Penultimate Draft

Routledge Companion to Thought Experiments

Eds. J.R. Brown, Y. Fehige, & M. Stuart

On the Identity of Thought Experiments: Thought Experiments Rethought

Alisa Bokulich and Mélanie Frappier

1. Introduction

From Lucretius's spear modification of Archytas's 4th century BC thought experiment about the size of the universe to David Bohm's modification of the EPR gedanken experiment about the completeness of quantum mechanics, historical research has revealed that, similarly to physical experiments, thought experiments can be replicated, reworked, and even retooled for very different ends by different thinkers at different times.[1] This historical insight raises the philosophical question of what is to count as the *same* thought experiment. Can one identify identity conditions for thought experiments? How similar do two narratives need to be in order to count as the same thought experiment? Can two thinkers perform the same thought experiment, and yet reach different conclusions?

One of the challenges in trying to answer such questions is that there is no consensus on what exactly thought experiments are. Are thought experiments *experiments* (Mach [1926] 1976; Brown [1991] 2011; Sorensen 1992a, Rowbottom 2014), *arguments* (Norton 1996, 2004a,b; see Brendel this volume), *cognitive models* (Nersessian 1993, 2008; see Nersessian this volume), mere *intuition pumps* (Dennett 1984, 1995), *props* for the imagination (Meynell 2014; see also Meynell or Becker this volume), or something else? As we shall see, what position one adopts on the nature of thought experiments has implications for one's view on the identity question. Second, as Roy Sorensen (1992a, p. 163) has noted, as with any discussion about identity, which aspects one identifies as being the most salient often depends on the purpose of one's inquiry: hence, for some projects historical continuity might be the most important, whereas for others it might be propositional content. Rather than trying to determine one set of identity conditions for thought experiments, a more fruitful approach, perhaps, is to explore how different ways of delimiting the identity of thought experiments enable different sorts of inquiries to unfold and different insights into the nature and use of thought experiments to be gained.

Some scholars, such as Ian Hacking, have outright denied that thought experiments can evolve and be retooled:

> I think of [physical] experiments as having a life: maturing, evolving, adapting, being not only recycled but also, quite literally retooled. But thought experiments are rather fixed, largely immutable. (Hacking 1993, p. 307).

Other scholars, when confronted with prima facie evidence about the historical evolution of a particular thought experiment, argue that this plasticity is only an illusion, and that thought experiments, when *properly construed*, are immutable. John Norton (1996, 2004a,b), for example, has repeatedly argued that thought experiments are nothing but arguments, hence if the premises and conclusions of two instances of a thought experiment are not identical, then they are not in fact the same thought experiment. Similarly, Roy Sorensen (1992a) sets out a set theoretic conception of thought experiments that also imposes stringent identity conditions on thoughts experiments.

On the other hand, scholars such as Alisa Bokulich (2001) and Darrell Rowbottom (2014) have adopted a lenient construal of the identity of two thought experiments, appealing to criteria such as a resemblance of the hypothetical or counterfactual states of affairs and continuity through historical connection. Bokulich argues *pace* Hacking that thought experiments do have a life of their own and that they can be rethought from the perspective of different and even incompatible theories (Bokulich 2001, p. 286). In order to better understand these debates between those who think thought experiments can evolve and those who think they cannot, it is helpful to begin by briefly reviewing some historical examples.

2. The Historical Evolution of Thought Experiments: Some Examples

When examining the history of various thought experiments, one can find a wide spectrum of variation: some thought experiments are relatively stable over time, such as Galileo's falling bodies thought experiment; others have undergone a series of minor modifications, such as Lucretius's spear modification of Archytas's hand at the edge of the universe; and some thought experiments have been given radically different interpretations by different thinkers even during the same period of time. At this latter extreme we find what have come to be known as "thought experiment–anti-thought experiment pairs".[2] These are defined as any two thought experiments that lead to opposite conclusions.

A particularly interesting sub-class of these for our discussion here is thought experiments pairs involving the *same* scenario and yet having analyses that lead to opposite conclusions.[3] A classic example of such a case is Newton's ([1687] 1999, pp. 412-413) bucket thought experiment, which invites us to imagine a bucket full of water hanging from a rope that has been twisted tight. When the bucket is released and the rope begins to unwind, the water in the spinning bucket will initially be flat, even though there is a relative motion between the bucket and water. Eventually as the spinning motion is communicated to the water, it will begin to climb the sides of the bucket, forming a concave surface even though the water is now stationary relative to the bucket (as it was initially before it began to unwind). On one reading, Newton can be seen as arguing that this establishes that the water's spinning motion is absolute, that it is motion in reference to absolute space, and that such motions have observable effects that are distinguishable from mere relative motion. Ernst Mach, in his *Science of Mechanics* ([1883] 1919), reanalyzes this same thought experiment and argues 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).[4] Here we seem to have the same thought experiment narrative, and yet the conclusions drawn from this thought experiment are diametrically opposed.

Another sort of case involves thought experiments with the same central narrative and that reach the *same* conclusion, but do so by appealing to different underlying theories and analyses. An example of a thought experiment of this sort is the rockets and thread thought experiment, discussed by Bokulich (2001). The scenario involves two rockets initially at rest in frame S, 100 meters apart and connected by a thread of the same length; both rockets fire simultaneously in this frame and accelerate to 4/5ths the speed of light, stop accelerating, and are moving with a uniform velocity relative to S. Will the thread break? Dewan and Beran (1959) use special relativity to show that according to an observer in S, the thread is Lorentz contracted to a length of 60 meters and so can no longer span the 100 meters between the rockets. Dewan and Beran show how this breaking thread is in fact a consistent result for an observer that is at rest in the rockets final inertial frame, S', because from this frame of reference the accelerations were not in tandem and the distance between the rockets grew, hence also leading to the breaking thread. Dewan and Beran's aim is to use this thought experiment to show that, despite intuitions of many to the contrary, Lorentz contraction can cause measurable stresses on moving bodies. Almost twenty years later John S. Bell (1976) rethinks this same thought experiment scenario from the perspective of Lorentz's aether theory, arguing that one can arrive at this same conclusion by means of this very different, though arguably more intuitive, theory.

Yet a third example of the ways in which thought experiments can be rethought is Albert Einstein's and Niels Bohr's development and analyses of the clock-in-the-box thought experiment to test the completeness of quantum mechanics. Bohr (1949) recounts the evolution of this thought experiment in his discussions with Einstein, beginning at the 1927 Solvay conference, undergoing modifications in their discussions at the 1930 Solvay conference, and yet further changes during their meeting in Princeton shortly after the 1933 Solvay conference. When Einstein first proposed this thought experiment, it consisted simply of a particle passing through a narrow slit in a diaphragm that was placed some distance before a photographic plate. Bohr recounts how the first modification involved inserting another diaphragm with two slits, in between the single-slit diaphragm and photographic plate. Bohr then makes what he describes as his "pseudo-realistic" modifications to the thought experiment, first imagining the two diaphragms firmly bolted to the table, then with the diaphragm suspended by springs and allowed to move up and down as the particle passes through (representations which he uses to underscore the incompatibility of various experimental arrangements and their role in the production of the phenomenon). On the next iteration of the thought experiment, Einstein replaces the slit with a shutter in a box that could be opened or closed, allowing a particle to escape, and then attaches that shutter to a clock. Once again, Bohr presents a "pseudorealistic" diagram of Einstein's clock-shutter device, with the support bolted to the table and the box suspended from the support by a spring allowing the box to move up and down. As Bohr recounts, this thought experiment appeared to show a violation of the energy-time uncertainty relations:

> [A] single photon was released through the hole at a moment known with as great accuracy as desired. Moreover, it would apparently also be possible, by weighing the whole box before and after this event, to measure the energy of the photon with any accuracy wanted, in definite contradiction to the reciprocal indeterminacy of time and energy quantities in quantum mechanics. (Bohr 1949, pp. 225-226)

Bohr recounts his initial surprise over this result and then reanalysis of the thought experiment leading to the opposite conclusion: according to general relativity, the rate of the clock would change as the box shifted its position in the Earth's gravitational field, introducing an indeterminacy on the time measurement that would satisfy the uncertainty principle.

The first thing to notice about this brief history of the clock-in-the-box thought experiment is that this thought experiment underwent a long evolution through a series of stepwise modifications. Each of these steps helped to refine the thought experimenters' thinking and bring to light different aspects of the phenomenon. Although the final version of the thought experiment looks quite different from the initial set up, the intervening steps show a strong pairwise similarity, and moreover there is a clear intention on the part of the thought experimenters to be discussing and modifying the *same* thought experiment. The situation here seems to be quite analogous to two experimenters making several runs of the same physical experiment, while tinkering with and swapping out components of their table-top apparatus as they try to refine their setup.

The second feature to note about this history of the clock-in-the-box thought experiment is how Bohr and Einstein were led to differing conclusions regarding the validity of the uncertainty relations, illustrating a thought experiment-anti-thought experiment pair, analogous to the Newton's bucket case. As will be discussed in more detail in Section 4, Michael Bishop, uses this fact, that the same thought experiment can be used to reach opposing conclusions, to challenge Norton's (1996b) claim that thought experiments are nothing but arguments. Conversely Bishop argues that while Bohr and Einstein were discussing the same thought experiment, each was drawing a different argument from the scenario (Bishop 1999, p. 535).[5] It is only if we admit that thought experiments can be analyzed through different arguments that we can, Bishop concludes, make sense of such historical episodes (see Brendel, this volume, for a thorough discussion of this debate).

Further historical research suggests that Bohr and Einstein didn't just disagree about the conclusion of the clock-in-the-box thought experiment, rather they more fundamentally disagreed about what this thought experiment was supposed to show. Citing a letter from Ehrenfest to Bohr, Don Howard, for example, has cogently argued that Einstein did not in fact intend the clock-in-the-box thought experiment as an argument against the limits of the uncertainty relations:

> He [Einstein] said to me that, for a very long time already, he absolutely no longer doubted the uncertainty relations, and that he thus, e.g., had BY NO MEANS invented the "weighable light-flash box" (let us call it simply L-F-box) "contra uncertainty relation," but for a totally different purpose. (Ehrenfest to Bohr, 9 July 1931, Bohr Scientific Correspondence, Archive for History of Quantum Physics; quoted in Howard 2007).

Howard argues that Einstein instead intended the clock-in-the-box experiment as a proto-EPR thought experiment to show the incompleteness of quantum mechanics: the photon is allowed to leave the box and is later reflected back, say, half a light-year away; we then choose to either weigh the box or check the clock, hence ascribing a different physical state to the photon depending on which distant measurement we choose to perform, even though no signal from our measurement could reach the photon and alter its state. If this historical reading is right, then Bohr and Einstein were not simply disagreeing about the conclusion of the clock-in-the box thought experiment, rather they were disagreeing about what phenomenon the thought experiment was supposed to explore.[6]

In the three examples briefly reviewed here, we can recognize a strong historical continuity in these thought experiments and there is at least the intention on the part of these thought experimenters to be discussing the *same* thought experiment. On the other hand, the details of the scenarios, the underlying theories and analyses, and even the purposes and conclusions drawn can change. Hence these examples bring out the opposing intuitions that such cases both do and do not involve the same thought experiment. Moreover, these examples are by no means anomalous; many more examples of the evolution of thought experiments can be found, such as in Carla Palmerino's (2011) work identifying medieval precursors to many of Galileo's thought experiments, the different version of the chain-around-a-double-incline-plane thought experiments attributed to Stevin (Rowbottom 2014) and other examples discussed in the historical entries of this handbook (e.g., Ierodiakonou or King this volume; for important qualifications on the various epistemological roles a given thought experimental scenario can play in different contexts, see McAllister, this volume also.)).

As noted earlier, one's intuitions about whether thought experiments can evolve and be retooled are strongly shaped by one's views on what thought experiments are. Hence, let us turn to four key competing views of the nature of thought experiments and see what implications they have for the question of whether thought experiments can be rethought.

3. Thought Experiments on the Intuitionist Approaches

Many intuition-based accounts of thought experiments underscore the continuity between physical experiments and thought experiments, but argue that the latter work by harnessing physical or modal intuitions to give us new insights into nature. There are two broad classes of intuitionist approaches: “a priori” intuitionist approaches, such as Brown’s (1991) that argue thought experiments give us a priori access to the laws of nature and “naturalist” intuitionist approaches, such as Sorensen's (1992a,b), that argue thought experiments allow us to harness intuitions of laws, intuitions which are refined through natural selection. Intuitionist approaches are fallibilist, though they assume our intuitions are sufficiently reliable to explain the successful uses of thought experiments (Rowbottom 2014).

A key question for this approach, then, is whether one needs to harness the same intuitions in a thought experiment in order for it to count as the same thought experiment. One worry is that there is growing empirical evidence that people’s intuitions can vary greatly depending on their culture, training, socio-economic status, etc. (Stich and Bulkwater 2011; Weinbert *et al.* 2008; Reiss 2002; Stich and Tobias this volume). Hence on this construal thought experiments might only be replicable for similar members of a social group. If, on the other hand, the intuitions invoked in a thought experiment are not essential to the type identity of the thought experiment, then on intuitionist approaches thought experiments can be replicated by different people at different times. This latter approach is adopted by Brown ([1991] 2011), who seems to identify a continuity in the central narrative as all that is required to replicate a thought experiment.

As we saw earlier, this is not the case on Sorensen’s intuitionist approach, which imposes stringent conditions on the identity of thought experiments. For Sorensen a thought experiment is defined by a set of propositions, and any modification of a thought experiment amounts to changing at least one proposition in the set. He writes, "since a set is defined in terms of its members, you no longer have the same set. Therefore we are forced to say that any alteration of a thought experiment yields a new thought experiment" (Sorensen 1992a, p. 160). He justifies this more stringent approach based, first, on the fact the context of theoretical classification demands maximally fine-grained criteria, and, second, on the fact that our ability to compare two versions of a thought experiment leads us to conclude, when pressed, that they are not really identical.[7]

4. Thought Experiments as Arguments

Like Sorensen, Norton thinks that a modified thought experiment constitutes a new thought experiment, hence thought experiments cannot be rethought and evolve. According to Norton, thought experiments are nothing but arguments disguised in picturesque or narrative form (e.g., Norton 2004a). Norton defends what he calls the Reconstruction Thesis:

> All thought experiments can be reconstructed as arguments based on tacit or explicit assumptions. Belief in the outcome-conclusion of the thought experiment is justified only insofar as the reconstructed argument can justify the conclusion (Norton 1996, p. 339).

More specifically, Norton takes the Reconstruction Thesis to be true precisely because thought experiments just are arguments.[8] Hence, if any of the premises or conclusion of a thought experiment changes, then it is not the same thought experiment.

Strictly speaking, Norton's criteria for the identity of a thought experiment are slightly less stringent than Sorensen's set-theoretic criterion in that they allow for inconsequential modifications—i.e., those that don't affect the premises or conclusions of the argument. For example, Norton would presumably allow that one could change the colors of the balls that Galileo is dropping off his tower, since this wouldn't affect the corresponding "t-argument" (which Bishop (1999) defines as the argument reconstructed from a thought experiment). Whereas for Sorensen, asserting that the ball being dropped is purple, rather than red, for example, would presumably change the propositional content of the thought experiment, and hence be a different thought experiment, even if it doesn't change the corresponding argument.

Norton's conditions for the identity of thought experiments, namely, that they must involve the same premises and conclusion, can be clearly seen in the Norton-Bishop debate about whether Einstein and Bohr were discussing the same clock-in-the-box thought experiment. According to Norton (2004b), the fact that they reached different conclusions regarding the validity of the uncertainty principle shows that they were executing different thought experiments, that is, different t-arguments: Einstein had an implicit premise assuming a non-relativistic space-time, whereas Bohr had an explicit premise assuming a relativistic space-time: non-identical premises therefore non-identical thought experiments.

A challenge for Norton's account is that it is not always clear how precisely to reconstruct the argument associated with a thought experiment. As we saw earlier, for example, this common reconstruction of Einstein's thought experiment has been disputed by Howard (2007). Often thought experiments are under-specified in such a way that their corresponding arguments are underdetermined. Rather than seeing this as a liability or weakness, however, one might argue that it is precisely their constrained-but-still-underdetermined status that makes them such fruitful tools in scientific investigation (an aspect that is captured by a dialectical approach to thought experiments like that of Roux and Goffi, this volume).

5. Thought Experiments as Simulative Model-Based Reasoning

Nancy Nersessian (1993; this volume also) has argued that this conception of thought experiments as being nothing but arguments is impoverished and that it fails to appreciate both the non-algorithmic forms of reasoning that take place in a thought experiment and the broader role that thought experiments play in helping scientists to change their conceptual structures (Nersessian 1993, p. 291). According to Nersessian, thought experiments are a form of what she calls simulative model-based reasoning. On this view, to perform a thought experiment is to construct a mental model and then simulate what would happen in that model under various manipulations.

Nersessian allows that the thought experiment narrative will underspecify the corresponding mental model, and thus the mental models in the minds of various thought experimenters will likely be different. As long as the mental models constructed by the various thought experimenters were generated by the relevant narrative and embody the same constraints with respect to the target phenomenon, then they should be treated as replications of the same thought experiment. She writes,

> On a mental modeling account, then, a thought-experimental model is a conceptual system representing the physical system that is being reasoned about. More than one instantiation or realization of a situation described in the narrative is possible. The constructed model need only be of *the same kind* with respect to salient dimensions of target phenomena. (Nersessian 2008, p. 179 emphasis original)

Thus despite worries over the potentially private and unique nature of mental-model simulations, on Nersessian's view thought experiments can be rethought. Moreover, in the process of constructing and manipulating these models, thought experimenters can observe how various constraints interact with one another and how neglected constraints can perhaps become salient. There are, as she explains, "cycles of construction, simulation, evaluation, and adaptation of [the] models" (Nersessian 2008, p. 184). Hence it is a central feature of this approach that thought experiments can be replicated, evolve, and be retooled.

6. Thought Experiments as Props for the Imagination

Some, however, have questioned whether such mental modeling accounts can provide adequate identity conditions for thought experiments. Letitia Meynell, for example, writes,

> given the explicit internalism of mental modeling accounts we inevitably come up against the problem of other minds in the guise of the problem of decisively identifying private mental content. If two thought experimenters report that they have run GFTE [Galileo's falling bodies thought experiment] and they report the same result how can we authoritatively assess whether they have actually experienced the same mental process or even a relevantly similar one? (Meynell 2014, p. 4157)

By contrast, Meynell seeks to provide an account of the content of thought experiments that is objective and intersubjectively accessible (Meynell 2014, this volume also). Her account draws on Kendall Walton's (1990) work *Mimesis as Make-Believe* and treats thought experiments as props for imagining fictional worlds.

On this Waltonian approach, the written or spoken text of the thought experiment (along with any diagrams, etc.) functions as a “prop”, which along with “principles of generation” (which include things such as explicit stipulations, implicit cultural conventions, or background knowledge) tell the thought experimenter what to imagine and generate the “fictional truths” within that imagining. The set of fictional truths that the work (e.g., the painting, novel, or thought experiment) generates is called the “work world.” According to Walton, there is an objectivity to work worlds in so far as they are “out there” to be investigated and explored (Walton 1990, p. 42). An individual’s subjective imaginings, based on the props and principles of generation, are called the “game world.” So there is one set of fictional truths in the “work world” and another presumably overlapping set of fictional truths in the “game world.” Walton notes that we can think of the game world as a subjective expansion of the objective work world (p. 216).

Meynell uses this Waltonian framework to provide an account of thought experiments and their identity conditions. The work world, which is fixed by the props of the thought experiment (i.e., the text narrative and diagrams) along with the appropriate principles of generation, is objective and intersubjectively accessible. When a person performs a thought experiment he or she generates a game world expansion that shares certain fictional truths with the work world. This distinction, she argues, can then explain how it is that two people can perform the same thought experiment:

> We perform the same TE [thought experiment] when we imagine two game worlds which share the same fictional truths as the work world of a TE. The fictional truths of the work world are what confer identity, which allows that rather different descriptions of a TE . . . are still importantly the same TE. (Meynell 2014, p. 4166)

She goes on to note that this framework can also explain what is happening when two thought experimenters reach different conclusions: they have generated two different game worlds on the basis of the same work world. The question then becomes one of determining whether one of the thought experimenters used the wrong principle(s) of generation, whether the principles of generation are unclear, or whether they are controversial, etc.

In some cases, the disagreement between two thought experimenters might just reflect that fact that one of them incorrectly applied the relevant principle of generation, such as in the rockets and thread thought experiment where Bell reports that the majority of physicists in the Theory Division at CERN answered incorrectly that the thread would *not* break (Bell [1976] 1993, p. 68). Such cases can easily be dismissed as mistakes on the part of those thought experimenters.

In other cases of disagreement, the two thought experimenters may arrive at importantly different game worlds because there is an underdetermination of the generating principles; that is, more than one set of generating principles can be legitimately applied to the same narrative, leading to divergent results. So, for example, when Newton and Mach performed the rotating bucket thought experiment, they shared the same "work world"—they both imagined a bucket hanging from a rope, the bucket rotating as the rope unwound, and the water rising up the sides of the bucket—but they used different principles of generation in the creation of their respective "game worlds". The thought experiment allowed two different principles of generation: one related to Newton's absolute space and the other to Mach's relationist account.

At first sight, Meynell's account seems to readily explain how thought experiments are rethought. When first presented with the clock-in-the-box experiment, Bohr wrongly concluded that it offered a counter-example to the uncertainty principle because he imagined the device in a Newtonian spacetime. Once he changed his generating principles to include a relativistic setting, he reached the opposite conclusion. Similarly, Bell reconceptualized the rockets-and-thread experiment by abandoning special relativity in favor of a Lorentzian theory of motion.

The challenge, however, for a Waltonian account of thought experiments is to articulate exactly how the narrative and generating principles work together to determine the content of the thought experiment. Thinking back to the one-slit thought experiment that started the Bohr-Einstein debate, it seems that the original one-slit thought experiment was an invitation from Einstein to Bohr to attempt to develop an imaginary experimental setup that would point to quantum mechanics' incompleteness. In their exploration of the completeness problem, Einstein and Bohr freely tinkered with the experiment, adding a shutter, adding a clock, getting rid of the photographic plate, suspending the diaphragm, etc. On this reading, the Einstein-Bohr conversation offers a great example of the evolution of a thought experiment.

On a Waltonian approach, however, it seems we should instead conclude that Bohr is simply offering a series of distinct thought experiments, insofar as a thought experiment is completely determined by its narrative, images, and allowable generating principles. The images and corresponding narratives in each permutation of this thought experiment have changed: hence, the work world of the one-slit experiment is limited to Einstein's initial scenario and the simple diagram that accompanied it; the subsequent iterations where a shutter is added, the photographic plate is removed, or the diaphragm is transformed into a box would be different work worlds, and hence different thought experiments. In other words, Meynell's Waltonian demand that the content be objectively determined by its scenario, pictures, and generating principles may mean that, ultimately, on her account, thought experiments do not organically evolve. Instead—in a fashion reminiscent to Norton's argument—they would lead to similar, yet strictly speaking, different thought experiments. More work is needed to show that such an account does allow for a tinkering and evolution of thought experiments, as Meynell's describes.

7. Replicating Physical Experiments and Replicating Thought Experiments

On many accounts, thought experiments are viewed as being on a continuum with ordinary physical experiments (see for example, Palmieri or Buzzoni, this volume), with the key difference that they are carried out in the "laboratory of the mind", to use James Robert Brown's (1991) phrase. If this is right, one might hope that some insight into the identity of thought experiments can be gleaned by thinking about the replicability of physical experiments. It is widely accepted that physical experiments can be replicated and that in replicating an experiment there will inevitably be some differences, though presumably inconsequential ones (Mulkay and Gilbert 1986). Hacking for example writes: “[experiments] develop, change, and yet retain a certain long-term development which makes us talk about repeating and replicating experiments” (Hacking 1993, p. 307).

As Hans Radder (1993) and Jutta Schickore (2011) have pointed out, however, there are different things that can be meant by ‘replicating an experiment,’ and these have changed over time. For example, one can talk of replicating the “material realization” of an experiment, replicating it from a given theoretical point of view, or replicating the outcome of an experiment by a different theoretical or material means (Radder 1993, pp. 64-65). Hence what is to count as a replication of a physical experiment is itself also likely to depend on the context and purpose of the inquiry. Radder defines material realization as the whole of the experimental actions carried out, and which can be described as a set of actions in ordinary language. This is then contrasted with the theoretical interpretation of those actions. Radder illustrates this dual description of an experiment with the following example:

> Suppose we want to determine experimentally the mass of an object that is at rest in relation to the measurement equipment. Two scientists each carry out such an experiment in the same way. Nevertheless, one interprets the actions performed as a measurement of the Newtonian mass; and the other, as a determination of the

> Einsteinian mass. But both performed “the same” actions and thus—in a certain sense—the same experiment. . . . [yet] concrete experimental action is always action on the basis of certain theoretical ideas: without theoretical ideas there can be no experiments. (Radder 1996, p. 13)

Hence, although both the material realization and the theoretical interpretation are instantiated in an experiment, they are conceptually distinct. In other words, there is a kind of underdetermination of the theoretical interpretation of an experiment by the material realization (Radder 1996, p. 20). One can reproduce the material realization of an experiment without reproducing a particular theoretical interpretation of that material realization.

Tim De Mey (2003) and Letitia Meynell (see above) have extended Radder's approach to the case of thought experiments, arguing that they too have a dual structure. In the case of thought experiments De Mey describes as the counterpart to "material realization" the actions to be performed by the thought experimenter, as phrased in everyday language. In his discussion of the clock-in-the-box thought experiment (on the standard interpretation as being about the uncertainty relations), De Mey writes,

> [T]hought experiments like that of the clock-in-the-box have a dual structure: they involve (1) the description of an imaginary situation and (2) the description of its settlement or winding up. . . . [O]n the experiment view of thought experiments

> [e.g., Bishop's], sameness of or difference between thought experiments is identified on the basis of (1). On the argument view of thought experiments [e.g., Norton's], by contrast, sameness or difference between thought experiments is identified on the basis of (2). (De Mey 2003, p. 71).

While this dual structure approach helps to clarify some of these debates, it still leaves open the identity question. There are three obvious possibilities: To be the same thought experiment it must be the case that one has (1) only the same material realization, (2) only the same theoretical interpretation, or (3) both the same material realization and the same theoretical interpretation. Prompted by some prominent historical examples, one might conclude that either condition (1) or condition (2) are sufficient for counting as the same thought experiment. It is still not clear, however, that cases such as the evolution of the clock-in-the-box thought experiment, as described here, can be adequately handled on either approach, in that it involved both changes in the material realization of the thought experiment and changes in its theoretical interpretation. But as those who emphasize the continuity between thought experiments and physical experiments will be quick to point out, these sorts of challenges are not unique to thought experiments (see Miščević, this volume).

References

Arthur, R. T. W. (2013). Can Thought Experiments Be Resolved by Experiments? The Case of Aristotle's Wheel. In M. Frappier, L. Meynell, and J.R. Brown (Eds.), *Thought Experiments in Science, Philosophy and the Arts* (pp. 107-122). New York: Routledge.

Bell, J. S. ([1976] 1993). How to Teach Special Relativity. Pp. 67-80 in *Speakable and Unspeakable in Quantum Mechanics*. Cambridge: Cambridge University Press.

Brown, J. R. ([1991] 2011). *The Laboratory of the Mind: Thought Experiments in the Natural Sciences*. 2nd edition. London and New York: Routledge.

Brown, J. R. (2007). Counter Thought Experiments, *Royal Institute of Philosophy Supplement 61* (82): 155–177.

Brown, J. R. (2013). What Do We See in a Thought Experiment? In M. Frappier, L. Meynell, and J.R. Brown (Eds.), *Thought Experiments in Science, Philosophy and the Arts* (pp.53-68). New York: Routledge.

Bishop, M. (1999). Why Thought Experiments Are not Arguments. *Philosophy of Science 66*(4), 534-541.

Bohm, D. (1951). *Quantum Theory*. Englewood Cliffs, NJ: Prentice Hall. Republished by Dover Publications, New York in 1989.

Bohr, N. (1949). Discussion with Einstein on Epistemological Problems in Atomic Physics. In P. Schilpp (Ed.), *Albert Einstein: Philosopher-Scientist* (pp. 199-242). Evanston, IL: The Library of Living Philosopshers

Bokulich, A. (2001). Rethinking Thought Experiments. *Perspectives on Science 9* (3), 285-307.

Dennett, D. C. (1984). *Elbow Room*: *The Varieties of Free Will Worth Wanting*. Cambridge, MA: MIT Press.

Dennett, D. C. (1995). Intuitions Pumps. Pp. 181-197, In *The Third Culture: Beyond the Scientific Revolution*, Ed. J. Brockman, New York: Simon & Schuster.

De Mey, T. (2003). The Dual Nature View of Thought Experiments. *Philosophica 72*, 61–78.

Dewan, E. & Beran, M. (1959). Note on Stress Effects Due to Relativistic Contraction. *American Journal of Physics 27*, 517-518.

Hacking, I. (1993). Do Thought Experiments Have a Life of Their Own? Comments on James Brown, Nancy Nersessian and David Gooding. Pp. 302-308 in *PSA 1992*. Volume 2. Edited by D. Hull, M. Forbes, and K. Okruhlik. East Lansing: Philosophy of Science Association.

Howard, D. (2007). Revisiting the Einstein-Bohr Dialogue. *Iyyun: The Jerusalem Philosophical Quarterly 56*, 57-90.

Huggett, N., and Hoefer, C. (2009). Absolute and Relational Theories of Space and Motion, *The Stanford Encyclopedia of Philosophy* (Fall 2009 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/fall2009/entries/spacetime-theories/>.

Ierodiakonou, K. (2011). Remarks on the History of an Ancient Thought Experiment. In K. Ierodiakonou and S. Roux (Eds.), *Thought Experiments in Methodological and Historical Contexts* (pp. 37-49). Leiden and Boston: Brill.

Laymon, R. (1978). Newton's Bucket Experiment. *Journal of the History of Philosophy*. 16 (4), pp. 399-413.

Mach, E. ([1883] 1919). *The Science of Mechanics: A Critical and Historical Account of its Development*. (T. J. McCourmack, Trans.). Chicago and London: Open Court Publishing.

Mach, E. ([1926] 1976). *Knowledge and Error: Sketches on the Psychology of Enquiry*. Dordrecht: D. Reidel Publishing Company.

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

Mulkay, M., & Gilbert, G. N. (1986). Replication and Mere Replication. *Philosophy of the Social Sciences 16*(1): 21-37.

Nersessian, N. (1993). In the Theoretician's Laboratory: Thought Experimenting as Mental Modeling. In D. Hull, M. Forbes, and K. Okruhlik (Eds.), *PSA 1992, Volume 2* (pp. 291-301). East Lansing: Philosophy of Science Association

Nersessian, N. (2008). *Creating Scientific Concepts*. Cambridge, Mass: MIT Press.

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

Norton, J. (1996). Are Thought Experiments Just What You Thought? *Canadian Journal of Philosophy 26*(3), 333-366.

Norton, J. (2004a). “On Thought Experiments: Is There More to the Argument?” *Proceedings of the 2002 Biennial Meeting of the Philosophy of Science Association, Philosophy of Science, 71*: 1139–1151.

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

Palmerino, C. R. (2011). Galileo’s Use of Medieval Thought Experiments. In K. Ierodiakonou and S. Roux (Eds). *Thought Experiments in Methological and Historical Contexts* (pp.101-125). Leiden and Boston: Brill.

Radder, H. (1993). Experimental Reproducibility and the Experimenters' Regress. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association, Volume One: Contributed Papers:* 63-73.

Radder, H. (1996). *In and About the World: Philosophical Studies of Science and Technology*. Albany, N.Y.: State University of New York Press.

Reiss, J. (2002). *Causal Inference in the Abstract or Seven Myths about Thought Experiments*. Technical report (CPNSS Research Project), CTR 03/03. Centre for Philosophy of Natural and Social Sciences, London School of Economics and

Political Science. London, UK.
http://www.lse.ac.uk/CPNSS/pdf/DP_withCover_Causality/CTR03-02-3.pdf

Rowbottom, D. P. (2014). Intuitions in Science: Thought Experiments as Argument Pumps. In A. Booth and D. Rowbottom (Eds). *Intuitions*. Oxford University Press.

Schickore, J. (2011). The Significance of Re-Doing Experiments: A Contribution to Historically Informed Methodology. *Erkenntnis 75*(3), 325-347.

Sorensen, R. (1992a). *Thought Experiments*. Oxford: Oxford University Press.

Sorensen, R. (1992b). Thought Experiments and the Epistemology of Laws. *Canadian Journal of Philosophy*, 22, 15-44.

Stich, S., and W. Buckwalter. (2011). Gender and the Philosophy Club. *The Philosophers' Magazine 52*, 60–65.

Walton, K. (1990). *Mimesis as Make-believe*. Cambridge, MA: Harvard University Press.

Weinberg, J. M.; Nichols, & S. Stich, S. P. (2008). Normativity and Epistemic Intuitions. In J. M. Knobe and S. Nichols (Eds), *Experimental Philosophy* (pp. 17-45). Oxford and New York: Oxford University Press

---

[1] For the history of Lucretius' spear thought experiment see Ierodiakonou (2011) and for Bohm's modification of EPR see Bohm (1951, pp. 610-623) and references therein.

[2] Norton (2004a,b) defines as thought experiment–anti-thought experiment pairs any two thought experiments that reach opposite conclusions (one of his examples is Aristotle's

thought experiment involving a rotating pointer to show that space is finite and Archytas's thought experiment involving going to the edge of the universe and sticking out your arm to show space is infinite).

[3] As James Robert Brown notes, this can occur for a variety of reasons: even if they accept the same premises, thought experiments can fail to reach the same conclusion either because they disagree about what phenomenon will be observed, or they agree on the outcome, but disagree on the proper lessons to draw from that outcome (2007, p.158).

[4] This brief summary glosses over some historical interpretive issues that are for example more fully dealt with in Huggett and Hoefer (2009) and Laymon (1978).

[5] A t-argument is just the argument that is supposed to be identified with a thought experiment (Bishop 1999, p.535).

[6] Arthur's (2013) analysis of the *Rota Aristotelica* and Brown's (2013) study of the Terrell effect offer further examples of cases where thought experimenters disagree about which phenomenon is being explored in the thought experiment.

[7] As we saw before, Sorensen allows that in other contexts, such as in establishing the priority of historical authorship, more lenient standards for identity may apply.

[8] Note that one can endorse the Reconstruction Thesis, without further maintaining that thought experiments are *identical* to these arguments.