Alice Murphy

alicemlmurphy@gmail.com

The Aesthetic and Literary Qualities of Scientific Thought Experiments

1. Introduction: The Aesthetics of Thought Experiments

The discussion of aesthetic value in science has primarily focused on the evaluation of theories or of mathematical proofs (Breitenbach 2015, Ivanova 2017a, 2017b). Questions then arise regarding the role of these evaluations, such as whether they have an epistemic function, that is, as an indicator of the truth of the theory and/or whether they can aid understanding, as well as the usefulness of such value in science. One aspect of scientific practice that has an obvious aesthetic dimension, but is currently overlooked in the literature, is the use of thought experiments. Thought experiments are a popular device in science and philosophy that are used to justify, undermine, or clarify theories. Thought experiments take the form of a description of an imaginary scenario, followed by a judgement of what might happen if the scenario occurred in reality. We then draw conclusions in order to say something about other cases. Examples include Galileo's falling bodies, Thomson's violinist, Maxwell's demon, Schrödinger's cat, brain in a vat scenarios, and Einstein's use of thought experiments in the development of special and general relativity.

Thought experiments are often referred to as beautiful or elegant. Take, for example, Galileo's famous falling bodies thought experiment used to undermine Aristotle's physics, what Brown refers to as 'the most beautiful thought experiment ever devised' (2004, 24).[1] The thought experiment begins by considering Aristotle's theory that heavier bodies fall faster than light ones. Galileo asks us to imagine taking a heavy cannonball, and a lighter musket ball, and dropping them from a tower. According to Aristotle, the heavy ball will fall much faster than the lighter ball. But Galileo points out that if we then imagine attaching the two balls together,

---

[1] 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 (Crease 2002).

and then dropping them, Aristotle's theory leads us to a contradiction. Because it states both: 1. That the compound object will fall faster than the heavy ball on its own, because the compound object is heavier, and 2. That the compound object will fall slower as the lighter ball is inclined to fall slower than the heavy ball, and so it will drag it back—making it fall slower. Both of these predictions cannot be true, and so, Aristotle's theory is contradictory and has to be rejected. From this thought experiment, Galileo establishes a new theory that moving bodies fall at the same speed.

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 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 why are certain thought experiments beautiful or elegant? As Sibley points out in his influential paper on aesthetic concepts, 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).[2] 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

---

[2] It is important to note that Sibley argues that aesthetic terms cannot be defined in terms of non-aesthetic concepts. 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).

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). A clear feature that is 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.

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, 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 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 aesthetically 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 the experience of chasing a beam of light.

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' (409). Since the nineteenth-century, 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 is a move away from admiring the workings of nature which can be done irrespective of whether we have a grasp of the theoretical framework

involved. Parsons and Rueger note that since this shift, a common way of characterising aesthetically pleasing 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' (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, where the potential heuristic fertility of thought experiments is emphasised. For Brown, 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. Thought experiments, of course, differ here; economy is achieved through the particulars that we are prescribed to imagine.[3]

To further illuminate these features in the case of thought experiments, it might also help to contrast these with cases of thought experiments that do not fulfil their function, and 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.

---

[3] Bringing thought experiments into the aesthetics of science discussion leads to some further considerations. So far I have presented reasons to suggest that the aesthetic 'canons' we form for theories, experiments and thought experiments will be the same in that each generate a positive aesthetic response if they are elegant and simple. But there may be reasons to argue that a different set of aesthetic properties could be valued in the three contexts. Further, we can ask whether the force of the aesthetic qualities of thought experiments change from era to era, throughout the history of science. On McAllister's (1996) view, in theory change our aesthetic canons change and consequently what counts as beautiful will be revised. Finally, I expect there to be variation with regards to the force of aesthetic qualities of a thought experiment according to the role it is playing in science.

Szilard's thought experiment involves a misuse of idealisations; 'an inconsistent muddle of improper idealizations', and leads to an incorrect generalisation (2018, 466). Another example is Darwin's whale thought experiment, which attempts to explain natural selection by demonstrating how whales could have evolved from bears. This was described by Louis Agassiz as "truly monstrous", and 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 (Stuart 2016, 31).

Now, it could be argued that when a thought experiment is described as beautiful, or when Darwin's whale is described as monstrous, what is really being said is that the thought experiment is successful or unsuccessful.[4] This worry has been raised in the context of the aesthetics of theories by Todd who 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, 77). Any account of aesthetics in science will have to provide reasons as to why we should take these descriptions as genuinely aesthetic (for discussion and response, see O'Loughlin and McCallum 2019). Another concern is beautiful theories that turn out to be false, or ugly theories (or at least, ones that are not considered to have aesthetic merit) that are successful, such as the standard model of particle physics (Ivanova 2017a). There may be problems, then, with regards to the epistemic value of aesthetic qualities when it comes to theories. There is clearly more to be said here, but I want to suggest that the function of a thought experiment is closely tied to aesthetic features.[5] The value of thought experiments is that they provide us with a scenario that makes something complex easy to visualise and to grasp. And so, the aesthetic virtues of thought experiments are significant in that a thought experiment that is cluttered or clumsy would be less useful or insightful due to the role that thought experiments play in science, in particular, their role in understanding and pedagogy.

---

[4] I outline Norton's view that thought experiments are arguments below. Given his account, it appears 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 their role in contributing to an argument and consequently, they are not genuinely aesthetic.

[5] This may also apply to concrete experiments in that a complex experiment offers more possibilities for being contested. The aesthetics of science may offer a further way to explore connections between thought experiments and ordinary, concrete experiments.

The beauty of thought experiments appears to lie in their ability to evaluate, explain or help us understand something profound based on a simple scenario. Furthermore, beautiful or elegant thought experiments are those that have carefully selected particulars. This leads us to the two ways we can think about the aesthetic dimension of thought experiments. The first, a narrow sense, is to do with their assessment as "beautiful" or "elegant" as presented above. In this sense, as we have seen, an aesthetic evaluation of thought experiments may be similar to that of theory and experiment with regards to the "optimal use of minimal material", that is, simplicity combined with significant consequences. The second, a broader sense, also includes the features that thought experiments have in common with artworks. The rest of the paper takes the aesthetics of thought experiments in this second, broader sense. In particular, I look to the similarities thought experiments share with works of literary fiction. Despite what has been indicated so far, this second way highlights the ways in which thought experiments are distinct from experiments and theories with regard to their aesthetic qualities.

2. Literature as Thought Experiment

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 this is inessential—and narrative form.[6] Further, the use of thought experiments in learning has been offered as a way of defending the cognitive value of literature. 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. The interesting issue is whether artworks can teach us in a way that is less constrained than this. Describing fiction 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 Graham 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

---

[6] 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.

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, which is similar to the role of thought experiments in addressing questions about our conceptual schemes (1998, 332).

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. It is 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 concrete experiment, thought experiment, and literary fiction. 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 the Galileo example presented in the last section 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 experiment, 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. Newton is taken to be describing an actual

experiment in *The Principia*, but in order for the theoretical conclusion regarding the existence of absolute space to follow, there are 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'. 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' (2014, 232). As with thought experiments, works of literature differ in terms of how much, and in what way, they depart from reality.

And so, there are views that maintain that (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. Here, I put to one side the question of whether literary works can be thought experiments, and instead want to think about how the comparisons can be drawn the other way as well, that is, from aesthetic and philosophy of art in order to illuminate the science cases. 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. Now I'm going to consider accounts that argue this way. This set of views claim that there are disanalogies between the art and science cases that undermine purported connections between how we learn from scientific and artistic representations, and the role of aesthetic considerations in science. In the next section, I consider two disanalogies between literature and thought experiments. The first is to do with the importance of the formulation of literary

works, and the second is to do with the ways in which we interpret artistic fictions on one hand, and literary works on the other.

3. Disanalogies Between Literature and Thought Experiment

a) Formulation

Any account that discusses the aesthetic or literary qualities of thought experiments is going to face opposition from Norton. Norton analyses thought experiments as arguments, insisting that they work by transforming our existing knowledge through a logical process. He 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). For Norton, this view is the only alternative to the claim that thought experiments, in virtue of providing us with knowledge about the world without new empirical data, are cases of "epistemic magic". Norton has in mind Brown's platonist account, whereby some thought experiments work by giving us access to the laws of nature that exist in a platonic realm (Brown 2011, 2004). 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 aesthetic qualities are irrelevant to the conclusion and therefore dispensable.

Egan's view is less strict than Norton's in that he is not committed to the claim that thought experiments are arguments, or that the actual conduct of a thought experiment consists in the execution of an argument. But for Egan, there is an essential connection between scientific and philosophical thought experiments and argument that leads to a crucial disanalogy between thought experiments and literary works. Thought experiments are always a part of a larger, argumentative structure.[7] Otherwise, Egan states, they would be merely 'intriguing narratives' (2016, 142). Egan allows that a work of literature could be used as a thought experiment, that is, a philosopher could use a work of literary fiction or sections of those works as part of a thought experiment or an argument. The same excerpt can function differently in put to work in this way, for example, a work of philosophy that argues in favour of the importance of freedom in our choices even when this may lead to us making morally reprehensible decisions might cite an excerpt from Burgess's *A Clockwork Orange.* However, in a case like this, Egan

---

[7] A counterexample to this 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.

stresses that the work is no longer being read as literature. Instead, we read it with a particular purpose in mind, that is, as part of an argument for a philosophical position (2016, 143).

Further to this, Egan claims that the purpose of a thought experiment 'is *exhausted* in making or contributing to an argument'. Like Norton, Egan maintains that any aesthetic virtues of a thought experiment narrative are irrelevant to the purpose of a thought experiment; 'we can, as it were, throw away our thought experimental narrative once its work is done' (2016, 142). Because of this, Egan claims that 'their role in arguments is also fungible'. In the case of say, Nozick's experience machine, a different thought experiment could have been used in its place if it did the same argumentative work—nothing in the argument against hedonism requires the story of the experience machine. Whereas we do not approach literature in this way. We do not 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. This means that literary works are not replaceable in the way 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 equally amusing story could be in its place—he states, 'the uniqueness of just *this* story remains' (2016, 143).

For Egan, a key part of this is that unlike literary works, thought experiments are not concerned with particulars. When presenting the experience machine thought experiment, Nozick (1974) is not concerned with the idea of floating in an experience machine with electrodes attached to our heads, nor is Galileo interested in musket balls and cannonballs. These particulars are a means to an end: they serve as a way of exploring more abstract problems. 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'. Literary works can of course prompt more abstract or general reflections, but these (compared to thought experiments) have not got a defined limit and so, we cannot abstract way from the 'concrete particularities of [for example] Kafka's narrative if we want to think about it as a work of literature' (2016, 144). In addition, Egan states that literature does not have to prompt this type of reflection, '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).

The issue here appears to be about what Stecker (2010) calls the unique value of artworks, and the indispensability and irreplaceability in the case of literary works which Egan contrasts with thought experiments. The aesthetic experience produced by say Kafka's *Metamorphosis* could not be produced by anything else, and so it cannot be replaced by another work of art of comparable worth. Further, the work is not dispensable once we have read it, that is, it is not a means to an end. Rather, we come back to artworks with the expectation that we can gain something new from them and that they can offer new experiences. Consequently, they offer new value each time.

For Egan and Norton, then, the role of narrative, aesthetic qualities and use of particulars merely make a thought experiment more interesting or compelling. As Norton summarises, 'it is 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 visualised can be supplied in many ways that will not affect the outcome' (2004, 60).[8]

b) Interpretation

The first disanalogy between thought experiments and literary works emphasised the importance of the format of artistic representations. A second, related disanalogy that has been raised focuses on the ways in which we interpret artworks such as literature 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, where representation consists in denotation, exemplification, keying-up and imputation—to both scientific and artistic cases. However, they 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 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). In works of literature, the interpretation is not fixed and attending carefully to the work and its features in order to come

---

[8] These thoughts are echoed in Currie's discussion of the comparison between scientific models and literary fiction: '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 aims' (2016, 305).

up with interesting and sometimes conflicting interpretations is part of engaging with and appreciating artistic works.

Similarly, when it comes to thought experiments, Hacking argues that in contrast with experiments, thought experiments do not have a life of their own. He states: 'I think of [concrete] experiments as having a life: maturing, evolving, adapting, being not only recycled but also, quite literally being retooled. But thought experiments are fixed, largely immutable' (1992, 307). By this he means that thought experiments do not evolve—they cannot be adapted or reworked for different purposes. For Hacking, this feeds into the epistemic privilege of experiment over thought experiment. In the next section, I address these arguments, and argue for the importance of the literary qualities of thought experiments.

4. The Literary Qualities of Thought Experiments

I agree that there are significant differences between our engagement with art on one hand, and our engagement with science on the other, and 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 above claims. I begin by addressing the importance of formulation in thought experiments before demonstrating how thought experiments can be interpreted in different ways.

a) Formulation and Concrete Particulars

As seen, Norton states that thought experiments can be reconstructed as arguments without any epistemic loss. I argue that while we can, of course, rationally reconstruct thought experiments into argument form, this will lead us to miss important features involved in their practice, and this is what I am interested in examining. What appears to be implicit in Norton and Egan's view is a distinction between the context of discovery and the context of justification. The idea would be that the narrative form of a thought experiment and their use of particular elements (the use of ordinary or familiar objects in thought experiments, for example, Newton using buckets and ropes to show the existence of absolute space) may play a role in the context of discovery, that is, in the conduct of a thought experiment, but it is limited to this—they have no part to play in the context of justification.

There are two ways of thinking about this distinction in the case of thought experiments: the first is to do with what went on in the mind of the scientist who initially conducted the thought

experiment, which brought them to a new insight. Take, for instance, Einstein's recollection in his autobiography, of him imagining chasing after a beam of light when he was 16, which was an important part of the development of special relativity as mentioned in footnote 6. We can label this a private thought experiment. A second way to think about thought experiments is at the stage in which they have been presented to a public or a scientific community in order to communicate an idea—such as in the Galileo and Newton cases presented above. In these cases, the context of discovery can be taken to embrace the cognitive processes of agents when performing a thought experiment, and what the thought experiment asks us to do. It is the latter that I focus on.

The discovery/justification distinction has received a lot of attention, and many have undermined attempts to draw a clear-cut distinction or have argued that the context of discovery is a legitimate topic for philosophers of science (for an overview see Schickore 2018). The distinction appears especially difficult to apply in the case of thought experiments, as it is in the performance of a thought experiment that they get their justificatory force. Against Norton's reconstruction thesis—that thought experiments can be reconstructed as arguments without epistemic loss—Gendler has shown that the demonstrative force of the thought experiment depends upon its formulation, that is, its narrative form and reference to concrete particulars, making these features indispensable. We can see this in the case of Galileo's falling bodies thought experiment. There are various logical "ways out" for the Aristotelian (who holds that heavier bodies fall faster than lighter ones) that are not available when the thought experiment is presented in its original, unreconstructed form. For example, considering the difference between the two bodies attached together as a new, unified object, compared to seeing it as a composite object made of two objects. Gendler states in the argument version, we lose 'the way in which, by evoking tacit knowledge about how the falling bodies actually behave, the thought experiment pre-emptively precludes such ways out' (1998, 407).

We can see the importance of formulation in other scientific representations, such as scientific models. Frigg and Nguyen (2017) argue that the 'very same model, when presented under a different format, can yield different predictions and offer different explanations. Formulation matters' (2017, 58). They cite the work of Vorms who argues that different formats of the same information impacts agents' reasoning processes, and that different formats allow access to different information. One case Vorms discusses is representing the results of a temperature survey in different ways. While it is the same information in each representation, 'the map

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'. Whereas drawing this from a different representation of the same information, say a list of numerals, would involve many inferential steps (2011, 289).

Given this focus on the practical features of representing in the scientific context, it is difficult to see the force of the claim that the particulars do not matter in the case of thought experiments, and that they can be dispensed with. It is through the introduction of the particulars and ordinary and/or familiar objects—Newton using buckets and ropes, Galileo towers and balls and so on—that we engage with the thought experiment, and therefore come to understand what the thought experiment and the relevant theorising is about. The particularities are important and carefully chosen. Thought experiments are devices of the imagination, and depend upon our imaginative capacities. The use of objects give our imagination something to latch onto, to help us work through the problem. In Galileo's case, it is the two balls attached together and imagining their behaviour that is an integral part of the presentation.[9] Furthermore, if we look at the context in which *The Two Dialogues* was written, Galileo was writing to appeal to not only a scientific, but a public community. The particulars used are suited to that audience, and enhance the accessibility of the thought experiment scenario, consequently contributing to the cognitive force of the thought experiment.

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 we have seen, Egan claims that in the case of literature, 'the concrete elements of the narrative remain irreducibly a part of our imaginative engagement' (2016, 144). They are ineliminable and irreplaceable in that they are not a means to an end—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' but when treating it as a work of literature, it is not exhausted or replaceable in this way (2016, 143). It seems true that there is a particular experience of reading a work of Tolstoy that cannot be had another way, and that the details of

---

[9] Similarly, for Einstein's 1905 On the Electrodynamics of Moving Bodies—the theory is formulated in terms of the behaviour of clocks and rods; this is what we are invited to imagine (French 2018).

a thought experiment could be changed to some degree without impacting its effectiveness, but there are a couple of worries here. Firstly, it appears difficult to identify exactly what changes in the features of a thought experiment are permissible without altering the thought experiment, or its appeal and force and therefore usefulness in scientific practice. Further, it appears that at least some of the details of a work by say, Tolstoy could be changed without altering the novel, or effecting the interpretations we draw from the novel.

Secondly, we can ask whether all artworks uniquely valuable in this way, that is, are they all irreplaceable and indispensable? Stecker is critical of the irreplaceability claim, he states that this 'is much less likely to be true for a certain fishing rod model, but it is also not true for many lesser works. In both cases, these sorts of things are constantly going out of existence or becoming unavailable without a great loss of value in the world' (2010, 89). It seems plausible that some artworks could be replaced, or that some elements of artworks could be replaced, without a change in aesthetic value and/or experience, and so a sharp divide between thought experiments and literary works cannot be drawn in virtue of this. For example, generic, formulaic pieces of music or novels, or certain forms of kitsch art. In the section 5, I suggest we ought to look closely at the types of literary fictions used when drawing analogies and disanalogies with thought experiments.

b) Flexibility of Interpretation

The second issue I will offer a response to is to do with the flexibility of interpretation in artistic and scientific representations as raised by Frigg and Nguyen. Firstly, we can address this on the philosophy of art side. Frigg and Nguyen's example of interpretation in the scientific context is highly regimented, someone 'who doesn't interpret the large sphere as the sun simply doesn't understand the Newtonian model' is true, but this seems too simple and has obvious parallels in the interpretation of artworks. If I don't take the *Mona Lisa* to be a representation of a woman, or at least of a person, then I have clearly misunderstood the painting. Here, it might help to appeal to a distinction between a description of a work (or representation more generally), and an interpretation of that work (Matthews 1977). This distinction can apply to both the art and the science cases. That the largest sphere is the sun is part of the description of the model, similarly, that the *Mona Lisa* is a portrait is part of the description of the painting. This differs from offering an interpretation of the representation. In addition, there is huge debate in the philosophy of art regarding the interpretation of artworks, and whether there are several correct or acceptable interpretations of a work (pluralism), or that there is a single

correct or acceptable interpretation (monism). 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 (for discussion, see Stecker 1994).

We can also address this on the philosophy of science side, and show that thought experiments are not limited to a single interpretation. As Bokulich shows with the case of the Einstein-Podolsky-Rosen thought experiment which asked whether quantum mechanics is complete, 'thought experiments are no more bound to any one particular theory than ordinary physical experiments are' (2001, 293). There can be disagreement regarding what exactly a thought experiment shows and it can be analysed from the perspective of different theories, in this case, quantum mechanics and hidden variable theories. The same goes for the Newton's bucket outlined in section 2. Mach presents an alternative analysis of the thought experiment, stating that it 'does not establish the existence of absolute space or absolute motion, 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 2017). This demonstrates his questioning of Newton's omission of certain particulars that Mach took to be crucial—namely the fixed stars, which provides the relata within his Leibnizian conception of space. Further, some thought experiments are still being revised and in some cases, their impact remains a contentious issue. An example is Einstein's clock in the box thought experiment that sets out to falsify Heinsenberg's uncertainty principle. It remains unclear how the thought experiment will be resolved; it 'continues to be debated in the same context as it was originally presented, namely, concerning whether and how the uncertainty principle maintains itself in the face of certain possible experimental arrangements', and different formulations are still provided (Stuart 2016, 28).

While the interpretation of thought experiments has limits, it is underdetermined— there can be disagreement on what would happen in the scenario presented, or what conclusion should be drawn. And this is part of thought experiments' value in science. Further, considering alternative interpretations of what phenomena would occur in a thought experiment setup, or what the thought experiment actually demonstrates can also be a part of the style of presenting thought experiments. We see this in Galileo's dialogues where differences amongst the interlocutors' interpretations are presented, and these interpretations depend on their different theoretical commitments (Palmerino 2018).

5. Selecting the Right Examples

I want to end by suggesting that we ought to look closely at the literary examples used when thinking about the qualities that thought experiments share with works of literature. If we have in mind the likes of Tolstoy, then similarities are going to be thin. Further, such works are not representative of all narrative art. Perhaps some short stories would be a better comparison, particularly those that are revolved around a certain theme or a clear idea. One example is the 2010 short story collection *Machine of Death* where each story centers on the exploration of the same premise; a device that identifies how a person will die based on an analysis of a blood sample.[10] Further, Cameron's discussion of speculative fiction as moral and metaphysical thought experiment shows that works of literary fiction vary with regards to their similarities with thought experiments. Unlike works of 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:

> '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 that realistic fiction by its very nature cannot, since the latter must inevitably not focus on totalitarianism *per se* but rather totalitarianism *within* the broader social-historical context that is the setting of the novel' (Cameron 2015, 33).

And so, we should be careful when selecting our points of comparisons when discussing the aesthetic qualities of thought experiments; some literary fictions will be more relevant than others. One form of short literary work that provides a useful comparison are fables and parables. Take an example of a fable from Aesop (from Hunt 2009, 370):

> 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

---

[10] Thanks to Jamie Cawthra for directing me to this example.

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."

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 persuade 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. 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.

Given these similarities, we can think about the connection between thought experiments, and fables and parables through genre. Genres are categories of artworks, and a work being in a particular genre influences the way in which it is experienced, understood and evaluated by the reader or audience (Friend 2012). The idea of thought experiments as a genre has been explored by Weinberg:

'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. And there are surely other, and more subtly articulable, rules for the proper performance of thought experiments still to be detailed' (Weinberg 2008, 214).

In section 1, we saw through Norton's discussion of Szilard's thought experiment, "the worst thought experiment", that there are certain conventions involved in the creation and engagement of thought experiments with regards to idealisations and generalisability. Moreover, when teaching students how to engage 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. Given the similarities presented, we could hold that thought experiments are a sub-genre of a larger genre that also includes fables and parables. 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", say. There is a shared convention to how we engage with these stories, they have a point or moral, and the intended point guides the reader's interpretation of the story.[11]

In the scientific context, Cartwright (1991, 2010) has drawn parallels between models on one hand, and fables and parables on the other: 'Fables transform the abstract into the concrete, and in so doing, I claim, they function like models in physics…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 described scenario, we may have to "climb up the ladder of abstraction", that is, express the conclusion or the moral in more abstract terms. The relation between moral and the fable (or a model/thought experiment and its conclusion) is 'that of the general to the more specific' and the moral is "fitted out" by the fable. That is, '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). In another case, the "fitting out" of the abstract moral will be very different from say, the sun and the wind in Aesop's fable above, or falling bodies from a tower in the Galileo thought experiment.

A key difference between parables, such as the good Samaritan, and fables such as Aesop's fables, 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

---

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

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 is flexible, that theoretical commitments alter the conclusions drawn, and that there is debate around what exactly a thought experiment shows. In Cartwright's terms, the moral of a thought experiment or parable is not part of the parable itself; outside work needs to be done to show which "ladder of abstraction" must be climbed to reach the result that can be generalised.

6. Conclusion

To sum up, we have seen that some have drawn comparisons between thought experiments in science and philosophy as a way of defending the cognitive value of literature. Here, thought experiments have been utilised to explore connections between scientific and artistic representations in a different way: thought experiments in science are a good case study for thinking about aesthetic features in the scientific context. I hope to have shown that when thinking about thought experiments in science, the difference between representations in art and science is not as stark as some have made it out to be, and that science is more heterogeneous than has been allowed. The differences between scientific and artistic representations raised in current discussions fail to adequately account for the use of thought experiments in scientific practice, and part of their value in this context includes the qualities that they share with literary works.

Bibliography

Bokulich, Alisa. "Rethinking Thought Experiments." *Perspectives on Science* 9, no. 3 (2001): 285–307.

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)

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

Brown, James Robert. "Why Thought Experiments Transcend Experience." In *Contemporary Debates in Philosophy of Science*, Blackwell, (2004)

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

Budd, Malcom. *Values of Art: Pictures, Poetry Music*. London: The Penguin Press, (1995)

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

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

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

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)

Currie, Gregory. "Models As Fictions, Fictions As Models" The Monist 99, no. 3 (2016): 296–310.

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

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

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

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

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

French, S., "Imagination in Art and Science" presentation as part of *Imagination in Science* symposium, PSA biannual meeting, Seattle, (2018)

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

Frigg, R. and J. 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, London and New York (2017): 41-61.

Gendler, Tamar Szabó. "Intuition, Imagination, and Philosophical Methodology", *Oxford University Press,* (2010).

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

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 (1992): 302–308.

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

Ivanova, Milena. "Aesthetic Values in Science." *Philosophy Compass* 12, no. 10 (2017a)

Ivanova, Milena. "Poincaré's Aesthetics of Science." *Synthese* 194, no. 7 (2017b): 2581–2594.

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

Kapitsa, Peter. "Reminiscences About Professor Ernest Rutherford" [1937], in A. Parry (ed.), *Peter Kapitsa on Life and Science,* New York: Macmillan, (1968): 75–99

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

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

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

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

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

Norton, John D. "Why Thought Experiments Do Not Transcend Empiricism." In *Contemporary Debates in the Philosophy of Science*, edited by Christopher Hitchcock, Blackwell, (2004): 44–66.

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)

Nozick, R. *Anarchy, State and Utopia*, Blackwell, (1974)

O'Loughlin, Ian, and Kate McCallum. "The Aesthetics of Theory Selection and the Logics of Art." *Philosophy of Science* 86, no. 2 (2019): 325–43.

Palmerino, Carla Rita. "Discussing What Would Happen: The Role of Thought Experiments in Galileo's Dialogues." *Philosophy of Science* 85, no. 5 (2018): 906–918.

Parsons, G. G., and A. Rueger. "The Epistemic Significance of Appreciating Experiments Aesthetically." *British Journal of Aesthetics* 40, no. 4 (2000): 407–423.

Schickore, Jutta. "Scientific Discovery." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Summer 2018. Metaphysics Research Lab, Stanford University, 2018. https://plato.stanford.edu/archives/sum2018/entries/scientific-discovery/.

Sibley, Frank. "Aesthetic Concepts." *Philosophical Review* 68, no. 4 (1959): 421–450.

Stecker, Robert. "Art Interpretation." *The Journal of Aesthetics and Art Criticism* 52, no. 2 (1994): 193–206.

Stecker, Robert. "Do All Valuable Artworks Possess Aesthetic Value?." *Annales Philosophici* 1, no. 1 (2010)

Stuart, Michael T. "Taming Theory with Thought Experiments: Understanding and Scientific Progress." *Studies in History and Philosophy of Science Part A* 58 (2016): 24–33.

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)

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 (2008): 61–79.

Vorms, Marion. "Representing with imaginary models: Formats matter". *Studies in History and Philosophy of Science Part A* 42 (2) (2011): 287-295.

Weatherson, Brian. Accessed March 27, 2019. "Surveys and Thought Experiments." *Thoughts Arguments and Rants* (blog), August 24, 2010. http://tar.weatherson.org/2010/08/24/surveys-and-thought-experiments/.

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)