# Introduction

**Yiftach Fehige**
**Michael T. Stuart**
*University of Toronto*

*This special issue is the outcome of a workshop entitled, "Thought Experiments in Science: Four Blind Spots," held at the University of Toronto, March 23rd, 2012. The recent revival in philosophical study of thought experiments has mostly been limited to fields like epistemology, science studies, and meta-philosophy. With this issue we hope to facilitate a discussion about how some other disciplinary perspectives might bear on the subject; specifically, the history of philosophy, literary studies, phenomenology and cognitive science.*

Most people are aware of at least one thought experiment, whether it's Archimedes in the bathtub, Schrödinger's cat, or Einstein chasing after a beam of light. Thought experiments are used extensively in the so-called "hard sciences," but they are also common in philosophy, history, economics, political science, and mathematics. They are important to a number of fields for a number of reasons, including entertainment, pedagogy and conceptual exploration.

Apparently, thought experiments also help us to make scientific *progress:* they provide evidence for claims and play a role in theory choice. These are among the most philosophically interesting, and by far the most controversial uses of thought experiments. Whether thought experiments are used controversially or not, however, the historian of science at least "must recognize them as an occasionally potent tool for increasing man's understanding of nature" (Kuhn 1964, 240); and the philosopher of science should acknowledge that thought experimentation "is part of accepted scientific practice" (Sorensen 1992, 19).

We would like to thank Catherine Z. Elgin, Walter Hopp, and Paul Thagard for contributing original papers, as well as Geordie McComb, Mark Shumelda, and Harald Wiltsche for thoughtful responses to these papers. James R. Brown is to be thanked for his relentless support and generosity in kindness and collegiality.

Perspectives on Science 2014, vol. 22, no. 2
 doi:10.1162/POSC_e_00126

But, did Galileo's imaginary falling bodies really refute Aristotle's theory of motion in favor of a successor theory? Was there a role to play for Newton's bucket in his proof for absolute space? How can it be that Einstein took support for his special theory of relativity from chasing a beam of light in his imagination? These questions have led to a profound epistemological challenge: If thought experiments can yield evidence powerful enough to influence theory choice, we must explain the source and means by which they do so, given the task of enlarging our understanding of science.

This epistemological challenge has guided the philosophical debate about thought experiments for more than 25 years now. Since the papers in this issue are meant to stimulate discussion on *new* avenues for addressing the topic, we think a short summary of the most often discussed positions would be helpful.

James R. Brown's 1986 paper, "Thought Experiments since the Scientific Revolution," is partly responsible for the attention thought experiments are now receiving, for at least two reasons. First, back in 1986 it was possible to become an expert on thought experiments in just a long weekend (Brown 2011, 66). This void of philosophical interest is certainly a reason why Brown's initiative resulted eventually in an explosion of papers, special journal issues and books. Brown's phenomenology of the cognitive power of some thought experiments tempted many to take a closer look at them. Contributions by others coming out in print at about the same time informed those inquiries. The initiative of Tamara Horowitz and Gerald Massey in putting together a conference on thought experiments in 1986 also deserves special mention. Their volume *Thought Experiments in Science and Philosophy* (1991) resulted from it, and contains a number of articles that to this day inform the discussion of crucial aspects of thought experiments, including the very first assessment of thought experiments in Medieval thought by Peter King, an analysis of the role of thought experiments in Darwin's *On The Origin of Species* by James G. Lennox, and John D. Norton's first defense of his view that thought experiments are nothing but arguments.

Second, since 1986 Brown has been defending a Platonic theory of thought experiments. According to Brown, some thought experiments can help us even to "see" new laws of nature. That is, thought experiments can facilitate an intellectual perception of abstract entities with the mind's eye. Brown commits himself to a strong metaphysical realism about universals, a non-epistemic theory of truth, and a non-causal theory of knowledge to support his Platonism about thought experiments. "Absurd," "highly implausible," "provocative," and "funny" are some of the keywords that have appeared in response to his account of thought experi-

ments. Yet Brown's ideas are still stimulating and *alive* (in the sense of James 1897). His *Laboratory of the Mind,* a monograph that has been a mainstay in the literature, just received a second edition about twenty years after the first (Brown [1991] 2010).

Since the publication of *The Laboratory of the Mind,* thought experiments have received the philosophical attention we think they deserve. What is especially intriguing is that Brown could not convince anyone, but neither could Brown's main opponent, John D. Norton. This is surprising given the austere appeal of Norton's account of thought experiments, which identifies thought experiments as arguments (see Norton 1996).

Norton's account has three main sources of support. First, it is *empiricist,* which is important if Norton is right that empiricism is "overwhelmingly the predominant epistemology in philosophy of science" (2004b, 50). Second, Norton argues that all thought experiments (without exception) can be reconstructed as arguments. Finally, every thought experiment is always justified to the exact same degree as its reconstructed argument. Few are convinced by Norton's account. For a sample of the arguments against it, see Cooper 2005; Davies 2007; Gendler 1998; Häggqvist 2009, Miščević 2007; Nersessian 1992, 2007; and Hopp [this issue].

Given the widespread rejection of both Norton's and Brown's accounts, the discussion turned to how we might be able to explain the power of thought experiments in a way that captures what is correct in the two views. What follows is an outline of some of the compromises to be found in the literature.

Timothy Williamson (2007, 2009) emphasizes the modal character of thought-experimental knowledge. Williamson agrees with Norton that thought experiments can be represented as valid arguments, but he notes that to achieve soundness we also need true premises. Very often, those premises involve modal claims, which require the use of counterfactual reasoning to confirm. This is where thought experiments become useful. For example, philosophers believe that if knowledge is justified true belief, it is so necessarily. In this case, simply finding a *possible* case like Gettier's is enough to refute it. Thus, while thought experiments play a role in arguments, they are not necessarily arguments themselves.

This connects to the earlier thesis of Sören Häggqvist (1996, 2009), who agrees with Norton that insofar as thought experiments play a justificatory role, they are argumentative. That is, thought experiments are often used to contest the claims of some theory by providing (usually modal) evidence that counts against it. For Häggqvist, the thought experiment plays a justificatory role in the same way that a real experiment

does: by contradicting a claim made by the theory. This is argumentative, but only in a more general sense. Häggqvist is not committed to the claim that the performance of a thought experiment is the performance of an argument, or that a thought experiment could be formally valid or invalid. The insight, taken up by Tim De Mey (2003), is that we should investigate the epistemic impact of the thought experiment's conclusion in one way, and the epistemic nature of the thought experiment itself, in another.

Roy Sorensen (1992), like Williamson, sees thought experiments as a type of modal reasoning, and places thought experiments on a continuum with real experiments. Along with Thomas Kuhn (1964), Tamar Gendler (1998, 2002, 2004, 2007), Alisa Bokulich (2001) and Richard Arthur (1999), Sorensen argues that thought experiments mostly eliminate irrationalities in our systems of thought. And like Ernst Mach (1905) he claims that they work using innate ideas and structures that are preprogrammed into us by evolution, and also drawing upon the stores of empirical and conceptual knowledge that we've accumulated in our personal lifetimes.

Tamar Gendler brings us even further from Norton by focusing on the performative aspect of thought experiments. Gendler argues that in a thought experiment, quasi-perceptual knowledge is gained via the construction and manipulation of "mental models." This is a technical term stemming from cognitive science (see Johnson-Laird 1983). The term was first applied to thought experiments independently and simultaneously by Nenad Miščević (1992, 2004, 2007) and Nancy Nersessian (1992, 1993, 2007, 2008), who agree that thought experiments are used to mobilize special skills of the experimenter which can be vaguely characterized as "knowledge how."

Marco Buzzoni introduced arguably the first Kantian account of thought experiments (developed in 2008, 20011a, 2011b, and 2013) to mediate between Brown and Norton. According to Buzzoni, thought experiments should be analyzed on two levels. First, looking at the actual performance of thought experiments within a discipline, Buzzoni agrees with Sorensen that thought experiments work much like real experiments. Namely, they use idealizations and representations, they put a question to nature, anticipate an answer, and employ the method of variation. However, we should not leave the issue here, according to Buzzoni, otherwise thought experiments could be eliminated in favor of real experiments or vice-versa. To see what separates them, we must consider thought experiments from a transcendental level, that is, we must see them as the condition of the possibility of real experiments. What he means by this is that thought experiments are instances of counterfactual reasoning, and such reasoning makes thought in general possible. Without our ability to ab-

stract and reason counterfactually, we could never devise hypotheses, experiments or instruments to test those hypotheses. Without ideal gases, frictionless planes, perfectly smooth surfaces and rigid bodies, we wouldn't know how to design and calibrate instruments. On the other hand, we would never have thought experiments without the knowledge gained from real experiments. This is how Buzzoni is able to express himself in Kantian terms by saying, "The (empirical) thought experiment without a real world experiment is empty, the real world experiment without a thought experiment is blind" (Buzzoni 2011, 102).

This summary leaves out many positions that deserve mention (see Brown and Fehige 2011); however, these are some of those that stand out as definitive in the growing debate about thought experiments. They form important rallying points in the discussion that contextualizes the papers of this issue. But there is still much to be done. In what remains of this introduction, we will provide a few specifics about the particular blind spots that motivated our workshop, and how the five papers address them.

### 1. Origins of the Philosophy of Thought Experiments

The history of the practice of thought experiments in all disciplines is becoming increasingly well known (see Kühne 2005, Rescher 1991, and Roux and Ierodiakonou 2011 for an excellent recent collection). However, a comprehensive discussion of the history of the *investigation* into thought experiments is still relatively absent in the literature. Several philosophers look back to Mach as the founder of this field, although we now know that the term originates with Hans-Christian Ørsted.

What remains is to look back at Ørsted's historical and philosophical environment, and examine the kinds of related notions that were in use at the time to see how the coming discussion would be shaped by them. And indeed, while at the end of the 18th Century the term "thought experiment" is not yet in circulation, a family of very similar notions are.

Yiftach Fehige and Michael T. Stuart find this family of notions to be embedded in reflections that *prima facie* seem highly relevant for a philosophical discussion of thought experiments. In their contribution they focus on three thinkers: Georg Christoph Lichtenberg (1742–1799), Georg Philipp Friedrich Freiherr von Hardenberg (Novalis) (1772–1801), and Immanuel Kant (1724–1804).

Kant was arguably the first to use, understand, and attempt to justify many of the types of thought experiments that we are familiar with today. Lichtenberg and Novalis developed some of Kant's ideas by working at the intersection of science and literature, using one to understand the other and vice-versa. Lichtenberg argued that the careful use of the subjunctive tense could be used experimentally to break apart conceptual connections

that limit scientific creativity. Novalis's Romanticism gave him a way to see experiment and art as two faces of the same coin. All three point to a greater emphasis on the connection between nature, reason and imagination through hypothetical or literary reasoning.

The key to understanding this connection lies with the power and unique position of the mind as a mediator. For Lichtenberg, Novalis and Kant, one way to unlock this power was by the use of fiction. This 18th century idealism and romanticism is reconstructed by Fehige and Stuart in relation to several modern accounts of thought experiments, showing that it can do equally well, and in some cases, better, to explain the cognitive efficacy of thought experiments.

### 2. Thought Experiments and Literary Fiction

Little attention has been paid to the literary character of thought experimental narratives, and how this feature links literary theory to science and philosophy. Catherine Elgin's contribution establishes an interesting analogy between thought experiments, fictions, and real experiments. Each of these devices idealize, abstract, narrate, and most importantly, provide new understanding by what Elgin calls "exemplification." Exemplification is "the relation of a sample, example, or other exemplar to whatever it is a sample or example of." Something that exemplifies is dually referential in that it refers to the pattern or property or relation that it exemplifies, but also to other things that instantiate that pattern, have that property, or stand in that relation. A splotch of red paint on a paint sample card exemplifies the color red, if that's how we choose to use it, and thereby represents all instances of the color red. By acquainting ourselves with this paint sample card, we've gained an ability to deal with red things.

Here is an example of how it works in a real experiment. The Miller-Urey experiment created organic compounds from inorganic ones in a controlled laboratory setting which we believe more or less accurately represented the conditions on early Earth. Elgin argues that this argument exemplifies a path from inorganic to organic compounds under those conditions. By doing this, it also represents and informs us about all the other instances in which such a transformation takes place. Namely, since the experiment tells us which chemicals were sufficient and in what proportions, and how they interacted, and what molecules were produced, it tells us about these features of the other instances in which life emerged from similar conditions.

Thought experiments work in a similar fashion: we try to instantiate and manipulate exemplars, using constraints that we think accurately reflect those in the real world (the laws of nature). But unlike real experiments, we cannot exemplify physical things in these imaginary scenarios.

Rather, what thought experiments exemplify are *abstract* patterns, properties and relations. These might be theoretical, conceptual, modal, moral, metaphysical or psychological. This is not to limit our understanding to matters about theories and concepts, however, since it is through these that we say something about the world.

This brings us to fiction. In both thought experiments and literary fictions you have suspension of disbelief. You grant some propositions, even though you know they are false or even impossible. And while the thought experiments in science share much in common with real experiments, thought experiments in philosophy share much in common with fictions. For example, they have less background assumptions, and they are to that extent less theory-dependent. But fictions also enhance our understanding by exemplification. Since exemplification can draw attention to things that are normally concealed by the messiness of real life, we can directly confront things that we normally could not. In real life, for example, we can never be sure of the motive underlying an action, but in literature we can. Thus we are exposed to unequivocal examples of moral actions, which real life cannot provide.

With reference to *Huckleberry Finn, To Kill a Mockingbird, Oedipus Rex, A Tale of Two Cities, Lolita, Middlemarch, The Matrix, 2001: A Space Odyssey, Henry V, King Lear, Hamlet, Animal Farm,* and *Uncle Tom's Cabin,* Elgin explains how fictions and thought experiments can be used reliably to expand our imaginations, develop our ability to entertain the same scenario from different perspectives, and learn what is, might be, must be, or ought to be the case.

### 3. Phenomenology and Thought Experiments

During the last couple of years, phenomenology has attracted more and more attention, particularly from cognitive scientists and philosophers of mind. In these fields, phenomenology is regarded as a viable expansion of existing analytical approaches. Perhaps a similar broadening of perspective is also possible, maybe even necessary, in the case of the philosophical debate on thought experiments.

While not the first to look at the relationship between phenomenology and thought experiments, Walter Hopp provides arguably the first genuinely phenomenological approach in accounting for the cognitive power of thought experiments. He begins his discussion by considering the following questions: what sorts of intentional acts must one perform in order to carry out a thought experiment? What sorts of objects are such acts directed toward, and how are those objects made present, or not, in carrying out those acts? Hopp argues that in order to find out, we must begin with a careful and initially metaphysically unbiased phenomenological descrip-

tion of what happens when we perform a thought experiment. He argues that in every case of knowledge, we have a judgment that something is the case, and a presentation that it is as we think it is. He calls this "fulfillment." One type of fulfillment that often takes place in science occurs when scientific judgments are fulfilled by presentations that stem from perception. These presentations may be checked against the corresponding scientific judgments with relative ease. But with thought experiments we face an important dilemma.

Suppose that thought experiments can provide evidence for propositions. Since they do not present (or try to present) actual states of affairs via perception, we have two options for what they might be doing. First, thought experiments have some *other* way of presenting us with actual states of affairs. Second, thought experiments are not directly concerned with actual states of affairs. The first route would be like Norton's, or the one taken by those who identify thought experiments with mental models, in which the empirical content comes from memory which is grounded in experience. After arguing against this solution to the dilemma, Hopp adopts the second: thought experiments are about (and provide evidence for) claims involving abstract universals and the relations between them. In this he agrees with Elgin and Brown. For example, when we follow Gettier's reasoning, we are not aiming at the *concept* of knowledge, but *knowledge* (the abstract universal) itself. We wouldn't care that Galileo's thought experiment applied to this rock or that one, except insofar as those rocks instantiated the relation between the universals *speed of free-fall* and *weight.*

Hopp ends by tackling the causal theories of knowledge and reference, and distancing himself from Brown's Platonism. As Hopp's is arguably the first genuinely phenomenological approach to thought experiments, it remains to be seen whether future accounts will follow his.

Elgin and Hopp both push back against Norton's claim that if thought experiments aren't arguments, they must be something mysterious or magical. Here we have two accounts, each of which offers to explain our epistemic access to something like Platonic universals, without siding completely with Brown's Platonism. This is an exciting new path for anyone dissatisfied with Norton's empiricism, yet unwilling to go as far as Brown.

### 4. Cognitive Science and Thought Experiments

Several philosophers have already pointed out a potential connection between cognitive science and thought experiments, suggesting that thought experiments can be portrayed as mental models. But a great deal of work still needs to be done. For instance, what are the mental and neu-

ral mechanisms that underlie our use of mental models? Can cognitive science explain the seemingly a priori nature of thought experimental knowledge? What exactly is the role of memory and experience in the derivation of a thought experimental conclusion? What philosophical assumptions need to be in place before we can straightforwardly apply what we have learned in cognitive science to thought experiments?

In his paper, Michael T. Stuart claims that a way of sorting out these assumptions is to address the new skeptical claim made by Paul Thagard, which derives from cognitive science. In *The Brain and the Meaning of Life,* Thagard claims that "the made-up thought experiments favored by many philosophers are not evidence at all" (2010, 209). Rather, "Philosophical attempts to establish truths by a priori reasoning, thought experiments, or conceptual analysis have been no more successful than faith-based thinking has been. All these methods serve merely to reinforce existing prejudices" (2010, 41). If this is what cognitive science proclaims, then one of the most hopeful avenues for understanding thought experiments is closed. Further, most of the people who have worked on thought experiments would turn out to have been misguided for concerning themselves to explain a tool of reasoning that is not only ineffective, but misleading and harmful. Stuart's paper aims to refute Thagard's arguments against the use of thought experiments as evidence, and show, at the same time, how Thagard's work might be used to set the stage for future research in a constructive way.

In his reply, Paul Thagard reaffirms his position and provides additional support coming from research on human intuition. He argues that philosophers overestimate the significance of thought experiments by supposing that they can provide evidence that supports the acceptance of beliefs. This is because thought experiments rely principally on intuitions, and intuitions are neural processes that are poorly suited to provide evidence for beliefs. Results in cognitive science are marshaled to show that through a combination of representation, binding, and interactive competition, people make sense of a story in a way that generates a reaction in the form of the intuitive sense that a claim is true or false, or a state of affairs is good or bad. This reaction is more or less sensitive to the real world depending on whether the thought experimenter has the right kind of experience with the subject of the experiment. Thagard argues that in the realm of philosophy at least, the relevant sort of experience is almost always missing.

Thagard finishes by putting forward an ameliorative proposal for philosophers to work within the proper bounds specified by cognitive science—without thought experiments—which he dubs "natural philosophy."

The most important outcome of the debate between Stuart and Thagard is probably greater clarity with respect to the philosophical assumptions that would need to be true in order for cognitive science to be fruitfully used to explain the use of thought experiments in science. For instance, there is much that needs to be said about the connection between patterns of neural activity in populations of neurons and the creation and use of abstract mental models; the role of memory, perception and emotion in mental modeling; narrative and a cognitive "order of operations"; and how cognitive science bears on the status of the a priori, innate ideas, intuitions, and the conditions under which these are to be treated as reliable.

**References**

Arthur, Richard. 1999. "On Thought Experiments as A Priori Science." *International Studies in the Philosophy of Science* 13:215–229.

Bokulich, Alisa. 2001. "Rethinking Thought Experiments." *Perspectives on Science* 9:285–307.

Brown, James R. 1986. "Thought Experiments since the Scientific Revolution." *International Studies in the Philosophy of Science* 1:1–15.

——— (1999) 2010. *The Laboratory of the Mind,* New York; London: Routledge.

——— 2011. "Über das Leben im Labor des Geistes." *Deutsche Zeitschrift für Philosophie* 59:65–73.

Brown, James R. and Yiftach Fehige. 2011. "Thought Experiments." In *Stanford Encyclopedia of Philosophy.*

Buzzoni, Marco. 2008. *Thought Experiment in the Natural Sciences.* Würzburg: Königshausen & Neumann.

——— 2011a. "Kant und das Gedankenexperimentieren." *Deutsche Zeitschrift für Philosophie* 59:93–107.

——— 2011b. "Mathematical Thought Experiments." *Epistemologia* XXXIV:61–88.

——— 2013. "Thought Experiments from a Kantian Point of View." Pp. 90–106 in *Thought Experiments in Science, Philosophy and the Arts.* Edited by James R. Brown, Letitia Meynell, and Mélanie Frappier. New York: Routledge.

Cooper, Rachel. 2005. "Thought Experiments." *Metaphilosophy* 36:328–347.

Davies, David. 2007. "Thought Experiments and Fictional Narratives." *Croatian Journal of Philosophy* 19:29–45.

De Mey, Tim. 2003. "The Dual Nature View of Thought Experiments." *Philosophica* 72:61–78.

Elgin, Catherine. 2011. "Making Manifest: Exemplification in the Sciences and the Arts." *Principia* 15:399–413.

Gendler, Tamar. 1998. "Galileo and the Indispensability of Scientific Thought Experiment." *British Journal for the Philosophy of Science* 49:397–424.

——— 2002. "Personal Identity and Thought-Experiments." *The Philosophical Quarterly* 52:34–54.

——— 2004. "Thought Experiments Rethought—and Reperceived." *Philosophy of Science* 71:1152–1163.

——— 2007. "Philosophical Thought Experiments, Intuitions, and Cognitive Equilibrium." *Midwest Studies in Philosophy* XXXI:68–89.

Häggqvist, Sören. 1996. *Thought Experiments in Philosophy.* Almqvist & Wiksell International, Stockholm.

——— 2009. "A Model for Thought Experiments." *Canadian Journal of Philosophy* 39:55–76.

James, William. 1897. *The Will to Believe and Other Essays.* London: Longmans, Green and Co.

Johnson-Laird, Philip N. 1983. *Mental Models.* Cambridge: Harvard University Press.

Kuhn, Thomas S. 1964. "A Function for Thought Experiments." Pp. 240–265 in *The Essential Tension.* Chicago: University of Chicago Press.

Kühne, Ulrich. 2005. *Die Methode des Gedankenexperiments.* Frankfurt: Suhrkamp.

Mach, Ernst. 1905. "On Thought Experiments." Pp. 134–147 in *Knowledge and Error.* Translated by Thomas J. McCormack. Dordrecht: Reidel.

Miščević, Nenad. 1992. "Mental Models and Thought Experiments." *International Studies in the Philosophy of Science* 6:215–226.

——— 2004."The Explainability of Intuitions." *Dialectica* 58:43–70.

——— 2007. "Modelling Intuitions and Thought Experiments." *Croatian Journal of Philosophy* VII:181–214.

Nersessian, Nancy. 1992. "How Do Scientists Think? Capturing the Dynamics of Conceptual Change in Science." Pp. 3–44 in *Cognitive Models of Science.* Edited by Ronald N. Giere. Minneapolis: University of Minnesota Press.

——— 1993. "In the Theoretician's Laboratory: Thought Experimenting as Mental Modeling." *Proceedings of the Philosophy of Science Association* 2:291–301.

——— 2007. "Thought Experiments as Mental Modelling: Empiricism without Logic." *Croatian Journal of Philosophy* VII:125–161.

——— 2008. *Creating Scientific Concepts.* Cambridge: MIT Press.

Sorensen, Roy A. 1992. *Thought Experiments.* Oxford: Oxford University Press.

Rescher, Nicholas. 1991. "Thought Experiments in Presocratic Philosophy." Pp. 31–42 in *Thought Experiments in Science and Philosophy.* Edited by T. Horowitz and G. Massey. Lanham, MD: Rowman & Littlefield.

Norton, John D. 1996. "Are Thought Experiments Just What You Thought?" *Canadian Journal of Philosophy* 26:333–366.

——— 2004. "Why Thought Experiments Do Not Transcend Empiricism" Pp. 44–66 in *Contemporary Debates in the Philosophy of Science.* Edited by Christopher Hitchcock. Somerset: Wiley-Blackwell.

Ierodiakonou, Katerina, and Roux, Sophie (eds.). 2011. *Thought Experiments in Methodological and Historical Contexts.* Leiden: Brill.

Williamson, Timothy. 2007. *The Philosophy of Philosophy.* Oxford: Blackwell.

——— 2009. "Replies to Kornblith, Jackson and Moore." *Analysis Reviews* 69:125–135.