Organon F 31 (1) 2024: 2–21
https://doi.org/10.31577/orgf.2024.31101

ISSN 2585-7150 (online)
ISSN 1335-0668 (print)

RESEARCH ARTICLE

A Systematic Account of the Argumentative Role of Thought Experiments

Michael Agerbo Mørch* – Atle Ottesen Søvik**

Received: 10 October 2023 / Revised: 10 November 2023 / Accepted: 13 November 2023

Abstract: What is the role of thought experiments in scientific exploration? Can they provide us with new knowledge about the world? In a recent article, Lorenzo Sartori argues that thought experiments function like ordinary (material) experiments: Both material experiments and thought experiments are made in a specific context, which must then be extrapolated and generalized to say something true about the world. This article discusses and criticizes Sartori's proposal. It suggests a new theoretical framework for understanding thought experiments, their argumentative role, and how they provide new knowledge about the world. The framework presented is a coherentist framework, where coherence has three aspects: consistency, cohesiveness, and comprehensiveness. The proposal is that the argumentative role of thought experiments is to demonstrate the presence or absence of consistency, cohesiveness, and comprehensiveness, thereby strengthening a theory, weakening a theory, or showing one theory to be better than another. This is the way thought experiments

* Fjellhaug International University College
iD https://orcid.org/0000-0002-5712-208X
✎ Leifsgade 33.6, 2300 Copenhagen S, Denmark
✉ mam@dbi.edu

** Norwegian School of Theology, Religion and Society
iD https://orcid.org/0000-0002-8616-7105
✎ MF Norwegian School of Theology, P.O. Box 5144 Majorstua, 0302 Oslo, Norway
✉ Atle.O.Sovik@mf.no

provides new knowledge about the world, since the way we learn something new about the world is by discovering which theories about the world are most coherent.

Keywords: Coherence theory; Lorenzo Sartori; Scientific epistemology; Thought experiments.

1. Introduction

A thought experiment is an imagined scenario, often presented in the form of a narrative, conducted in mind with functions similar to scientific experiments but without collecting new empirical data from the world. But what is the role of thought experiments in scientific exploration? Can they provide us with new knowledge about the world? These are old and fascinating questions that are still discussed in the philosophical literature.

In a recent article, Lorenzo Sartori (Sartori 2023) argues that the discussion on this topic lacks an overarching theoretical framework. He provides a useful classification of positions but claims that no one has succeeded in giving a clear answer on how thought experiments provide us with new knowledge about the world. He then presents his own theory, which is that thought experiments function like ordinary (material) experiments, which we can see by distinguishing between internal and external validity. Both material experiments and thought experiments are made in a specific context, which must then be extrapolated and generalized to say something true about the world.

In this article, we first present an overview of the debate and Sartori's position before criticizing it. We propose an alternative theoretical framework for understanding thought experiments, their argumentative role, and how they provide new knowledge about the world. The framework we present is a coherentist framework, where coherence has three aspects: consistency, cohesiveness, and comprehensiveness. The proposal is that the argumentative role of thought experiments is to demonstrate the presence or absence of consistency, cohesiveness, and comprehensiveness, thereby strengthening a theory, weakening a theory, or showing one theory to be better than another. We argue that this is the way thought experiments provides new knowledge about the world, since the way we learn something new about the world is by discovering which theories (broadly understood) about the world are most coherent.

2. An overview of the debate and Sartori's position

Lorenzo Sartori provides a useful overview of the philosophical debate on thought experiments, based on the following question from Thomas Kuhn: Do thought experiments provide us with new knowledge about the empirical world? If so, how do they do that when no observation is involved? If not, why not? (Sartori 2023, 2; Kuhn 1977).

Sartori uses Kuhn's question to categorize various positions into a yes-camp and a no-camp. Many answer yes because thought experiments have been so important in the history of science, for example, with Galileo, Newton, Einstein, and others. (Sartori 2023, 2) Sartori presents three different answers to how thought experiments provide us with new knowledge – Platonism, objectualism, and structuralism (Sartori 2023, 3).

Platonism is represented by James Brown (Brown 2004). He argues that thought experiments allow us to "see" abstract laws and structures that apply to the world through a priori intuitions. Objectualism is represented by Tamar Gendler and Nenad Miscevic (Gendler 2004; Miscevic 1992). They envision thought experiments as objects or images that allow us to see the world in a different way than through propositions. Structuralism is represented by Nancy Nersessian (Nersessian 1992). Her view is that thought experiments are a type of simulative model-based reasoning that reveals structural analogs to reality (Sartori 2023, 3).

Sartori raises objections to the three positions. Platonism is mysterious in its answer to how thought experiments work. Objectualism fails to explain why we gain new or different knowledge by thinking about objects instead of propositions. The problem with structuralism is that structural analogies can be either wrong or right. But then it seems that thought experiments do not help unless you already know the structures of reality (Sartori 2023, 3-4).

Other philosophers have argued that thought experiments do not provide us with new knowledge about the world (the no-camp). According to Daniel Dennett, thought experiments do not teach us anything new about the world but rather stimulate our intuition ("intuition pumps") (Dennett 1996). Ian Hawking believes that thought experiments can reveal inconsistency but do not tell us anything new about the world (Hacking 1993; Sartori 2023, 4). But, one could ask, why have thought experiments been so important in the history of science if they do not teach us anything new about the world?

A more detailed answer in the no-camp regarding why thought experiments do not teach us anything new comes from John Norton. He says we acquire knowledge through observation and logic, and since thought experiments do not provide us with new observations, their contribution must be logical. Thought experiments are like pictorially presented arguments. We gain *knowledge* from thought experiments because they are arguments with empirical knowledge embedded in the premises, but it is not *new* knowledge because the knowledge was already implied in the premises (Sartori 2023, 5; Norton 2004).

Sartori objects to Norton that some thought experiments have non-empirical and even impossible premises (such as running as fast as light, riding in an elevator with no gravity, etc.) (Sartori 2023, 5). Another position Sartori discusses is that of Rawad El Skaf (El Skaf 2018). He builds on Hacking but says that thought experiments reveal inconsistencies within or between theories (Sartori 2023, 5). Against this, Sartori argues that not all thought experiments are about inconsistencies. Examples of thought experiments that are not are Maxwell's demon, Newton's rotating spheres in an empty universe, or Einstein's elevator. Thought experiments like these seem to say something not only about old theories but also something new about the world (Sartori 2023, 6).

Sartori's overview is a helpful systematization, where one could easily insert other theorists based on whether they believe thought experiments provide us with new knowledge about the world and what role they believe thought experiments play in science. For example, while many have appealed to intuition as evidence for learning something new about the world, others reject it (Cappelen 2012).

After Sartori's presentation of various alternatives, his summary is that the yes-side provides vague answers, while the no-side does not explain the significance and success of thought experiments (Sartori 2023, 6). He believes much of the disagreement is due to lack of a common theoretical framework (Sartori 2023, 7). Sartori then presents his own proposal, which is to think of thought experiments as ordinary experiments in science (material experiments) but to distinguish between internal and external validity (Sartori 2023, 7).

The distinction between internal and external validity comes from Donald Campbell (Campbell 1957) and deals with material experiments (Sartori 2023, 8). Internal validity is about whether the specific experiment in the specific setting is correct, while external validity is about whether one correctly extrapolates the results outside the specific setting of the experiment. It is important to distinguish between these two forms of validity since specific experiments are conducted under a set of assumptions that do not necessarily apply to other contexts (Sartori 2023, 8).

Sartori then applies this distinction to thought experiments. Internal validity for thought experiments is to understand thought experiments as a "game of make-believe" (Walton 1990), while external validity is to interpret thought experiments as an accurate representation of the world (Sartori 2023, 9 and 16). As an example, one can think of Galileo, Newton, and Einstein first conducting a thought experiment in their minds, describing a specific context and specific assumptions, and then deducing a general statement for all contexts from it (Sartori 2023, 9-11). When generalizing from both material experiments and thought experiments, one must make a series of assumptions. This process is the same in both material experiments and thought experiments and is best understood as a transition from internal to external validity (Sartori 2023, 11-12).

Regarding the external validity of thought experiments, it means asking whether the result of a thought experiment provides a true representation of the world (Sartori 2023, 19). Then one must check if the representation of the world is actually correct, but this is the same in material experiments as well (Sartori 2023, 23-25). The way thought experiments tell us something true about the world is then similar to material experiments (Sartori 2023, 27).

According to Sartori, not all thought experiments fit perfectly into this model, for example, if a thought experiment only points out an inconsistency in another theory (Sartori 2023, 25). Sartori has no recipe for determining the external validity of a thought experiment (Sartori 2023, 25), but he believes there are no universal criteria for it (Sartori 2023, 26).

In contrast to Sartori, we believe there are some universal criteria that can and should be used when establishing the validity of thought experiments. The coherence criterion with its various aspects can be used to explain how we establish validity and how thought experiments work. It provides an alternative answer compared to Sartori regarding how thought experiments teach us something new about the world. Thought experiments can have both a destructive and a constructive function by weakening some theories and strengthening others. They provide new data even if these are not observational data, and they clarify connections or lack of connections in our theories of the world, thereby teaching us which understandings of the world are most likely true.

A central insight from coherence theory is that the way we learn something new about the world is by discovering which theories about the world are most coherent. Observations are just one of many types of data that we must combine in the most coherent way possible to discover how the world is. These claims will be further explained and defended in the next section.

3. An alternative understanding of thought experiments

In 1973, Nicholas Rescher defined the concept of coherence as having three aspects: consistency, cohesiveness, and comprehensiveness (Rescher 1973, 169). *Consistency* means that the elements of a theory cannot contradict each other. *Cohesiveness* means that the elements of the theory are connected. The more connections and the more precise and fine-grained they are, the more cohesive the theory. Connections should be thought of as including any kind of connection (spatial, temporal, logical, causal, etc.): describing relations between elements in a theory makes it more cohesive. *Comprehensiveness* is a measure of how many elements a theory manages to integrate consistently. The ideal is an integration of an optimal number of relevant elements.

The previous paragraph speaks of coherence between the *elements* of a theory. More precisely, the elements of a theory are *data*, where the concept of data is understood broadly to include any truth candidate, i.e., anything somebody has reason to hold as true (Rescher 1973, 39-40), including scientific laws.[1] There are many advantages to having a broad definition of data, as opposed to a narrow understanding of data as, for example, strictly empirical observations. It clarifies the relation between data and theory, since all data are interpreted, and might be interpreted differently in the future in light of new theories. There are in fact many elements of theories that are not empirical observations, and many empirical observations can be interpreted in different ways that are not consistent with each other, such as the different interpretations of quantum mechanics. This broad understanding of data makes good sense of actual scientific praxis.[2]

All of our experiences with the world are interpretations of how the world is and happens in our mind. There is no experience and no access to the world that is not given to us as content in our minds. If you say to someone "Do not tell me how you think the world is, but how it actually is," this is an impossible order, since nobody can say other than how they think the world is (Rescher, 2010, p. 5). Our understanding of all situations is interpreted and can be thought of as theories about the situation in a broad sense of theory, where a theory is meant to be a true understanding of how things are related. Our understanding of the world is constantly revised in light of observation and thinking. To learn something new about the world, means that new understanding of the world in your mind has replaced an old understanding.

We now proceed to present a theoretical framework for understanding the argumentative role of thought experiments. While Sartori focuses on thought experiments in natural science, our account is meant to cover both natural sciences and the humanities. From now on, the term "science" is used broadly to include the humanities. We focus on the argumentative function of thought experiments and how they can teach us something new,

[1] This means that no data are "raw." All data is interpreted, and laws are also data, because they are fallible truth candidates. But in a coherentist understanding, the data are placed in a theoretical framework, which means that they are related to each other, so that the theory has both data and structure. The theory is expressed in language, and there are rules for how things should be related in the theory.

[2] For a more extensive discussion of this notion of data, see Puntel (2008, 11 et passim).

but we acknowledge that thought experiments can have many functions beyond that, for example illustrative, pedagogical, or heuristic functions (Cohnitz 2000; Corcilius 2018, 69).

Scientific work progresses by *strengthening* theories, *weakening* theories, or *comparing* theories to show that one is better than another. One can *strengthen* a theory by demonstrating or increasing the *presence* of consistency, cohesiveness, or comprehensiveness. One can *weaken* a theory by demonstrating *absence* of consistency, cohesiveness, or comprehensiveness. One can *compare* two theories by showing that one is *more* consistent, cohesive and/or comprehensive than the other.

We argue that all thought experiments used in science have the function of demonstrating either the presence or absence or a relatively better score of consistency, cohesiveness, or comprehensiveness. In the following, we substantiate this claim by testing it with regard to some examples. We comment on how the examples fit the theory by showing that they are examples of goals 1, 2, 3 (strengthening, weakening, or comparing) or means A, B, C (consistency, cohesiveness, or comprehensiveness) in our theory.

Galileo made a famous thought experiment to show that bodies fall to the ground at the same speed regardless of their weight (unless hindered by other forces such as air resistance). Aristotle had claimed that a heavier object will fall faster than a lighter object, but Galileo then suggested the following thought experiment: Imagine that we combine a heavy object A with a light object B and drop the combined object to the ground. Now the lighter object B should make the heavier object A fall more slowly if Aristotle is right. But the combined objects A+B can also be seen as one heavier object C, which should now fall faster than both A and B. Aristotle's theory implies that A should fall both faster and slower in this scenario, which is inconsistent (Palmieri 2018; Brown 1991, 1-3).

In this thought experiment, we see how Galileo demonstrates the presence of an inconsistency in Aristotle's theory. Galileo's alternative theory—that objects fall at the same speed regardless of their weight—does not have this inconsistency. In comparison, then, Galileo's theory is more coherent than Aristotle's, and thus a better theory. Galileo compares Aristotle's theory with his own (cf. goal 3) by means of demonstrating the presence of inconsistency in Aristotle's theory and consistency in his own theory (means

A). He thus weakens Aristotle's theory (goal 2) and shows his own theory to be better in comparison (goal 3).

Galileo is famous for introducing Galilean relativity, which says that the laws of nature are the same for all observers regardless of whether they are standing still or moving at uniform speed. He defended the heliocentric worldview, but understandably people had problems believing that the earth should be moving through space at a very high speed. After all, we experience standing still and seeing the sun move—would we not have noticed if we were moving at more than 100 000 km/h?

Galileo answers with another thought experiment: Imagine sitting below deck in a boat with the curtains pulled. In this scenario you would not know whether you were moving at a uniform speed or sitting still in still water. We can conclude from the thought experiment that if the earth moves at a uniform speed, we will not notice the difference between standing still and moving at high speed. If, in addition, the earth rotates around itself, it will seem like the sun is rising and setting.

This thought experiment has the function of demonstrating the cohesiveness of a theory (goal 1, means B). The theory that the earth orbits the sun seems unable to explain several data, like our experience of standing still and watching the sun move. Galileo uses the thought experiment of the boat to demonstrate how these data are nevertheless coherently connected since we would not notice any difference between the earth standing still or the earth moving at uniform speed.

Galilean relativity seems to imply that there is no objective answer to who is moving and who is standing still. Newton famously disagreed, arguing that there is an absolute space making it true that some objects are actually standing still while others are moving. He introduced the famous thought experiment of the bucket to argue this point. Imagine a bucket of water, hanging by a twisted cord, and then released. First the surface of the water is flat, but when the bucket starts spinning, the surface of the water turns concave in shape. Even if the water is immobile relative to the spinning bucket, we know from the shape of the water that the bucket is in fact spinning and not hanging still. According to Newton, this cannot be explained if motion and immobility are considered relative matters. Instead, we need the concept of an absolute space to explain the difference between the two scenarios of bucket spinning and bucket hanging still (Brown 1991, 8-10).

With this thought experiment, Newton introduces a datum to a specific discussion and argues that it can be explained by his own theory, but not by Galilean relativity. By means of showing his own theory more comprehensive (means C), he tries to weaken Galilean relativity (goal 2) and strengthen his own theory (goal 1) to show it to be comparably better (goal 3).

Note the broad use of the concept of data. Data are truth candidates (Rescher 1973, 39-40). When scientists make an observation in a traditional scientific experiment, the data are interpreted (e.g., that the dots on the screen are in fact Higgs' boson). They are thus truth candidates and can be wrong. Thought experiments also deliver truth candidates that can be wrong (e.g., that there could be a zombie like humans in all respects, but without consciousness). Some data from thought experiments are new in the sense of being truth candidates nobody has thought about, like philosophical zombies, twin earths, etc. Other data from thought experiments are based on empirical data that are not new (like Newton's bucket), but they are used in a new context where they are relevant in deciding what is most coherent. In searching for the truth, researchers must take data (in the sense of truth candidates) and combine them as coherently as possible, and some of the data must then also be rejected as false.

Einstein later expanded Galilean relativity into his own theory of special relativity. This theory is based on two fundamental principles. The first is the principle of Galilean relativity, that the laws of nature are the same for all observers in uniform motion. The second and new principle is that all observers measure the same speed of light in a vacuum. According to Einstein, he was led to this insight in his youth, pondering various thought experiments of himself moving at the speed of light.[3]

Einstein imagines sitting on a train at the speed of light, looking into a mirror. Would he see nothing in the mirror? That would contradict Galilean relativity, which says that you cannot know whether you are standing still or moving from data inside your own frame of reference. Light should

[3] It is not important for our purposes what historically preceded what—we are only interested in the argumentative function of Einstein's thought experiments.

instead be measured as moving at the same speed regardless of your motion relative to light. From the insight that everyone measures the same speed of light, Einstein drew the consequences that measurements of time, distances and simultaneity are relative, as shown by various thought experiments involving light, trains and embankments.

Here is a thought experiment providing us with a new datum (truth candidate: light speed is the same for all), from which further new data can be deduced by means of thought experiments—for example that simultaneity is relative. To integrate these new data, Einstein needed to develop more concepts for describing relations more precisely, such as distinguishing between proper time and coordinate time or between rest mass and relativistic mass. This makes the theory more cohesive by showing more precise connections between the data, which then strengthens the theory.

Of course, Einstein's theory of relativity has been confirmed by empirical observations and would have been weaker without those observations. Hypothetically, observations could be made that would contradict the theory, but possibly the theory could also be adjusted to fit new observations. The point is that both thought experiments and empirical experiments can strengthen and weaken theories and are open to different interpretations.

We find that all common examples of thought experiments are easy to fit into our model. So far, we have focused on natural science, but in what follows we add some more examples for support, many coming from other disciplines than natural science, since our theoretical framework is meant to work for thought experiments in all disciplines of natural science and the humanities. The examples are categorized as examples of the means of consistency, cohesiveness, and comprehensiveness.

We start with consistency. Many thought experiments are created to show that a theory is inconsistent, thus *weakening* the theory. Since consistency is an either/or issue, if the thought experiment is successful the theory (in its present form) is destroyed, but it may be rescued later by introducing new distinctions or clarifications. Unless such repairs are ad hoc, the thought experiment which first points out inconsistency can help to improve the theory by making it more cohesive.[4] Here are some examples.

[4] An ad-hoc repair means adding a claim where the only reason for believing the claim to be true is that it would solve the problem. The repair is not ad-hoc if we

Bertrand paradoxes, such as first presented by Joseph Bertrand in *Calcul des probabilités* from 1889, suggest that all understandings of probability are inconsistent. Here is an example offered later by Bas van Fraassen: A factory produces cubes with side lengths between 0 and 1 meter. The probability that a randomly selected cube should have a side length of less than ½ meter seems to be ½. But the probability that a randomly selected cube should have a face area of less than ¼ square meter seems to be ¼. The problem is that we then get two different probabilities describing the same event, since a cube with a side length of ½ meter also has a face area of ¼ square meters (van Fraassen 1989, 303).[5] When a thought experiment thus points to inconsistency in all theories, the thought experiment can be understood as a new datum that a new or any theory must integrate. In this case, the truth candidate is that all theories of probability are inconsistent, and thus a coherent theory of probability must be able to reinterpret Bertrand paradoxes or show why they are wrong and can be discarded.

Sometimes a thought experiment is created to defend a theory against the critique of inconsistency. The thought experiment can then support the view that the theory is consistent after all. One example is from the philosophy of time, in which different views are presented in modern philosophy. The Platonic view says that time itself can move even though everything else in the universe stands still, while the Aristotelian view says that if everything else in the universe stands still, time stands still too. The critic of the Platonic view challenges the Platonists to explain how it could make sense to imagine that time moves even though everything else stands still. Sydney Shoemaker took on the challenge of demonstrating how the Platonic view could be consistent: Imagine people living in three zones—A, B and C—where each of the zones sometimes experiences a local freeze—everything stops moving for an hour. This happens every other year in A, every three years in B, and every five years in C. For the people who experience the freeze, it just feels like going from one second to the next,

have other reasons to believe that the claim is true. This means that the coherence is very low in ad-hoc repairs, and that is why they are not a good thing.

[5] Van Fraassen uses 2 cm cubes, but we found the example easier to understand using 1 meter.

but after every freeze period, there is a red glow on things for a short while. The people in the different zones know about the freezes in the other zones. The inhabitants realize that every thirty years, all three zones should experience a freeze at the same time, and they do experience the usual red glow at all places. They conclude that they have probably had an hour of global freeze, meaning that one hour has passed, even if nothing has moved (Shoemaker 1969).

A unique type of demonstration of consistency is to show that the alternative is inconsistent such that the theory is necessarily correct. Sometimes this consistency can be proved by a thought experiment. The most well-known example stems from Descartes, who describes the possibility that an evil demon deceives our perceptions. But the demon cannot deceive us when it comes to the question whether we think, since we need thought in order to be deceived. You cannot be inconsistent in thinking that thoughts exist, since even being wrong requires that thoughts exist. In conclusion, we can know for sure that thoughts exists (Descartes 1641/1986, 12-15).

With these examples concerning consistency, we now proceed to the second aspect of coherence—cohesiveness. To recapitulate, cohesiveness refers to the connections between the data in a theory. The more connections, the better, since connections increase the plausibility that the data are relevant and needed in a theory. Thought experiments can be created to show a lack of relevant connections or clarify existing connections in a theory. In the following, we look at some examples.

In *Reasons and Persons* from 1984, Derek Parfit discusses the condition for personal identity over time. Is it physical continuity or is it psychological connectedness and continuity, or maybe different combinations of these? Parfit creates some thought experiments connected to teleporting and to a possible split between brain and body halves (Parfit 1984, ch. 10). Take the latter first: Imagine that you are in an accident. You are heavily injured, but the doctors manage to save half your brain and half your body. You have a lot of memories in the remaining part of the brain, and it is connected to a new brain hemisphere. The surviving half of your body is then successfully sewn together with a new half body. You therefore think that you survived the accident and that you are still yourself. But then the doctors inform you that they also managed to save the other halves of your brain and body, and that these parts are now sewn together with new halves, and that they also have memories of the past. Now, suddenly, there are two persons with physical and psychological coherence and continuity with the former person, but which of these is you? What should be the reason that only one of them is you? Is it the case that you survived first, but then ceased to exist when two new persons appeared—but how could a double success be a failure? Or can we say that two persons can be identical to one person—but how can one be identical to two?

Another example of absence of cohesiveness is the well-known trolley problem (Foot 1967, 4): A person has tied five persons to a rail track and a runaway trolley is approaching, about to kill them. You can make the trolley change tracks by pulling a lever, but there is another person tied to that track who will then be killed instead. Should you pull the lever? Most people say "yes." But what if a trolley is about to run over five persons, and you are standing on the bridge with a big man leaning over to see—is it then acceptable to push the big man over the bridge to stop the trolley and save five persons? This time most people would say "no," and then the challenge is to explain the morally relevant difference in the two cases. The challenge here is to unite two moral intuitions with an overall principle that explains them both. This is lack of cohesiveness because we lack an explanation for why two descriptively similar events nevertheless are morally different.

While many thought experiments show connections lacking between data, thought experiments can also show how data are connected (as shown by Shoemaker above). Sometimes you have elements that you wish to connect or to give a specific justification. John Rawls, for example, wants to connect social democracy and justice by showing how social democracy yields a just society, and he does so through a thought experiment where people design a society behind a veil of ignorance. He argues that people would choose to create a kind of social democracy if they had to make a society where they had to live afterwards, not knowing what position or role they would have in the society. This is then meant to show that such a way of organizing society is fair (Rawls 1999, 118-123). Thomas Hobbes wants to connect the use of violence by the king with an ethical justification of it, and he does so through the thought experiment that a social contract is written where the right to violence is consigned to the king in return for the protection this gives to all (Hobbes 1651/2017).

In the following, we discuss the aspect of coherence theory that concerns the amount of data a theory seeks to integrate, i.e., comprehensiveness. To recapitulate, the aspect of comprehensiveness refers to the amount of data that a theory integrates. The more relevant data that are integrated, the better. A thought experiment can be created to demonstrate that theory A lacks specific relevant data or that theory B integrates important data, but most often it is demonstrated that one theory is superior to another because it manages to integrate a larger amount of relevant data. In the following, we run through some examples.

Jonathan Schaffer has an interesting examination of different views on the concept of causality and the connection between cause and effect (Schaffer 2007). The philosophical discussions on causality are full of thought experiments that are used to test different views (Schaffer 2007). There are two main views on what constitutes causality. The first is causation as probability-raising and the other is causation as process linkage. If Pam throws a stone at a window, for example, so that it breaks, the probability-raising view will say that Pam's stone-throwing was the cause since it increased the probability of a broken window, while the process-linkage view will say that Pam's stone-throwing was the cause because a process linked her arm, the stone, and the window. Thought experiments can be used as arguments against both views by describing events that none of the theories manage to integrate.

A thought experiment against causality as probability-raising, on the one hand, is the following: Pam is standing with a stone in front of the window, while at the same time the more reliable vandal, Bob, holds his throw waiting to see if Pam throws instead. When Pam throws, the probability that the window will break decreases, since there would be a higher probability of a broken window if Bob were the thrower, and he would have thrown if Pam had not.

A thought experiment against causation as process-linkage, on the other hand, is the following: Pam uses a catapult to throw a stone at the window. Pam pulls a lever to release a spring, and then the catapult throws a stone on the window, and it breaks. Pam is process-linked to the lever and the catapult is process-linked the window, but there is no energy, force, momentum or other link between Pam and the window. Yet we want to say that Pam was the cause of the broken window.

Here we can see that thought experiments can be used to point to data that a theory does not manage to integrate. Defenders of the different theories could use these examples against each other to argue for the superiority of their own theory.

4. Conclusion

In the previous section, we described how theories can be strengthened, weakened or compared by use of coherence and provided examples from existing thought experiments. Strengthening a theory can be understood as giving an argument for a theory. Weakening a theory can be understood as giving an argument against a theory. Comparing two theories to show that A is better than B, can be understood as giving an argument for A being better than B.

A deductive argument clarifies what is entailed in the premises. A deductive argument can clarify connections in a theory, and thus make it more coherent and better justified as true. It can also demonstrate the presence of an inconsistency or lack of coherence, thus weakening a theory. An inductive argument is an argument where the conclusion is not necessarily true even if the premises are true. How good the argument is depending on how relevant (“relevant” in the sense of logical strength) the premises are, if true. It is contested what makes inductive arguments relevant. We argue that the relevance of an inductive argument is the degree to which it makes one theory more coherent than the alternatives (or less coherent if it is a counterargument).

Given this understanding, thought experiments can obviously be both deductive and inductive arguments, used to strengthen, weaken, or compare theories. But scientific theories are not only strengthened and weakened by arguments, they are also strengthened and weakened by new data that we discover. Thought experiments can also be data, when we use a broad understanding of data—as we have good reasons to do.

We have already given examples of how thought experiments can be understood as new data. An area where they can obviously be new data is when the mind is the topic of scientific exploration. Thought experiments can teach us about things that are impossible to think or things where the negation cannot be thought.: For example, you cannot imagine an event separate from time and space (cf. Kant). You cannot consistently think that thoughts themselves are illusions (cf. Descartes).[6] Thought experiments can give us data about modal facts of possibility, impossibility, necessity, or transcendental conditions.

In many cases, thought experiments employ knowledge we have by empirical means. But empirical knowledge is also interpreted by thoughts. Thought experiments and empirical experiments are interwoven and have very similar and overlapping functions in science. One might think that thought experiments are mainly about deducing inconsistencies. But in this article, we have shown that pointing out inconsistencies very often has the inductive function of showing one theory to be more cohesive and comprehensive than another, while the theories can also be reconfigured and further nuanced to deal with the thought experiments. In other cases, the function of thought experiments is not about deducing inconsistency, but instead demonstrating consistency, cohesiveness or comprehensiveness. The goal of this article was to show this rich use of thought experiments and their close argumentative link to normal experiments owing to the fact that thought experiments are also data for theories to integrate.

The coherence theory we have here presented uses a broad understanding of data and of theory. We do not have access to the world in itself outside of our mind. All data are like small theories: interpretations of the world that can be wrong. Very often we have good reason to believe that what we observe is true, especially if many people observe it, and there is no coherent alternative explanation but to believe that what we observed was true. But many observations are also uncertain, contested and open to many interpretations. Both observations and thought experiments are truth candidates and thus data that theories should consider when trying to make

[6] Some have contested these claims, which we think is unfeasible given proper definitions of the terms—but there is not room for that discussion here.

the most coherent theory of the world. Some observations and thought experiments will be included and some will be discarded even in the most coherent theory.

When we learn something new about the world, it is not the case that the world in itself is revealed to us. What in fact happens is that observations, thought experiments, reflections on language and definitions, understandings of connections etc. help us understand that one theory of the world is more coherent than another. We then replace our earlier understanding with a more coherent understanding – often by integrating new data, but sometimes also by rejecting old data as false. This is how thought experiments teach us something new about the world, namely by strengthening, weakening or comparing theories, thus making us reconsider which understanding of the world is most likely to be true.

Sartoris theory of moving from internal to external validity is not wrong, but very narrow, focusing on a subset of thought experiments and not explaining how the external validity is established. Given coherentism, external validity is established by showing that a theory is more coherent than alternative theories. In this article, we hope to have contributed with both a broader and deeper understanding of how thought experiments function and give us new knowledge about the empirical world.

References

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

Brown, J. R. 2004. "Why thought experiments transcend empiricism." In C. Hitchcock (Ed.), *Contemporary Debates in Philosophy of Science*, 23–43. Oxford: Blackwell.

Campbell, D. T. 1957. "Factors Relevant to the Validity of Experiments in Social Settings." *Psychological Bulletin*, 54, 297–312. https://doi.org/10.1037/h0040950

Cappelen, Herman. 2012. *Philosophy without Intuitions.* Oxford: Oxford University Press.

Cohnitz, Daniel. 2006. *Gedankenexperimente in der Philosophie.* Leiden: Brill.

Dennett, D. 1996. "Intuition pumps." In J. Brockman (Ed.), *Third Culture: Beyond the Scientific Revolution*, 181–197. New York: Simon and Schuster.

Descartes, René. 1641/1984. *Meditations on First Philosophy.* Translated by John Cottingham. Cambridge: Cambridge University Press

El Skaf, R. 2018. “The Function and Limit of Galileo’s Falling Bodies Thought Experiment: Absolute Weight, Specific Weight and the Medium’s Resistance.” *Croatian Journal of Philosophy*, 18(52), 37–58.

Foot, P. 1967. “The Problem of Abortion and the Doctrine of the Double Effect.” *Oxford Review*(5), 5-15.

Gendler, Tamar Szabo. 2004. “Thought Experiments Rethought–and Reperceived.” *Philosophy of Science,* no. 71, 1152-1163. https://doi.org/10.1086/425239

Hacking, I. 1993. “Do thought experiments have a life of their own? Comments on James Brown, Nancy Nersessian and David Gooding.” In Hull, D., M. Forbes, & K. Okruhlik (Eds.), *Proceedings of the Philosophy of Science Association Conference 1992*, Volume 2, 291–301. Chicago: University of Chicago Press.

Hobbes, Thomas. 1651/2017. *Leviathan.* New York: Penguin.

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

Miscevic, N. (1992). “Mental Models and Thought Experiments.” *International Studies in the Philosophy of Science*, 6(3), 215–226. http://dx.doi.org/10.1080/02698599208573432

Norton, J. D. (2004). “Why Thought Experiments Do Not Transcend Empiricism.” In C. Hitchcock (Ed.), *Contemporary Debates in the Philosophy of Science*, 44–66. Oxford: Blackwell.

Palmieri, Paolo. 2018. “Galileo’s Thought Experiments: Projective Participation and the Integration Of Paradoxes.” In Michael Stuart, Yiftach Fehige, and James Robert Brown (eds.). *The Routledge Companion to Thought Experiments*, 92–110. London: Routledge.

Parfit, Darek. 1984. *Reasons and Persons.* Oxford: Oxford University Press.

Puntel, Lorenz B. 2008. *Structure and Being. A Theoretical Framework for a Systematic Philosophy.* University Park: The Pennsylvania State University Press.

Putnam, Hilary. 1975. “The Meaning of ‘Meaning.’” In Hilary Putnam. *Mind, Language and Reality. Philosophical Papers*, Vol. 2. https://hdl.handle.net/11299/185225

Rawls, John. 1999. *A Theory of Justice.* Oxford: Oxford University Press.

Rescher, Nicholas. 1973. *The Coherence Theory of Truth.* Oxford: Clarendon.

Rescher, Nicholas. 2010. *Reality and its Appearance.* London: Continuum.

Sartori, L. (2023). “Putting the ‘Experiment’ back into the ‘Thought Experiment.” *Synthese* 201:34. https://doi.org/10.1007/s11229-022-04011-3

Schaffer, Jonathan. 2016. “The Metaphysics of Causation.” *Stanford Encyclopedia of Philosophy.* https://plato.stanford.edu/entries/causation-metaphysics/. Visited November 15, 2023.

Shoemaker, Sydney. 1969. "Time without Change." *The Journal of Philosophy,* vol. 66, no. 12, 363–381. https://doi.org/10.2307/2023892

Van Fraassen, Bas C. 1989. *Laws and Symmetry.* Oxford: Oxford University Press.

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