaking, but it is also different. That is, we (at least usually) recognise the difference between having a belief and having a propositional imagining. And similarly for perception; our objectual imaginings are phenomenally similar to when we perceive something, but we recognise the difference between actually seeing, and having a mental image of, the vase of flowers. Thus, the qualifier that it is "quasi" sensory. There is therefore a distinctive phenomenology to imaginings compared to their counterparts.

Imaginings are also often differentiated from these other attitudes by highlighting how it has a voluntary nature. That is to say, there is a certain freedom to our imagination (as seen in the creative use of imagination presented at the beginning of the chapter). We may not be able to choose to imagine whatever we want to, but we are free to imagine things that we know are not the case.[45] For example, I can happily imagine that right now there is a pink elephant sat next to me as I write this. This clearly differs from perception in which I do not have the same kinds of choice regarding what I want to see and hear and so on.[46] And neither can I choose what I believe; belief (when it is working properly at least) is evidence sensitive. It makes no sense to say that I choose to believe that there is a pink elephant sat next to me.

This freedom is clearly part of the value of our imagination and is what allows us to create and engage with the types of scenarios that scientific models and thought experiments present to us. But it is also part of why there might be scepticism regarding the use of our imaginations to inform us about the world. If the imagination is not by nature reality-orientated (in that I can choose to imagine things that I know are not the case) then we might worry that it just is not in the business of providing us with knowledge. As a response to this, those who have defended the role of imagination in acquiring knowledge have pointed out how we ought to (and can successfully) constrain our imaginings if it is to produce anything of epistemic worth. As seen, the need for constrained imaginings in science is part of the motivation for a Waltonian view. As Amy Kind puts it:

[45] There are various involuntary constraints on our imagination. For example, embodied constraints such as in Nagel's (1974) classic example of it being impossible to imagine what it would be like to be a bat (see Jones and Schoonen (2018) for a discussion of embodied constraints on imagination). Another case is the phenomenon of imaginative resistance as mentioned above in which we cannot imagine certain things to be true. For instance, as indicated in the above footnote, Gendler (2000) gives cases of difficulties in imagining morally deviant worlds as morally good. Finally, there are discussions regarding the imagination's role in modal epistemology which takes imagination to be constrained by possibility in a significant way (see Gendler and Hawthorne (2002)).

[46] There are of course cases in which we can shift our perception to see say, the same image in different ways, such as in duck-rabbit images, or Necker cubes.

> ‘when we constrain our imaginings to fit the facts of the world as we know them, we are using an epistemic procedure that is much more akin to scientific experimentation than it is to mere flights of fancy. Although our imaginative experimentation will not be fool proof, neither is scientific experimentation. But in both cases, when we proceed cautiously, the beliefs that we arrive at will...usually be justified’ (2018, 244).

Now that we have an idea of the different forms the imagination can take, in particular the difference between propositional and objectual imaginings, we can now turn to existing views of the imagination in scientific thought experiments.

3.5 Thought Experiments and Imagery

It is clear that thought experiments appeal to our imaginative capacities. Thought experiments often begin with an invitation for us “imagine”, “consider” or “suppose” followed by a description of a state of affairs. However, little attention has been paid to the question of what kind of imagination is involved in the conduct of a thought experiment. It has typically been taken to be like forming a picture in the mind’s eye; an objectual, in particular, a visual form of imagination.[47] We can see a commitment to this view in the works of (to take just a few examples) James R. Brown, Tamar Gendler, Nenad Miščević and David Gooding:

> ‘Thought experiments are carried out in the mind and they involve something akin to experience; that is, we typically “see” something happening in a thought experiment’ (Brown 2004a, 25)
>
> ‘in the case of imaginary scenarios that evoke certain sorts of quasi-sensory intuitions, their contemplation may bring to us new beliefs about contingent features of the natural world that are produced not inferentially, but quasi-observationally; the presence of a mental image may play a crucial cognitive role’ (Gendler 2004, 1154)
>
> ‘When a reader encounters a description of a situation, she builds a model, a quasi-spatial “picture” of it’ (Miščević 1992, 220)
>
> ‘visual perception is crucial because the ability to visualise is necessary to most if not all thought experiments’ (Gooding 1992, 285)

[47] This observation has been the starting point for Arcangeli (2010) and Salis and Frigg (2020).

We can take Gendler's account as an example. As indicated in the quote above, Gendler focuses her analysis of the imagination in thought experiments on mental images. Not only is mental imagery involved in their conduct for Gendler but in some cases, it plays a key role. She acknowledges that the imagination (in the imagistic sense) might not play an essential role in all thought experiments—in some cases, the role may be merely heuristic. But against Norton's view that thought experiments are just arguments, she presents cases in which visualisation or "quasi-observation" is key to the success of the example. Following Mach, Gendler presents Stevin's inclined plane to illustrate her view.

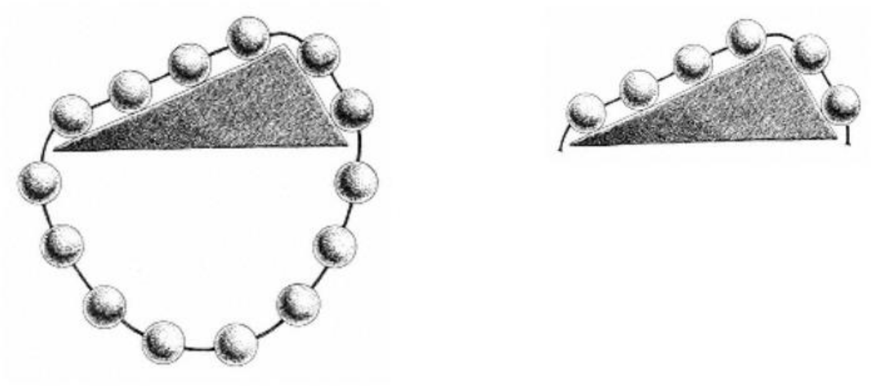

Figure 4: Stevin's Chain (from Gendler 2004)

Stevin used the thought experiment in order to establish the force that is required to prevent an object on a frictionless, inclined plane from sliding down. We are asked to imagine fourteen balls (of equal weight and size) threaded on a string and laid on top of a triangular prism of unequal sides. What would happen? If we imagine this scenario in accordance with the diagram on the right, then we can see that there are three options: i) the chain remains in a state of static equilibrium, ii) the chain moves to the right (because the incline is steeper) or iii) the chain moves to the left (because there are more balls and so, the string is heavier on that side). Stevin argues that the answer is (i), the chain remains in a state of static equilibrium and shows how the diagram on the left shows this to be the case. As we can see, in this diagram the chain is a closed loop. If the string moved (as in (ii) or (iii)) and was not in a state of static equilibrium, then it would be in a state of perpetual motion. Given the impossibility of perpetual motion machines, the force on the string must be balanced in each scenario.[48] From

[48] As Sorensen (1992, 54) discusses, in Mach's outline of how this thought experiment works, he highlights how a key assumption, that there can be no perpetual motion, strikes us not just logically but also psychologically; 'He feels at once, and we with him, that we have never observed the motion of the kind referred to, that a thing of such character does not exist' (Mach 1976, 34). Given that Gendler operates within a Machian tradition, it is plausible that she would agree that this example works in this way.

this, Stevin concludes that the force needed to keep an object in place along an inclined plane is inversely proportional to the length of the plane.

Gendler argues that cases such as this show that not all thought experiments work by deductive or inductive reasoning. She states:

> 'presumably there's a way of *reconstructing* this reasoning process as an argument: I will leave that task to others… Contemplation of an imaginary scenario (the cut string laid atop the prism) evokes certain quasi-sensory intuitions, and on the basis of these intuitions, we form a new belief about contingent features of the natural world (that the weight of four balls offsets the weight of three balls). This belief is produced not inferentially, but quasi-observationally: the presence of the mental image plays a crucial cognitive role in its formation' (2004, 1162).

We can see then that the imagination in thought experiments is typically associated with mental imagery and for some, this is crucial in understanding how thought experiments perform their function. And as Salis and Frigg (2020) note, those who discuss imagination in the context of modelling, whilst not offering an explicit analysis, tend to associate it with mental imagery. For instance, on Levy's Waltonian view of models, imagining 'typically involves having a visual or other sensory-like mental state—"a seeing in the mind's eye"' (2015, 785).

It is worth noting that there are some thought experiments that might be said to employ other sensory modalities. One example, that comes from philosophy, is Strawson's thought experiment that asks us to imagine a purely auditory world. But as we can see from the above quotations that the type of imagery relevant to scientific thought experiments is typically taken to be visual. This may be because many have drawn comparisons with ordinary experiment (as seen in chapters 1 and 2). For example, Sorensen states that a 'thought experiment is a limiting case of experiment in which the question is to be answered by reflection on the experimental design rather than by execution. Imagination substitutes for perception. The reliability of the thought experiment's answer depends on how well imagination can fill a role originally intended for perception' (2018, 31). Brown explains that it is possible to have thought experiments that rely on imagined sounds or smells as what is important is 'that it be experiencable in some way' (2011, 17). However he highlights that sight is our most important sense, and in experiments, the primacy of visual perception is typically assumed. For Brown, this extends to thought experiments: the 'only difference [between a thought experiment and an ordinary experiment] is that the perception is not a sense perception but, rather, is an intuition, an instance of seeing with the mind's eye' (2004a, 35).

Miščević, quoted above, as well as Nersessian (2007, 2018) apply the cognitive science literature on mental modelling to thought experiments.[49] Given that this offers a detailed account of the reasoning involved in conducting thought experiments, and that Salis and Frigg takes this as the most developed defence of the imagistic view, it is worth spending some time on their approach. "Mental model" is a technical term used in cognitive science and technology. It can be traced back to Craik (1943) and has been developed as an account of reasoning by Johnson-Laird amongst others. Johnson-Laird summarises their function in the following way:

> '…mental models play a central and unifying role in representing objects, states of affairs, sequences of events, the way the world is, and the social and psychological actions of daily life. They enable individuals to make inferences and predictions, to understand phenomena, to decide what action to take to control its execution, and above all to experience events by proxy; they allow language to be used to create representations comparable to those deriving from direct acquaintance with the world; and they relate to the world by way of conception and perception' (1983, 397)

Mental models are temporary structures that are constructed in working memory to carry out a specific reasoning task. Like physical models and diagrams, they are iconic representations which means that their structure corresponds to the structure of the state of affairs that they are taken to represent. As Johnson-Laird notes above, mental models are used in many contexts. An example that Nersessian discusses is creating mental models that help us produce reliable beliefs about whether a piece of furniture will fit through a doorway. The idea is that we tackle this problem by rotating a mental token that approximates the shape of the object, such as a sofa or a table, in a way that is 'constrained by the boundaries of the doorway-like token' (2007, 128). And as a result, we gain new information about whether the furniture will fit through the door. Similarly, operations 'on thought experimental models require transformations be consistent with the constraints of the domain, which can be tacit or explicit for the experimenter' (Nersessian 2018, 319). In order to be successful, not only does the state of affairs need to be represented accurately, but so do the transition rules for how the system

[49] Gendler (2004) also links her account with developments in cognitive science and psychology but is not committed to the mental models framework.

changes over time: ‘Causal coherence, spatial structure, and mathematical consistency are examples of kinds of constraints’ on what transformations are legitimised (ibid.).

The proponents of the view recognise the importance of language in constructing the model. And they attend to the narrative presentation of many thought experiments.[50] But they stress that the operations used in the performance of a thought experiment are on the model that is constructed, rather than on linguistic representations. Iconic representations therefore contrast with linguistic or formal representations that are used in logical and mathematical operations. In the latter cases, we reason using formal rules of inferences which refer to the objects or states of affairs descriptively. Iconic representations, on the other hand, are taken to represent *demonstratively.* The thought is that in the case above, the model “shows” us in that it allows us to “see” that the furniture will or will not fit through the door.

As mentioned, Salis and Frigg locate the mental model accounts within the imagistic group and it is clear that the view differs significantly from their propositional account that I turn to in the next section. However, it is important to note that there is some variation with regards to the role of mental imagery in different versions of the views. For Miščević, thought experiments consist in building a quasi-spatial “picture” of the scenario described ‘in front of one’s inner eye’ (1992, 220). Nersessian’s account is slightly different in that it does not involve ‘pictures in the mind’ (1992, 294); it can be a matter of forming more abstract analogical representations. In making this point, Nersessian refers to Bohr who utilised thought experiments but also claimed he was unable to visualise well.[51] Yet, Nersessian highlights, the reasoning involved is of a non-propositional nature; ‘inferences subjects make are derived from constructing a mental model of the situation, rather than by applying rules of inference to a system of propositions representing the content of the text’ (*ibid,* 293).

While I agree that the mental model accounts have offered important insights on the types of reasoning involved in thought experiments, and have offered a perspective that lies somewhere between the extremes of Norton and Brown, there are some difficulties with the position. The view often goes from examples of reasoning in everyday contexts—say, the

[50] In light of their connection with narrative, mental models are also said to be involved in our engagement and understanding of (fictional or non-fictional) stories (Matravers 2014).

[51] It is interesting to consider what to make of such statements regarding the testimony of particular scientists when it comes to offering an account of the imagination in science. It may well be that certain scientists (including those who produced thought experiments) are not able to visual well or lack conscious mental imagery altogether, or do not find mental imagery useful. I show in the next chapter how my pluralist view can accommodate such instances.

mental rotation task as presented above—and then generalises to thought experiments. This is because, the argument goes, they similarly involve problem solving tasks. I do not want to claim that mental models are not involved in the conduct of (any) thought experiments. However, as it stands, there is a lack of a detailed outline of how the mental model accounts explain various examples from science. Because of this, a core problem for these views is that the theory lacks an explanation as to whether all scientific thought experiments utilise mental models in the way set out for the more general tasks, and if not all scientific cases do, which ones rely on mental models and why.

As a consequence, the fruitfulness of the approach is undermined. Furthermore, within cognitive science, the mental model framework is highly contested and many aspects are largely speculative, which Nersessian herself acknowledges (see also Arcangeli 2010, Cooper 2005, Meynell 2014). Because of this, I leave the mental model views to one side and speak in terms of the imagination and the various forms it can take. As the mental model theorists often refer to imagination, I will pick up on some of their claims in the next chapter. But avoiding talking in terms of mental models and instead focusing on imagination means that we can steer clear of providing an account that relies upon a commitment to a specific model of the mind.

We can see, then, that many accounts of thought experiments that highlight a crucial role for the imagination have characterised it in broadly imagistic terms. I'll now turn to a view of the imagination that departs from such accounts.

3.6 The Propositional Alternative

Salis and Frigg (2020) begin their paper with the observation that I have outlined above. This is that while there is talk of imagination in science, there is a lack of detailed analysis regarding what type of imagination is involved in the scientific realm and usually it is equated with mental imagery. Salis and Frigg develop an account of the scientific imagination that they apply to both scientific models and thought experiments. Although they agree that both models and thought experiments involve the imagination, they propose that imagery (of any sensory modality) is unnecessary for their performance. For them, it is only the propositional form of the imagination that is relevant in the scientific domain. Salis and Frigg take it that thought experiments and models 'involve the same kind of imagination' (ibid., 19). As thought experiments are my central concern, and the defenders of the imagistic view that Salis and Frigg discuss are part of the thought experiment literature, I will focus on the details of their account as applied to thought experiments.[52]

Salis and Frigg note that it is not only advocates of the imagination in science that assume an imagistic form of imagination. They also demonstrate how those who have been critical have taken the imagination to consist in mental images and argue that this has been central to any scepticism regarding the imagination in science. Take, for example, Norton's argument view. Salis and Frigg argue that Norton has no place in his system for imagination, but when he makes any references to imagination, or dismisses the "picturesque" qualities of thought experiments, it appears that he has in mind an imagistic form of imagination.[53] They argue that their propositional account can circumvent such opposition.

While I have outlined the features of the propositional imagination earlier in this chapter, it is worth mentioning some of the details of Salis and Frigg's conceptualisation of this kind of imagination.[54] The core features that the propositional imagination exhibits, that is, the features that are necessary and sufficient conditions for "imagining-that", are:

(i) Freedom. As Salis and Frigg state that 'we are not free to believe whatever we want, but typically we are free to imagine whatever we want' (2020, 30). As previously mentioned, we may not be able to imagine whatever we want to imagine, but the important point is that we are free to imagine things that we know are not true. Whereas to 'believe that *p* is to hold *p* as true at the actual world, and whether the actual world makes *p* true or false is not up to us' (ibid.).

[52] In the next chapter, however, I do offer a discussion of the differences between models and thought experiments. This includes how the nature and role of imagination may differ in the two practices.

[53] They also attribute such a view to Weisberg who rejects the fiction view of scientific models. Weisberg describes the imaginings in such views as 'mental pictures' (2013, 51). I'll go through Weisberg's objections in the next chapter. See also McLoone (2019) for a defence of the propositional imagination in modelling as a response to Weisberg.

[54] Salis and Frigg include counterfactual reasoning, supposition and dreaming as species of propositional imagination. Dreams are set to oneside, and Salis and Frigg favour make-believe over counterfactual reasoning and supposition. In the case of the former, this is because possible worlds are complete whereas models are not: 'Claims about the date of the Battle of Waterloo…are neither true nor false in, say, Einstein's elevator TE' (2020, 42). See also Friend (2017) for an outline of the differences between possible worlds and the Waltonian framework. Further, Salis and Frigg highlight the tensions in the literature regarding gaining knowledge from counterfactuals (see Williamson 2020 for discussion). In the case of supposition, Salis and Frigg argue that it fails to be constrained enough to be appropriate in the scientific context: 'One can suppose anything, and as long as no further restrictions are imposed, one can conclude almost anything from certain assumptions' (2020, 41).

(ii) Mirroring. The inferences that we carry out when we propositionally imagine are similar to those operating in belief (ibid). For example, when we believe someone is human, we also believe that they have blood in their veins. Similarly, to recall an earlier example, when we imagine Hermione, we also take it as fictionally true that she has blood in her veins.

(iii) Quarantining. Given that imagining that *p* does not mean that we believe *p,* this feature highlights how we take any effects of our imaginings to only apply within the imagined context, that is to say, our imaginings do not motivate action. For example, if I imagine that there is a ferocious animal chasing after me, I am not going to run away as I would do if I believed there to be. Further, if we learn anything from our imaginings, then this is only when we export them to the real world. Salis and Frigg give the example of reading a work of Dickens which prescribes us to imagine that many people were treated badly in the Victorian era. We might also believe this to be true, but such '"exports" are, however, one step removed from the imagination' (ibid., 31).

To demonstrate their view, Salis and Frigg provide an example from Galileo, namely his thought experiment which was used to answer the question: 'is a force required to keep an object moving with constant velocity?' Galileo asks us to imagine a frictionless U-shaped cavity. If a ball is dropped from one side, it will continue to move until it recovers the original height it was dropped from on the other side of the cavity due to the law of equal heights. We then imagine that the second side of the U-shaped bend is flattened, so that the ball is now being dropped from a height and then travels along a straight line. The law of equal heights still applies yet the ball can never recover its original height and so, it will continue moving. This thought experiment exposes a contradiction in Aristotle's theory that moving objects will come to a stop. From this, Galileo establishes the law of inertia; no force is needed to keep an object moving with constant velocity.[55]

Against the imagistic views, most notably, the works of Gendler and Nersessian, Salis and Frigg argue that the importance of mental imagery has been overstated. They state that mental imagery is insufficient: 'When considering Galileo's cavity we do not seem to have a perception-like representation of the cavity being frictionless or the lack of air resistance.

[55] In order to draw the connection between thought experiments and models, Salis and Frigg offer a model version of Galileo's thought experiment in which we can derive a 'mathematically rigorous justification of the law of inertia' (2020, 21).

Likewise, we cannot form a perception-like representation of the concept of force without having a theoretical definition, which is usually given in linguistic and formulaic symbols' (ibid., 29). Here, Salis and Frigg highlight how many steps in a thought experiment are best captured in terms of the propositional imagination, as opposed to as mental images.

Focusing on Galileo's thought experiment, it can be asked: What would it be like to have a perception-like representation of the cavity being frictionless? It seems correct to say, as Salis and Frigg do, that we cannot have a visual image of frictionlessness.[56] But we could consider how we might represent absences in our imaginations. In this case, we would imagine the state of affairs but in such a way that we subtract fiction from the visual mental image we produce. In the *Two New Sciences,* Galileo through Salviati, provides descriptions of experiments previously performed to help convince Simplicio that certain mathematical results apply to nature and to make certain theoretical claims plausible. In one section, he describes cutting a channel along a piece of wood and states 'having made this groove very straight, smooth and polished as possible, we rolled along it a hard, smooth, and very round bronze ball…' (1914, 178). Here, it is plausible to suggest that Galileo helps us imagine the effects of frictionlessness, that is, what it would look or feel like, by providing us with descriptions whereby friction is reduced as much as possible.[57]

Turning to Arcangeli's *The Two Faces of Mental Imagery* (2019) provides a way of considering this difference.[58] Arcangeli draws a distinction between two different senses of mental imagery which she argues have often been conflated in the imagination literature. Mental imagery can be understood as sensory imagination which is a psychological attitude or mode that re-creates perception as on, for example, Currie and Ravenscroft's view as previously discussed (2002). In this sense, mental imagery is a perception-like attitude such as when we imagine seeing or smelling flowers. But the relationship between mental imagery and perception can also be understood in a different way. As Arcangeli explains, on

[56] Going back to Brown's statement above, that should not necessarily consider the mental imagery thought experiments produce as being limited to visual imagery, it could be suggested that Salis and Frigg are too quick to conclude that no kind of mental imagery is irrelevant in imagining frictionlessness. It is difficult to describe what frictionlessness might look like, but perhaps that is because another sense modality is more important when detecting friction. For example, we could imagine what frictionlessness might feel like, thus having an image of a different sense modality in this case, namely a tactile one.

[57] The importance of what Galileo might be asking us to do will become apparent when I outline my pluralist account.

[58] Thanks to Mike Stuart for directing me towards this paper.

definitions such as Gaut's, where imagery 'is a matter of having sensory perceptions' (Gaut 2003, 272), mental imagery is placed 'on the content, rather than on the attitudinal level of our mental life' (Arcangeli 2019, 7). Here, mental imagery is a type of sensory content. On this understanding having 'a mental image of a flower would be bringing to our mind a sensory (i.e. visual, olfactory) presentation of a flower without the stimulation of our vision or olfaction by an external flower' (ibid., 8).

Arcangeli states that the two senses of mental imagery often go together. In such cases, sensory imagination has mental imagery as its content. But Arcangeli offers one way in which it is possible for them to come apart. She demonstrates this through the example of recalling or desiring (that is, having an attitude that is different from sensory imagining) food that is currently cooking in the oven. Here, a desiring of the food, say, may involve a mental image of the food inside the oven, that is to say, such a mental image is the content of the attitude.

If we return to Galileo's thought experiment, we can also see how the distinction that Arcangeli offers can allow for an instance of sensory imagination without a mental image of a particular aspect of the imagined content, in this case, frictionlessness. In the second sense of mental imagery, that is, as a type of mental content, it does appear that we cannot imagistically imagine frictionlessness. That is to say, we cannot form an image in our minds of frictionlessness, and this is what Salis and Frigg point out. However, what Arcangeli's distinction opens up is a way in which we could consider how we can imagistically imagine frictionlessness in the first, attitudinal sense. The thought here is that we can entertain the concept "frictionlessness" in a way that is perception-like. We do this by imagining how a certain state of affairs would look in the absence of friction and so on.

Putting to one side potential issues with this particular point, we should acknowledge that Salis and Frigg are correct to say that in order for thought experiments to be successful, we must have certain conceptual knowledge and thought experiments are going to include steps that are best captured as propositional reasoning.[59] Further to maintaining that mental imagery is insufficient, Salis and Frigg take it to be unnecessary. They state 'it would be implausible to argue that individuals with a poor imagistic ability could not derive the correct outcome of Galileo's TE (or, for that matter, of any TE). Presumably, one could perform the TE and draw

[59] This point has also been developed by Arcangeli (2010) who, states that the imagination in thought experiments is not limited to a pictorial kind, and argues that a broader notion of the imagination is present in the work of Mach. See also Reiss (2002) for an argument against the view that thought experiment scenarios are importantly visualisable.

the relevant conclusion by understanding the propositional content of the argument underlying it' (2020, 40).[60] They argue that it is a propositional form of the imagination, imagining *that* something is the case, that is necessary for conducting thought experiments, insisting that we 'need to grasp the relevant concepts, with or without forming a mental image of the objects and the transformations they stand in for' (ibid.). They generalise this claim to cover all cases of the scientific imagination in modelling and thought experiments.

Salis and Frigg address head on the question of what kind of imagination is involved in scientific thought experiments and models and offer a useful taxonomy, drawing on philosophy of mind and aesthetics, in order to shed light on this question. I agree with them that we should think beyond mental imagery when discussing the imagination in science and they successfully illuminate the essential role of the propositional imagination. In the next section, however, I set out worry with their approach. And in chapter 4, I develop my own view of the nature of the imagination in science.

3.7 Propositional Imagination and Argumentation

An initial problem with Salis and Frigg's account is that there are close connections between the propositional view of the imagination in thought experiments that they set out, and Norton's argument view. To recap, Norton analyses thought experiments as arguments and maintains that all thought experiments can be reconstructed into argument form without any epistemic loss, and that the 'actual conduct of a thought experiment consists in the execution of an argument' (2004, 50). Norton has reconstructed many thought experiments into arguments and holds that there are no examples that cannot be handled in such a way. Consequently, their typical narrative form and any of their creative, or to use Norton's terminology, "picturesque" qualities, are deemed epistemically redundant.

Salis and Frigg argue that Norton's view misses the importance of the imagination and the use of imagined particulars. They state that characterising thought experiments as arguments 'presupposes a propositional kind of imagination' (2020, 25) and 'the arguments leading to the general conclusions involve imagined scenarios and particulars' (ibid., 37). However, it is

[60] Salis and Frigg do, however, acknowledge in a footnote that whether any thought experiments would be difficult for those who have limited imagistic capacities 'would be an interesting empirical question' (2020, 50 fn 24). In the next chapter, I will say more about aphantasia, that is, the inability or reduced ability to experience conscious mental imagery.

difficult to see how their account departs in any significant respect from an argument view when it comes to the types of reasoning involved in thought experiments. In their analysis of Galileo's case, which is the only thought experiment they discuss, they state: 'Galileo deliberately imagines a certain hypothetical scenario, he develops a deductive reasoning leading to a contradiction, and he quarantines its content since he explicitly invites us to imagine a non-actual situation' (ibid., 38).

Their view actually comes out stricter than Norton's in certain ways. Norton has expanded his notion of argument involved in thought experiments over the years. His position is broad; valid logical inferences go beyond deduction and induction to include informal reasoning and even reasoning with imagery. For example, a picture can be a premise in an argument as seen in some mathematical cases. Thought experiments are governed by a 'very general notion of logic' (Norton 2004, 54). As we saw in the introductory chapter, there is a worry with Norton's view that the notion of argument becomes so broad that his position ends up trivial—the claim amounting to the idea that there is some reasoning or inference involved (Stuart 2016b, Brendel 2018). In light of this, I understand the argument view of thought experiments, as many have done in the literature, to be restricted to a narrower sense of argument.

The worry is that either Salis and Frigg's propositional account, although insisting that there is a role for imagination, is stricter than Norton's (broad) argument view, or it is difficult to see what the distinction between Salis and Frigg's propositional account, and Norton's (narrow) argument view is.

Salis and Frigg do of course situate their account within a broader framework of modelling and representation in general. They endorse the claim that models and thought experiments are examples of make-believe, that is, they are representations that function as 'props' that prescribe imaginings. However, given my claim that their account is strongly aligned with Norton's narrow view, what is the role for make-believe? And what is the relation between propositional imagination, argument, and make-believe? This is not to say that characterising thought experiments as arguments necessarily means that they cannot be fictions in the Waltonian sense. As Friend demonstrates, Walton's category (what Friend calls *Walt fiction*) is broader than how we usually apply "fiction". She states that 'Walton is not interested in our ordinary conception of fiction. He is concerned with representational art more generally, including both texts and pictures… On his view, any work that prescribes imaginings counts as fiction' (2011, 164). But if this is all that is meant, what do the advocates of this view think they are getting from treating thought experiments as make-believe?

The problem is that it is hard to see the benefit of treating thought experiments as fictions in the Waltonian sense over Norton's narrow view if the nature of the imagination is propositional (belief-like) and thought experiments involve deductive reasoning.[61] Meynell gave an earlier account of thought experiments as make-believe, and raises a similar concern. She argues that if we were to maintain that the imaginings prescribed follow the logical form of beliefs, and the relations between beliefs, then it is not clear how they would differ from Norton's background assumptions (2014, 4165). As a consequence, Salis and Frigg will be subject to the types of worries that have been aimed at Norton's account. Meynell instead emphasises the importance of experiential language, and the use of pictures as aids for thought experiments which gives insight into the nature of the imagination involved which I return to in the next chapter.

Perhaps this worry can be responded to by further considering which principles of generation are involved in thought experiments. That is, by thinking about how we get from those truths that are explicitly stated to those that are implied. Salis and Frigg do not offer any discussion of thought experiments beyond the Galilean case, but they do suggest that the principles of generation involved in scientific models and thought experiments need not be restricted to the reality principle (where we "fill in" what is given as fictional truths based on the real world) or the mutual belief principle (where we constrain our implied truths in line with the beliefs of the community in which the fiction originated). At the end of the paper, they suggest, with reference to Meynell's work, that other principles could be involved depending on the context in which the model or thought experiment is being used.

However Meynell's Waltonian view of thought experiments differs significantly from Salis and Frigg. When Meynell raises the possibility of a variety of principles of generation—the 'conventions, aliefs, habits of mind, and primed expectations' that work with 'background

[61] A parallel worry is raised by French who considers a version of the models as fictions view that takes the imaginings involved to be of a conceptual nature in order to get around worries concerning our abilities to visualise. French states 'in the context of this review of fictionalism, how is *conceptual imagination* to be demarcated from the kinds of conceptual explorations that advocates of the Semantic Approach or Weisberg's 'mathematical models' approach will also acknowledge?' (French 2020a, 173).

beliefs to create TE content'—she states that these are not reducible to belief-like reasoning.[62] She states 'the content of these imaginings and the ways in which they are produced and provide insight simply does not have a propositional or argumentative form' (ibid., 4165). It would thus be interesting to see how a more flexible approach to the principles of generation involved in scientific thought experiments would be fleshed out on a propositional view of imagination, and whether this offers a way of highlighting how the view departs from Norton's picture.

And so, a defender of the models as make-believe view that takes the imagination as propositional might be able to overcome the worry presented in this section and demonstrate how their account differs from, or is a better version, than Norton's view, through offering a more detailed account of the principles of generation involved. But this remains to be seen. In the next chapter, I will not say much more on the make-believe aspect of Salis and Frigg's account. My focus instead is on the type of imagining involved in science.

3.8 Conclusion

In this chapter, I have outlined some of the existing discussions regarding the imagination in science. Despite the increasing attention given to the imagination's role in science, especially in the practice of modelling as seen on the models as fictions view, the question regarding what the nature of the imagination in this domain has been largely overlooked. Focusing on thought experiments, I demonstrated that most have assumed the imagination as taking an objectual form; we have pictures in our minds of a state of affairs described by the thought experiment. I then turned to a recent account that is set against this assumption, and explicitly addresses the type of imagination involved in models and thought experiments. This is Salis and Frigg's propositional view. While I agree with them that we should also include propositional imagination in our analysis of thought experiments, I presented an initial worry with the view which arises given the close connections between this propositional picture and Norton's argument view which states that thought experiments involve deductive and inductive reasoning. In light of this, the account will be open to the worries raised against the

[62] Aliefs are 'associatively linked content that is representational, affective and behavioural' (Gender 2008, 642). The connections between aliefs are not logical as with beliefs (or make-beliefs), but rather come in "associated chains". Aliefs can contradict a person's beliefs. For example, I might believe that I am very safe on a flight, but I have an alief that I am in danger. See Currie and Achino (2012) for a critical overview of aliefs.

Nortonian picture. Whether or not this particular problem can be overcome, I will now turn to my pluralist account of the imagination in science. I argue that this can better accommodate the variety of ways in which thought experiments call upon our imaginative capacities.

4. A Pluralist Account of the Imagination in Science

4.1 Introduction

In the previous chapter, we saw how it is increasingly accepted that the imagination has some role in science, yet the question of what is meant by the imagination in this context is under explored. I outlined how the imagination in thought experiments has been characterised in terms of the objectual imagination, in particular, in forming mental pictures of the scenario described. I then discussed Salis and Frigg’s recent objections to this assumption, and I set out their alternative view. They argue that mental imagery is neither a sufficient nor a necessary type of imagination for the conduct of scientific models and thought experiments. Instead, it is the propositional form of imagination that is required. I agree with Salis and Frigg in that we should not equate the scientific imagination with mental imagery, and they are correct to emphasise importance of the propositional imagination in this context.

However, in this chapter, I present a different view of the nature of imagination in thought experiments. I argue that while we ought to acknowledge the role of the propositional imagination in science, and more generally, attend to the nature of the imagination more closely, it is incorrect to take the objectual imagination out of the picture altogether. Instead, I propose a pluralist account. I begin with considering the differences between thought experiments and models which has implications for the ways in which the imagination is used in their conduct. I then restrict my focus to thought experiments and argue that when we consider various examples and their role in scientific practice, we can see that different thought experiments invite different types of imagination.

4.2 Thought Experiments, Models and Mental Images

Although my main focus in this chapter concerns thought experiments, I will begin by suggesting that there are relevant differences between thought experiments and models that can impact the nature and role of the imagination in their use. This thus undermines a “one size fits all” account of the scientific imagination and puts pressure on those that seek to simply extend discussions of modelling in order to accommodate the imagination in thought experiments.

Salis and Frigg are right to stress similarities between thought experiments and models and they show that bringing the two together offers a fruitful way of thinking about the imagination in science. Both involve creating and engaging with idealized hypothetical scenarios, and because of this, some have argued that 'models are often experiments *in thought'* (Cartwright 2010, 19), or that models are cases of "extended cognition" (Thoma 2016).[63] Further, it is often claimed that they share important similarities with experiment: They offer a description of an initial set up which is then manipulated for us to consider what would happen. Yet they each depart from experiment in that they do not involve intervention in the world. Salis and Frigg also acknowledge some distinctions between thought experiments and models, such as the fact that the former does not include the 'formal apparatus', that is, mathematical reasoning to provide a formal proof, which is present in the latter. I agree that scientific models involve the imagination in some way, and refer to imaginary objects or systems, and I do not want to insist on a strict distinction between models and thought experiments. However, I want to note that we should attend to their differences when giving an account of the nature of imagination involved in their use.

A key strength that Salis and Frigg claim for their account is that moving the focus away from mental imagery can aid in avoiding scepticism around the imagination in science. They state that the imagination is 'dismissed because of its allegedly imagistic character' (2020, 18). As Salis and Frigg cite Weisberg (2013) as holding this view, and because he offers the most developed form of the worry surrounding "mental pictures" in the models as fictions view, we can focus on his discussion.[64] Weisberg presents examples of models that are unimaginable, in the sense that we cannot hold images in our mind of what they describe. He argues that although 'it is relatively easy to imagine [to form "mental pictures" of] the content of finite, deterministic, individualistic models like a population of genes undergoing assortment, it is unclear that this could generalize to more complex cases' and importantly, this 'rules out the possibility of equating such models with imagined fictional scenarios and undercuts Waltonian versions of the fictionalist position' (2013, 63).

[63] These points echo the ways in which some have drawn connections between thought experiments and computer simulations as presented in chapter 1.

[64] McLoone (2019) also defends a propositional account of imagination in modelling, and his paper is set up against Weisberg's criticism that some models are about "unimaginable" (in the sense of unvisualisable) systems.

Take the case of the Lotka-Volterra model which describes the dynamics between predator-prey populations. Weisberg states that on a fictionalist picture, this model is an imaginary system of two interacting, concrete and discrete organisms. But the problem is that the equations that describe the model do not refer to individual organisms but rather to populations. And the size of these populations are not a discrete quantity; they may vary continuously. How could we visualise this? Other examples that are presented as problematic for fictionalists are cases of models that involve probabilistic elements. The thought is that we could imagine, that is, form an image in our mind, of a single instantiation, say a predator dying, but this does not represent the probability of the death (ibid., 63). And we can also include models that have properties far removed from concrete systems, such as models in statistical mechanics that describe infinite degrees of freedom, or models in physics that have more than three spatial dimensions as discussed by French (2020a, 173).[65]

And so, the question is whether or not such instances of unvisualisablity in the case of models also occurs in the case of thought experiments. To answer this, it is helpful to consider some of the differences between models and thought experiments. As already noted, models have an underlying formal apparatus. But the imagination is sufficient for conducting thought experiments in that no calculations on paper nor a computer are required when we carry out, for example, Galileo's thought experiments.[66] To use Thoma's phrase, 'the phenomenon is established purely in the imagination' (2016, 136). Thoma discusses this in the context of what she labels "hidden" thought experiments in economics. The example she focuses on is Banerjee's (1992) model of herd behaviour which is preceded by a short story. Thoma argues that stories such as this have typically been analysed by philosophers as a part of the formal model that follows them. However, Thoma argues that this is mistaken; these "stories" are

[65] Although Weisberg rejects the fictionalist view of models (and in part, because of their worries regarding the limits of imagination), it is important to note that he does not argue that imagination (characterised in an imagistic sense) has no role in the scientific domain. For Weisberg, mental pictures help when thinking about models. Weisberg regards these imaginings not as the model, but as the "folk ontology" of models, which is essential to modelling and therefore must be included in philosophical accounts of modelling. They can aid the development of models, and are useful in cases of using multiple models that are mathematically different, but share many features in common when described at the concrete level (2013, 68-70, see also French 2020a, 174). As my pluralism does not commit to the Waltonian framework, what I go on to say about the imagination in thought experiments could be adapted to accommodate models taken as mathematical structures as on Weisberg's view. Todd (2020) offers a defence of imagistic imagination in modelling against Salis and Frigg.

[66] It might be the case that there are certain instances in which scientists could see what happens in their model by just thinking or imagining about the assumptions of their model and so, some models will be more like thought experiments than others in this respect.

actually instances of thought experiments that have independent evidential significance, and differ in significant respects, from the models that are introduced after them.

Banerjee sets out to explain “herd behaviour”. This is where a group of people all act in the same way despite individuals having different information regarding what is the best decision to make. Banerjee shows that this can occur even in cases where there is no preference for following what others are doing (which is how herd behaviour has typically been explained). The thought experiment goes as follows:

> ‘Most of us have been in a situation where we have to choose between two restaurants that are both more or less unknown to us. Consider now a situation where there is a population of 100 people who are all facing such a choice.
>
> There are two restaurants A and B that are next to each other, and it is known that the prior probabilities are 51 percent for restaurant A being the better and 49 percent for restaurant B being better. People arrive at the restaurants in sequence, observe the choices made by the people before them, and decide on one or the other of the restaurants. Apart from knowing the prior probabilities, each of these people also got a signal which says either that A is better or that B is better (of course the signal could be wrong). It is also assumed that each person’s signal is of the same quality.
>
> Suppose that of the 100 people, 99 have received signals that B is better but the one person whose signal favors A gets to choose first. Clearly, the first person will go to A. The second person will now know that the first person had a signal that favored A, while her own signal favors B. Since the signals are of equal quality, they effectively cancel out, and the rational choice is to go by the prior probabilities and go to A.
>
> The second person thus chooses A regardless of her signal. Her choice therefore provides no new information to the next person in line: the third person’s situation is thus exactly the same as that of the second person, and she should make the same choice and so on. Everyone ends up at restaurant A even if, given the aggregate information, it is practically certain that B is better’ (1992, 798-99).

Thoma argues that in this thought experiment ‘we can imagine ourselves in the relevant choice situation, and visualize what it would be like to stand in front of the two restaurants, one of which is gradually filling with people’ (2016, 9). Whereas the formal model that follows the thought experiment ‘does not ask us to visualize the situation like this, while it is crucial for the story to be convincing' (ibid., 10). I do not want to say that all thought experiments need to be visualised, as I have set out, I agree with Salis and Frigg that some thought experiments, or some aspects of their performance, can be conducted in the absence of mental imagery. I will say more about this in the next section and consequently what to make of statements such as visualisation being "crucial" for a thought experiment. But I do think that attending to the differences models and thought experiments highlights how these two practices differ when it comes to the ways in which the imagination is involved in their use.

We can also see this difference via attending to the type of idealisations involved in models and thought experiments. For example, Thoma highlights how Banerjee's thought experiment does not ask us to imagine perfectly rational agents. Rather, the thought experiment works because we can imagine ourselves in this situation and consider the types of choices we would make, and therefore we see it as likely that agents would act in this way. Whereas the model assumes perfectly rational agents that have common knowledge of the other agents' rationality. How could we imagine this? We cannot intuitively access the preferences of this type of agent nor could we put ourselves in their shoes and think what it would be like to be such an agent. And as Thoma states, common knowledge 'is shorthand for an infinite number of mental states. And it is at least not realistic that humans can have an infinite number of mental states' (2016, 137). Further, in the model, the agents do not have a simple choice between A and B. Rather, there is a continuum of options.

Reiss (2018) also discusses idealisations in thought experiments in comparison with other practices in science. Thought experiments involve what Reiss labels "Galilean idealisations", where these are taken to be causal idealisations, that is, those that isolate a single causal line to consider the question "what would happen if?". For example, in Galileo's falling bodies, air resistance and other forces that would affect the speed of the bodies are idealised away in order for us to see how the scenario would unfold. Mathematical idealisations, which are introduced on pragmatic grounds to make problems mathematically tractable, cannot be incorporated into thought experiments whereas they are essential in many types of modelling. This is because the former succeed without formal apparatus, that is, without 'mathematization or calculation' (ibid., 472). While I do not think that all thought experiments require us to form mental images, I do think that in light of these differences between models and thought experiments, a weaker version of Thoma's claim could be provided. This is that thought experiment scenarios are at least *possible* to imagine in an objectual sense. Along with Thoma's example of Banerjee's thought experiment on herd behaviour, we can also consider Stevin’s chain, Galileo’s falling bodies, or Newton’s bucket to name just a few examples.[67] These scenarios describe elements that we are able to picture in our minds, and we can manipulate our visual imaginings as the scenario unfolds.

However, it is important to recall Salis and Frigg’s point that certain elements of the scenario described in a thought experiment cannot be visualised and instead involve only the propositional imagination. I suggested a way around their example of forming a mental image of frictionlessness. But perhaps there are other cases that provide stronger support for the point that Salis and Frigg want to make. Let’s consider a case discussed in chapters 1 and 2; Dewan and Beran’s rocket and thread thought experiment. This example includes a step in which we have to imagine that the two rockets reach four-fifths of the speed of light (as Lorentz contraction is only noticeable when objects travel near the speed of light). It seems clear that in this step, we imagine that, as opposed to form a visual imagining of the rockets travelling at that speed. Similarly, we can consider Einstein’s thought experiment in which he gets us to imagine chasing a beam of light and to consider what we would see. Given that we cannot see beams of light, then it may appear difficult to make sense of imagining this scenario in an imagistic way.[68]

I take it that there are cases of thought experiments that are more difficult to visualise than others and the set up of thought experiments do require certain background and conceptual knowledge. But unlike the cases of models that Weisberg and French present, the scenario described in these examples can still be largely visualised. In the Dewan and Beran example, we can clearly form a mental image of the setup of the thought experiment, that is, of the two rockets, one in front of the other, linked together by a thread. We can form an image of the rockets accelerating, and then moving at a uniform velocity. We can also imagine the perspective of the two different observers—one at the starting point, one at the finish point. And in Einstein’s thought experiment, we may not be able to form an accurate imagining of what we would see, and so the example differs from say, Galileo’s falling bodies or Newton’s bucket in which the concrete elements of the scenario are familiar objects that we have experience with. However, Einstein’s thought experiment does not require a detailed or accurate image of such a scenario, and mental images can be more or less substantive. The

[67] I look to Newton’s bucket in the next chapter.

[68] Stuart provides a discussion of the light beam example in the context of “accurate” imaginings in science. Clearly, this scenario does not involve an accurate representation of the target system. Not only can we not “see” lightwaves, but travelling at that speed would cause someone to explode (and hence we couldn’t experience the lightwave) (forthcoming, 13).

images we form when we imagine travelling next to a beam of light depend upon background beliefs concerning Newton and Maxwell’s theories, but when we have these concepts, we can then form an objectual imagining of what we would see according to each of the theories.

Thought experiments describe scenarios that can be imagined in a visual way, (granted that there might be degrees to this visualisability), even if this is not always required. It is difficult to think of a scientific thought experiment that involves imagining a scenario that include any of the problematic features that Weisberg and French highlight above in the context of modelling such as continuous populations. This is at least partly due to the fact that thought experiments rely on our imaginative capacities and do not involve complex calculation and so on as in the case of models. Thus, there are differences with regards to the type of imagination involved in various areas of scientific practice and therefore, we should not be too quick to assume that an analysis of the nature of imagination in modelling could easily carry over to thought experiments, or vice versa. In the remaining sections, I turn to the varied nature of the imagination across scientific thought experiments.

4.3 Invitations to Imagine

In the previous chapter, we saw that the imagination in thought experiments has typically been taken to be an objectual form. Salis and Frigg propose a different, propositional view. I agree with Salis and Frigg that the imagination in thought experiments should not always be characterised in terms of imagery, and it can be a matter of entertaining propositions. We can also hold that some aspects of conducting individual thought experiments will not require sensory imagination. But I disagree with the scope of this claim. While we can attempt to rationally reconstruct thought experiments into a propositional or argument form, the idea I wish to defend is that at least sometimes, this will lead us to miss important features involved in their use in scientific practice. This includes the ways in which scientists call upon our imaginative capacities to convince us of an outcome or help us understand a theory or problem.

As we saw in the previous chapter, Salis and Frigg note that ‘it would be implausible to argue that individuals with a poor imagistic ability could not derive the correct outcome of Galileo’s TE (or, for that matter, of any TE)’ (2020, 40). What should we make of this? As Salis and Frigg recognise in a footnote to this point, this would be an interesting empirical question. That is, do those who do not experience mental imagery (or who have poor imagistic abilities) have problems with conducting certain scientific thought experiments? Given that no such research has been done, we cannot say either way. And detailed study of "aphantasia", that is, a condition in which there is an inability (or a reduced ability) to have voluntary mental imagery, has only recently emerged. While discussions of the phenomenon can be traced back to Galton (1880), the term was coined just a few years ago in 2015 by Zeman et al. It is said to affect around 2% of people. The term is applied to a range of cases; most people surveyed by Zeman et al still had involuntary mental imagery, for example, when dreaming (2015, 378-379). Some people did not experience any mental imagery, whereas for others, mental imagery was significantly less vivid than the controlled group. Further, there was variation with regards to whether imagery in all sensory modalities is affected. In addition to this, there is discussion around whether aphantasiacs still have unconscious (as opposed to non-existent) mental imagery (Nanay 2018, 127).

Given that this research is still in its infancy, and that we cannot answer whether there are scientific thought experiments that those who have aphantasia could not successfully carry out, it is going to be difficult to argue for the necessity of imagery in (some) thought experiments, or deny it altogether.[69] And further to this, it is plausible that some people will find reasoning via visualising more useful than others, and may even require this form of imagination in order to arrive at an outcome. Whereas for others, accompanying imagery will not or cannot be present, or if it is, it will not always be necessary.[70] In light of this uncertainty, I argue that we should shift our attention towards a different, but related, issue. This is the question of: What do thought experiments ask us to do? What kind of imaginative engagement do they invite? This will help us see cases in which objectual imagination may play a significant cognitive role.

Here, I draw on Balcerak Jackson (2016) who argues that imagining, conceiving and supposing are ways of thinking about often hypothetical objects or scenarios, but they are

[69] There are studies on aphantasia and reasoning and problem solving. For example, Zeman et al have looked into the effects of aphantasia and hyperphantasia (an abundance of mental imagery) on memory, face recognition abilities, as well as differences in methods (for example, whether conscious visual imagery was utilised) when responding to questions such as "how many windows are in your house?". Those with aphantasia reported using 'a range of alternative strategies including the use of avisual spatial imagery, kinaesthetic imagery and amodal 'knowledge'' (2020, 9). See also Watkins (2018) for a perspective on mental imagery (or a lack of mental imagery) in scientists and Jacobs et al (2018) on aphantasia and working memory.

[70] This therefore allows my account to accommodate instances of individuals such as Bohr who utilise thought experiments yet do not experience mental imagery.

three different cognitive activities that each play distinct epistemic roles.[71] I will stick to the broader distinction introduced in the previous chapter, between objectual imaginings and propositional imaginings. There is debate surrounding whether suppositions should be included under the umbrella term "imagination" which Balcerak Jackson denies. For some, such as Frigg and Salis, although they distinguish make-believe from supposition, they hold that both fall under the category of imagination. For my purposes, I need not go into this. My aim is to show that in the spirit of Balcerak Jackson's account of the difference between imagination, supposition and conceiving, different thought experiments invite different types of cognitive activity. It is also important to keep in mind that often 'suppose', 'conceive' and 'imagine' appeared to be used in an interchangeable way. One of Balcerak Jackson's aims is to demonstrate how terminological choice ought to matter.[72]

We can now see what Balcerak Jackson takes the difference to consist in. She gives the example of a thought experiment from ethics, Judith Jarvis Thomson's violinist case in defence of the right to abortion, as an instance of inviting us to imagine a situation obtaining (as opposed to merely supposing or conceiving). Or in the terminology I am adopting, an example that invites us to form an objectual imagining. In this thought experiment, you are asked to imagine waking up in the morning and finding yourself back to back in bed with an unconscious famous violinist. He has a fatal kidney ailment and it turns out that you are the only person who has the right blood type to help. You have been kidnapped, and the violinist's circulatory system has been plugged into yours so that your kidneys can be used to extract poison from his blood as well as your own—to unplug you would be to kill him. But never mind, you are informed by the doctors and medical staff, this is a temporary state; he needs to be plugged into you for just nine months.

[71] A central aim in Balcerak Jackson's paper is to get clearer on the nature of conceiving. She provides Chalmers's zombie thought experiment as a case that asks us to conceive which on Balcerak Jackson's view, is a matter of simulating what a reasoner would be rationally committed to in the situation described. I am yet to find a scientific case that asks us to "conceive" in the particular sense that Balcerak Jackson defines and so I leave this part of her discussion out for the purposes of this chapter.

[72] Thought experiments are introduced using a variety of terms—'imagine', 'picture', 'suppose', 'consider' and so on. It seems that these terms are also used interchangeably. Does it matter that scientists do not always ask us to 'imagine' the scenario described? Weisberg (2013) notes that sometimes 'imagine that' or 'visualise the following' and so on is not in science, often modelers will just stipulate that certain expressions stand in for elements of the target system. But as Odenbaugh (2015, 287) notes, even in spheres that we more readily accept as engaging imagination—literary fiction, games of make-believe and so on—'imagine that' can be replaced by other terms. See also French (2020b) for a discussion of terminology in invitations to imagine in science.

Here, Balcerak Jackson argues, ‘unless you vividly represent the scenario from the perspective of the experiencing subject, you do not really follow the invitation to *imagine* it’ (2016, 45), we can specify that this is in the sense of ‘objectual’ imagining. The correct engagement with this scenario is a matter of putting ‘yourself in the shoes of the person waking up in the hospital’; imagining seeing the violinist attached to us, feeling a certain way—surprised, scared, anxious and so on. If instead, Balcerak Jackson states, we were invited to suppose (or more generally, propositionally imagine) that the scenario had obtained ‘it would have been perfectly in accordance with the request for you simply to take the situation as obtaining, without representing it as being experienced from the first-person perspective, and in fact without representing it in any particular way at all’ (2016, 46).[73]

In contrast to the case of imagining the violinist, Balcerak Jackson gives the example of supposing that there are finitely many prime numbers. This request does not expect us to imagine this is the case in the same way the violinist thought experiment does. We are not asked to picture or to simulate an experience of finitely many prime numbers, or to adopt a certain perspective so that we see or feel or sense that there are finitely many prime numbers. Instead, we are asked to use our ‘ability to think a thought with a particular content’ (ibid., 51). To take another example of a thought experiment used in philosophy, we could consider Howell’s “Google Morals” in the context of moral deference:

> Suppose those wizards at Google come out with a new app: Google Morals. No longer will we find ourselves lost in the moral metropolis. When faced with a moral quandry or deep ethical question we can type a query and the answer comes forthwith (Howell, 2014).

As with the case of supposing finitely many prime numbers, this also seems to invite merely a propositional form of imagination, with no mental imagery or experiential component required. Even if we did imagine ourselves using the Google Morals app, or picture what it might be like, it seems that this would be unnecessary for the thought experiment to achieve

[73] While I draw on Balcerak Jackson in order to develop my pluralist account, I actually have reservations regarding the role of objectual imagination in the violinist thought experiment. What seems to be important in this case is that we have some link to affect, and this contributes to the persuasiveness of the thought experiment. However, I take it that we could propositionally imagine that we are attached to the violinist and so on, which would also give us the link to affect. Further, the outcome of the thought experiment isn’t one that seems tied to us “seeing” something in our imaginings. Rather, we form a moral evaluation. However, this issue is not significant for my argument and I take my examples below to be more convincing cases of invitations to form objectual imaginings.

its function. What we are asked to do is to propositionally imagine that there is a Google Morals app, and to consider its moral implications.

This emphasises the importance of paying close attention to particular examples. Salis and Frigg rely on one case and then generalise to all other scientific thought experiments. Similarly, the mental model theorists often talk in very general terms about reasoning in problem solving contexts to then make claims about scientific thought experiments. I argue for a pluralist view: if we look to a range of cases, and think about what their function is, we can see that there are different requirements of our imaginative capacities when engaging with different thought experiments. The focus of the discussion here then is the point at which the thought experiment has been designed by scientists in order to communicate, put pressure on, or explain an idea or theory. This is in contrast to theorising about the processes that went on in the minds of a scientist who initially conducted a thought experiment that was part of an important discovery. And so, the issue I want to focus on is to do with how scientists might invite a certain type of imagination in order to aid readers or listens of thought experiments to arrive at an outcome.

4.4 The Case for Pluralism

4.4.1 Propositional Imagination

Let’s begin with a case of a scientific thought experiment that seems to invite us to propositionally imagine only, where any objectual or experiential component is unnecessary. We can look to one of Darwin’s “imaginary illustrations” presented in *On the Origin of Species* as an example:

> ‘Let us take the case of a wolf, which preys on various animals, securing some by craft, some by strength and some by fleetness; and let us suppose that the fleetest prey, a deer for instance, had from any change in the country increased in numbers, or that other prey had decreased in numbers, during that season of the year when the wolf is hardest pressed for food. I can under such circumstances see no reason to doubt that the swiftest and slimmest wolves would have the best chance of surviving, and so be preserved or selected’ (Darwin 1859a, 42).

The function of the thought experiment is to demonstrate Darwin's theory's explanatory potential, as opposed to provide evidence in support of natural selection (Lennox 1991). Here, as with the 'Google Morals' case, it seems that we are not required to picture the wolf and the properties it is described as possessing, nor does the thought experiment ask us to adopt a perspective of the scenario. The language is descriptive, the thought experiment refers to concrete objects and processes, and it is of course possible to visualise aspects of it. But I think a Salis and Frigg-type analysis would be right in this case: to succeed, the thought experiment merely requires that we grasp its propositional content—we imagine *that* there is a wolf and so on, with no mental imagery of a wolf or phenomenal component of seeing or feeling a wolf required.

Another example that does not appear to invite the objectual imagination, this time from economics, comes from Hume:

> 'For suppose, that, by miracle, every man in GREAT BRITAIN should have five pounds slipt into his pocket in one night; this would much more than double the whole money that is at present in the kingdom; yet there would not next day, nor for some time, be any more lenders, nor any variation in the interest [rate]' (1777, 299).

As with the wolf and deer example, we could form an image of this in our minds by visualising the money being slipped into people's pockets and so on. But the thought experiment does not seem to invite such an imagining nor can it be said that forming an image of the scenario would help the reader or listener of the thought experiment to understand Hume's point. Rather, what is key to this thought experiment is that 'we understand the doubling of the money stock is indeed simultaneous, that people with more money in their hands will take some time to change their behaviour, that institutions are rigid' and so on (Reiss 2002, 27).

And so, there are clear examples of thought experiments that seem to invite propositional imagination (even if we can visualise the scenario described). Now we can think about cases in which the invitation also includes the objectual imagination.

4.4.2 Objectual Imagination

There have been many examples of thought experiments, or thought experiment-type reasoning, that cannot be reduced to a (narrow) argument form. For instance, those that involve spatial reasoning such as seeing that a square object will not fit through a circular hole

(Cooper 2005, 223) or rotating mental images of shapes as discussed by the mental model accounts.[74] But I will focus on examples from the scientific domain. To begin, we can consider Maxwell's demon. Maxwell introduced his 'neat fingered being' to show that the Second Law of Thermodynamics has only statistical validity (Earman and Norton 1998).

At the point in which Maxwell introduces the thought experiment in *A Theory of Heat*, he has already established his statistical theory of heat. He states: 'We have already shown that heat is a form of energy that when a body is hot it possesses a store of energy, part at least of which can afterwards be exhibited in the form of visible work' (1871, 308). As Stuart (2016a) discusses, the thought experiment is not used to offer support to the theory. Rather, its aim is to aid our understanding of the theory. Understanding is a topic of significant interest in current philosophy of science, and there has been a huge amount of research in recent years highlighting the value of understanding in science, as well as its relation to knowledge (De Regt et al 2009, De Regt 2017, Elgin 2017, Khalifa 2017). As noted in the introduction, the ways in which thought experiments contribute to understanding has largely been overlooked and instead, the focus has been on how they bring about new knowledge.

There are many ways in which we could utilise this literature to shed light on thought experiments (see Stuart 2016a, 2018 and Meynell 2020 for discussion) and I will come back to scientific understanding in the final chapter when thinking about thought experiments' aesthetic qualities. But one way of drawing on this literature is through attending to how some philosophers of science have drawn links between scientific understanding and visualisation. This includes a prominent account of understanding as developed by de Regt (2014, 2017) in the context of theories. In particular, he discusses visualisation in the context of twentieth-century theoretical physics.[75]

[74] For example, there are studies in which participants are presented with a pair of images and asked whether the images are identical. The two figures are presented from different angles and participants report answering the question through forming mental images that allowed them to rotate the figure in their minds to see if they could make it match the other image. And these types of tasks often take longer when the images needed rotating to a higher degree (Nersessian 2018, 314).

[75] Visualisation is of significant import to De Regt. Other prominent accounts of understanding also mention mental images. For instance, Khalifa discusses the explanation of scattering phenomena. Feynman's model added a mechanical interpretation to the existent purely mathematical explanation which allows for a simple mental image. But unlike for De Regt, this is not central for Khalifa; 'the value of qualitative insight (a shared language, visual images, etc.) is exhausted by its facilitation of good old-fashioned hypothesis testing' (2017, 45). See also Elgin (2017, chapter 7) for a discussion of images and diagrams (that can aid the imagination) in scientific understanding.

A key notion for De Regt is intelligibility; ‘the value that scientists attribute to the cluster of qualities of a theory (in one or more of its representations) that facilitate the use of the theory’ (2017, 40). While not essential for achieving understanding, one of the qualities included in this cluster is visualisability which de Regt argues is an effective tool for understanding. Briefly put, de Regt’s view is that scientists often prefer visualisable over abstract theories and further, find pictorial representations useful in understanding. He further states that this is to be expected given that ‘seeing is for humans plausibly the most important way of grasping the world around us’ and so ‘when we want to extend our grasp of the world beyond what we observe directly, we prefer to rely on our well-developed visual skills and employ visualization as a tool for understanding’ (ibid., 257).[76] Thought experiments are one way in which theories can be made intelligible. And it appears that in the demon case, Maxwell deliberately engages our non-propositional imagination to help us understand the statistical basis of the second law of thermodynamics.

Let’s see how visualisation or more broadly, objectual imagination, may help in this case. Maxwell’s thought experiment is outlined as follows:

> ‘[I]f we conceive of a demon whose faculties are so sharpened that he can follow every molecule in its course, such a being, whose attributes are as essentially finite as our own, would be able to do what is impossible to us. For we have seen that molecules in a vessel full of air at uniform temperature are moving with velocities by no means uniform, though the mean velocity of any great number of them, arbitrarily selected, is almost exactly uniform. Now let us suppose that such a vessel is divided into two portions, A and B, by a division in which there is a small hole, and that a being, who can see the individual molecules, opens and closes this hole, so as to allow only the swifter molecules to pass from A to B, and only the slower molecules to pass from B to A. He will thus, without expenditure of work, raise the temperature of B and lower that of A, in contradiction to the second law of thermodynamics’ (Maxwell 1871, 338-339).[77]

[76] The problems raised by Weisberg in the context of unvisualisable models may apply to some theories. Here, I am just concerned with De Regt’s account in so far as it can be useful for considering how thought experiments increase understanding and so, I need not respond to such worries here.

[77] Some of the language may be confusing for our aims here as Maxwell uses both ‘conceive’ and ‘suppose’, but as highlighted above, these are often used interchangeably in the context of thought experiments. Because of this, I take it to be best to understand the invite to be a general one of ‘imagine’, and then we can consider whether it is best thought of as a propositional or an objectual kind. See French (2020b) for a discussion of the language used in invitations to imagine.

The thought experiment describes a demon who can control a door separating a box of hot gas with faster moving molecules, and a box of cold gas with slower moving molecules. The demon can selectively open the door so that heat flows from the cold gas to the hot gas, making the hot side hotter, and the cold side colder. This violates the second law and therefore demonstrates that the Second Law can only have statistical validity.

We can adopt the position of the demon who has this capacity greater than our own, and we can form a visualisation of the box and the molecules from the demon's perspective. Alternatively, we can imagine this without taking on the perspective of the demon, but rather by visualising the demon and his movements in the vessel. It is clear that in this thought experiment, although we could never do what the demon does (given his "sharpened faculties") and we may have a problem visualising seeing molecules in the way the demon does, we do not have a problem with visualising the state of affairs described. And in fact, much of the value of thought experiments lies in how they provide us with scenarios that makes something difficult easy to grasp, and as mentioned, do not appeal to mathematical reasoning to succeed. This therefore contrasts to the difficulties with forming mental pictures of some model descriptions as outlined above, that is, those that involve probabilities or perfect rationality.

Stuart states that a crucial part of how Maxwell's demon aids our understanding is because it relates the second law of thermodynamics to experiences that we already have: 'We may have trouble imagining a being that can see molecules, but if we imagine *ourselves* in an analogous position say, in control of a sliding door, surrounded by molecules which act like medium sized rubber balls, we understand the scenario perfectly' (Stuart 2016a, 27). This results in us understanding that the fact that we will not experience violations of the second law is down to our lack of capacity to do what the demon does, that is, track the individual molecules. And thus, the theory is made intelligible to us; we understand how such violations could be possible.

And so, reflecting on the role that thought experiments can play in understanding allows us to see how the objectual imagination can play a significant role in their usage. What other candidates are there for thought experiments that ask us to do more than to consider a set of propositions? That is, those that ask us to put ourselves in a particular situation, visualise a state of affairs, and/or imagine what we would observe.

Firstly, we can consider thought experiments in which we are asked to imagine different perspectives on the same scenario. For example, Einstein's elevator which revealed the connection between gravity and accelerated frames of reference underpinning his general theory of relativity. In Einstein's and Infeld's *The Evolution of Physics* they introduce 'idealized experiments created by thought'. These 'may sound very fantastic' but they 'help us understand as much about relativity as possible by our simple methods' (1938, 226). In this thought experiment, we are to imagine an inertial co-ordinate system (c.s.) in which an elevator is being pulled upward with a constant force. 'Since the laws of mechanics are valid in this c.s., the whole lift moves with a constant acceleration in the direction of the motion' (ibid., 231). We then consider what is going on inside the elevator from the perspective of two different people, one inside the elevator, and one outside:

> '*The outside observer:* My CS is an inertial one. The elevator moves with constant acceleration, because a constant force is acting. The observers inside are in absolute motion, for them the laws of mechanics are invalid. They do not find that bodies, on which no forces are acting, are at rest. If a body is left free, it soon collides with the floor of the elevator, since the floor moves upward toward the body . . . .
>
> *The inside observer:* I do not see any reason for believing that my elevator is in absolute motion. I agree that my CS, rigidly connected with my elevator, is not really inertial, but I do not believe that it has anything to do with absolute motion. My watch, my handkerchief, and all bodies are falling because the whole elevator is in a gravitational field' (ibid, 231).

Einstein and Infeld then consider a way of determining which observer is right. We are asked to imagine a light ray entering the elevator horizontally through a window which reaches the opposite wall. The outside observer (who believes in the accelerated motion of the elevator) argues that the light ray would enter horizontally but as the elevator moves upwards, it would travel (relative to the elevator) not in a straight line towards the opposite wall, but rather in a slightly curved line. Whereas the inside observer (who believes that the whole elevator is in a gravitational field) states 'there is no accelerated motion of the observer, but only the action of the gravitational field. A beam of light is weightless and, therefore, will not be affected by the gravitational field' (ibid, 233). The light beam will enter horizontally and then travel in a straight line to the opposite wall.

The thought experiment thus gets us to shift our imagination between the two perspectives, allowing us to consider what they would observe and to consider the differences between their descriptions of the light beam's path. Einstein and Infeld go onto explain that there is a mistake in the inside observer's description: 'A beam of light carries energy and energy has mass' (ibid., 234). Because of this, the light beam will bend 'exactly as a body would if it was thrown horizontally with a velocity equal to that of light' (ibid.). Thus, we see how 'the problem of general relativity theory is closely connected with that of gravitation and why the equivalence of gravitational and inertial mass is so essential for this connection' (ibid., 235).

Secondly, we can consider thought experiments that involve spatial reasoning. One example as discussed in the previous chapter is Stevin's chain, where we can imagine the movement of the balls on the thread. Further, Starikova and Giaquinto (2018) discuss how mathematicians reliably imagine using visual mental imagery (that differs from applying mathematical rules) in examples of thought experiments in knot theory, graph theory and geometric group theory, and these contribute to mathematical knowledge. Like De Regt, they highlight how visualisation allows us to utilise our past perceptual experiences in order to problem solve. For example, some thought experiments require us to consider whether various knot diagrams are diagrams of the same knot. To do this, 'one must perform one or more trials, a trial being a finite sequence of steps, each of which consists of (a) visualizing a deformation in 3-space of a knot as represented by one seen diagram and (b) drawing (or otherwise producing) another knot diagram corresponding to the projection of the knot at the end of the visualized deformation so far' (2018, 260).[78]

Thirdly, we can attend to the fact that many thought experiments are presented alongside images. Again, think of Stevin's chain which is often accompanied by a diagram. The picture of the fourteen balls on a thread hanging over the prism is an aid for our imagination—the image allows us to grasp immediately what the described scenario looks like, what its most essential features are, and thus helps constrain our imaginings. Further, we can combine our imaginations and what the image presents to us in order to manipulate the static image and to reason through the thought experiment, considering the ways in which the balls on the thread will move. Thought experiment descriptions, then, can point us towards certain points of the picture that they are presented alongside which directs our attention in the right way. Treating

[78] While these cases are dependent upon background knowledge regarding how mathematical definitions link with the physical objects imagined, once we have this, we can carry out visual thought experiments 'without adverting to foundational definitions' (Starikova and Giaquinto 2018, 262).

this thought experiment as relying solely on propositional reasoning fails to recognise the ways in which the image—whether in our minds or presented in front of us, or both—has a significant cognitive role. We can compare this with the Darwin example outlined above. It would be difficult to see how an image of what the thought experiment describes—that is, the interaction between wolves and deer—would contribute to the imaginative exercise that the thought experiment narrative prompts.

Meynell (2018, 2020) offers a detailed account of the function of images that accompany scientific representations, including thought experiments.[79] She highlights how the use of pictures could be especially useful when the audience of the thought experiment is not a scientific expert; in her discussion of Einstein's train, she argues that the 'imaginings are counterintuitive from a common sense perspective…Many thought experimenters will have to work against their own habits of mind and quite possibly implicit beliefs in order to imagine as they are directed' (2018, 506). And so, perhaps for some, especially those who are more familiar with the theoretical background of a thought experiment, the image will be superfluous in obtaining the outcome. Whereas for others, an image or diagram will play an important role in guiding the imaginative activity.

More generally it can be highlighted that readership may impact what kind of imagination (or whether an external image) is appropriate. Returning to the example of Maxwell's demon, it is interesting to note that Maxwell first presented the demon thought experiment in various letters to other scientists including physicists Tait and Strutt (Klein 1970, 86). Later, it was published (in the form presented above) in *The Theory of Heat.* The end of the book included a caption of the work by the publishers who described the series it was a part of as 'text-books of science adapted for the use of artisans and of students in public and science schools'. Further, they were meant to be comprehensible to the 'working man', and theories should be 'reduced to the stage of direct and useful application' (Maxwell (1871), discussed by Klein (1970, 89)). The book itself was considered very difficult, even by those trained in physics and therefore can be regarded as ill-suited to the purpose the publishers had in mind. But if we attend to the context of the demon thought experiment, namely that it is situated in a text that is meant for a wide readership, then we can see the benefit of offering an example that

[79] Meynell argues that thought experiments that are 'inherently spatial, relational and causal' are those that are best displayed pictorially (and thus mere linguistic representations are limited). This also explains why many physics thought experiments, yet not many philosophical ones, are presented alongside pictures (2018, 502). See also Sheredos and Bechtel (2020) for a discussion of graphics and diagrams in biology and how these interact with imagination.

has a visualisable character; it can aid students to better engage with the otherwise complex subject matter. Similarly, consider Galileo's thought experiments as presented in the *Dialogues.* This was not just aimed towards other scientists but rather to a more general community. As a consequence, when Salis and Frigg and Norton talk of what the conduct of a thought experiment consists in, and argue that it is a matter of propositional and/or argumentative reasoning, this is too simple; the performance of a thought experiment might be different depending on who it is that is imagining the scenario.

And so, while there might be thought experiments that appear to merely invite our propositional imagination, there are clearly also those that invite objectual imaginings as well, some of which may be presented with images which aid our imaginations. I want to end by emphasising that I do not think this is a case of either or. In fact, we can think about how the different types of imagination may work together when we are conducting thought experiments. The propositional imagination may play a crucial role in presenting the scenario which then directs certain objectual imaginings that follow. And imagining in an objectual way can bring our attention to important features of this set up, or can illuminate the consequences of these imagined propositions. Another option is that we can be presented with an image and/or form a visualisation that works as, to use a term from French, a "hook". In the context of theorising, French discusses the quasi-sensory elements that 'function as the 'hooks' on which we can hang the 'belief-like' features' of imaginings (2020b, 27). We can then deliberate, using propositional imagination, the ways in which the scenario could unfold.[80]

For example, in Stevin's chain, the image we are presented with, and the way in which we use our imagination to manipulate the image, functions as the hook. Within this, we propositionally imagine the options; the balls either stay still or they move to the left or to the right, before then utilising our objectual imagination again to "play out" these options. And these objectual imaginings play a part in the demonstrative force of the thought experiment, that is, they aid our realisation that the balls must remain static. And in Maxwell's demon, the thought experiment calls upon our propositional imagination in order to set up the scenario and to define what the demon's capacity is. Within this set up, we then form objectual

[80] French discusses Einstein's 'On the Electrodynamics of Moving Bodies' (1905) in which Einstein asks us to imagine certain 'quasi-sensory' elements, in this case, clocks and rods which are then stipulated as 'perfect' (in the case of the clocks) and 'practically rigid' (in the case of the rods) (French 2020b, 27). See also Todd (2020) for a discussion of how propositional and non-propositional imagination can work together in science.

imaginings of the controlling of the door and the movement of the molecules. This helps us understand how such a scenario could obtain which in turn, aids our inference that the second law of thermodynamics could be violated.

As mentioned, a propositional view of the imagination in science does capture some of the reasoning involved in thought experiments, and a proponent of that view may insist that we could *reconstruct* all of the above examples within the propositional framework. But it is evident that scientists utilise their visual imaginations while engaging with certain thought experiments and have reason to invite members of the community they are communicating with, whether scientific or public, to do the same. A reconstruction into propositional form will therefore distort cognitively important features. A complete account of the imagination in science needs to accommodate such instances, and a pluralistic view allows us to do so.

4.5 Conclusion

We have seen that philosophers have typically taken the imagination in scientific thought experiments to consist in mental images. A recent challenge insists that it is only a propositional form that is required. I have argued that while I think this offers an important insight into how some thought experiments work, or how aspects of thought experiments work, I disagree with the scope of the claim. If we ask: ‘what do thought experiments ask us to do?’ it becomes evident that they appeal to a variety of our imaginative capacities, and some demand a type of imaginative activity that goes beyond the consideration of propositions. Consequently, we should embrace the richness of the imagination and the different resources it can bring into play when thinking about how scientists construct thought experiments for different purposes. I have argued that we should adopt a pluralist stance rather than limiting an account of the imagination in thought experiments to one type, whether imagistic or propositional.

In the next two chapters, the ways in which thought experiments employ our imaginations, and how far this can be said to be similar to our engagement with literary fiction, will be further discussed in the context of the aesthetic features of thought experiments.

5. The Aesthetics of Science: The Case of Thought Experiments

5.1 Introduction

Chapters 3 and 4 looked at the nature of the imagination in scientific thought experiments. I defended a pluralist view of the imagination in science, arguing that different thought experiments call upon different types of imagination and therefore, we should not limit our analysis of the imagination in the scientific context to one particular type, whether propositional or imagistic. In this chapter and the one that follows, I develop a new approach to the aesthetics of science. My account demonstrates how aesthetic features contribute to a thought experiment being a suitable prompt for, or aid to, the imagination and are thereby informative for useful imaginations. Further, my view offers a way of advancing the comparisons drawn between thought experiments and literary fictions.

The discussion of aesthetic value in science has primarily focused on the evaluation of theories or of mathematical proofs (McAllister 1996, Breitenbach 2015, Ivanova 2017a, 2017b). Questions then arise regarding the role that aesthetic values play: How can beauty motivate scientific research? Do aesthetic properties have an epistemic function? Can they indicate the truth of a theory? And in what way do aesthetic features aid scientific understanding and guide theory choice in instances of under-determination (where empirical evidence is insufficient to choose between competing theories)? While aesthetic value includes, but is not limited to beauty, philosophers have taken as their starting point claims from mathematicians and scientists emphasising the import of beauty in their domain. To take just a couple of examples, Hardy states that ‘beauty is the first test: there is no permanent place in the world for ugly mathematics’ (1940, 14), and for Dirac, ‘one has a great confidence in a theory arising from its great beauty, quite independently of its detailed successes’ (1980, 40).

And it is not just scientists and mathematicians who have highlighted these features. The discussion of the beauty of theories can be seen in a seminal work in aesthetics, Hutcheson’s *Inquiry Concerning Beauty, Order, Harmony, Design*. Hutcheson notes that theories ‘and universal Truths, in General causes’ (1726/2004, 24) including Newton’s gravitational principle and Euclidean geometry often strike us as beautiful. And this observation is not just a passing comment. Rather, it plays an important role in Hutcheson’s account; it motivates the need for a theory of aesthetics that does not characterise beauty as a sensation that is akin to those correlated with the five senses.[81]

While there has been attention given to the aesthetics of theories, thought experiments are another important part of scientific practice which have an aesthetic dimension, but they are currently overlooked. This is surprising given that they are designed to engage the imagination and have been compared to other aesthetically appreciated objects, namely works of art. In particular, thought experiments are said to share qualities with literary fiction as they invite us to imagine a fictional scenario and often take a narrative form (Elgin 2014, Davies 2007). Given these similarities, many have used thought experiments to defend the cognitive value of literature. The idea is that literary fictions are extended, more complex thought experiments and hence we can learn from engaging with them.

In this chapter, I look to the aesthetic and literary qualities of thought experiments. I begin with a discussion of how thought experiments have been used as a way of defending the cognitive value of literature, including reasons why we might be sceptical of such comparisons. I then outline how thought experiments are also evaluated aesthetically, that is, how they are considered beautiful and elegant and so on. Consequently, thought experiments ought to be a part of the discussion surrounding the aesthetics of science. I go on to raise some worries for any aesthetics of science project. The problem is that it appears that either the application of aesthetic terms to scientific cases are really disguised epistemic features, or that the application of aesthetic terms is literal, but is scientifically irrelevant.

In the next chapter, I propose my own account which looks to an overlooked source of aesthetic value in science; the fit between form and content. This offers a way of defending the role of aesthetic features in science without reducing them to epistemic features. This is because the interrelation between form and content is a source of aesthetic value that has important epistemic pay offs. I flesh out these pay offs in terms of the ways in which thought experiments can enhance understanding and play a persuasive role in an argumentative context. To end, I come back to the comparisons between thought experiments and literary fictions in light of the theory of aesthetics of science that I develop, and I argue that we should

[81] In discussing non-perceptual art, Shelley (2013) offers an outline of Hutcheson's aesthetics, including his argument that the sensation of beauty is an 'internal' or 'mental' sensation (as opposed to an 'external' one, i.e. one that is perceived by the five senses). See also Kivy (1992) for an overview of Hutcheson's aesthetics.

look closely at the literary examples used when thinking about the qualities they share with scientific thought experiments.

5.2 The Cognitive Value of Literature

Thought experiments can be characterised as taking the form of short, fictional narratives that have the purpose of instructing the reader to evaluate the described scenario in a certain way. In the philosophy of art, comparisons have been drawn between thought experiments and artworks, particularly works of literary fiction, as they share (at least some of) the key features of thought experiments, namely their fictionality—the events have not actually taken place, or at least, whether they have or not is inessential—and narrative form.[82] Further, the use of thought experiments in learning has been offered as a way of defending the cognitive value of literature.

A central issue in philosophy of art is to do with whether and how artworks can provide knowledge or understanding, as well as whether a work's cognitive value contributes to its aesthetic value. That is, do cognitive merits or flaws of an artwork affect the value of art as art?[83] Given the comparisons made with thought experiments, I'll focus on literary works. We can, of course, learn about art from engaging with artworks; reading Kafka's *The Metamorphosis* teaches us something about the novel and about Kafka's literary style. Similarly, through literature we could learn about historical, geographical or scientific truths. Rooney's *Normal People* might teach us facts about the city of Dublin and its various parks and buildings of interest, for example. While there might be interesting things to say about the ways in which works of literature can teach us in this sense, the focus of the debate concerns the ways in which literary works can be said to teach us deep truths concerning human nature, morality, relationships, politics and so on. For instance, many come away from Shakespeare's *Othello* with a sense that they have learned something about human psychology, including the corruptive power of jealousy. Importantly, these insights are part of why we value *Othello* so highly. This contrasts with the case of Rooney's *Normal People*

[82] A further issue is whether all scientific thought experiments count as narratives. Here, I am just going to focus on cases that do take a narrative form.

[83] For example, Lamarque argues: 'To value a work of art as a work of art is *not* to value it for its truth or the knowledge it imparts or its capacity to teach. In short, truth is not an artistic value' (2006, 326).

teaching us something about the geography of Dublin; this is not part of why we value the novel as a work of art.

There is a compelling set of views that take seriously the ways in which literature can inform us about the world, and our social, political and psychological lives by pointing out how thought experiments invite us to engage in an imaginative exercise and ask: 'What would happen if the following was to occur?". Typically, we take thought experiments as having epistemic value and so, characterising literary fictions as a kind of extended, more complex thought experiment allows us to maintain that engaging with narrative art can lead us to new insights about the world and ourselves. For example, Carroll argues that Greene's *The Third Man* is a thought experiment that presents a powerful counterexample to the maxim "When loyalty to a friend conflicts with loyalty to a cause, one ought to choose in favour of the friend" (2002, 10). And for John, fictions such as Paley's short story *Wants* function like a philosophical thought experiment 'in which problematic imagined cases are used to prompt responses relevant to philosophical problems'. John argues that in our engagement with *Wants,* we are led to explore the concept of desire. This is therefore similar to the way in thought experiments can address questions about our conceptual schemes (1998, 332). Further, John takes it to be literarily valuable for a work to challenge us in this way, that is 'to push us to examine what our concepts mean and what we use them to do: the capacity to inspire that kind of activity is one of the things that can make a work of fiction interesting and good' (ibid., 331).

While many discussing thought experiments in the philosophy of literature have focused on examples in philosophy, Elgin (2014) and Davies (2007) have used thought experiments as a way of bringing together issues in philosophy of art and philosophy of science. And it is, of course, scientific thought experiments that are my focus here. Let's look at another example, this time from Newton. The thought experiment sets out to undermine Descartes' relational account of motion. We are asked to imagine a bucket hanging from a long rope that is twisted tight. The bucket is then filled with water and the rope is released, making it unwind and the bucket spin. At first, the water and the bucket are in relative motion, and the water is still flat in the bucket. But after some time, the water will pick up the motion of the bucket, forming a concave shape as the water rises up the sides of the bucket. Now there is no relative motion between the water and the bucket—which is how it was in the initial starting point, before the rope was released. So how do we explain the observed difference between the first state (before the bucket was released) and the final state (when the water forms a concave shape), when in each case, there is no relative motion? Newton explains that the motion of the water is absolute and not relative to the bucket, and has to be represented as such in absolute space (Brown 2011, 8).

Elgin argues there is a continuity between physical experiments, thought experiments such as this case from Newton, and literary fictions. Experiments and thought experiments involve studying an object or a system that stands in for a target system, and they each require us to control our (real or imagined) set up, ensuring that we carefully isolate the features that we are interested in investigating. We can note here a key difference between, say, Galileo's falling bodies and Newton's bucket thought experiment, that demonstrates how much scientific thought experiments can vary with regards to their departure from how things are. As Elgin notes, like ordinary experiments, thought experiments involve the study of simplified and distorted versions of nature. In the Galileo case, there is a more straightforward idealisation; the abstraction of air resistance. Further, we can easily imagine ourselves going to the top of a tower and performing the thought experiment. Whereas Newton's bucket places more demands on our imagination. We are required to imagine that there is nothing in the universe except the bucket filled with water hanging from a rope. Although the rope remains, it is not tied to anything, and even the earth (whose gravity keeps the water in the bucket) does not exist.

For Elgin, this control of our scenario, and the use of idealisation carries over to literary fiction: 'a work of fiction selects and isolates, contriving situations and manipulating circumstances so that patterns and properties stand out' (2014, 232). On this view, fiction functions as a thought experiment that provides us with new insights or understanding, it 'may frame or isolate mundane features of experience so that their significance is evident. It may defamiliarize the commonplace, making us aware of how remarkable normal behavior can be' (ibid).[84] As with thought experiments, works of literature differ in terms of how much, and in what way, they depart from reality.

[84] Elgin discusses these connections in the context of her account of understanding. In particular, through her account of "exemplification". An experiment, thought experiment and a work of fiction exemplify certain properties—they both have and refer to those properties. Scientists and artists thus manipulate their representations with the aim of making certain features salient which directs us towards a certain way of understanding the target. She offers the example of an Austen novel in which 'relations among the three or four families are sufficiently complicated and the demands of village life are sufficiently mundane that the story can exemplify something worth noting about ordinary life and the development of moral personality' (2014, 233).

However, many have highlighted the difficulties in pinning down what it is that artworks teach us, and have argued that reducing the cognitive value of literature to the ways in which they can be said to be similar to philosophy and the sciences (including thought experiments in these domains) fails to fully capture what is distinctive about literary works and the insights they can offer (John 2013, Gibson 2008). Further, there is the potential issue that on the thought experiment approach, we are no longer locating the cognitive value of art within the artwork itself, but rather in an activity that falls outside of the work; in reading a novel in a particular way, selecting certain features to make broader argumentative points (I return to this below). Take the example of Morrison's *Beloved.* The novel opens as follows:

> '124 was spiteful. Full of a baby's venom. The women in the house knew it and so did the children. For years each put up with the spite in his own way, but by 1873 Sethe and her daughter Denver were its only victims. The grandmother, Baby Suggs, was dead, and the sons, Howard and Buglar, had run away by the time they were thirteen years old — as soon as merely looking in a mirror shattered it (that was the signal for Buglar); as soon as two tiny hand prints appeared in the cake (that was it for Howard). Neither boy waited to see more; another kettleful of chickpeas smoking in a heap on the floor; soda crackers crumbled and strewn in a line next to the doorsill. Nor did they wait for one of the relief periods: the weeks, months even, when nothing was disturbed. No. Each one fled at once — the moment the house committed what was for him the one insult not to be born or witnessed a second time' (Morrison 1988, 3).

The novel beings in the late 1870s, in the aftermath of the American Civil War, and tells the story of Sethe and her family who live in a haunted house, avoided by others in the neighbourhood. The novel goes back and forth from this present to the time that Sethe was enslaved, piecing together how she escaped and when found, killed her own daughter to spare her being captured. When Sethe begins a relationship with Paul D, he exorcises the ghost living at 124, which then results in the arrival of Beloved, a young woman in her twenties. Sethe comes to realise that this is the return of her daughter.

Morrison is explicitly political in her work and in her analysis of it, and it is evident that many come away with a sense that the novel has enlightened them. But as the opening passage indicates, the novel's exploration of the haunting impacts of slavery, and its lasting trauma for those who were enslaved and the generations that follow, as well as of the struggle of making sense of choices made in heinous circumstances, are difficult to reduce to cases of testing or exploring some particular concept or proposition. Further to this, works of literature do not pose what can be regarded straightforwardly as arguments (unlike say works of philosophy), and they often contain ambiguous and contradictory claims that are not meant to be taken as true (in the fiction) in any straightforward sense such as the opening sentences above: ‘124 was spiteful. Full of a baby’s venom’.

It might be possible to approach *Beloved* as a thought experiment that explores certain questions regarding the psychological impacts of slavery post-abolition, for example. But such an approach does not appear to do justice to the work. This is not to undermine all comparisons drawn with thought experiments in order to defend how some literary fictions can offer insights. But it is worth noting that in light of these features of novels and other narrative fictions, others have proposed a different approach, expanding what it means to say that art has cognitive value. It has been argued that we should instead look to how literary fiction gives us the opportunity to gain access to experiences that we otherwise would not have had through rich descriptions of others’ thoughts and feelings, therefore broadening our perspective beyond our own. Others have highlighted how literature provides an emotional experience through the imagination which can in turn bring about an understanding of ourselves and others. Proponents argue that a benefit of this approach is that it offers an argument in defence of the cognitive value of art as art, that is, this is something that is internal to the practice of reading works of literature (see Gibson (2008) and Vidmar (2010) for helpful discussions on the debates over art and knowledge).

And so, there are on-going debates in philosophy of art surrounding how and what exactly we can learn from literature, as well as how far the way in which we learn can be compared to the pursuits of science and philosophy which are seen as more apparent epistemic domains. But the key point is that there are views that maintain that (at least some) narrative art functions like thought experiments, and that this helps explain how we can learn from engaging with these works. Consequently, an analysis of thought experiments and their epistemological value has offered a way of drawing fruitful analogies between philosophy of science and philosophy of fiction. In the next chapter, I will come back to the comparisons between thought experiments and literary fictions and think about which examples from narrative art are most appropriate.

For now, I will put to one side the question of whether literary works can be said to function as thought experiments, and instead I will consider how the comparisons can be drawn the other way as well, that is, from aesthetic and philosophy of art in order to illuminate the cases in science. My aim is to address how the aesthetic choices scientists make in the design of thought experiments contribute to the function of the thought experiment: to communicate, convince, or explain a theory or phenomena to a scientific or a public community. I argue that when thinking about the commonalities between scientific and artistic representations, thought experiments are a fruitful case study for philosophers of science. Part of their value in science are the features they share with literary works.

The key issue is whether the aesthetic qualities provide anything beyond catching and maintaining our attention or at best, are a mere heuristic aid. In the next section, I am going to consider accounts that argue this way. I will consider some of the proposed disanalogies between literary fictions and thought experiments that are said to undermine purported connections between how we learn from scientific and artistic representations, and the role of aesthetic considerations in science.

5.3 Disanologies Between Literature and Thought Experiments

Egan (2016) offers a recent and thorough discussion of the comparisons between philosophical and scientific thought experiments and works of literature as outlined above, and sets out to undermine the force of such comparisons. Egan's primary concern is with the claims regarding how the cognitive value of thought experiments can explain literature's cognitive value, but the disanalogies he draws are relevant for my purposes in this chapter, that is, for the function of the aesthetic and literary qualities of thought experiments. And so, it is worth spending some time on his account.

Egan allows that the similarities between thought experiments and literary fictions entails that a work of literature, or a section of it, could be *used* as a thought experiment or part of one within, say, a philosophical paper. For example, a work of philosophy that argues in favour of the importance of freedom in our choices even when this might lead to morally reprehensible decisions might cite an excerpt from Burgess's *A Clockwork Orange.* But in making this point, Egan echoes the concern raised above that the cognitive value of art needs to lie in our engagement with art as art. Egan states that while literature could be effectively put to use in this way (and some philosophers clearly refer to literary fiction in their work), this is not adequate for a defence of the view that cognitive value of literary fictions lies in its similarities with thought experiments (or the stronger claim that literary fictions are thought experiments) (2016, 141).

Egan's account is different from Norton's in that he is not committed to the claim that thought experiments are arguments that work through transforming our existing knowledge through a logical process. Nor is he committed to the claim that the actual conduct of a thought experiment consists in the execution of an argument (Norton 2004, 50). Egan does, however, argue that there is an essential connection between scientific and philosophical thought experiments and argument that leads to crucial disanalogies between thought experiments and literary works. Egan grants that thought experiments can have a diverse role in arguments, but the important point is that thought experiments are always a part of a larger, argumentative structure.[85] Otherwise, Egan states, they would be merely 'intriguing narratives' (2016, 142). For example, Thomson's violinist thought experiment (as outlined in chapter 4) is part of an argument surrounding the legitimacy of abortion that spans a whole paper, and for Egan, the thought experiment is 'essentially tied' to this wider discussion; it is 'only within their contribution to this larger argumentative structure that they get their distinctive cognitive payoff' (ibid.).

Further to this, Egan claims that the purpose of a thought experiment 'is *exhausted* in making or contributing to an argument. Whatever other aesthetic qualities that narrative might contain are irrelevant to the purposes of the thought experimenter' (ibid., 142). He goes onto qualify the claim, stating that such features may 'make the thought experiment more vivid or more compelling, but this vividness and compellingness is useful to the thought experimenter only insofar as it contributes fruitfully to the argument' (ibid.). Like Norton, then, whose argument view of thought experiments renders their typical narrative form and any of their aesthetic qualities as irrelevant to the conclusion and therefore dispensable, Egan maintains that 'we can, as it were, throw away our thought experimental narrative once its work is done' (ibid., 142). Because of this, Egan claims that thought experiments' role in arguments is "fungible". In the case of Thomson's violinist, a different thought experiment could have been used in its place if it did the same argumentative work: 'Nothing in her argument requires this particular story about the violinist' (ibid., 143).

[85] A counterexample to this (as discussed in chapter 1) is Einstein's recollection in his autobiography of imagining himself chasing a beam of light, and considering what he would observe. This thought experiment, in its initial stage, was exploratory and separate from any larger argumentative structure.

Egan contrasts the essential connection between thought experiments and their situation within a larger argument with works of literary fiction. We do not approach works of literature in the same way as philosophical or scientific thought experiments, nor do we have a sole purpose in mind for an artwork, such as making an argument. Rather, there are a plurality of reasons for engaging with literary fiction, and when we come to literary works, we do not assign it a particular purpose nor do we take the work as addressing a certain set of questions. If we did, Egan argues, we would massively limit what we could gain from the work than if we approached it with a more open mind. And this means that literary works are not replaceable in the way that thought experiments are. Egan gives the example of deriving amusement from a literary work. We do not regard the work simply as a way of deriving amusement, nor do we think that any other amusing story (however funny) could be in its place—he states, 'the uniqueness of just *this* story remains' (2016, 143). Similarly, we would not want to replace *A Clockwork Orange* with a work that could equally be said to demonstrate the importance of freedom in choices, for instance.

Egan takes it that there is value in say reading novels that could not be produced by anything else, and so a particular novel cannot be replaced by another work of art. Further, the work is not dispensable once we have read it, that is, it is not a means to an end.[86] Rather, we come back to works of literature with the expectation that we may gain something new from them that on our previous readings we overlooked. This is in contrast to a thought experiment which, the argument goes, we would only re-read in order to remind ourselves of its role in an argument or to consider possible criticisms regarding how far the thought experiment can be said to achieve its purpose. We do not re-visit Thomson's violinist thought experiment with the expectation that we may learn something beyond this.

For Egan, a key part of this difference is that unlike literary works, thought experiments are not concerned with particulars. When presenting the violinist thought experiment, Thomson's concern does not really lie with the idea of keeping alive a famous violinist by attaching them to our bodies, nor is Galileo ultimately interested in musket balls and cannon balls when he describes dropping them from a tower. Egan grants that the 'concrete elements in a thought experiment make an arguably ineliminable contribution to exploring these relations: they make salient the relations under examination with a clarity that would arguably be impossible in a purely abstract treatment'. But these particulars are merely a means to an end: they serve

[86] For critical discussion of the unique value of artworks see Stecker (2010).

as a way of exploring more abstract problems (ibid., 144). In the Galileo case, this would be relation between the weight of a body and the speed at which it falls. But Egan highlights that the same cannot be said for works of literature, where ‘the concrete particularities of narratives are irreducible parts of what we attend to when we read a narrative as literature’ and so, ‘they cannot be straightforwardly reduced to abstract ideas’ (ibid.). In addition to this, Egan argues that while literary works can of course bring about more abstract or general reflections, they do not need to. He states ‘I can derive great interest and pleasure from reading *Pride and Prejudice* as a story about Elizabeth Bennett’ and so on, without drawing more general conclusions about say, class or marriage (2016, 145). Whereas engaging with Galileo’s thought experiment without making the more general conclusions beyond the particularities of the case would be to fail to engage with it as a thought experiment.

Further to this, the ways in which we can go beyond the particularities in the text and explore more abstract and generalisable conclusions when reading works of literature does not have a defined limit. This is compared to the singular purpose of a thought experiment as contributing to a specific argument. Because of this, we cannot abstract away from the ‘concrete particularities of [for example] Kafka’s narrative if we want to think about it as a work of literature’ (2016, 144). This is related to a discussion presented by Frigg and Nguyen on the ways in which we interpret artworks on one hand, and scientific representations on the other. Frigg and Nguyen (2017) highlight the similarities between representation in art and science, and they apply their own framework, the DEKI account—in which representation consists in denotation, exemplification, keying-up and imputation—to both scientific and artistic cases.[87] However, they also outline an important difference which is to do with what they call ‘the flexibility of interpretation’ in artistic representations compared to scientific ones. In the case of scientific models, they claim that the interpretation ‘is usually fixed by the context and the interpretation highly regimented. Someone who doesn’t interpret the large sphere as the sun simply doesn’t understand the Newtonian model [of the solar system]’ (2017, 57). Thus the interpretations of models are ‘regimented and controlled’ (ibid., 58).

[87] Frigg and Nguyen draw on Goodman (1976) and Elgin’s (2010) notion of ‘representation-as’ to outline their account. They focus on the example of the Phillips-Newlyn machine, a hydraulic model used to represent the Guatemalan economy as a Keynesian economy. As I do not take a position on representation in this thesis, I discuss their view only insofar as they make the claim regarding the differences between interpretation of artistic and scientific representations. I return to this in the next chapter.

Frigg and Nguyen grant that the flexibility of interpretation varies depending upon the work in question, and that there can also be variation due to the genre of a work. But they highlight how in works of literature (and artworks more generally), the interpretation is not fixed in this way, and attending carefully to the work and its features in order to come up with interesting and even conflicting interpretations is part of engaging with and appreciating artworks. This point can be illustrated through returning to the case of Morrison's *Beloved.* While there are obvious topics being explored in the novel, there is also a variety of ways we can view the work. We could highlight the way in which the novel explores the dangers of giving meaning to pain and suffering, or we could focus on how it brings to the fore the complexities in mother-daughter relationships. Further, scholars have studied the work via Black feminist theory, as well as through a Marxist lens, for example.

A final disanology of Egan's that I will present here is to do with the way in which thought experiments and literary fictions are criticised. In the former, the criticism takes the form of how well the case at hand contributes to the argument it is taken to support. In the latter, the criticism is of an aesthetic nature. For example, in response to Thomson's violinist thought experiment, philosophers might point out relevant ways in which the fictional scenario departs from cases of unwanted pregnancies that undermines the use of the example as a way of defending a persons' right to abortion. Egan argues that Thomson's thought experiment is not criticised aesthetically; no one responds to the thought experiment by claiming 'that the character of the hospital director lacks clear motivation or is cliched or two-dimensional. No one is inclined to spill much ink over praising or criticizing the elegance or compactness or precision of Thomson's prose' (2016, 147). I take Egan as stating here that at least if they were to criticise it in such a way, they would be displaying a misunderstanding of what thought experiments set out to do and thus, how we should assess their value.[88]

In this section, I have outlined arguments that argue against the usefulness of drawing analogies between literary fiction and thought experiments and which dismiss the import of aesthetic qualities as well as the use of concrete particulars in the case of the latter. As Norton summarises, such features in the case of thought experiments are 'merely rhetorical window dressing that, for psychological reasons, may well ease acceptance of the result. In many cases, this superfluity is easy to see, since the elements visualized can be supplied in many

[88] I will come back to the way in which there is an appropriate way of engaging with thought experiments in the next chapter, and I explore this in connection with the view that thought experiments should be thought of as a distinctive (sub-)genre.

ways that will not affect the outcome' (2004, 60).[89] I will now turn to the aesthetics of thought experiments which demonstrates how, contrary to Egan's final claim, thought experiments can be, and indeed, are, evaluated on aesthetic grounds, and this motivates the account of aesthetics of science that I develop in the next chapter. While I think Egan taps into some key differences between thought experiments and literary fictions, and poses some important worries for those who want to defend the cognitive value of literature via thought experiments, the account that I develop will allow for a way of resisting the full extent of Egan's claims as well as showing the importance of the literary and aesthetic qualities of thought experiments.

5.4 The Beauty of Thought Experiments

As set out in the introduction, accounts of the aesthetics of science has focused primarily on theories. Thought experiments are also often referred to as beautiful or elegant. Take, for example, Galileo's famous falling bodies thought experiment used to undermine Aristotle's physics, which Brown refers to as 'the most beautiful thought experiment ever devised' because it is 'brilliantly original and as simple as it is profound' (2004, 24). Brown also cites the results of a poll conducted by *Physics World* of the ten most beautiful experiments of all time. Galileo's thought experiment is second on the list (the double slit experiment is number one) (Crease 2002).

Galileo's thought experiment is considered beautiful, and there is evidence that other thought experiments are also considered to have aesthetic value. In 2012, edge.org conducted a survey which asked 192 people, including scientists and philosophers, what their "favourite deep, elegant or beautiful explanation" is. As Stuart points out, 21 of the answers given were thought experiments, and a further 8 were 'imagination-based inferences that any broad-minded characterization of thought experiments should include'. This means nearly 1/6 of all replies provided a thought experiment as their answer (2018, 530). Physicist Sean Carroll's favourite deep, elegant or beautiful explanation, for instance, is Einstein's thought experiment used as part of his explanation of why gravity is universal, what Einstein called the "happiest thought" of his life. While Stuart is interested in how this supports the claim that thought

[89] Egan allows that the particulars can be ineliminable in the sense that without them, we may not have been able to reason about the more abstract features that we are ultimately interested in. So it appears that for Egan at least, this is not merely psychological in the dismissive sense that Norton advocates.

experiments provide good explanations and can contribute to scientific understanding, I am interested in the widespread view that thought experiments have aesthetic value.

It is clear, then, that thought experiments are often taken to have aesthetic value. But what is the basis for taking them to have such value, or more specifically, to be regarded as beautiful or elegant? To consider this question, we can turn to Sibley's influential paper on aesthetic concepts. Sibley points out that when we describe something using aesthetic terms such as "unified", "serene", "dynamic", "vivid", "balanced", "graceful", or "elegant" (to take just some of his examples of aesthetic concepts), we often point to non-aesthetic features to explain our application of an aesthetic term. Sibley offers the following examples, "delicate because of its pastel shades and curving lines", or "it lacks balance because one of one group of figures is so far off to the left and is so brightly illuminated" (1959, 424).[90]

We can think about this in the science case as well, and identify the non-aesthetic features of theories, models, thought experiments and so on, that help explain our application of aesthetic terms. Firstly, what are these features in the case of theories? This is not always clearly set out, and often beauty is described by referring to other aesthetic terms such as simplicity, symmetry or harmony. But to take one example, Poincaré reduces beauty to simplicity and unity. A theory is simple because of the 'number of hypotheses and axioms of the theory. Syntactic elegance or simplicity can be understood as the lack of complexity, adhocness, or free parameters in a theory' (Ivanova 2017b, 2585). The unity of theories is a matter of finding "hidden relations" between phenomena that appears disconnected (2017b, 2588). And so when we consider this association of aesthetic qualities with other non-aesthetic epistemic virtues, we can see that a clear feature that is often considered part of the beauty of theories is economy; the theory postulates a small number of hypotheses and axioms which provides many successful predictions, or can explain a wide range of phenomena.

[90] It is important to note that Sibley argues that aesthetic terms cannot be defined in terms of non-aesthetic concepts. And further, he claims that aesthetic concepts such as graceful are not condition-governed. That is to say, there are no set of (non-aesthetic) features that always count towards something being, say, graceful. He contrasts this with, for example, the fact that being a good chess player will always count towards someone being intelligent (1959, 425). Todd suggests that whether or not Sibley is correct, it seems that the conditions under which an aesthetic term is employed are clearer in the case of science and mathematics than in art (2008, 71) and as we shall see, this forms part of his argument against accounts of aesthetics in science. Further, the aesthetic terminology utilised in science cases as discussed by philosophers is far more limited than in cases of art (ibid.). There is a lack of a comparison of aesthetic terminology in art and science by philosophers. However, O'Loughlin and McCallum (2019) discuss how a broader array of aesthetic terms are utilised in the scientific context that have gone ignored which, unlike the terms often discussed, are more similar to those employed in the context of art.

While most accounts of the aesthetics of science have focused on theories, there has been some consideration of the aesthetics of experiments. This is perhaps a more useful comparison to thought experiments than theories are, as although the question of whether thought experiments classify as genuine experiments is disputed (as discussed in chapters 1 and 2), they share some important features with ordinary experiments. Unlike theories, both thought experiments and experiments involve (real or imagined) particulars, and there is an initial set up of the experiment, or description of the scenario, which is then manipulated to see or consider what would happen. The main difference, as already noted, is that unlike experiment, thought experiments take place in the imagination or what Brown (2011) calls “the laboratory of the mind” rather than intervening on the world.

Parsons and Rueger have discussed scientists’ aesthetic responses to certain experiments. Experiments may be considered beautiful because they produce phenomena that is pleasing to experience, for example, Canton’s electric *aurora borealis* experiment (2000, 408). Some thought experiments can also be seen as beautiful in this sense, such as Einstein’s thought experiment that gets us to imagine what it would be like to chase a beam of light. It may be fairly usual for people to get some kind of aesthetic pleasure from scientific (or philosophical) thought experiments in that they get us to imagine a phenomena or a set of circumstances that is aesthetically interesting. And this may be part of their value in making say certain philosophical ideas engaging.[91] But Parson and Rueger claim that such aesthetic evaluations are not relevant to the experiment’s success.

Parsons and Rueger note that the prevailing eighteenth-century view was that an experiment was beautiful when it ‘made visible particular aspects of the beauty of nature itself’ (ibid., 409) such as in the *aurora borealis* experiment. Since the nineteenth-century, however, another way of thinking about the aesthetics of experiments has emerged. Parsons and Rueger show this through the example of Rutherford and the artificial disintegration of atomic nuclei, described by Peter Kapitsa (1968) as a ‘most simple experiment’, that led to ‘striking results’. Here, an understanding of what is being tested becomes central to aesthetic appreciation. This contrasts with admirations of the workings of nature, such as in the *aurora borealis* case in which an aesthetic appreciation can be irrespective of whether or not we have a grasp of the theoretical framework involved. Parsons and Rueger note that since this shift, a common way

[91] I’ll come back to the various ways in which we can appreciate a scientific thought experiment in the next chapter.

of characterising aesthetically valuable experiments is to say they involve "an optimal use of minimal material":

> 'An experiment now is aesthetically valuable because it shows 'aptness' in relation of result and tools, of plan and success; it is a beautiful artefact, a manifestation of human ingenuity, an instrument optimally suited to achieve its purpose. What is appreciated is, for instance, the simplicity of the arrangement, its economy, or its ability to unify several tasks in one display' (ibid., 411-412)

We have already seen this idea of beauty because of "optimal use of minimal material" in the discussion of theories. It is also present in the aesthetic judgement of thought experiments in the cases above. For Brown, as we saw, Galileo's thought experiment is beautiful because it is 'brilliantly original and as simple as it is profound' (2004, 24). Similarly, Carroll states 'Einstein, in his genius, realized the profound implication' of the situation described in the thought experiment. In the experiment case, the material is concrete objects.[92] Thought experiments, of course, differ here; economy is achieved through the particulars that we are prescribed to imagine.[93]

To further illuminate these features in the case of thought experiments, it might also help to contrast these with cases of thought experiments that could be regarded as cluttered or clumsy and so on. Norton describes Szilard's version of Maxwell's demon as "the worst thought experiment". Why is this? Norton offers a detailed account of the thought experiment and its flaws, but we can focus on a couple of reasons he provides. Thought experiments are illuminating when they provide us with a simple scenario that allows us to focus on the most essential features, and when the scenario can serve as a representative case. Szilard's thought experiment involves a misuse of idealisations; 'an inconsistent muddle of improper idealizations', and leads to an incorrect generalisation (2018, 466).[94] Another example is

[92] There are discussions of profundity as an aesthetic quality such as in the context of music (Dodd 2014). Dodd explains that a necessary condition for profundity is that 'the work has a profound subject matter, and that the work handles its profound subject matter in such a way to elicit a deeper understanding of it (or a fuller grasp of its significance) in the suitably situated and prepared appreciator'. This is apt for thought experiments and as will see, my proposal in the next chapter accords well with this sense in which a work "handles" its profound subject matter in a particular way.

[93] In his account of the beauty of theories, McAllister (1996) argues that in theory change our aesthetic canons change and consequently what counts as beautiful will be revised. Bringing thought experiments into aesthetics of science prompts questions regarding whether their aesthetic qualities have also changed from era to era, throughout the history of science.

[94] As outlined, it appears that Norton would maintain that any aesthetic qualities of thought experiments are irrelevant to their epistemology, or that the aesthetic qualities can be reduced to claims regarding

Darwin’s whale thought experiment, which attempts to explain natural selection by demonstrating how whales could have evolved from bears:

> ‘In North America the black bear was seen by [the explorer] Hearne swimming for hours with widely open mouth, thus catching, like a whale, insects in the water. Even in so extreme a case as this, if the supply of insects were constant, and if better adapted competitors did not already exist in the country, I can see no difficulty in a race of bears being rendered, by natural selection, more and more aquatic in their structure and habits, with larger and larger mouths, till a creature was produced as monstrous as a whale’ (1859a, 184).

Picking up on the use of ‘monstrous’ in Darwin’s description, Louis Agassiz described the thought experiment itself as “truly monstrous” (Stuart 2016, 31). The example was dropped by Darwin in later editions of *On the Origin of Species.* In attempting to explain the morphology of whales by referring to an existing creature, the example invokes ‘needlessly strange evolutionary explanations’ and fails to aid our understanding of natural selection (ibid.).

5.5 Problems for Aesthetics of Science

We have seen both positive and negative evaluations of thought experiments, and thus, I argue they ought to be included in philosophers’ discussions of aesthetic evaluations in science. However, this discussion brings us to a set of worries for any aesthetics of science project.

It could be argued that when scientists or philosophers of science describe or evaluate certain thought experiments aesthetically, what is really being said is that the thought experiment is successful or unsuccessful. Take Norton’s view that Szilard’s thought experiment is the worst thought experiment. It is clear that given his argument account of thought experiments, Norton would maintain that any aesthetic evaluations of this or any other thought experiment are irrelevant to their epistemic value, or that the aesthetic qualities can be reduced to claims regarding their role in contributing to an argument and consequently, they are not genuinely aesthetic. In Norton’s terms, it is “merely rhetorical window dressing” (2004, 60). Szilard’s

their role in contributing to an argument and consequently, they are not genuinely aesthetic. I discuss this below.

thought experiment, then, is just a bad argument on this reasoning.[95] As I have already outlined, Norton has many opponents and I will not go into the worries with his view here. But we do not need to commit to an argument view of thought experiments to see the worry that when say, Galileo's thought experiment is being described as "beautiful", or Darwin's as "monstrous", the aesthetic language is being used in a mere metaphorical way.

This issue has been raised by Todd (2008) in the context of theories. Todd discusses a number of views which claim that the beauty of a theory can be an indicator of the truth of a theory. To flesh out Todd's argument, we can focus on his discussion of McAllister (1996) which he takes to be one the most developed accounts of the role of aesthetics in the assessment of theories. McAllister sets out to defend the view that we can hold that aesthetics has value in science, whilst maintaining a rationalist image of science, that is, one in which preference for scientific theories is dictated by their logical consistency and empirical adequacy. McAllister explains the connection between the beauty of a theory, and its truth or empirical success, with reference to the aesthetic canon. On this view, scientist's aesthetic preferences have been shaped over time to match the features of successful theories. And so, the connection is based on aesthetic induction; when a theory is regarded as beautiful this is because it is similar to existing, successful theories (i.e. it accords with the aesthetic canon) and thus it is more likely to be true or empirically successful (ibid., 33-34).[96]

McAllister reduces beauty to features such as simplicity, symmetry, elegance, harmony and visualisable structures. Because of this reductionist approach, Todd is sceptical that "beauty" is really being used in an aesthetic sense when applied to scientific theories. He motivates this by highlighting how some of these terms are used in theory assessment without a link made to beauty or aesthetics. And so, he calls into the question whether scientists are expressing something about the aesthetic value of theories rather than saying something about the kind of intellectual pleasures that might arise from the construction or use of simple or symmetrical theories, for example. Because of this, Todd argues, we should understand such terms as being

[95] Or on Egan's view, the thought experiment does not fulfil its intended role in a larger, argumentative structure.

[96] As Ivanova (2017a) points out, McAllister utilises a phenomenon which has been labelled the 'mere exposure effect' in which people develop a preference or a liking for something based on their increasing familiarity with it. For example, a psychology study by Cutting (2003) demonstrated that the mere exposure to certain Impressionist paintings led to an increase in the participants' positive response to the paintings. The limits of Cutting's research has been discussed by Meskin et. al. (2013) who carried out an alternative study that demonstrated that exposure to what is taken to be bad visual art, Kinkade's paintings, did not lead to an increase in the participants' liking of the works. Rather, exposure decreased liking.

used in a way that actually tracks epistemic features. He claims that ‘there are strong grounds for suspecting that what appears to be aesthetic claims may often be, if perhaps not always are, really masked ‘epistemic’ functional ones’ (2008, 72).[97]

Todd further claims that the part of the challenge for those who want to defend the genuine aesthetic nature of the use of beauty and so on in science is to provide a theory of ‘aesthetic value, appreciation, or properties’ that will demonstrate ‘how theories and proofs might fit the general contours of more paradigmatic examples of objects of aesthetic appreciation, such as artworks and natural objects’ (ibid., 63). For instance, Todd argues that the use of ‘simplicity’ in the case of describing a theory or an idea is a different use than when describing a Mondrian painting, thus ‘it is the context of use that determines whether they are being used to signify aesthetic interest or value, or not, and this McAllister and others generally fail to notice’ (2008, 70). Again it can be highlighted how there might be some dispute regarding the strength of the purported connections between thought experiments and literary fictions and the way in which we critically evaluate them as seen in section 3.

On the other hand, it can be emphasised that if aesthetic judgements of “beauty”, “elegance” and so on are genuinely aesthetic (or are perhaps “merely” aesthetic) and cannot be reduced to epistemic features or do not have any kind of epistemic role, then it is difficult to see how they have an important part to play in the scientific context and consequently, why they are interesting to philosophers of science.

To clarify, the dilemma is as follows:

a) The application of aesthetic terms to science (such as to theories, experiments, and thought experiments) is merely metaphorical, and should instead be regarded as a proxy for epistemic features or

b) The application of aesthetic terms to science (theories, experiments, and thought experiments) is literal, but their aesthetic features are scientifically irrelevant

Therefore, any account of aesthetics in science will have to provide reasons as to why we should take these descriptions as genuinely aesthetic whilst maintaining that they play a

[97] As Todd recognises, delineating what counts as “genuinely aesthetic” and what falls outside of that domain is not a straightforward issue. And as already mentioned, there is great debate regarding whether or not the cognitive value of artworks is part of their aesthetic value.

meaningful role in science.[98] I want to avoid a reductionist view, and I think we will ignore important aspects of scientific practice if we reconstruct aesthetic language in such a way. And so, we should take seriously the aesthetic claims of scientists and try to make sense of them. In the next chapter, I propose a new account of the aesthetics of science which meets Todd's criteria.

5.6 Conclusion

Discussions of aesthetic values in science have focused mainly on theories and mathematical proofs. My aim in this chapter was to demonstrate why thought experiments should also be included in the literature on aesthetics in the scientific realm. The first reason is that thought experiments are often compared with artworks. In particular, thought experiments are often compared with works of literary fiction given that each invite us to imagine some hypothetical state of affairs and have a narrative form. The second reason is that thought experiments are also described using aesthetic language as seen in Brown's description of Galileo's falling bodies as well as Agassiz's response to Darwin's "monstrous" whale, for example. I outlined existing views on the connections between thought experiments and literary fiction, before turning to the aesthetic evaluation of thought experiments which I compared to theories and experiments.

However, this chapter also raised ways in which connections between thought experiments and works of literary fiction can be undermined and I ended with a dilemma for aesthetics of science projects. In the next chapter, I respond to these worries and develop a new approach to aesthetic values in science which solves the dilemma.

[98] For recent responses to Todd's claims, see Dutilh Novaes (2019) on beauty in mathematical proofs, and O'Loughlin and McCallum (2019) on aesthetics in theory evaluation.

6. The Aesthetic and Literary Qualities of Thought Experiments

6.1 Introduction

In chapter 5, we saw that present discussions of aesthetics in science have overlooked thought experiments. I outlined how this is surprising, given the ways in which connections between thought experiments and artistic fictions, namely works of literary fiction, have been drawn. In particular, those involved in debates regarding the cognitive value of literature have highlighted the similarities between the two in order to defend the view that at least some literary works function like (or are) thought experiments and therefore we can learn from them. In addition, thought experiments are also considered beautiful or elegant, and I discussed the aesthetic evaluation of thought experiments by calling on the attribution of beauty to theories and experiments.

However, the chapter also included a discussion of the ways in which thought experiments are importantly different from works of literary fiction which, it has been argued, undermines the view that the use of concrete particulars and narrative play a similar role in each. And I ended with a dilemma for any aesthetics of science project: On one hand, it appears that when scientists are utilising aesthetic language, they are meaning it in a merely metaphorical sense. That is to say, what is really being picked out are epistemic features of thought experiments (or experiments or theories). On the other hand, if the language is genuinely aesthetic, then it appears difficult to defend the usefulness of aesthetic qualities in the scientific domain. A challenge for an aesthetics of science project, then, is to defend the view that the aesthetic language is being used literally, whilst demonstrating how such evaluations play an important role in science. As well as to show how the terminology, such as “simplicity” or “elegance” and so on, is being used in a way that is comparable with the use of such terms in art and other more obviously aesthetic domains.

I agree that there are significant differences between our engagement with art on one hand, and with scientific theories/experiments/thought experiments on the other, and that these differences need to be taken into consideration when drawing comparisons between scientific and artistic representations and how we learn from them. But I want to resist the force of the arguments against a meaningful connection between thought experiments and literary fictions, and our aesthetic evaluation of them.[99] I begin with a proposal for a way of characterising an aspect of the aesthetics of science that is currently neglected in the literature. This is by attending to the relation between form and content in scientific thought experiments. On this view, aesthetic value has to do with the way in which epistemic content is expressed.[100] I then come back to the comparisons between thought experiments and works of literature and end with some further issues that my discussion raises.

6.2 Form and Content in Aesthetics

A focus on the fit between form and content is common in accounts of the aesthetic appreciation of artworks. For example, in *The Pleasures of Aesthetics* (1996) Levinson states that when we attend to an artwork or an aspect of it, the pleasure we derive from it is aesthetic when 'there is also attention to the *relation* between content and form—between what a work represents or expresses or suggests, and the means it uses to do so' (1996, 10).[101]

Take, for example, Picasso's *Guernica* (1937) created in response to the bombing of Guernica, a village in the Basque region of Spain, during the civil war. The painting explores the horror of war and is composed of injured children, women and animals, and it expresses this in a distinctive way—the distorted lines and fragmented composition, the way in which the animals and humans are positioned as jumbled together, such as a bull's head directly above a woman's screaming face, and the colours of the work being restricted to black, white, and grey which adds to the starkness of the work, allowing us to focus on the structure of what is depicted. And thinking again of Morrison's *Beloved,* when we consider what the work conveys and its cognitive and political import, we can think about how that is bound up with

[99] I take the account that I develop here to be adaptable to other areas of science, but given my focus in this thesis and the discussions of literary works presented in the previous chapter, here I am predominately concerned with thought experiments.

[100] At the end of his paper, Todd (2008) suggests (but does not flesh out) a view that focuses on the relation between form and content as a possible direction for aesthetics of science given his worries as presented at the end of chapter 5.

[101] Levinson grants that we could take pleasure in the formal features of a work (particular arrangement of lines, brushwork, colour and so on in painting, or the use of alliteration and rhythm in poetry) but distinguishes this from *aesthetic pleasure.* He states: 'It is clear that aesthetic pleasure as I have characterized it in this essay comprises more than pleasure in aesthetic qualities per se—that is, those that Frank Sibley has famously identified—and equally more than pleasure in mere appearances' (1996, 9). Similarly, while we can take pleasure in the content of a work—say considering a *Madonna and Child* beautiful because it depicts motherly love—this would not constitute *aesthetic pleasure* for Levinson.

how it is expressed; the particular events described regarding the lives of Sethe and her family, the use of rich imagery, the style of Morrison's writing which incorporates elements of African-American folklore, and in the structure of the work such as the presentation of both the past and present which conveys how the two are intertwined in the characters' inner lives. Through the example of literature, Levinson explains that aesthetic satisfaction in artworks is 'precisely when such symbolic or moral content is apprehended in and through the body of the literary work itself—its sentences, paragraphs and fictive events—and not as something abstractable from them… Aesthetic appreciation of art thus always acknowledges the vehicle of the work as essential, never focusing only on detachable meanings or effects' (Levinson, 1996, 7).

Similarly, Carroll characterises the form of an artwork as 'whatever functions to advance or to realize whatever the artwork is designed to bring about. The form of an artwork is what enables the artwork to realize its point or its purpose' (1999, 142).[102] Further, this is part of our aesthetic appreciation of artworks: 'What we appreciate in an artwork is how the forms function as means to bring about the ends of the artwork. Where these forms are well suited to the ends of the artwork, we generally take satisfaction in their design' (ibid., 150). Carroll's view thus entails that artworks have a purpose or a point to make, but he emphasises that this should be understood in a broad sense. A purpose of an artwork could be to arouse certain feelings or responses in their audience, to advance a particular point or view or communicate an idea, or to explore a theme and so on. Additionally, a particular work of art can have more than one purpose or more than one point to make. What is especially useful about Carroll's conceptualisation for the consideration of scientific thought experiments is that he makes explicit that our analysis of artworks depends on having a conception of the point(s) or purpose(s) of the work. In the case of thought experiments, then, our aesthetic analysis conceived in this way is dependent upon what the example is used for. Carroll highlights that in the case of artworks, there will be variation with regards to how easy it is to determine the point(s) or purpose(s) of a work. He notes that this is why 'formal analysis also usually comes hand-in-hand with interpretations or explications of the work' (1999, 145).[103]

[102] See also Thomson-Jones (2005) and Eldridge (1985) for accounts of the interrelation between form and content in art. Carroll is also interested in artworks that do not have content such as much orchestral music or "pure dance". Given that my focus is on thought experiments which are clearly about something (that is, they have content) I need not go into these details here (see Carroll 1999, chapter 3 part II for discussion).

[103] I discuss interpretation of artworks (and thought experiments) below.

While the link to the current aesthetics of science literature has not been made, there are some existing discussions of the significance of the ways in which scientific representations are formulated. Vorms (2011) and Frigg and Nguyen (2017) emphasise that "formulation matters" in scientific models; varying the formulation of the same model can provide different explanations and predictions (Frigg and Nguyen 2017, 58). And Frigg and Nguyen link this claim to comparisons between artworks and scientific models. They argue that Currie is mistaken when he uses the import of formulation in artistic contexts to undermine connections drawn between scientific models and artistic fictions. Currie claims that scientific models 'are not dependent for their value in learning on any particular formulation; rather they depend on their capacity to get good predictive or explanatory results or to achieve some other epistemic aim' (2016, 305).[104]

In her discussion of accounts that defend the imaginary status of models (as presented in chapter 3), Vorms argues that such views are right to highlight, against the semantic view, that modelling involves scientists' creative and imaginary skills.[105] However, she argues that they have failed to attend to the ways in which models crucially involve 'the concrete inferences agents perform' when reasoning with models (2011, 288). She sets out to show how in order to capture the "representational power" of models, 'one has to focus on the cognitive interactions between agents and the representational devices they reason with and manipulate' (ibid.). Vorms demonstrates that the way in which a model is formulated is crucial to these interactions.

One case that Vorms discusses to motivate the view concerns the representation of the results of a temperature survey. These results are represented in different ways; as a list of numerals and on a map. While it is the same information in each representation (thus, it is the same content that is being conveyed), and agents can draw conclusions from any of them, 'the map

[104] Currie makes this point in the context of his scepticism regarding the epistemic value of scientific and artistic representations. Currie is sceptical of claims regarding literature's value in learning. In contrast to the clear roles of models in the growth of scientific knowledge and understanding, there is 'no more than the vague suggestion that fictions sometimes shed light on aspects of human thought, feeling, decision, and action; a proposition that no one has found a way to test' (2016, 307). Currie further states that the "feelings of truth" that we gain from artworks is dependent upon the way they are formulated. For example, alliteration might contribute to a feeling of truth. Which he takes to undercut their epistemic value; the persuasiveness of their content is dependent upon their style and so on (see Thomson-Jones 2005 for a criticism of this view).

[105] While I draw on Vorms, I am not committed to the fiction view of models as emphasised in chapters 3 and 4. Because of this, the view of aesthetic values in science that I develop in this chapter (as with my pluralist view of imagination) could be incorporated into a variety of positions on modelling, including semantic accounts.

makes some information much more easily available: for instance, if warm shades stand for high temperatures and cold shades for low temperatures, one can quickly conclude that the southern part of the represented area is warmer than its northern part' (ibid., 289). Whereas drawing this from a list of numerals that stand for the coordinates of the place and its temperature value, would involve many inferential steps (ibid.).[106] She highlights how different formats of the same information impacts agents' reasoning processes, and can allow access to different information.

Now that we have a sense of the import of form and content in the aesthetic appreciation of artworks, as well as of the existing discussions of formulation in the scientific realm, I will turn my attention to the interrelation between form and content in thought experiments.

6.3 The Formulation of Thought Experiments

In this section, I will demonstrate that the usefulness of a thought experiment in scientific practice is impacted by the way in which its content is conveyed. Following Vorms, I emphasise how formats matter when we attend to the context in which a thought experiment is used, as well as the particular skills and interests of those engaging with the example.

What do I mean by the formulation of thought experiments? In presenting Egan's arguments in the previous chapter, we saw how he emphasises that thought experiments are ultimately concerned with abstract ideas that can be generalised beyond the particular details of the thought experiment narrative. For example, in Galileo's falling bodies thought experiment, Galileo is ultimately interested in exploring the relation between the weight of a body and the speed at which it falls. This I take to be the content of Galileo's thought experiment, or in Carroll's terminology, the point or purpose of the example. More specifically, we can say that this is part of the "thematic content" (in the scientific context, we can label this the "scientific content"). As Lamarque explains in the context of literary fictions, the content at the thematic level is the more general, overarching reflections that go beyond the particularities of the text. Importantly, it is the 'perspective or vision or general reflection that informs the subject matter

[106] An example of an imaginary model that Vorms discusses is the simple pendulum. This can be conveyed in different formats and which format is the most appropriate is dependent upon what information the user wants to obtain. She outlines the differences in the description of the pendulum in Newtonian mechanics compared with the Hamilton formulation. While these formats are mathematically equivalent, Vorms explains that they are inferentially different and hence do not involve the 'same kind of cognitive operation' (2011, 292).

and moves beyond the immediate events portrayed' (2009, 150). Take Egan's example of Austen's *Pride and Prejudice.* At the thematic level, the work is an exploration of gendered expectations, reputation and class. This informs the choices regarding the particularities of the text; the story of Elizabeth Bennett and her romantic relationship with Mr Darcy and so on.[107]

The form of Galileo's thought experiment includes the way in which the relation between speed and weight is expressed through the particulars used in the scenario; the musket ball and cannon ball dropped from a tower. Another sense in which we can focus on the formal features of thought experiments relates to Vorms' discussion above, in the sense that we can think about the style in which thought experiments are presented, including the particular construction of the narrative as well as the use of diagrams and images. As with artworks, I argue that a source of the aesthetic value of thought experiments comes from the way in which their formal features function as a way of bringing about its overarching content. When the form is well-suited to the purpose of the thought experiment, we may regard it as aesthetically valuable. And so, while we could take pleasure in the content of say, Galileo's falling bodies or in its formal features alone, what I am interested in here is how the interrelation of these two aspects of the example is a source of aesthetic value that has epistemic pay offs.

In the last chapter, I discussed one of Darwin's thought experiments (or "imaginary illustrations") used to, as Darwin states, 'make it clear how, as I believe, natural selection acts' (1859a, 90). The example I outlined was a case in which we are asked to imagine a bear catching insects in the water. By arguing that wider mouths would be advantageous, Darwin describes the bear's mouths becoming wider and wider until 'a creature was produced as monstrous as a whale' (ibid., 184). This thought experiment was described as "truly monstrous" by Agassiz, and others agreed. In Darwin's letters, we can see that the example caused controversy. While Darwin intended to show how selective pressure could lead to the

[107] Lamarque (2009) distinguishes two levels of content, or aboutness, of literary works: the thematic content and the narrative or immediate content. On Lamarque's view then, the tower and the balls in Galileo's thought experiment would be a part of the narrative content of the work, as would the events described in *Pride and Prejudice*. Whether or not we want to label this a relation between two levels of content, or a relation between form and content, is an interesting question. However, the important point for my account is that we have this relation between two things—between the thematic or scientific content and how that is expressed in the particularities of the example. I will therefore stick to Levinson and Carroll's ways of drawing the distinction but note that this implies that choices regarding narrative content can count as part of the form of a work. If the point of a thought experiment is to, for example, convince us of something, then the form comprises the choices made to realise that purpose, some of which will include choices at the level of narrative content.

widening of the bears' mouths, some took the thought experiment as intending to convey that the mouths could widen over time due to the bear using it to catch insects (Letter to Murray, 1860). In a review of *On the Origin of Species,* Owen described the example as "gross" but in correspondence with Darwin, noted that the passage stood out to him. As Darwin explained in another letter, Owen had also misunderstood the example; taking it to state that it was attempting to show how a bear could become a whale, or as Darwin puts it, how 'a sort of Bear was the grandpapa of Whales!' (Darwin, 1859b).

There are multiple ways in which we can think about the aesthetics of this example. One way is to consider the imagery that the thought experiment produces; the image of the bear's mouth becoming wider and wider is at best peculiar and perhaps even ugly. And so, the thought experiment could be considered "monstrous" in this sense. This would be similar to regarding Einstein's light beam thought experiment aesthetically valuable because of the imagery of chasing a beam of light. A related way is that due to the bizarreness of the bear-whale imagery, the thought experiment can be considered captivating. This might accord with how the passage was "gross" to the reviewer, yet also stood out. An analogy of this in the case of art are instances of "good-bad" artworks as discussed by Dyck and Johnson (2017). In such cases, the artistic failure of such works makes them bizarre, which, it is argued, is aesthetically valuable. As Walton puts it, we enjoy 'something like awe and amazement at how awful the thing turned out to be *despite* the efforts of its creators' (2008, 21).[108]

But another way, and the way in which I will consider the case, is to focus on the relation between the thought experiments' form and content. To do this, it is helpful to compare it with another of Darwin's "imaginary illustrations". In particular, Darwin's thought experiment that is used to explain how an organ as complex as the eye could have developed through spontaneous mutation and natural selection. In the 1872 edition, he states: 'To suppose that the eye with all its inimitable contrivances for adjusting the focus to different distances, for admitting different amounts of light, and for the correction of spherical and chromatic aberration, could have been formed by natural selection, seems, I freely confess, absurd in the highest degree' (1872., 143). The thought experiment is outlined as follows:

> '[W]e ought in imagination to take a thick layer of transparent tissue, with spaces filled with fluid, and with a nerve sensitive to light beneath, and then suppose every part of

[108] Examples of "good-bad" art as discussed by Dyck and Johnson include the 2003 film *The Room* which is 'a confusing mix of a bizarre storyline, terrible acting, very little plot cohesion, and a script that consists almost exclusively of clichés. But this mix makes for an enjoyable film' (2017, 279).

> this layer to be continually changing slowly in density, so as to separate into layers of different densities and thicknesses, placed at different distances from each other, and with the surfaces of each layer slowly changing in form. Further we must suppose that there is a power, represented by natural selection or the survival of the fittest, always intently watching each slight alteration in the transparent layers; and carefully preserving each which, under varied circumstances, in any way or in any degree, tends to produce a distincter image. We must suppose each new state of the instrument to be multiplied by the million; each to be preserved until a better one is produced, and then the old ones to be all destroyed. In living bodies, variation will cause the slight alterations, generation will multiply them almost infinitely, and natural selection will pick out with unerring skill each improvement' (ibid., 146).

As Stuart highlights, the two thought experiments are functionally similar. In each, Darwin is attempting to explain how a complex part of nature—the intricacies of the eye or the morphology of whales—could have come about through a series of steps (2016a, 31). This I take to be the "thematic" or "scientific" content of Darwin's two examples, that is, its point or purpose. While they involve different applications, each is used to demonstrate the explanatory power of the theory of evolution by natural selection. Throughout the book, Darwin invokes such illustrations to show how the theory can account for a variety of natural phenomena.

In chapters 3 and 4, I discussed how thought experiments work by calling upon our imaginative faculties, and I offered a pluralist account of the nature of the imagination in their conduct. Part of this account included the role of thought experiments in scientific understanding. I called upon the work of De Regt (2017) to discuss how thought experiments can contribute to making certain theoretical content intelligible, as outlined via the use of Maxwell's demon to enhance our understanding of the second law of thermodynamics. Given that thought experiments are performed in the imagination, we can use them most effectively when they are formulated in an appropriate way; the particular details of the thought experiment narrative give our imaginations something to latch onto. Thought experiments thus work by presenting more generalisable and abstractable problems (that is, their scientific content) in more vivid terms. When they are formulated well, thought experiments can help us focus on the structurally relevant features of the case.

Another key aspect of De Regt's account of understanding was that he focuses on features of theories that enhance its usability. This is also relevant to the examples considered here: While Darwin is interested in explaining the complexity of the eye, and the morphology of whales, his imaginary illustrations are also intended to be generalizable beyond the particular details of the case. That is to say, they are used in order to illuminate the theory in such a way that it allows those who engage with the thought experiment to see how natural selection can also explain other complex organs and species. With this in mind, we can explore how the formulation of Darwin's examples impacts their effectiveness at increasing understanding of natural selection.

Darwin's whale thought experiment can be easily misinterpreted. In explaining how whales could have evolved, Darwin chooses to invoke an existing animal. This makes the example more convoluted than is necessary, and makes it difficult to focus in on the key features of the example—the step-by-step changes of a species that results in it becoming something else—and instead can easily lead the reader astray, such as understanding the bear as integral to the example and thus, taking it to suggest that whales have evolved from bears. This is a stark contrast to the eye illustration. In this example, we are given a solution to the problem of the "absurdity" of the prospect of natural selection accounting for complex and intricate phenomena through the description of the eye. The example is easy to imagine; we can see how by various steps and modifications, we end up with an increasingly complex organ which at each stage, becomes more useful. Further to this, the description offered gives the reader a way of going beyond the particulars of the case of the eye; engaging with the thought experiment enhances our ability to apply the process of step-by-step changes to the evolution of other complex phenomena. Finally, along with the described scenario above, Darwin utilises a helpful comparison with a telescope 'which could be built up in stages from a single lens, where each stage involves a small improvement on the last' (Stuart 2016a, 29).[109]

De Regt of course offers just one account of understanding, but we can see an emphasis on usability in other accounts. Take for example Levy's discussion of the role of metaphors in scientific understanding. A good metaphor, Levy argues, should represent the relevant facts in a usable way. He discusses how metaphors "frame" their targets, that is, they direct attention in a way that is 'striking and illuminating' to particular properties of their target

[109] These comparisons may be particularly helpful in cases such as Darwin's, since his thought experiments were used as an explanatory device of a radical new account, rather than serving a heuristic purpose.

subject via a more familiar subject matter (2020, 293; see also Camp 2009). Similarly, in thought experiments, the particular fictive events described—often familiar objects such as balls and towers in Galileo's example, or buckets in Newton's example—are utilised in order to realise the more general, abstract content.

Stuart takes the whale thought experiment as a failed thought experiment, arguing that it does not increase understanding. I agree that the thought experiment has problems and can easily lead to confusion (as it clearly did at the time) regarding what Darwin was trying to explain. While Darwin removed the thought experiment from later editions due to the ridicule he received, he stood by its potential explanatory power. I think it is possible that the thought experiment could aid our understanding of natural selection, but it is far less apt to do so than the eye example. Hence, this affects its usefulness as a thought experiment for Darwin's purposes.

Furthermore, an appropriate fit between form and content may be more or less significant depending on who the example is designed for. In discussing the imagination in chapter 4, I highlighted that certain contextual features of a thought experiment will impact the nature of the imagination that it invites. Similarly, as noted before, Vorms's account of formulations in science emphasises that we ought to attend to the relations between the representation and the cognitive interactions of those utilising the example. We have already seen that both Galileo and Maxwell were communicating with not only a scientific community, but also a public one. Similarly, in the *Origin of Species,* Darwin was appealing to a broad audience. The particular elements of how the scientific content is expressed are thus chosen to be suited to the audience of the text in order to contribute to its cognitive force. For example, while someone who may be familiar with the theory of natural selection could navigate Darwin's whale example, someone less familiar with the theory is more likely to get confused by the case. Again, returning to chapter 4, we can think about how the use of diagrams and so on can form part of the presentation of a thought experiment which acts as aids to our imaginations. The value of such images may vary depending on the level of expertise of the user.

A final consideration is that many of the examples that I have discussed are not only used to communicate certain scientific content, but to persuade. In this sense, I take it that Norton focuses on the wrong way in which thought experiments are closely related with arguments; thought experiments are utilised in an argumentative context in scientific practice. As seen in chapter 2, Gendler argued against the view that thought experiments can be reconstructed into arguments without losing anything of epistemic worth by showing how the argument form of Galileo’s falling bodies lacks the demonstrative force of its original, narrative form. And we can also think again about Einstein’s clock-in-the-box thought experiment as presented in chapter 1. In this case, we see that the formulation of the example (including the addition and modification of diagrams presented alongside the description of the scenario) was revised at the hands of Einstein and Bohr, in order to become a better, more refined thought experiment and thus more appropriately formulated for their purposes.

And so, Egan downplays the role of the particular ways in which scientific content is expressed when comparing scientific thought experiments with literary fictions. Similarly, Norton is too quick to regard the way in which the overarching point of a thought experiment is expressed is merely a use of “irrelevant particulars”. Egan could still maintain that although the role of concrete elements in thought experiments are indispensable in some sense, this does not fully address his concern. As outlined, Egan claims that in the case of literature, ‘the concrete elements of the narrative remain irreducibly a part of our imaginative engagement’ (2016, 144). The worry was that treating literature as thought experiment suggests that, for example, Tolstoy’s *The Death of Ivan Ilyich* ‘exhausts its purpose once it has made a particular distinction salient, and we could just as well used some other thought experiment provided it made the same distinction equally salient’. As a work of literature, the argument goes, it is not exhausted or replaceable in this way (2016, 143). It seems true that there is a particular experience of reading say a work of Tolstoy that cannot be had another way, and that the details of (scientific) thought experiments could be changed to some degree without impacting its force, but there are a couple of worries here. Firstly, it is difficult to identify exactly what changes in the formal features of a thought experiment are permissible without altering the thought experiment’s effectiveness and hence, its usefulness in scientific practice. Further, it appears that at least some of the details of a work of literature could be changed without altering the novel or affecting the interpretations we draw from the novel. This might depend on what works of literature we are focusing on. I come back to the issue of what examples from art are relevant in comparisons with thought experiments in section 5.

6.4 Interpretation of Scientific and Artistic Representations

The final issue that was raised by Egan in the previous chapter that was connected to the formulation of literary works, and echoed by Frigg and Nguyen, was to do with the flexibility of interpretation in artistic and scientific representations. Frigg and Nguyen argue that interpretation of artworks is far less constrained than in the case of science. To demonstrate how interpretation in the case of scientific representations is highly regimented, we saw that Frigg and Nguyen provide the example of the Newtonian model of the solar system. They argue that if we did not interpret the largest sphere as the sun, then we would have misunderstood the model. I agree with Frigg and Nguyen about this case in that such an understanding of the model would constitute a misunderstanding. However, this does not allow for a sharp contrast with artistic representations. This is because there are clear parallels in the case of artworks. For example, if I do not take say the *Mona Lisa* to be a representation of a woman, then I clearly have misunderstood the painting. Or if I do not take *Beloved* to be about a formerly enslaved woman and her family, then I have misunderstood the novel.[110]

In light of this, it is helpful to consider a distinction that is widely appealed to in the philosophy of art between a *description* of a work and an *interpretation* of a work. Matthews (1977) motivates the distinction through the example of three critics discussing James's *The Turn of the Screw.* One critic was asked to describe the work, and offered the following: the story 'is told by a governess, who lets us know how she saw two children under her charge, a little boy and a little girl...corrupted by the ghosts of two evil servants' (Matthews 1977, 6). The other two critics agree with the description in so far as the novella is about two children who are under the care of a governess, but they object to the inclusion of the ghost. They argue that the ghost is not part of the description of the work, but rather, is part of an interpretation of the work. This is because, Matthews explains, such an understanding of the novella goes beyond the features of the text itself. And this means that there can be (and indeed is) wide disagreement regarding the ghost in the story. For example, about whether it is actually seen, or if it is a figment of the governess's imagination (ibid.).

For Matthews, the distinction is rooted in the fact that descriptions can be known to be correct, whereas interpretations cannot. He states how a critic would not be able to know whether the children were corrupted by ghosts because there are no facts about the novella that would decide this. He explains that whatever 'the facts about James's story happen to be, they do not include either the children's being corrupted by ghosts or their not being corrupted by them. There is no fact of the matter here; hence, nothing that one could be in a position to know'

[110] Frigg and Nguyen make this point within the context of their view on representations in art and science. Here, as mentioned in the last chapter, I am just concerned with the way in which interpretation can be used to distinguish scientific thought experiments and artistic representations.

(ibid., 8). Matthew's claim that an interpretation of a work cannot be known to be correct is contentious and so, we might be wary of drawing the interpretation/description distinction in this way. In light of this, we can weaken the claim. For instance, we might take interpretations of work to be about the "non-obvious" (rather than essentially unknown) aspects of a work. As Lamarque puts it: 'Interpretation is called for whenever there is need to "make sense" of something that is initially puzzling or not open to any obvious construal' (2009, 148).

There is more than one way in which the difference can be pinned down, but in general, there is agreement that there is a distinction to be made between descriptions and interpretations.[111] Drawing on this, we can say that "the portrait is of a woman" is part of the description of the *Mona Lisa* or "the novel is about a formerly enslaved woman and her family" is part of the description of *Beloved,* and so on. And this can be extended to representations more generally, including scientific models and thought experiments, in order to see that "the largest sphere is the sun" is part of the description of the model of the solar system, rather than an interpretation of it.

It is also important to emphasise that while artworks such as Morrison's *Beloved* initiate conversations about the meaning of the work and has led to many different critical approaches to its content, there are various accounts proposed in the philosophy of art regarding the proper interpretation of artworks, and what features of a work and its history are relevant to its interpretation. This debate includes whether there can be several correct or acceptable interpretations of a work (pluralism), or if there is a single correct or acceptable interpretation (monism). And so, even if it is allowed that there can be multiple, inconsistent interpretations of a work, it is not the case that "anything goes" when interpreting works of art (Stecker 2010, chapter 7). For example, even though on Matthews's view, interpretations can never be known, this does not mean that anything can be offered as an interpretation of a work. Matthews states that interpreters must be able to justify their interpretations by offering evidence from the work that demonstrates why it is a plausible understanding of the text (1977, 8).

[111] For an alternative picture that engages with Matthews' view, see Goldman (1990). On his view, there can be (uniquely) correct interpretations and hence, the distinction cannot be drawn in epistemic terms as on Matthews's account. Instead, Goldman argues that the distinction is one between direct perception and inference. He explains that to 'directly perceive that elements of artworks have certain properties is not to interpret them…we can perceive without needing to interpret the expressive properties of works, say the sadness in a musical phrase, while at the same time we interpret the phrase as being there to express sadness' (ibid., 206).

We can also see problems with drawing a sharp distinction between how artistic and scientific representations are interpreted by turning to the discussions in chapter 1. In that chapter, I outlined how thought experiments can be situated within different theoretical contexts or arguments and thereby support different interpretations. I discussed the rocket and thread thought experiment as outlined by Bokulich (2001), but we can also see this by considering Newton's bucket as presented in chapter 5. Newton used the thought experiment to argue that the motion of the water is relative to absolute space. Mach presents an alternative analysis of the same thought experiment scenario, denying that it establishes the existence of absolute space or motion, and rather demonstrates 'only that motion relative to the Earth or fixed stars produces such effects (whereas the water's motion relative to the bucket does not)' (Bokulich and Frappier 2018, 546). While the interpretation of scientific representations such as thought experiments clearly has its limits, there can be disagreements on what would happen in the scenario presented or what conclusions ought to be drawn. In some cases, for example in Einstein's clock-in-the-box and Maxwell's demon, the debate regarding what the thought experiment demonstrates are ongoing.

Furthermore, considering alternative interpretations of what phenomena would occur in a thought experiment setup, or what the thought experiment actually demonstrates can even be a part of the style of presenting thought experiments. We see this in Galileo's dialogues where differences amongst the interlocutors' interpretations of the thought experiment scenario are presented. In some cases, they agree on what would happen but offer different explanations in light of their different theoretical commitments. In other cases, the issue does not revolve around how best to explain an agreed outcome. Rather, the different theoretical standpoints of the interlocutors influence judgments regarding what would happen in the thought experiment. As Palmerino puts it, Galileo's thought experiments function as 'magnifying glasses that render the different theoretical assumptions of the three interlocutors accessible' (2018, 907).

And so, thought experiments can be reanalysed and retooled from different theoretical standpoints. This means that the ways in which scientific and artistic representations can be interpreted is more complex than has been allowed (see also Elgin 2017, chapter 11). Despite this complexity, I do think that there is a key difference that should be highlighted. Part of the aesthetic appreciation of artworks can consist in how they can be "open-ended" in certain ways. Artworks are often valued (as artworks) for the conversations that they prompt regarding their meaning and proper interpretation(s), and so, their indeterminacy is often a virtue. This I take it is part of Egan’s point when he emphasises the richness of literary works in that we can come back to them and find new ways of understanding them. Thought experiments can similarly prompt such discussions, and as outlined in chapter 1, this can form part of their value. However, I take it that the best version of an argument that wants to emphasise the difference between artistic and scientific representations in light of their interpretation is one that highlights how this is not part of scientific representations’ *aesthetic* value.

In the previous chapter, we saw how thought experiments such as Galileo’s bodies, or Einstein’s “happiest thought” have been described using aesthetic language. Their beauty or elegance and so on lies in their ability to evaluate, explain or help us understand something profound based on a description of a simple scenario that allows for reasoning about complex, abstract relations through the introduction of concrete particulars. This chapter so far has offered a view of an aspect of aesthetics of science through a focus on how thought experiments and other representations in science are formulated in a way that fits their content. This resists a reductive approach to aesthetics in the scientific domain, as this interrelation between form and content is a source of aesthetic value. But it also shows why aesthetics should be of interest to philosophers of science. This is because well-formulated thought experiments aid our imaginations and contribute to their ability to effectively communicate, explain and convince. In this section, I returned to a purported difference between the interpretation of artworks on one hand, and scientific works on the other. By utilising an important distinction in philosophy of art, between descriptions and interpretations, I argued that interpretation cannot be used as a clear way of demarcating scientific and artistic representations. While I take the “open-endedness” of some thought experiments to be a part of their value, I do grant that unlike in the case of artworks, it is less clear that this is part of their aesthetic value. In the final section, I will further explore the comparison between thought experiments and literary fiction and offer a discussion that highlights which examples from art will be more appropriate in an attempt to avoid over-stating the shared qualities between the two.

6.5 Selecting the Right Examples

The discussion so far indicates that we ought to look closely at the literary examples that are used when thinking about the qualities that thought experiments share with works of literature.

If we focus on the likes of Morrison or Tolstoy, then the comparisons are going to be thin. Further, as I highlighted in the previous chapter, pinning down what it is that we learn from, say, Morrison's *Beloved* is difficult, and there is reason to think that reducing the novel to functioning like a thought experiment does not do justice to the work. However, such works are not representative of all narrative art and in this section, I will outline some case studies that are more relevant to thought experiments, whether scientific or philosophical. This includes speculative fiction and fables and parables. To do this, I begin with an outline of genre conventions which enables a consideration of the similarities between thought experiments and other genres of literary fiction.

6.5.1 The Thought Experiment Genre

Genres (including hybrid or sub-genres) are categories of artworks. Liao offers a broad characterisation of genre as 'simply groupings of narratives that are recognized by the relevant community as special' (2016, 469). Philosophers of art including Liao have offered rich discussions on how an artwork's inclusion within a certain genre influences our engagement with it (see also Currie 2004, Friend 2012), but the key point for my purposes is that recognising that a work fits within a particular genre generates certain expectations about the work, and affects the way in which the reader or audience understands and evaluates it. For example, the title alone indicates that *Buffy the Vampire Slayer* fits within a horror and/or fantasy genre which will lead us to expect that the series involves supernatural elements. And if we know that the series centers around a group of high school students, then we will also have the expectation that it will involve elements of teen dramas and coming-of-age stories, as well as the highs and lows of romantic relationships and friendships. Further to this, if someone was to watch the series and offer a criticism along the lines of "how stupid, demons could never exist!" then we would take it that, to use Weinberg's (2008) terminology, they have not mastered the genre that *Buffy* is a part of.

Weinberg has presented the idea of thought experiments as a genre (see also Peterson 2018):

> 'There are rules to engaging properly with a hypothetical scenario, after all. To make just some of the more obvious generalizations about our imaginative practices with thought experiments: one should embellish as little as possible; generally it is a practice conducted in an affectively 'cool' manner; and our inferential systems must often be brought to bear in this particular sort of imaginative project as well' (2008, 214).

While Weinberg’s focus is on philosophical thought experiments, this can be carried over to scientific ones. We have seen through Norton’s discussion of Szilard’s thought experiment, as well as through the comparison between Darwin’s whale and eye, that there are certain conventions involved in the creation and engagement of thought experiments with regards to their use of idealisations and how their results are taken to be generalisable. Further, in Galileo’s *Dialogue Concerning the Two Chief World Systems* (1632/1967), we see how Galileo puts boundaries on what properly constitutes a thought experiment. In the *Dialogues*, Galileo, through the interlocutors, offers examples of different hypothetical scenarios from different theoretical standpoints. This includes passages in which one of the interlocutor proposes a thought experiment, and it is rejected as meaningless. I will focus on an example that Palmerino (2018, 2011) discusses. Simplicio proposes a thought experiment through calling upon the work of Locher who denied the Copernican theory that the rotation of the earth is related to the circular motion of objects on earth as they fall. To support this, Locher provided some counterfactual scenarios in which we imagine, for example, that the earth was ‘reduced to nothing’, and argues that an absurd thing (that ‘experience and reason refute’) would happen if Copernicus was correct: ‘no hail or rain would fall from the clouds, but would only be carried naturally around; nor would any fire or flaming thing ascend, since in their view, probably, there is no fire above’ (Galileo 1632/1967, 243).

Saliviati does not offer an analysis of the scenario. Instead, he argues that it is nonsensical to consider such situations. This is because we cannot know what would happen if the earth was annihilated in the same sense that we cannot know ‘what was going to take place on it and around it before it was created’ (ibid., 245). Saliviati highlights how this thought experiment goes beyond what can be evaluated through calling on our experiences of the world. Further, the scenario itself is not adequately described in order for us to work through its consequences. For instance, Saliviati highlights how it is unclear as to whether the thought experiment assumes that when the earth vanished, it took gravity with it (which would effect what would happen).[112] A helpful way of understanding cases such as these is that they involve a violation of the thought experiment genre.

[112] In Galileo’s *Dialogue,* we can thus see the type of scepticism that has been directed towards some thought experiments in philosophy (Palmerino 2018, 916). For instance, Rescher discusses Putnam’s thought experiment in which we are asked to imagine that all domestic cats are actually robots. In ‘tearing apart what experience…binds together’ (in this case, the biology and the behaviour of a cat), Putnam creates a scenario in which ‘all bets are off’ (2005, 152-3). See also Wilkes (1988).

Moreover, as Weinberg discusses, our ability to master a genre may not come naturally and may take some time and experience with a number of examples within that genre. In teaching students how to engage with and criticise thought experiments, there is often some work to be done in order to get them familiar with the genre, such as explaining how thought experiments work, why they are used, and what are the right and wrong questions to ask about the imagined scenario. To discuss a philosophical case, if in response to Thomson's violinist thought experiment someone responded with "you could never attach two people in that way, that's ridiculous!" then it would be clear that they have not mastered thought experiments yet. Similarly, to get caught up in the colour of the balls that Galileo drops from the tower would be to misunderstand what is key to the thought experiment, and consequently what is an appropriate way to evaluate it. In the previous chapter, we saw that Egan argues that unlike artworks, thought experiments are evaluated aesthetically. Against Egan, I set out how thought experiments are often evaluated aesthetically, but what I do think that Egan gets right is that we critically engage with thought experiments based on how well they function in a larger argumentative structure. I add that this is part of the convention of their genre.

We can further understand a literary genre by comparing it other genres. The first genre I will consider is speculative fiction, and then I'll turn to fables and parables.

6.5.2 Speculative Fiction

Cameron (2015) offers an account of speculative fiction as moral and metaphysical thought experiment. Unlike works of say, realist fiction, speculative fiction as a genre describes worlds that depart radically from our own and are not constrained by the history of our world. In creating such "extreme worlds", relevant features can be isolated and exaggerated. Cameron gives the example of Orwell's *1984,* and compares it with works of realist fiction that also depict totalitarianism. Works of speculative fiction can '*aim at,* and to varying degrees be successful at, distilling the essence of what it is to be focused on in a way' (ibid., 33). And in some cases, a work of speculative fiction 'is truly fantastical but at the same time very simple, so that pretty much every feature of the imagined society is there as the result of its being a comment on totalitarianism' (ibid.). This therefore echoes Egan's characterisation of thought experiments in which each of the concrete particularities presented ultimately "stand in" for an element of a more abstract problem or phenomenon. Whereas, Cameron argues, works of realist fiction such as Koestler's *Darkness at Noon* cannot focus in on totalitarianism in the same way; such works necessarily situate their depiction of totalitarianism within a social-historical context. In this case, within Stalinist Russia (ibid., 32).

Another work of speculative fiction that can be considered is Le Guin's *The Left Hand of Darkness* first published in 1969 which Le Guin herself described, in the 1976 introduction to the book, as a thought experiment comparable to those in the scientific realm such as Schrödinger's cat. The novel explores the question: What would society look like in the absence of a presumed male-female gender binary? Le Guin states that the thought experiment's purpose is not predictive, in that it is not a commentary on what gender could be like in another world, but rather, its role is to 'describe reality, the present world'. She goes on to say that in the 'thought-experimental manner proper to science fiction' the work describes 'certain aspects of psychological reality in the novelist's way, which is by inventing elaborately circumstantial lies' (1976, 3). Despite being set on an alien planet, and therefore asking us to imagine a world far-removed from our own in many respects, we are presented with certain features and personalities that we are familiar with and can thus latch onto.

The novel describes Genly Ai's time spent on the planet Gethen where the inhabitants are genderless. The novel is more than an exploration of gender. It is also lauded for its descriptions of the harsh conditions of an extreme winter, its exploration of loyalty and betrayal and experiences of isolation and alienation, as well as its non-linear structure, amongst other things. But it is at least in part a study of this question, and this is clearly an intended aspect of the work.

Furthermore, works of speculative fiction are critically evaluated (again, at least partly) in light of how well they explore certain questions. This mirrors Egan's discussion of how thought experiments in science and philosophy are evaluated based on how well they serve the broader argumentative structure that they are a part of. For example, there has been criticism centered around the use of "he/him" pronouns in *The Left Hand of Darkness* which, it is argued, effects the success of the novel as an exploration of a society in which gender is eliminated and hence, its success as a thought experiment. Le Guin agreed, and later published another story also set on Gethen in which she changed the pronouns to "she/her", although never adopting "they/their" which arguably would have been a better choice. Similarly, there is criticism due to the fact that the gender of the inhabitants of Gethen is determined by their biology. Again, this is said to implicate its effectiveness as a thought experiment; even though the ‘book drives you to question all of our assumptions about male and female bodies, it never raises any questions about how gender shapes us independently of our biological sex’ (Anders 2019).

This indicates that we should be careful when selecting our points of comparisons when discussing the aesthetic and literary qualities of thought experiments; some literary fictions will be more relevant than others. Furthermore, this has implications for Egan’s view regarding the types of more abstract reasoning literary fictions will prompt. As already mentioned, the relation between aesthetic and cognitive value in art is very much an alive debate in philosophy of art. Consequently, there will be opposition to Egan’s claim that when engaging with literary fictions in general, we do not need to have more general reflections beyond the particularities presented in the work. Egan’s own example, of gaining pleasure from an Austen novel without reflecting on say, class and marriage and so on, will be contentious. Someone could of course greatly enjoy the story of the novel without grasping its thematic content, but it is plausible that such an engagement would count as a failure to fully appreciate the work. Similarly, someone could gain pleasure from reading *1984* or *Left Hand of Darkness* by appreciating the events told in the story and the formal features of the work and so on, without reflecting on totalitarianism or gender. And we could even gain pleasure from, say, a weird and wonderful thought experiment without grasping its thematic or scientific content.

So I have doubts about Egan’s examples and his view regarding the proper appreciation of literary works in general. But I do agree with him insofar as the value (cognitive and/or aesthetic) of, say a work of Austen, or Morrison’s *Beloved* to use my example, does not lie in the way in which they are similar to a thought experiment. However, I do not think this extends to all artworks. In particular, *1984* and *The Left Hand of Darkness,* insofar as they are an exploration of totalitarianism and gender, have cognitive value in virtue of the way they function as thought experiments.

6.5.3 Fables and Parables

Another form of literary work that provides a useful comparison to thought experiments are fables and parables. Take an example of a fable from Aesop:

> ‘Between the North Wind and the Sun, they say, a contest of this sort arose, to wit, which of the two would strip the goatskin from a farmer plodding on his way. The North Wind first began to blow as he does when he blows from Thrace, thinking by sheer force to rob the wearer of his cloak. And yet no more on that account did he, the man, relax his hold; instead he shivered, drew the borders of his garment tight about him every way, and rested with his back against a spur of rock. Then the Sun peeped forth, welcome at first, bringing the man relief from the cold, raw wind. Next, changing, he turned the heat on more, and suddenly the farmer felt too hot and of his own accord threw off the cloak, and so was stripped.
>
> Thus was the North Wind beaten in the contest. And the story means: “Cultivate gentleness, my son; you will get results oftener by persuasion than by the use of force”’ (from Hunt 2009, 370).

As with literary works and thought experiments, fables and parables such as this focus on examples of concrete objects and events. Like thought experiments, they are not merely a short fictional story, nor are they simply an articulation of a viewpoint, but are written with a purpose to convince or explain something to the reader. In this case, the fable’s moral is that persuasion is superior to force. Further, the characters and objects in a fable are idealised and depart from their real world counterparts and yet importantly make the scenario relatable. The situations are simple, they are not situated in a certain historical context or geographical location, and the conclusion is intended to be generalised in order to apply to a broader range of cases than depicted in the example. At the same time, although their relation to literature is contentious, their style is clearly literary rather than that of abstract argumentation.

In the scientific context, Cartwright (1991, 2010) has drawn parallels between models (including thought experiments) on one hand, and fables and parables on the other. She argues that fables ‘transform the abstract into the concrete’ in a way that is comparable to models in physics. She states that ‘the relationship between the moral and the fable is like that between a scientific law and a model’ (1991, 57). As Cartwright explains, in order to find a conclusion that is true in both the model or fable and in other cases beyond the particulars in the described scenario, we may have to “climb up the ladder of abstraction”, that is, express the conclusion or the moral in more abstract terms. We generalise the result of Galileo’s falling bodies thought experiment beyond the particularities of the case, that is, we take the result as applying to all bodies, not just the balls dropped from the tower in the thought experiment.

Cartwright further argues that the relation between the moral and the fable (or a model/thought experiment and its result) is ‘that of the general to the more specific’ and that the moral is “fitted out” by the fable. That is to say that, ‘the moral describes just what happens in the fable; but the fable fits it out in a special way—a way true to the moral but not necessarily shared by all cases of which the moral is true’ (2010, 26-27). The “fitting out” of the moral of the superiority of persuasion over force is done in a particular way in the example above, that is, through the comparison of the wind’s forceful blow versus the sun’s patient approach. In another case, the fitting out of the abstract moral will be “fitted out” in a different way, and the sun and the wind will be irrelevant. So too for the use of the tower and the falling balls in Galileo’s thought experiment. In this sense, then, Cartwright is appealing to the distinction I introduced above; that is, the difference between the thematic (or scientific) content of a thought experiment and the particular way in which that content is expressed through the narrative.

A key difference between parables, such as the good Samaritan, and fables such as Aesop’s, that Cartwright highlights in her later work, is that the former do not typically have the moral or lesson “built in”. Instead, defending a view of what the parable shows involves interpretative work, including attending to other parts of the text in which it is presented, as well as how the world operates. Cartwright argues that many of the highly idealised models utilised in physics and economics are more like parables than fables in this sense:

> ‘A variety of morals can be attributed to the models, expressed in a variety of different vocabularies involving abstractions of different kinds and at different levels. Importantly, these morals can point in different directions, implying opposite predictions for the real-life situations to which we want to apply them’ (2010, 21).

In the previous section, we saw how the interpretation of thought experiments can vary; different theoretical commitments alter the conclusions drawn from the example, and that there can be debate around what exactly a certain thought experiment shows. In Cartwright’s terms, the moral of a thought experiment or parable is not part of the parable or thought experiment itself. Rather, further work needs to be done to show which “ladder of abstraction” must be climbed to reach the result that can be generalised.

We can see, then, that there are connections between the ways in which thought experiments and fables/parables and works of speculative fiction explore certain questions as well as convince and/or explain. In each, we attempt to take on board the things that are stipulated by the author, and do not object to them on the grounds that they are "unrealistic" because they involve say, imagining the wind and the sun as having agency, or animals that can speak, or a lack of air resistance and so on. There is a shared convention to how we engage with these stories. We recognise that they have an intended aim or moral. Additionally, the role of their formulation and particular elements is similar. In each case, they are carefully chosen to effectively convey their more abstract message that can be generalised beyond the scenario given. This in turn guides our engagement with and critical evaluation of them.[113]

In one sense, thought experiments can be more straightforwardly compared with fables and parables than with speculative fiction in that they are more simply intended explorations of certain ideas and questions, whereas works of speculative fiction have other features that go into their evaluation as highlighted above. Another way in which thought experiments and fables and parables differ from speculative fiction can be seen by attending to our affective responses. Weinberg (2008) highlights how we conduct philosophical thought experiments in a "cool" manner which means that we do not have the kind of affective responses typically associated with our engagement of literary fictions. It would be odd (or at least not part of an appropriate engagement) to feel a sense of sadness for the violinist's illness in Thomson's thought experiment. Similarly, it would be odd to feel a sense of injustice for the shivering man in Aesop's fable because he is subject to a competition between the sun and the wind, for example.[114] More generally, it is evident that membership in a certain genre impacts our affective responses to a work. Take for example the ways in which we might respond to death and misfortune with amusement when watching black comedies, compared to our response to similar events in dramas. As Peterson (2018) discusses, feeling sad at the end of *Dr. Strangelove* would be to overlook the point of the film, which is dictated, at least in part, by its inclusion within a particular genre.

And so, considering the way in which we conduct thought experiments in a "cool" manner indicates that they are more like fables and parables than speculative fictions in one respect. However, in another sense, thought experiments are more similar to literary works (including speculative fiction) than fables and parables. Like great works of literature, thought

[113] Weatherson (2010) has also discussed this connection between fables, thought experiments and genre on his blog, under 'surveys and thought experiments'.

[114] Although we can highlight that at least for some, part of the persuasiveness of Thomson's violinist might lie in our feelings of distress and so on that occur when we imagine waking up and finding ourselves in that position. I won't say anything about affective responses with regards to scientific thought experiments, although see Todd (2017, 2020) for a discussion of aesthetics, imagination and affect in mathematics and science.

experiments are described as “profound” or “deep” (as we saw in chapter 5). Thought experiments are invoked to realise or communicate something that might have otherwise been difficult to comprehend or access, and they can offer surprising and novel insights. Whereas the case of fables and parables often seem to provide a message that is already known (albeit repackaged in a succinct and often elegant way) rather than contributing anything new of cognitive value. Consequently, there are various dimensions along which we can explore the connections between thought experiments and other literary genres.

In this section, I have argued that a discussion of the comparisons between thought experiments and literary works would benefit from a more careful selection of literary cases. Further to this, there are other elements from the artistic realm that can be brought to bear on (scientific) thought experiments, namely genre conventions and how membership in a genre affects our engagement with fictional scenarios.

6.6 Conclusion

The previous chapter ended with a dilemma for aesthetics of science. This was the problem that either aesthetic language in science is being used in a merely metaphorical way (and what is really being tracked are epistemic features) or the aesthetic language is literal, but such values are not useful in science. In this chapter, I have set out to solve this dilemma. I have characterised an aspect of the aesthetics of science which is currently overlooked in the literature; the interrelation between form and content. This offers a novel and promising way of defending the role of aesthetic features in science without reducing them to epistemic features. This is because an appropriate fit between form and content is a source of aesthetic value that can contribute to understanding through aiding our formation of useful imaginings, and can increase the demonstrative force of scientific thought experiments. The account I have developed also has the benefit of showing how aesthetic values in science can be compared with the aesthetic evaluation of artworks as the way in which artworks express their content via their formal features is a basis of their aesthetic appreciation. In light of this, I returned to some of the comparisons made between thought experiments and works of literature. While I agree that there are significant differences that ought to be attended to, and that not all works of literature are like thought experiments, I hope to have shown that part of the value of scientific thought experiments in scientific practice includes the qualities that they share with literary works

Conclusion

In this chapter, I will sum up the focal points of my thesis and indicate some future avenues of research that this project opens up.

Summary

This project began with an outline of the debate that shapes the literature on thought experiments. This contrasts Norton’s argument view with Brown’s platonist account. Both of these views are concerned with what is called the “puzzle” of thought experiments, that is, the question of how thought experiments can increase our knowledge of the world. In this thesis, I set out to show how a closer look at the imagination in thought experiments allows us to explore their value beyond the narrow confines of a Norton/Brown dichotomy in order to consider the use of thought experiments in scientific practice. In doing so, we realise that we do not need to commit to a platonist view of thought experiments if we want to maintain that there are severe limitations to a position that reduces thought experiments to arguments. At the same time, I aimed to show how focusing on thought experiments can illuminate the role of scientific imagination more generally in science. Thought experiments are a good starting point for an analysis of the scientific imagination as they rely on our imaginative faculties and are an accepted part of scientific practice.

My aim in the first part of the thesis was to get clearer on the sense in which thought experiments can be considered experiments in the imagination. Currently, the issue regarding thought experiments and ordinary, physical experiments has been primarily concerned with whether or not the two can be placed on a continuum. I argued that this is largely down to a matter of emphasis—we either focus on the features that they share (such as the method of variation) or we focus on ways in which they differ and hence, argue that thought experiments are something else (for example, arguments). In light of this, I drew on a parallel debate on computer simulations in science which can further the discussion surrounding the epistemic status of thought experiments compared with that of physical experiments. In the first chapter, I outlined how, against Hacking, thought experiments also have a “life of their own”. Thought experiments can evolve in order to become more suited to their purposes, thought experimental results can be “robust” in the sense that they withstand changes in the theoretical realm and further, thought experiments can precede theory. In the second chapter, I presented two arguments for the privileging of experiments over computer simulations and I argued that these can also be carried over to thought experiments. This is because, like computer simulations, they do not intervene on the world and instead can be said to explore hypothetical scenarios. The first of these arguments is the materiality thesis which states that there is a crucial difference between the object-target relation in an experiment compared with that in a computer simulation; experiments are materially continuous with their target, whereas computer simulations lack this “shared ontology”. While this view may be compelling, it ultimately fails. This is because there are difficulties in pinning down how far one system can be said to be materially similar to another and further, there are instances in which the relevant similarity between object and target is not of a material kind. The outcome of this was that materiality does not always matter. Hence, experiments cannot be automatically privileged for this reason.

I then turned to the argument from surprise which states that computer simulations do not have the capacity to surprise in the way that physical experiments can. We saw that Morgan (2003, 2005) distinguishes between mere surprise and confoundment, and argues that experiments only can achieve the latter. Confoundment can be characterised as “productive surprise” (Currie 2018) because it can force scientists to revise their theories in a fundamental way. The reason why only experiments can confound, according to Morgan, is related to the materiality claim; as physical experiments capture the part of the world that the scientist is ultimately interested in, the scientist is not in complete control over the experiment’s results. Whereas in a computer simulation, scientists are working with something that they themselves programmed and ultimately retain control over.

While surprise has not been discussed in the context of thought experiments, I set out how different views on thought experiments will have different consequences regarding how thought experiments generate surprise. Via a comparison with abstract objects in other domains, such as mathematical platonism and Popper’s view regarding theories as “World 3” entities, I demonstrated that on Brown’s view, thought experiments can confound in the same way as experiments. On Norton’s view, we saw that thought experiments can still productively surprise, but this is in a different way than experiments. This is because thought experiments (and computer simulations) involve inference rather than genuine discovery. Finally, I demonstrated that attending to the role of the imagination in thought experiments shows how thought experiments can productively surprise in a distinctive way. Thought experiments, like computer simulations and arguments, “probe” our existing knowledge, but the way in which they do so cannot always be characterised in terms of a process of logical steps from the initial set up of the scenario. Through the example of Galileo's falling bodies, I argued that there is a certain "freedom" to our imagination which allows for a source of surprise that is disruptive but cannot be captured as argumentative reasoning.

The upshot of chapter 1 and 2 was that we need to get a clearer idea of the function of the imagination in thought experiments. This was my core aim in the second part of the thesis.

In chapter 3, I gave an overview of attitudes surrounding imagination in science. While there has been scepticism regarding the imagination as a tool for learning (both in science and more generally), which can be at least in part explained by the imagination's link with creativity, more recently, philosophers of science have recognised that it is an important tool in science that cannot be dismissed as merely part of the context of discovery. In particular, there is a set of views that argue that models have important similarities with artistic fictions and our engagement with each relies on our imaginative capacities. Some of the most prominent versions of these accounts have drawn on Walton's (1990) account of representation as "make-believe". While models' imaginary status has received a great amount of attention, there is a crucial question that has been for the most part overlooked: What is the nature of the imagination in science?

We saw that many have equated the imagination with mental imagery, that is, an objectual form of imagination. I then set out ways in which we should broaden the notion of imagination to also include the propositional imagination. A recent proposal that has furthered the discussion of the type of imagination in science argues that other views have been wrong to assume the imagination consists in mental imagery. Furthermore, they argue that mental imagery is neither a sufficient nor a necessary form of imagination in the scientific domain. Rather, it is the propositional imagination only that has import in science. This view is presented by Salis and Frigg (2020) and they argue that it avoids some of the scepticism directed towards imagination in science. I argued that Salis and Frigg are correct to bring the propositional imagination into the picture. However, they are too quick to dismiss the objectual imagination, and further, we should be cautious when extending our analysis of the imagination in modelling directly to thought experiments.

In chapter 4 I developed a new pluralist view of the nature of imagination in thought experiments. In light of the uncertainty surrounding which forms of imagination will be "necessary" for the conduct of a thought experiment, I argued that we should instead consider an alternative set of questions: What do thought experiments ask us to do? What kind of imagination do they invite? On my view, we cannot take one example and generalise to all other cases of thought experiments. Instead, we need to consider a variety of thought experiments and the role they play in science. I set out two examples of thought experiments that I think a Salis-Frigg analysis gets right; they merely invite the propositional imagination and any mental imagery experienced would be superfluous to the conduct of the example. However, I then consider examples in which we are asked to do more than consider a set of propositions. By drawing on De Regt (2015, 2017), I discussed how imagining via mental imagery can contribute to scientific understanding, and I argued that this might vary depending on who is conducting the thought experiment or who the example was designed for. Further, there are cases of thought experiments that involve spatial reasoning and many thought experiments are presented alongside images which work as aids to our imagination as in the case of Stevin's chain. To end, I emphasised that we should not see our imagination as being restricted to either a propositional or an objectual type. Rather, these different forms of imagination can work together in order to allow us to conduct a thought experiment successfully.

The final part of the thesis turned to aesthetics, and I developed a new approach to the aesthetics of science that incorporates the significance of the imagination in thought experiments. In chapter 5, I outlined the connections that have been drawn between thought experiments and works of literary fiction. The cognitive value of literature has been defended through comparing literary works with thought experiments; each ask us to engage in fictional scenarios and if the latter can lead us to new insights about the world, then so can the former. The proposed similarities between thought experiments and artworks in part motivated my view that we ought to include thought experiments in our accounts of aesthetics of science. This literature currently focus on theories and mathematical proofs. But thought experiments are also often evaluated by scientists and philosophers using aesthetic terminology. While the comparisons between artworks and thought experiments are persuasive, I set out some important differences that have been used to undermine the connections drawn between the two. For example, thought experiments are always part of a larger, argumentative structure whereas literary fictions are not; thought experiments are ultimately about something more generalisable than the particulars of the case, whereas we are not required to engage in the more abstract themes of a work of fiction; finally, the interpretation of thought experiments is fixed more firmly than in the case of literary works. And I ended with a dilemma for those who seek to defend the value of aesthetics in science.

The thought is that either aesthetic terminology is being used in a metaphorical sense (and really tracks epistemic features). Or scientists' aesthetic evaluations are genuinely aesthetic, but such values do not form an important part of science.

In chapter 6, I set out a new way of considering aesthetic values in the scientific domain. I outlined views in aesthetics that argue that a source of aesthetic pleasure in artworks has to do with the relation between the overarching content of a work (the purpose of a work or the point that the work expresses) and the way in which it is formulated. In drawing on some existing work concerning the way in which formulations matter in science (Vorms 2011), especially when we attend to the interactions between the representation and its user, I outlined how the fit between form and content in thought experiments plays a significant role in its usability in science. To do this, I discussed two examples from Darwin. Each of these aim to demonstrate the explanatory power of the theory of natural selection (hence, there are similarities in their content) but the way in which they are formulated helps explains the success of one (the eye thought experiment), and the failure of the other (Darwin's "monstrous" whale). I argued that appropriate fit between form and content enhances the accessibility of a thought experiment scenario, and hence its usefulness as a prompt for our imagination as well as playing a role in effectively communicating some idea or theory.

Finally, I came back to some of the comparisons between thought experiments and works of literary fiction. I utilised a distinction drawn in the philosophy of art—between an interpretation and a description of a work—to complicate the view that interpretations of scientific works are "fixed" in a way that artistic fictions are not. Further, going back to chapter 1, I demonstrated how thought experiments can indeed be interpreted in different ways, and that this can even form part of their presentation in science. Finally, through accounts of genre in philosophy of art, including how membership in a genre can affect our engagement with a work, I argued that some of the disanologies between thought experiments and literature can be responded to by being more selective in the artistic works that we use as a comparison. In particular, thought experiments, and our evaluation of them, is similar to that of speculative fiction as well as fables and parables. I hope to have shown that there are fruitful analogies to be drawn between scientific and artistic representations and I indicated that there are various dimensions along which we can draw these comparisons.

Future Directions

To end, I will indicate some other areas of enquiry that can be pursued with the aid of the arguments developed in this thesis.

Firstly, as previously mentioned, surprise has not received much attention in the philosophy of science. My discussion of thought experiments and the imagination has already indicated that Morgan's (2003, 2005) distinction between mere surprise and confoundment does not capture all there is to say about surprise and its value in the scientific realm. This opens up other ways to consider surprise in science. One possible avenue is to consider Dutilh Novaes's work on mathematical proofs. What is especially fruitful about this account for the topics explored in this thesis is that she links surprise with aesthetic values. In her dialogic account of mathematical proofs, she states that a deductive proof 'corresponds to a dialogue between the person wishing to establish the conclusion (given the presumed truth of the premises), and an interlocutor who will not be easily convinced and who will bring up objections, counterexamples, and requests for further clarification and precision' (2019, 74). Most proofs are not actual dialogues and so Dutilh Novaes describes them as "fictive interactions" in the sense that they reproduce 'multi-agent communicative scenarios' (ibid., 76). Further, she argues that 'one of the main functions of deductive proofs (then as well as now) is to produce *persuasion*, in particular what one could call *explanatory persuasion*: to show not only *that* something is the case, but also *why* it is the case' (ibid, 74).

Dutilh Novaes claims that surprise in mathematical proofs also results in pleasurable aesthetic feelings. Crucially, she argues, such pleasure increases the mathematician's conviction regarding the proof in question, increasing its persuasiveness by eliciting positive affective responses towards it (ibid.). Whether or not a dialogic approach works well for proofs, it does appear to carry over well to thought experiments, granted of course that unlike Dutilh Novaes's interest, these go beyond deductive reasoning. Thought experiments are often used to convince as well as to increase knowledge and understanding. In some cases, they are presented as dialogues (as in Galileo's writings) whereas in others, we can appeal to the notion of "fictive interactions" and think about how thought experiments are typically engaged with and analysed. A further research topic is thus one which combines the value of the *psychological* sense of surprise that Dutilh Novaes highlights for mathematics with an exploration of the ways in which scientific thought experiments are formulated so as to bring about such feelings. In particular, this would require a closer look at the social aspects of communication and information sharing in science which would further my argument presented in the thesis that we need to attend more carefully to the context in which a thought experiment is used.

A crucial aim of my project was to get clearer on both the nature of the imagination in science, and on the comparisons between thought experiments and other more obviously aesthetic artefacts, namely, works of literary fiction. In developing my pluralist account of the imagination in science, I raised ways in which the limitations of mental imagery (used to undermine the fiction views of models) do not appear to apply to the case of thought experiments. While I did not say much more on the imagination in modelling, I suspect that my pluralist account can also apply in their case. Models form a hugely diverse group and my view can act as a framework within which we can consider the ways in which the imagination will also vary in their case, in accordance with different types of models. For example, one helpful point of comparison might be some cases of "toy models". These are very simplified and highly idealised representations of their targets and they investigate a small range of causal or explanatory factors (Reutlinger et. al. 2018, Nguyen 2019).

While theoretical models may not be so straightforwardly visualisable as in the case of (at least most) thought experiments, and some models may only invite a propositional form of imagination, I also discussed how the objectual and propositional imagination can interact in the conduct of a thought experiment. This can also be applied to models. For example, we can consider the ways in which we imagine a model system via a mental image, then within this, we draw certain conclusions regarding the behaviour of the system via the propositional imagination. As a consequence, a fruitful question to consider is which instances of models will rely more on the objectual imagination and why. As with thought experiments, I expect this to vary with the function of the model and may be particularly apparent in cases in which a model is used to increase understanding.

A further dimension that could be added to my pluralist account comes from considering the use of not only theoretical, but physical models. In chapter 4, I demonstrated how external pictures can be used with a thought experiment. Not only can they aid us in imagining the scenario correctly, but we can mentally manipulate the static image in our imaginations in order to consider how it can change as we reason through the thought experiment scenario. A feature of physical models such as those used in chemistry that Toon (2012) discusses is the tactile, bodily aspect of manipulating the model. This, Toon argues, 'allows scientists to investigate the properties of molecules in a kind of 'imagined analogue' of the way in which we discover the properties of normal, everyday objects' (ibid., 129).

Finally, while I anticipate that the account of aesthetics in science that I have outlined will be a promising defence of aesthetic value across the board in science, an obvious starting point for the extension of the view (given the typical presentation of thought experiments) would be to consider other areas in which narratives are used in science. In particular, it would be fruitful to compare my view with those that discuss narratives and narrative explanations in science such as in modelling (Morgan 2012), computer simulations (Wise 2011) and the historical sciences (Currie and Sterelny 2017) As Morgan and Wise explain, 'for some scientists, or at some sites of science, narrative works to create coherence between a variety of different elements that otherwise do not appear to hang together, but do need to be made to fit sensibly together whenever an investigator recognises that they are all elements that belong to the phenomenon to be described or explained' (2017, 1-2). They go onto state that presenting the same information in an argument form or as 'theory-led description' would be less advantageous; 'all too often theories are too thin to cover the problem or the ground and the answering account requires the construction of something like a mosaic' (ibid.). This thus chimes with my account in chapter 6 on how we can represent content in different ways, but some will be more fitting than others. Furthermore, there may be a further connection to be made to philosophy of art, with regards to accounts of narrative understanding in literary works in which the function of narratives in coherence making is also emphasised (Barwell 2009).

In a couple of places throughout the thesis, I called on some examples of thought experiments in philosophy. But I am yet to consider the beauty of philosophical thought experiments compared with scientific ones, nor have I said anything about how the appropriate fit between form and content may be a source of their aesthetic value that can contribute to their epistemic value. The use of thought experiments in philosophy is more of a contentious issue than in science (for a classic criticism, see Wilkes 1988). Something to consider would be the ways in which their aesthetic features and narrative form can be part of how they can lead us astray; such features can be seductive and hence, contribute to the force of misleading examples. One potentially helpful comparison comes from William's "The Self and the Future". Williams outlines two different presentations of the same thought experiment, the different presentations yield different judgments of the scenario. In light of this, it would be interesting to consider whether there are scientific cases in which their formulation plays a role in persuading us even when a closer analysis reveals that they are based on faulty reasoning. Further, it would be interesting to consider whether the nature of the imagination that the example invites plays a role in such instances. And so, the fruitfulness of the imagination and the value of aesthetics in scientific thought experiments and other areas of practice may simultaneously be possible source of errors. This relates to Levy's comments regarding metaphors in science for which, compared with more "direct" forms of representation, there is also a potential danger 'for the model or metaphor to be overapplied, reified, or otherwise taken "too seriously"' (2020, 295). This is an interesting question and my pluralist account, as well as my proposal for an source of aesthetic value in science, including my outline of the thought experiment genre and its conventions, provides the tools to begin diagnosing where such features can go wrong and hence, lead us astray.

Therefore, the core components of my thesis surrounding thought experiments and the scientific imagination—the way in which thought experiments can surprise, the pluralist nature of the imagination in their conduct, and the way in which their formulation can affect their value—can be usefully extended in order to make progress in other areas in philosophy of science.

Bibliography

Anders, Charlie Jane. “The Left Hand of Darkness at Fifty.” *The Paris Review* (blog). https://www.theparisreview.org/blog/2019/03/12/the-left-hand-of-darkness-at-fifty/. 2019.

Arcangeli, Margherita. “Imagination in Thought Experimentation: Sketching a Cognitive Approach to Thought Experiments.” *Model-Based Reasoning in Science and Technology*, edited by Magnani L., Carnielli W., Pizzi C. Springer. 2010.

Arcangeli, Margherita. “The Hidden Links between Real, Thought and Numerical Experiments.” *Croatian Journal of Philosophy*. 2018.

Arcangeli, Margherita. “The Two Faces of Mental Imagery.” *Philosophy and Phenomenological Research*. 2019b.

Balcerak Jackson, Magdalena. “On the Epistemic Value of Imagining, Supposing, and Conceiving.” In Kind, Amy, and Peter Kung, eds. *Knowledge Through Imagination*. Oxford University Press. 2016.

Banerjee Abhijit V. “A Simple Model of Herd Behavior.” *The Quarterly Journal of Economics* 107, no. 3: 797–817. 1992.

Barberousse, Anouk, Sara Franceschelli, and Cyrille Imbert. “Computer Simulations as Experiments.” *Synthese* 169, no. 3: 557–74. 2009.

Barwell, Ismay. “Understanding Narratives and Narrative Understanding.” *The Journal of Aesthetics and Art Criticism* 67, no. 1: 49–59. 2009.

Bedessem, B. and Ruphy, S. ‘Scientific autonomy and the unpredictability of scientific inquiry: The unexpected might not be where you would expect ’, *Studies in History and Philosophy of Science* 73: 1-7. 2019.

Beisbart, C. “Are Computer Simulations Experiments? And If Not, How Are They Related to Each Other?” *European Journal for Philosophy of Science* 8, no. 2. 2018.

Beisbart, C. “How Can Computer Simulations Produce New Knowledge?” *European Journal for Philosophy of Science* 2. 2012.

Beisbart, Claus, and John D. Norton. “Why Monte Carlo Simulations Are Inferences and Not Experiments.” *International Studies in the Philosophy of Science* 26, no. 4: 403–22. 2012.

Bishop, Michael A. “Why Thought Experiments Are Not Arguments.” *Philosophy of Science* 66, no. 4: 534–41. 1999.

Boge, Florian J. “Why Computer Simulations Are Not Inferences, and in What Sense They Are Experiments.” *European Journal for Philosophy of Science* 9, no. 1: 1–30. 2018.

Bohr, Niels. 1949. “Discussions with Einstein on Epistemological Problems in Atomic Physics.” Accessed September 6, 2020. http://www.physics.metu.edu.tr/~uoyilmaz/Physics/nielsbohr/Discussions%20with%20Einstein%20on%20Epistemological%20Problems%20in%20Atomic%20Physics.htm.

Bokulich, Alisa, and Mélanie Frappier. “On the Identity of Thought Experiments: Thought Experiments Rethought.” In *The Routledge Companion to Thought Experiments*, edited by Michael T. Stuart, Yiftach J. H. Fehige, and James Robert Brown. Routledge. 2018.

Bokulich, Alisa. “Rethinking Thought Experiments.” *Perspectives on Science* 9, no. 3: 285–307. 2001.

Boumans, M. “Mathematics as Quasi-Matter to Build Models as Instruments.” In *Probabilities, Laws, and Structures*, edited by Dennis Dieks, Wenceslao J. Gonzalez, Stephan Hartmann, Michael Stöltzner, and Marcel Weber, 307–18. *The Philosophy of Science in a European Perspective 3*. Springer Netherlands. 2012.

Breitenbach, Angela. “Beauty in Proofs: Kant on Aesthetics in Mathematics.” *European Journal of Philosophy* 23, no. 4: 955–77. 2015.

Brown, James Robert. “Peeking into Plato’s Heaven.” *Philosophy of Science* 71, no. 5: 1126–38. 2004b.

Brown, James Robert. “Thought Experiments since the Scientific Revolution.” *International Studies in the Philosophy of Science* 1, no. 1: 1–15. 1986.

Brown, James Robert. “Why Thought Experiments Transcend Experience.” In *Contemporary Debates in Philosophy of Science*, Blackwell. 2004a.

Brown, James Robert. “Counter Thought Experiments.” *Royal Institute of Philosophy Supplements* 61: 155–77. 2007.

Brown, James Robert. 1991. *The Laboratory of the Mind: Thought Experiments in the Natural Sciences*. New York: Routledge. 2011.

Bueno, Otavio, George Darby, Steven French and Dean Rickles, *Thinking about Science, Reflecting on Art: Bringing Aesthetics and Philosophy of Science Together.* Routledge. 2017.

Buzzoni, Marco. “Pierre Duhem and Ernst Mach on Thought Experiments.” *Hopos: The Journal of the International Society for the History of Philosophy of Science* 8, no. 1: 1–27. 2018.

Cameron, Ross P. “Improve Your Thought Experiments Overnight with Speculative Fiction!” *Midwest Studies In Philosophy* 39, no. 1: 29–45. 2015.

Camp, Elisabeth. “Two Varieties of Literary Imagination: Metaphor, Fiction, and Thought Experiments.” *Midwest Studies In Philosophy* 33, no. 1: 107–30. 2009.

Carroll, Noël. “The Wheel of Virtue: Art, Literature, and Moral Knowledge.” *The Journal of Aesthetics and Art Criticism* 60, no. 1: 3–26. 2002.

Carroll, Noël. *Philosophy of Art: A Contemporary Introduction.* Psychology Press. 1999.

Cartwright, Nancy. “Fables and Models.” *Aristotelian Society Supplementary Volume* 65, no. 1: 55–82. 1991.

Cartwright, Nancy. “Models: Parables v Fables”, In Frigg, Roman, and Matthew Hunter (eds.) *Beyond Mimesis and Convention - Representation in Art and Science,* Boston Studies in the Philosophy of Science, Springer. 2010.

Cartwright, Nancy. *How the Laws of Physics Lie.* Oxford University Press. 1983.

Chandrasekharan, Sanjay, Nancy Nersessian and Vrishali Subramanian, “Computational Modeling: Is this the end of thought experiments in science?”, in J. Brown, M. Frappier and L.Meynell (eds.) *Thought Experiments in Philosophy, Science and the Arts*, Routledge. 2012.

Cooper, Rachel. “Thought Experiments”. *Metaphilosophy* 36, no. 3: 328–47. 2005.

Craik, K. *The nature of explanation*. Cambridge: Cambridge University Press. 1943

Crease, R.P. “The Most Beautiful Experiment.” *Physics World* 15, no. 9. 2002.

Crowther, Karen, Niels S. Linnemann, and Christian Wüthrich. “What We Cannot Learn from Analogue Experiments.” *Synthese*. 2019.

Currie, A. “The Argument from Surprise.” *Canadian Journal of Philosophy* 48, no. 5: 639–61. 2018.

Currie, Adrian, and Kim Sterelny. “In Defence of Story-Telling.” *Studies in History and Philosophy of Science Part A*, SI: Narrative in Science, 62: 14–21. 2017.

Currie, Greg, and Anna Ichino. “Aliefs Don’t Exist, Though Some of Their Relatives Do.” *Analysis* 72, no. 4: 788–98. 2012.

Currie, Gregory and Ian Ravenscroft, *Recreative Minds: Imagination in Philosophy and Psychology*, New York: Oxford University Press. 2002.

Currie, Gregory. “Models As Fictions, Fictions As Models” *The Monist* 99, no. 3: 296–310. 2016.

Currie, Gregory. *Arts and Minds*. Oxford University Press. 2004.

Cutting, James, E. ‘Gustave Caillebotte, French Impressionism, and Mere Exposure’, *Psychonomic Bulletin & Review* 10: 319–43. 2003.

Dardashti, Radin, Karim P. Y. Thébault, and Eric Winsberg. “Confirmation via Analogue Simulation: What Dumb Holes Could Tell Us about Gravity.” *The British Journal for the Philosophy of Science* 68, no. 1: 55–89. 2017.

Darwin, C. R. “Letter to Lyell”. 1859b. *Darwin Correspondence Project.* https://www.darwinproject.ac.uk/letter/DCP-LETT-2575.xml. Accessed August 31, 2020.

Darwin, C. R. “Letter to Murray”. 1860. *Darwin Correspondence Project.* https://www.darwinproject.ac.uk/letter/DCP-LETT-2772.xml. Accessed August 31, 2020.

Darwin, C. R. 1859a. *On the Origin of Species by Means of Natural Selection.* London: John Murray.http://darwinonline.org.uk/content/frameset?itemID=F373&viewtype=text&pageseq=1. Accessed August 31, 2020.

Darwin, C. R. 1872. *The Origin of Species by Means of Natural Selection.* London: John Murray. 6th Edition; with Additions and Corrections. Eleventh Thousand. http://darwinonline.org.uk/content/frameset?pageseq=1&itemID=F391&viewtype=text. Accessed August 31, 2020.

Davies, David. "Thought Experiments and Fictional Narratives." *Croatian Journal of Philosophy* 7, no. 1: 29–45. 2007.

Davies, Stephen. "Authors' Intentions, Literary Interpretation, and Literary Value." *The British Journal of Aesthetics* 46, no. 3: 223–47. 2006.

Dennett, D. "Artificial life as philosophy", *Artificial Life,* 1(3), 1994.

Di Paolo et al. "Simulation Models as Opaque Thought Experiments". Conference: Artificial Life VII: The Seventh International Conference on the Simulation and Synthesis of Living Systems, Reed College, Portland, Oregon, USA, 1-6 August. 2000.

Dirac, P. a. M. "The Excellence of Einstein's Theory of Gravitation." *Einstein: The First Hundred Years*, Edited by Goldsmith, Maurice, Alan Mackay, and James Woudhuysen. Elsevier, 1980.

Dodd, Julian. "The Possibility of Profound Music." *The British Journal of Aesthetics* 54, no. 3: 299–322. 2014.

Douglas, Heather, and P. D. Magnus. "State of the Field: Why Novel Prediction Matters." *Studies in History and Philosophy of Science Part A* 44, no. 4: 580–89. 2013.

Dyck, John, and Matt Johnson. "Appreciating Bad Art." *Journal of Value Inquiry* 51, no. 2: 279–292. 2017.

Earman, John, and John D. Norton. "Exorcist XIV: The Wrath of Maxwell's Demon. Part I. From Maxwell to Szilard." *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics* 29, no. 4: 435–71. 1998.

Edge.org: ‘What is your favorite deep, elegant or beautiful explanation?’ https://www.edge.org/responses/what-is-your-favorite-deep-elegant-or beautiful-explanation. Accessed March 10, 2019.

Egan, David. “Literature and Thought Experiments.” *The Journal of Aesthetics and Art Criticism* 74, no. 2: 139–50. 2016.

Einstein, Albert and Leopold Infeld. *The Evolution Of Physics*. The Scientific Book Club And Company Limited., 1938.

Einstein, Albert, and Paul Arthur Schilpp. *Autobiographical Notes*. La Salle, Illinois: Open Court

Einstein, Albert. 1905. “Zur Elektrodynamik bewegter Korper”, *Annalen derPhysik*. 17: 891–921; English trans.: ‘On the Electrodynamics of Moving Bodies’, Translation by G. B. Jeffery and W. Perrett in *The Principle of Relativity*, London: Methuen and Company, Ltd. 1923.

El Skaf, Rawad and Cyrille Imbert, “Unfolding in the Empirical Sciences: Experiments, Thought Experiments and Computer Simulations”, *Synthese*, Vol 190, no. 16: 3451-3474. 2013.

Eldridge, Richard. “Form and Content: An Aesthetic Theory of Art.” *The British Journal of Aesthetics* 25, no. 4: 303–16. 1985.

Elgin, Catherine Z. “Fiction as Thought Experiment.” *Perspectives on Science* 22, no. 2: 221–241. 2014.

Elgin, Catherine Z. “Telling Instances”. *Beyond Mimesis and Convention - Representation in Art and Science*, edited by Roman Frigg and Matthew Hunter, Springer. 2010.

Elgin, Catherine Z. *True Enough*. MIT Press. 2017.

Fehige, Yiftach, and Michael T. Stuart. “Introduction to Special Issue of Perspectives on Science.” *Perspectives on Science* 22, no. 2: 167–178. 2014.

Franklin, Allan. *The Neglect of Experiment*. Cambridge University Press. 1986.

Franklin, Allan. The epistemology of experiment. *The uses of experiment*, edited by D. Gooding, T. Pinch, and S. Schaffer, 437–460. Cambridge: Cambridge University Press. 1989.

Franklin, Allan, and Ronald Laymon. *Measuring Nothing, Repeatedly: Null Experiments in Physics*. Morgan & Claypool Publishers. 2019.

Franklin, Allan. *Shifting Standards: Experiments in Particle Physics in the Twentieth Century*. University of Pittsburgh Press. 2013.

Franklin, L. R. "Exploratory Experiments." *Philosophy of Science* 72, no. 5: 888–99. 2005.

Franklin, Allan, and Slobodan Perovic. "Experiment in Physics." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Winter 2019. Metaphysics Research Lab, Stanford University. 2019. https://plato.stanford.edu/archives/win2019/entries/physics-experiment/.

Frappier, Melanie, Letitia Meynell and James R. Brown, eds. *Thought Experiments in Science, Philosophy, and the Arts* Routledge. 2012.

French, S., and P. Vickers. (2011). "Are There No Such Things as Theories." *British Journal for the Philosophy of Science* 62

French, Steven, and James Ladyman. "Reinflating the Semantic Approach." *International Studies in the Philosophy of Science* 13, no. 2: 103–121. 1999.

French, Steven. "Imagination in Scientific Practice." *European Journal for Philosophy of Science* 10, no. 3: 1–19. 2020b.

French, Steven. *There Are No Such Things as Theories.* Oxford University Press. 2020a.

Friend, Stacie. "Fiction as a Genre." *Proceedings of the Aristotelian Society* 112: 179–209. 2012.

Friend, Stacie. "II—Fictive Utterance and Imagining II." *Aristotelian Society Supplementary Volume* 85, no. 1: 163–80. 2011.

Friend, Stacie. "The Real Foundation of Fictional Worlds." *Australasian Journal of Philosophy* 95, no. 1: 29–42. 2017.

Frigg, Roman and James Nguyen, "Of barrels and pipes: representation - as in art and science", in Otávio Bueno, George Darby, Steven French & Dean Rickles (eds.), *Thinking about Science* *and Reflecting on Art: Bringing Aesthetics and the Philosophy of Science Together*. Routledge: 41-61. 2017.

Frigg, Roman and Matthew Hunter, eds. *Beyond Mimesis and Convention - Representation in Art and Science,* Springer. 2010.

Frigg, Roman, and James Nguyen. “The Fiction View of Models Reloaded.” *The Monist* 99, no. 3: 225–42. 2016.

Frigg, Roman, and Julian Reiss. “The Philosophy of Simulation: Hot New Issues or Same Old Stew?” *Synthese* 169, no. 3: 593–613. 2009.

Frigg, Roman. “Fiction and Scientific Representation”. *Beyond Mimesis and Convention - Representation in Art and Science*, edited by Roman Frigg and Matthew Hunter, Springer. 2010.

Galileo, Galilei. 1632. *Dialogue Concerning the Two Chief World Systems*, ed. and transl. Stillman Drake, 2nd revised edition. Berkeley: University of California Press. 1967.

Galileo, Galilei. 1638. *Dialogues Concerning Two New Sciences,* Macmillan. 1914.

Galison, P., *How Experiments End*. Chicago: University of Chicago Press. (1987).

Galton, Francis. “Statistics of Mental Imagery”, *Mind, 5,* 301-318. 1880.

Gaut, Berys. “Creativity and Imagination.” *The Creation of Art,* edited by Berys Gaut and Paisley Livingston. Cambridge University Press. 2003.

Gelfert, Axel. *How to Do Science with Models: A Philosophical Primer*. SpringerBriefs in Philosophy. Springer International Publishing. 2016.

Gendler, Tamar Szabó and John Hawthorne. *Conceivability and Possibility*. Clarendon Press. 2002.

Gendler, Tamar Szabó. “Alief in Action (and Reaction).” *Mind & Language* 23, no. 5 (2008): 552–85. 2008.

Gendler, Tamar Szabó. “Galileo and the Indispensability of Scientific Thought Experiment.” *The British Journal for the Philosophy of Science* 49, no. 3: 397–424. 1998.

Gendler, Tamar Szabó. “The Puzzle of Imaginative Resistance”, *The Journal of Philosophy*, 97(2): 55–81. 2000.

Gendler, Tamar Szabó. “Thought Experiments Rethought—and Reperceived.” *Philosophy of Science* 71, no. 5: 1152–63. 2004.

Gibson, John. “Cognitivism and the Arts.” *Philosophy Compass* 3, no. 4: 573–89. 2008.

Giere, Ronald N. “How Models Are Used to Represent Reality.” *Philosophy of Science* 71, no. 5: 742–52. 2004.

Giere, Ronald N. “Using Models to Represent Reality,” *Model-Based Reasoning and Scientific Discovery,* edited by L. Magnani, N.J. Nersessian, and P. Thagard. Dordrecht: Kluwer Academic/Plenum Publishers. 1999.

Gilbert, N. and K. Troitzsch, *Simulation for the Social Scientist,* Philadelphia, PA: Open University Press. 1999.

Goldman, Alan H. “Interpreting Art and Literature.” *The Journal of Aesthetics and Art Criticism* 48, no. 3: 205-14. 1990.

Gooding, David C. “What Is Experimental About Thought Experiments?” *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*: 280–290. 1992.

Goodman, Nelson. *Languages of Art,* Indianapolis and Cambridge: Hackett, 2nd edition. 1976.

Guala, Francesco. “Models, Simulations, and Experiments.” In *Model-Based Reasoning,* 59–74. Springer US. 2002.

Guin, Ursula K. Le. *The Left Hand of Darkness*. Ace Books. 1976.

Hacking, Ian. “Do Thought Experiments Have a Life of Their Own? Comments on James Brown, Nancy Nersessian and David Gooding.” *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association,* 1992: 302–308. 1992.

Hacking, Ian. *Representing and Intervening: Introductory Topics in the Philosophy of Natural Science*. Cambridge University Press, 1983.

Häggqvist, Sören. “A Model for Thought Experiments.” *Canadian Journal of Philosophy* 39, no. 1:55–76. 2009

Hardy, G. *A Mathematician’s Apology*. Cambridge: Cambridge University Press. 1940.

Horowitz, Tamara and Gerald J. Massey. *Thought Experiments in Science and Philosophy*. Rowman & Littlefield. 1991.

Howell, Robert J. “Google Morals, Virtue, and the Asymmetry of Deference.” *Noûs* 48, no. 3: 389–415. 2014.

Hume, David. 1777. “Essays Moral, Political, Literary (LF Ed.) - Online Library of Liberty.” Accessed August 24, 2020. https://oll.libertyfund.org/titles/hume-essays-moral-political-literary-lf-ed#lf0059_label_407

Humphreys, Paul. “Seven Theses on Thought Experiments”, *Philosophical Problems of the Internal and External Worlds,* University of Pittsburgh Press. 1994.

Hunt, Lester H. “Literature as Fable, Fable as Argument.” *Philosophy and Literature* 33, no. 2: 369–385. 2009.

Hutcheson, 1726. *An Inquiry into the Original of Our Ideas of Beauty and Virtue*. 2004. https://oll.libertyfund.org/titles/hutcheson-an-inquiry-into-the-original-of-our-ideas-of-beauty-and-virtue-1726-2004.

Ivanova, Milena and Steven French. *Aesthetics of Science: Beauty, Imagination and Understanding.* Routledge. 2020.

Ivanova, Milena. “Aesthetic Values in Science.” *Philosophy Compass* 12, no. 10. 2017a.

Ivanova, Milena. “Poincaré’s Aesthetics of Science.” *Synthese* 194, no. 7: 2581–2594. 2017b.

Jacobs, Christianne, Dietrich S. Schwarzkopf, and Juha Silvanto. “Visual Working Memory Performance in Aphantasia.” *Cortex*, The Eye’s Mind - visual imagination, neuroscience and the humanities, 105: 61–73. 2018.

John, Eileen. “Art and Knowledge”. *The Routledge Companion to Aesthetics*, 3rd ed., edited by Berys Gaut and Dominic Lopes, New York: Routledge. 2013.

John, Eileen. “Reading Fiction and Conceptual Knowledge: Philosophical Thought in Literary Context.” *The Journal of Aesthetics and Art Criticism* 56, no. 4: 331–48. 1998.

Johnson-Laird, P. N. Mental Models, Cambridge: MIT Press. 1983

Jones, Max and Tom Schoonen. 2018. “Embodied Constraints on Imagination.” *The Junkyard*. Accessed September 4, 2020. https://junkyardofthemind.com/blog/2018/8/19/embodied-constraints-on-imagination.

Kant, Immanuel. *Critique of the Power of Judgment*. Cambridge University Press, 2000.

Kapitsa, Peter. 1937. “Reminiscences About Professor Ernest Rutherford”, in A. Parry (ed.), *Peter Kapitsa on Life and Science*, New York: Macmillan. 1968.

Kenna, Margaret E. “Icons in Theory and Practice: An Orthodox Christian Example.” *History of Religions* 24, no. 4: 345–68. 1985.

Khalifa, Kareem. *Understanding, Explanation, and Scientific Knowledge*. Cambridge: Cambridge University Press. 2017.

Kind, Amy, and Peter Kung. “Introduction: The Puzzle of Imaginative Use.” *Knowledge Through Imagination*, edited by Amy Kind and Peter Kung. Oxford University Press. 2016.

Kind, Amy. “How Imagination Gives Rise to Knowledge.” In Perceptual Memory and Perceptual Imagination, edited by Fiona Macpherson and Fabian Dorsch, 227-246. New York: Oxford University Press. 2018.

Kind, Amy. “Imagining Under Constraints.” *Knowledge Through Imagination*, edited by Amy Kind and Peter Kung. Oxford University Press. 2016.

Kind, Amy. “The Heterogeneity of the Imagination.” *Erkenntnis* 78, no. 1: 141–159. 2013.

Kivy, Peter. “Hutcheson’s Idea of Beauty: Simple or Complex?” *The Journal of Aesthetics and Art Criticism* 50, no. 3: 243–45. 1992.

Klein, Martin J. “Maxwell, His Demon, and the Second Law of Thermodynamics: Maxwell Saw the Second Law as Statistical, Illustrated This with His Demon, but Never Developed Its Theory.” *American Scientist* 58, no. 1: 84–97. 1970.

Kuhn, Thomas. *The Structure of Scientific Revolutions*. Chicago: The University of Chicago Press. 1962.

Kuhn, Thomas. “A Function for Thought Experiments”. *The Essential Tension.* Chicago: The University of Chicago Press. 1977.

Lamarque, Peter. “Cognitive Values in the Arts: Marking the Boundaries.” *Contemporary Debates in Aesthetics and the Philosophy of Art*, edited by Matthew Kieran. Blackwell. 2006.

Lamarque, Peter. *The Philosophy of Literature*. Blackwell. 2009.

Leng, Mary. “Creation and Discovery in Mathematics.” In *Meaning in Mathematics*, edited by John Polkinghorne. Oxford University Press, 2011.

Lenhard, Johannes. “Thought Experiments and Simulation Experiments”. *The Routledge Companion to Thought Experiments*, edited by Michael T. Stuart, Yiftach J. H. Fehige, and James Robert Brown. Routledge. 2018.

Lennox, James. “Darwinian Thought Experiments: A Function for Just-so Stories.” *Thought Experiments in Science and Philosophy,* edited by G. Massey, T. Horowitz. 1991.

Levinson, Jerrold. “Defending Hypothetical Intentionalism.” *British Journal of Aesthetics* 50, no. 2: 139–150. 2010.

Levinson, Jerrold. *The Pleasures of Aesthetics: Philosophical Essays*. Cornell University Press. 1996.

Levy, Arnon, and Adrian Currie. “Why Experiments Matter.” *Inquiry: An Interdisciplinary Journal of Philosophy* 62, no. 9–10: 1066–1090. 2019.

Levy, Arnon. “Metaphor and Scientific Explanation.” *The Scientific Imagination*. Oxford University Press. 2020.

Levy, Arnon. “Modeling without Models.” *Philosophical Studies* 172, no. 3: 781–98. 2015.

Levy, Arnon. “Models, Fictions and Realism: Two Packages”. *Philosophy of Science,* 79 (5). 2012.

Liao, Shen-yi. “Imaginative Resistance, Narrative Engagement, Genre.” *Res Philosophica* 93, no. 2: 461–482. 2016.

Lusk, Greg. “Computer Simulation and the Features of Novel Empirical Data.” *Studies in History and Philosophy of Science Part A* 56: 145–52. 2016.

Mach, E. 1896. “Uber Gedankenexperimente”, *Zeitschrift für den Phys. und Chem. Unterr*. 10, 1–5 .Translated by W. O. Price, S. Krimsky: On thought experiments, Philosophical Forum 4(3), 446–457. 1973.

Matthews, Robert J. “Describing and Interpreting a Work of Art.” *The Journal of Aesthetics and Art Criticism* 36, no. 1: 5–14. 1977.

Maxwell, J. C. *A Theory of Heat.* London: Textbooks of Science. 1871.

McAllister, J. *Beauty and Revolution in Science.* Ithaca, NY: Cornell University Press. 1996.

McLoone, Brian. “Thumper the Infinitesimal Rabbit: A Fictionalist Perspective on Some ‘Unimaginable’ Model Systems in Biology.” *Philosophy of Science* 86, no. 4: 662–71. 2019.

Meskin, Aaron, Mark Phelan, Margaret Moore, and Matthew Kieran. “Mere Exposure to Bad Art.” *The British Journal of Aesthetics* 53, no. 2: 139–64. 2013.

Meynell, Letitia. “Getting the Picture: Towards a New Account of Scientific Understanding”. *The Aesthetics of Science; Beauty, Imagination and Understanding,* edited by Milena Ivanova and Steven French, Routledge. 2020.

Meynell, Letitia. “Images and Imagination in Thought Experiments.” *The Routledge Companion to Thought Experiments*, edited by Michael T. Stuart, Yiftach J. H. Fehige, and James Robert Brown. Routledge. 2018.

Meynell, Letitia. “Imagination and Insight: A New Account of the Content of Thought Experiments.” *Synthese* 191, no. 17: 4149–68. 2014.

Miščević, Nenad. “Mental Models and Thought Experiments.” *International Studies in the Philosophy of Science* 6, no. 3: 215–226. 1992.

Moran, R. “The expression of feeling in imagination.” *The Philosophical Review* 103, 75-106. 1994.

Morgan, Mary S. “Experiments Versus Models: New Phenomena, Inference and Surprise.” *Journal of Economic Methodology* 12, no. 2: 317–329. 2005.

Morgan, Mary S. “Experiments without Material Intervention: Model Experiments, Virtual Experiments, and Virtually Experiments” in Hans Radder (ed.) *Model-Based Reasoning. Science, Technology, Values,* New York: Kluwer Academic/Plenum Publishers, 2002

Morgan, Mary S. and Margaret Morrison, *Models as Mediators: Perspectives on Natural and Social Science,* Cambridge: Cambridge University Press. 1999.

Morgan, Mary S., and M. Norton Wise. “Narrative Science and Narrative Knowing. Introduction to Special Issue on Narrative Science.” *Studies in History and Philosophy of Science Part A*, SI: Narrative in Science, 62: 1–5. 2017.

Morrison, Toni. *Beloved.* Picador. 1988.

Nagel, Thomas. “What Is It Like to Be a Bat?” *Philosophical Review* 83, no. October: 435–50. 1974.

Nanay, B. “Multimodal Mental Imagery”. *Cortex: A Journal Devoted to the Study of the Nervous System and Behavior, 105,* 125–134. 2018.

Nersessian, Nancy J. “Conceptual Change: Creativity, Cognition, and Culture ” in J. Meheus and T. Nickles (eds), *Models of Discovery and Creativity*, Dordrecht: Springer, 127–66. 2009.

Nersessian, Nancy J. “In the Theoretician’s Laboratory: Thought Experimenting as Mental Modeling.” *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*: 291–301. 1992.

Nersessian, Nancy J. “Cognitive Science, Mental Modeling and Thought Experiments.” *The Routledge Companion to Thought Experiments*, edited by Michael T. Stuart, Yiftach J. H. Fehige, and James Robert Brown. Routledge. 2018.

Nersessian, Nancy J. “Thought Experimenting as Mental Modeling: Empiricism without Logic.” *Croatian Journal of Philosophy* VII, no. 20: 125–61. 2007.

Newton, I. *The Principia: Mathematical Principles of Natural Philosophy*, translated by I. B. Cohen and A. Whitman, Berkeley: University of California Press. 1999 [1687].

Nickles, Thomas, ed. *Scientific Discovery: Case Studies, Original Text*. Boston Studies in the Philosophy and History of Science. Springer Netherlands, 1980.

Norton, John D. “All Shook Up: Fluctuations, Maxwell’s Demon and the Thermodynamics of Computation.” *Entropy* 15, no. 10: 4432–83. 2013.

Norton, John D. “Are Thought Experiments Just What You Thought?” *Canadian Journal of Philosophy* 26, no. 3 (1996): 333–66.

Norton, John D. “The Worst Thought Experiment.” In *The Routledge Companion to Thought Experiments*, edited by Michael T. Stuart, Yiftach J. H. Fehige, and James Robert Brown. Routledge. 2018.

Norton, John D. “Thought Experiments in Einstein’s Work.” In *Thought Experiments In Science and Philosophy*, edited by Tamara Horowitz and Gerald J. Massey. Savage, MD: Rowman & Littlefield. 1991.

Norton, John D. “Why Thought Experiments Do Not Transcend Empiricism.” In *Contemporary Debates in the Philosophy of Science*, edited by Christopher Hitchcock, Blackwell. 2004.

Novaes, Catarina Dutilh. “The Beauty (?) Of Mathematical Proofs.” *Advances in Experimental Philosophy of Logic and Mathematics*, edited by Andrew Aberdein and Matthew Inglis, 63–93. London: Bloomsbury Academic. 2019.

O’Loughlin, Ian, and Kate McCallum. “The Aesthetics of Theory Selection and the Logics of Art.” *Philosophy of Science* 86, no. 2: 325–43. 2019.

Odenbaugh, Jay. “Semblance or Similarity?, Reflections on simulation and similarity”, Biol. Philos. 30: 277–291. 2015.

Palmerino, C. R. “Galileo’s Use Of Medieval Thought Experiments.” *Thought Experiments in Methodological and Historical Contexts*, January 1: 101–25. 2011.

Palmerino, Carla Rita. “Discussing What Would Happen: The Role of Thought Experiments in Galileo’s Dialogues.” *Philosophy of Science* 85, no. 5: 906–918. 2018.

Palmieri, P. “‘Spuntur lo scoglio più duro’: did Galileo ever think the most beautiful thought experiment in the history of science?”Studies in History and Philosophy of Science 36: 305–322. 2005.

Parke, Emily C. “Experiments, Simulations, and Epistemic Privilege.” *Philosophy of Science* 81, no. 4: 516–536. 2014.

Parker, Wendy. “Does matter really matter: Computer Simulations, experiments, and materiality”, *Synthese*, 169:483-96. 2009.

Parker, Wendy. “Franklin, Holmes, and the Epistemology of Computer Simulation.” *International Studies in the Philosophy of Science* 22, no. 2: 165–83. 2008.

Parsons, G. G., and A. Rueger. “The Epistemic Significance of Appreciating Experiments Aesthetically.” *British Journal of Aesthetics* 40, no. 4: 407–423. 2000.

Peterson, Eric. 2018. “Thought Experiments as a Kind of Genre.” *Junkyard of the Mind.* https://junkyardofthemind.com/blog/2018/5/21/thought-experiments-as-a-kind-of-genre. Accessed August 31, 2020.

Popper, Karl, *Objective Knowledge: An Evolutionary Approach,* Oxford: Clarendon Press, 1972.

Popper, Karl. 1934. *The Logic of Scientific Discovery*, London and New York: Routledge. 2002.

Post, Heinz. “Correspondence, Invariance and Heuristics.” *Correspondence, Invariance and Heuristics: Essays in Honour of Heinz Post*, edited by French, S., and H. Kamminga. Springer Science & Business Media. 2013.

Regt, Henk W. de. “Visualization as a Tool for Understanding.” *Perspectives on Science* 22, no. 3: 377–396. 2014.

Regt, Henk W. de. *Understanding Scientific Understanding*. Oxford Studies in Philosophy of Science. Oxford, New York: Oxford University Press. 2017.

Regt, Henk W. de., Sabina Leonelli, Kai Eigner (eds.), *Scientific Understanding: Philosophical Perspectives.* University of Pittsburgh Press, Pittsburgh. 2009.

Reichenbach, H., *Experience and Prediction. An Analysis of the Foundations and the Structure of Knowledge*, Chicago: The University of Chicago Press. 1938.

Reiss, Julian. “Causal Inference in the Abstract or Seven Myths about Thought Experiments.” *London School of Economics, Centre for Philosophy of Natural and Social Sciences, Causality: Metaphysics and Methods* Technical Report 03/03. 2002.

Reiss, Julian. “Thought Experiments and Idealizations”, *The Routledge Companion to Thought Experiments*, edited by Michael T. Stuart, Yiftach J. H. Fehige, and James Robert Brown. Routledge. 2018.

Rescher, Nicholas. *What if? Thought Experimentation in Philosophy*. New Brunswick, N.J./ London: Transaction Publishers. 2005.

Roush, Sherrilyn. “The Epistemic Superiority of Experiment to Simulation.” *Synthese* 195, no. 11: 4883–4906. 2018.

Salis, Fiora, and Roman Frigg. “Capturing the Scientific Imagination.” *The Scientific Imagination*, edited by Peter Godfrey-Smith and Arnon Levy. Oxford University Press. 2020.

Schickore, Jutta, and Friedrich Steinle, eds. *Revisiting Discovery and Justification: Historical and Philosophical Perspectives on the Context Distinction*. Archimedes. Springer Netherlands. 2006.

Schickore, Jutta. “Scientific Discovery.” In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Summer 2018. Metaphysics Research Lab, Stanford University. 2018.

Schindler, Samuel. “Theory-Laden Experimentation.” *Studies in History and Philosophy of Science Part A* 44, no. 1: 89–101. 2013.

Shelley, James. “Empiricism: Hutcheson and Hume.” *The Routledge Companion to Aesthetics*, 3rd ed., edited by Berys Gaut and Dominic Lopes, New York: Routledge. 2013.

Sherados, Benjamin and William Bechtel. “Imagining Mechanisms with Diagrams. *Scientific Imagination*, edited by Peter Godfrey-Smith and Arnon Levy. Oxford University Press. 2020.

Shinod, N. K. “Why Thought Experiments Do Have a Life of Their Own: Defending the Autonomy of Thought Experimentation Method.” *Journal of Indian Council for Philosophical Research* 34, no. 1: 75–98. 2017.

Sibley, Frank. “Aesthetic Concepts.” *Philosophical Review* 68, no. 4: 421–450. 1959.

Simons, P. ‘Wittgenstein on Surprise in Mathematics’. (Unpublished).

Sorensen, Roy A. “Thought Experiments and the Epistemology of Laws.” *Canadian Journal of Philosophy* 22, no. 1: 15–44, 1992b.

Sorensen, Roy. *Thought Experiments,* New York: Oxford University Press. 1992a.

Starikova, Irina, and Marcus Giaquinto. “Thought Experiments in Mathematics.” In *The Routledge Companion to Thought Experiments*, edited by Michael T. Stuart, Yiftach J. H. Fehige, and James Robert Brown, London: Routledge. 2018.

Stecker, Robert. “Art Interpretation.” *The Journal of Aesthetics and Art Criticism* 52, no. 2: 193–206. 1994.

Stecker, Robert. *Aesthetics and the Philosophy of Art: An Introduction*. Rowman & Littlefield Publishers, 2010.

Stock, Kathleen. “Sartre, Wittgenstein, and Learning from Imagination”, in Peter Goldie and Elisabeth Schellekens (eds.), *Philosophy and Conceptual Art,* Oxford University Press. 2006.

Stokes, Dustin. “Imagination and Creativity.” *The Routledge Handbook of Philosophy of Imagination,* edited by Amy Kind. Routledge. 2016.

Stuart, Michael T. “How Thought Experiments Increase Understanding.” In *The Routledge Companion to Thought Experiments*, edited by Michael T. Stuart, Yiftach J. H. Fehige, and James Robert Brown. Routledge. 2018.

Stuart, Michael T. “Norton and the Logic of Thought Experiments”. *Axiomathes* 26, no. 4: 451–466. 2016b.

Stuart, Michael T. “Taming Theory with Thought Experiments: Understanding and Scientific Progress.” *Studies in History and Philosophy of Science Part A* 58: 24–33. 2016a.

Stuart, Michael T., Yiftach Fehige, and James Robert Brown, “Thought Experiments: State of the Art”, In *The Routledge Companion to Thought Experiments*, edited by Michael T. Stuart, Yiftach J. H. Fehige, and James Robert Brown. Routledge. 2018.

Stuart, Michael T. “The Productive Anarchy of Scientific Imagination.” *Forthcoming.*

Suppes, Patrick. “A Comparison of the Meaning and Uses of Models in Mathematics and the Empirical Sciences”, *Synthese,* 12(2–3): 287–301. 1960.

Swan, L. S. “Synthesizing insight: artificial life as thought experimentation in biology”, *Biology & Philosophy,* 24(5). 2009.

Thagard, P. “Conceptual Combination and Scientific Discovery”, *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association,* (1): 3–12. 1984.

Thoma, Johanna. “On the Hidden Thought Experiments of Economic Theory.” *Philosophy of the Social Sciences* 46, no. 2: 129–146. 2016.

Thomson-Jones, Martin. “Missing Systems and the Face Value Practice.” *Synthese* 172, no. 2: 283. 2010.

Thomson-Jones, Katherine. “Inseparable Insight: Reconciling Cognitivism and Formalism in Aesthetics.” *The Journal of Aesthetics and Art Criticism* 63, no. 4: 375–84. 2005.

Todd, Cain S. “Unmasking the Truth Beneath the Beauty: Why the Supposed Aesthetic Judgements Made in Science May Not Be Aesthetic at All.” *International Studies in the Philosophy of Science* 22, no. 1: 61–79. 2008.

Todd, Cain. “Fitting Feelings and Elegant Proofs: On the Psychology of Aesthetic Evaluation in Mathematics.” *Philosophia Mathematica*. 2017.

Todd, Cain. “Imagination, Aesthetic Feelings, and Scientific Reasoning.” *The Aesthetics of Science; Beauty, Imagination and Understanding,* edited by Milena Ivanova and Steven French, Routledge. 2020.

Tooley, M. “The nature of laws.” *Canadian Journal of Philosophy*, 7: 667–98. 1977.

Toon, Adam. “Imagination in Scientific Modelling.” *The Routledge Handbook of Philosophy of Imagination,* edited by Amy Kind. Routledge. 2016.

Toon, Adam. *Models as Make-Believe - Imagination, Fiction and Scientific Representation,* Palgrace MacMillan UK, 2012.

Unruh, W. “Experimental Black-Hole Evaporation?”, *Physical Review Letters,* 46. 1981.

Vaihinger, H. 1911. *The Philosophy of As If*. Routledge, 2014.

Vickers, Peter. *Understanding Inconsistent Science*. Oxford University Press. 2013.

Vidmar, Iris. “Against the Cognitive Triviality of Art”. *Proceedings of the European Society for Aesthetics,* vol. 2. 2010.

Vorms, Marion. “Representing with imaginary models: Formats matter”. *Studies in History and Philosophy of Science Part A* 42 (2): 287-295. 2011.

Walton, Kendall. *Marvelous Images: On Values and the Arts*. Oxford University Press, 2008.

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

Watkins, Nicholas W. “(A)Phantasia and Severely Deficient Autobiographical Memory: Scientific and Personal Perspectives.” *Cortex*, The Eye’s Mind - visual imagination, neuroscience and the humanities, 105: 41–52. 2018.

Weatherson, Brian. 2010. “Surveys and Thought Experiments.” *Thoughts Arguments and Rants* (blog), http://tar.weatherson.org/2010/08/24/surveys-and-thought-experiments/. Accessed March 27, 2019.

Weinberg, Jonathan and Aaron Meskin. “Puzzling Over the Imagination: Philosophical Problems, Architectural Solutions”, in Nichols (ed.). 2006.

Weinberg, Jonathan M. “Configuring the Cognitive Imagination.” In *New Waves in Aesthetics*, edited by Kathleen Stock and Katherine Thomson-Jones, 203–23. New Waves in Philosophy. London: Palgrave Macmillan UK. 2008.

Weisberg, Michael. *Simulation and Similarity: Using Models to Understand the World*. Oxford University Press. 2013.

Wenzel, Christian Helmut. *An Introduction to Kant's Aesthetics: Core Concepts and Problems*. John Wiley & Sons. 2008.

White, Alan R. *The Language of Imagination*. B. Blackwell. 1990.

Wilkes, Kathleen. *Real People: Personal Identity without Thought Experiments.* Oxford University Press. 1988.

Williams, Bernard. “The Self and the Future.” *The Philosophical Review* 79, no. 2: 161–80. 1970.

Williamson, Timothy. Review of Levy and Godfrey Smith (eds.) *The Scientific Imagination.* 2020. https://ndpr.nd.edu/news/the-scientific-imagination-philosophical-and-psychological-perspectives/

Winsberg, Eric. “A Tale of Two Methods.” *Synthese* 169, no. 3: 575–92. 2009.

Winsberg, Eric. “Computer Simulations in Science.” In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Winter 2019. Metaphysics Research Lab, Stanford University, 2019. https://plato.stanford.edu/archives/win2019/entries/simulations-science/.

Winsberg, Eric. *Science in the Age of Computer Simulation*. University of Chicago Press. 2010.

Wise, M. Norton. “Science as (Historical) Narrative.” *Erkenntnis* 75, no. 3: 349–376. 2011.

Wittgenstein, Ludwig. *Remarks on the Foundations of Mathematics*. Blackwell.1978.

Wittgenstein, Ludwig. *Remarks on the Philosophy of Psychology,* II, edited by G.H. von Wright and H. Nyman Oxford: Blackwell, 1980

Yablo, Stephen. “Is Conceivability a Guide to Possibility?” *Philosophy and Phenomenological Research*, 53(1): 1–42. 1993.

Zeman, A., Dewar, M., & Della Sala, S. “Lives without imagery—Congenital aphantasia.” *Cortex: A Journal Devoted to the Study of the Nervous System and Behavior, 73,* 378–380. 2015.

Zeman, Adam, Fraser Milton, Sergio Della Sala, Michaela Dewar, Timothy Frayling, James Gaddum, Andrew Hattersley, et al. “Phantasia–The Psychological Significance of Lifelong Visual Imagery Vividness Extremes.” *Cortex*. 2020.