
On the Origins of the
Philosophy of Thought
Experiments:
The Forerun

Yiftach Fehige
Michael T. Stuart
University of Toronto

The history of thought experiments is now gaining a great deal of attention, and this is due to the renewed interest of philosophers on the subject. This paper inquires into the history of the philosophy of thought experiments. We name the period to be examined in this paper the “forerun.” Its main stakeholders are Georg C. Lichtenberg, Novalis, and Immanuel Kant. We will present and discuss the work of each of them in order to characterize the period, and then reveal parallels and lessons that apply to more recently proposed accounts of thought experiments.

Philosophical debate about the nature and function of thought experiments would be impoverished without good historical sources. And while valuable work is being done on the history of thought experiments, a comprehensive discussion of the history of *philosophical investigation* into thought experiments is still absent in the literature (but see Kühne 2005; Moue et al. 2006). In what follows we take the first steps towards providing a more complete picture of the diverse attempts to shed light on thought experiments.

The term “thought experiment” made its first appearance about 200 years ago in 1811. The most prolific period in the history of the philosophical investigation into thought experiments is the current one, which commenced about 175 years later, more or less in 1986. The lack of focus and direction in the ongoing debate over the nature and usefulness of thought experiments is our main motivation to revisit the history of philosophical investigation into thought experiments. We are confident that this enterprise can help to correct some of the shortcomings in current approaches to thought experiments, including a lack of appreciation of the literary features of thought experiments, the normativity of scientific instances in assessing the nature of thought experiments, and a ten-

dency to marginalize the historical contingency of the evidential significance any thought experiment may provide. Such an enterprise is also important because philosophical theories about thought experiments can impact the work of historians, and we think historians should have as many interpretative tools made available to them as possible.

This paper will focus on the period leading up to Ørsted's baptism of the term "thought experiment" in 1811. We begin by characterizing the period, and then introducing and discussing in subsections its main stakeholders. The discussion of each figure will be structured in the same way. First, we will provide any necessary historical or philosophical background information at the beginning of that figure's subsection. Then we will identify the concept of thought experimentation that is being proposed (*The Notion of Thought Experiment*), before summarizing the crucial elements of the particular account of thought experiments (*The Nature of Thought Experiments*). At the end, we will highlight the strong and weak points in each reconstructed account, and establish what we deem to be the most important parallels to other periods and thinkers, both earlier and later (*Discussion*).

We have called the period under discussion in this paper the *forerun* because the term "thought experiment" was not yet in circulation. However, as we are about to show, a family of very similar notions was in circulation during this period, and they are embedded in reflections that *prima facie* seem highly relevant for a philosophical discussion of thought experiments. We should also make mention of the fact that, between 1750 and 1830, the use of experiments became (a) an essential part of university curriculum, (b) popularized in science books written for the laity, and (c) a primary source of evidence for or against scientific theories (see Daiber 2001, pp. 294–295). Yet, at this stage, pure empirical science limited to observation and experiment had not yet fully emerged (see Beiser 2003, p. 156). These are important facts to be considered when studying thought experiments, given (a) the important pedagogical role thought experiments can play, (b) the heuristic value of thought experiments for popularizing science, and (c) the role of thought experiments in science.

There are three thinkers of the forerun, Georg Christoph Lichtenberg (1742–1799), Georg Philipp Friedrich Freiherr von Hardenberg (better known as Novalis) (1772–1801), and Immanuel Kant (1724–1804). Novalis is most intelligible if we first discuss Lichtenberg. And while both philosopher-scientists are actually influenced by the Kantian enlightenment, we discuss Kant last to ensure a smooth transition to the Baptismal period, which is heavily indebted to Kant and will have to be the focus of a different paper.

I. Lichtenberg's Experiments with Thoughts and Ideas

Lichtenberg began his career in science and later turned to literature (see Stern 1963, p. 110). During his lifetime, he was recognized as an important figure in European science, and was admired as a man of letters by some as influential as Goethe (see Mautner and Miller 1952, p. 223). Yet, Lichtenberg faded from the consciousness of scientists in the first decades of the 19th century. "The reason for Lichtenberg's comparative failure as a scientist is . . . that he is ceaselessly engaged in giving accounts of the ultimate validity of his investigations" (Stern 1963, p. 76). As a result, he survived as a man of letters and philosophical insight who attempted a synthesis of both "the imaginative and the scientific modes of thought . . . Chronologically speaking, it is one of the last attempts in the literature of the West" (Stern 1963, pp. 100–101).

Lichtenberg's fame is due to the influence of what he called his "Waste Books"¹ (see E46²). These posthumously published notebooks are filled with highly entertaining and insightful aphorisms. Lichtenberg's "best thoughts are contained in, and due to, his aphorisms and reflections" (Stern 1963, p. xii), and he "is among the last scientists to use the scientific aphorism for recording their findings" (Stern 1963, p. 101). There is a transition in Lichtenberg from scientific aphorisms, whose history goes back to Hippocrates (see Hollingdale 1990, pp. ix–x; Stern 1963, pp. 104–110), to the aphorism as a genre in literature (see Stern 1963, p. 110). In the 20th century, Lichtenberg's aphorisms continued to arouse admiration in philosophers like Ludwig Wittgenstein, about whom it was said "the resemblance between some of his statements and those of Lichtenberg is astounding" (Mautner; Hatfield 1959, p. 3). The main reason, however, that he is of interest in the present context is that we find in Lichtenberg an *experimenter of thoughts and ideas*.

I.1 The Notion of Thought Experiment: Lichtenberg

Here are three examples of Lichtenberg's "experiments with thoughts and ideas" (see K308):

1. The anglophile Lichtenberg used the English phrase "waste book," an eighteenth-century mercantile term for an account book in which transactions are roughly recorded, to be posted afterwards into the more official books of the same kind.

2. The Waste Books are a collection of aphorisms, or better: "a very disorderly series of notebook jottings extending over thirty years" (Hollingdale 1999, p. xxiv). They are quoted by the letter assigned to a book, namely A–L, and the number of the aphorism within that book. So "E46" can be found in Book E, aphorism 45. Unless otherwise indicated, citations have been translated by the authors. For the original text, see Lichtenberg (1968) and (1971).

If there was a shaft drilled through the center of the earth, then without difficulty you could jump inside, and if the air did not kill you as you traversed the core of the earth, you would accelerate in such a way that you would fall through the shaft to its other end and arrive there comfortably. (A200)

What would happen if I put an ounce of alcohol in a pressure cooker and then ignited the vaporized alcohol? It is probably best to try this first in your yard, in the open! (J1733)

What kind of movement would a planet have if its center of gravity changed in accordance with a certain law? (A201)

To begin our brief discussion of these examples, the second thought experiment can be realized while the remaining two cannot. And there is also an interesting difference between the first and the third in terms of the type of impossibility considered: Only in the latter do we see Lichtenberg invoking a *theoretical impossibility* for the sake of *cosmological speculation*, while the former uses a *practical impossibility* to illustrate a law of nature. More than the other two thought experiments, the third exemplifies Lichtenberg's conviction that it takes experiments with thoughts and ideas to make *scientific progress*. As we will see, scientific progress for Lichtenberg consists in the elimination of errors, not the confirmation of truths. Thought experiments and real-world experiments alike play a role in such elimination, and therefore in scientific progress as well.

I.2 The Nature of Thought Experiments: Lichtenberg

There are two main uses of experiments for Lichtenberg. First, real-world experiments matter for ruling out errors in reasoning. Second, experiments with thoughts and ideas matter for hypothetical speculation, which, according to Lichtenberg, is an important part of scientific practice. The two kinds of experimentation go hand in hand. Lichtenberg understood science to be about *explaining* the phenomena of nature, and he granted that these explanations could emerge by the use of real-world experiments. This is not to say that real-world experiments simply demonstrate the truth of theoretical knowledge, as suggested by some of his contemporaries (like Johann-Georg Walch in his influential philosophical dictionary of 1775. See Schöne 1982, p. 56). Instead, they help to eliminate errors. But there will be no scientific progress at all if we do not “think up new errors” (L886). We come up with new errors by means of hypothetical speculation. So, while a hypothesis might be nothing more than a bold speculation, which is nullified in the moment that it contradicts the phenomena (see J1521), without it there is nothing to falsify.

And for Lichtenberg, the way we “think up new errors” is by means of experiments with thoughts and ideas.

The necessity of *hypothetical speculation* for *experimentation* can be illuminated by considering Lichtenberg’s example of observing and predicting the moves of a chess game:

One who merely accumulates observations and experiments is like someone who only records which chess figures are moved or removed by the players of a game. He or she is successful merely by noticing the moves made by the players . . . Yet it will take a long time before he or she realizes the intentions of the players that motivate their various moves, namely, to capture the king. Without hypotheses of this kind nothing can be accomplished. (J1521)

Without a system of intelligent guessing, we cannot understand what is going on around us. Thought experiments are probably at the heart of scientific creativity thus described. They are not a method in the strict sense of a path to knowledge, but the best we can do without a method of scientific creativity in order to escape habits of thought, or what Francis Bacon called the “Idols of the Mind” (Bacon 1901). In Lichtenberg’s case, the most damning Idol of the Mind is the stagnation of the intellect by inflexible, static concepts. Thought experiments work by “unfreezing” our concepts, and as such contribute significantly to the acquisition of knowledge.

The “freezing” of images in concepts is a by-product of the associations and categorizations established by the mind, which enable us to make meaningful observations. While this “freezing” of images is inevitable, it is not irreversible. It can be reversed by thinking subjunctively, and also by occasionally breaking the rules of grammar. Both help to “melt” the concepts that create the necessities in thought that can inhibit scientific progress. Thus, according to Lichtenberg, the use of complex subjunctives allows us to remove inhibitors to scientific progress by helping us overcome restrictive necessities in thought. In fact, Lichtenberg used complex subjunctives to explore what lay beyond the domain defined by accepted theories. They even helped him in the development of new scientific instruments and procedures.

Experimenting with thoughts and ideas is what it takes, according to Lichtenberg, to break habits of thought and to find new and perhaps better ways of bringing order into nature. It means to force things together that naturally don’t seem to fit, or sunder things apart that naturally come united. Things are not given to us without concepts, and this means that we must occasionally force concepts in new directions by “melting” them down in order to unleash the images that they contain.

The aim of experiments with thoughts and ideas is to create new images that help to break the habits of thought that inhibit scientific progress. Experiments with thoughts and ideas help us to see “what nobody before has seen or thought of” (J1363).

I.3 Discussion: Lichtenberg

Given that Lichtenberg used fire as a metaphor for the Enlightenment, we might compare thought experiments to fireplaces. In them the fire of the Enlightenment burns and refines all ideas subjected to their heat: “[Fire] gives light and warmth, it is indispensable for everything that lives in order to grow and progress, only—if handled without caution it can also burn and destroy” (J971). We must be careful to use thought experiments appropriately.

Both real-world and thought experiments serve Lichtenberg’s Enlightenment-inspired skepticism. Thought experiments allow us to question knowledge claims in a more controlled and systematic manner than does random criticism. Scientific creativity requires spontaneity, but also control. The combination of these two qualities makes criticism of entrenched knowledge possible, without leading to global skepticism. In the case of real-world experiments, such criticism takes the form of a dialectical relationship between hypothetical speculation concerning the possible outcome of an experiment, and the experimental operations performed physically to test hypothetical speculations. In other words, for Lichtenberg, real-world experimentation is always guided by hypothetical speculation concerning possible outcomes. If a predicted outcome does not obtain, then the hypothetical speculation was erroneous.

However, given this interpretation, we face a difficulty. If thought experiments are mere hypothetical speculations, why did Lichtenberg think that what he called “experiments with thoughts and ideas” were *experiments* at all? Lichtenberg claimed that the aim of real-world experiments is to rule out errors, yet thought experiments only point towards possible outcomes of real experiments. Some of the examples above nicely illustrate this fact. Assuming that Lichtenberg is coherent, his experiments with thoughts and ideas are either not genuinely experimental, or there is some other factor that gives them and real-world experiments alike their experimental character. We will explore the latter option.

One potential candidate for this “something else” is what may be called “forceful combination.” As we saw above, Lichtenberg argues that thought experiments allow us to forcefully combine ideas and thoughts that are not normally combined. In physical experiments, too, one “forces together in a single day phenomena which, if one were passively to wait for them, would take a thousand years of careful observation” (in the trans-

lation of Stern 1963, p. 90). With respect to “forceful combination,” thought and physical experiments are analogous. Yet this cannot be the factor that makes experiments with thoughts and ideas experimental. This is because while both real and thought experiments employ forceful combination, they do so for different reasons. Real experiments use forceful combination for the sake of falsification—to expose scientific error. This is not their purpose in thought experiments.

To approach the problem under consideration from a different angle: If thought experiments do not falsify, and that is the main function of real-world experiments, how can thought experiments be experimental *in a way that carries any evidential weight?* We wish to discuss this objection in the following way. While there are hints in Lichtenberg that he was not a strict proponent of falsificationism (a good example is Lichtenberg’s outspoken and engaged endorsement of Georges-Louis Lesage’s highly speculative and unfalsifiable mechanistic explanation of gravity—see Hermann 1974, pp. 50–51), let us assume for the sake of argument that he *was* a falsificationist. Now we can ask if falsificationism is irreconcilable with the idea that thought experiments can provide evidence against a theory, and thereby share the aim of physical experiments, without falsifying anything.

Our answer is that falsificationism *is* consistent with this idea. Several reasons to think so are given by the arch-falsificationist Karl Popper, who discusses the ability of thought experiments to provide genuine evidence for or against a theory (and therefore play a role in theory choice) without falsifying anything. A quick look at Popper’s writing on the subject will clarify the relationship between thought experiments and falsificationism, which might illuminate the role Lichtenberg envisions for them.

Popper (1959) argues that thought experiments are fruitful in science, and he distinguishes among those employed *apologetically*, *heuristically* and *critically*. The apologetic use of imaginary experiments in defense of a theory is only admissible in science whenever it employs idealizations that are either concessions to someone who challenges the theory or are at least acceptable to the opponent. Admissible in any case are thought experiments that are used heuristically or critically. Critical thought experiments provide evidence against a theory by bringing to light the fact that the author of a theory overlooked certain possibilities. Such expansions of the space of possibility can falsify theories that attempt to explain phenomena by appeal to the only (or most likely) mechanism. This role allows thought experiments to be hypothetical speculations about the outcomes of physical experiments, but at the same time, provide falsifying evidence against a theory. We take this to address the objection that thought experiments lose their experimental nature if they do not falsify directly. What is miss-

ing in Popper is anything on how thought experiments can accomplish the feats he ascribes to them. Lichtenberg offers more in this respect, and to see what this is, we are brought to a parallel between his views and those of Thomas S. Kuhn.

Like Lichtenberg, Kuhn is committed to conceptual constructivism. Accordingly, Kuhn argues that thought experiments in science allow us to learn about the world and our concepts at the same time (see Kuhn (1964) 1977, p. 253). He supports this claim with four arguments, of which we name three: First, thought experiments present us with paradoxes. These paradoxes occur when we attempt to deploy a familiar concept in the context of an idealized but typical situation represented in a thought experiment. These thought experimental scenarios are designed to reveal contradictions in the criteria that are meant to guide the competent use of our concepts. A thought experiment is therefore more effective the more clearly it brings this conflict to light, perhaps by presenting a situation where several conflicting criteria seem to apply equally. The thought experiment exercises a force meant “to give incompatible answers to one and the same question” (Kuhn (1964) 1977, p. 254).

Second, thought experiments do more than simply bring to light confused concepts, since concepts have no context-independent standards for quality. The quality of a concept is a matter of applicability, which is partially determined by the theory in which the concept plays a part. What is exposed by a thought experiment is therefore not an intrinsic or logical defect of the concept itself, but a defect inherent in the combination of many concepts and their relation to the theory.

Finally, the reconceptualization resulting from the performance of a thought experiment can be the same as the reconceptualization resulting from a recurrent and persistent anomaly leading to a scientific revolution. Scientific revolutions are not about new data, but changes in paradigm. Scientific progress is a matter of reconceptualization. As such, reconceptualization means progress in our understanding of the world, because “when paradigms change, the world itself changes” (Kuhn (1962) 1996, p. 111). This view echoes Lichtenberg’s, that there is no mental access to the world but through our concepts. Or said another way, concepts have the imprint of our paradigms. In Kuhn’s famous words, “the proponents of competing paradigms practice their trades in different worlds” (Kuhn (1962) 1996, p. 150). Hence, for Lichtenberg and Kuhn alike, reconceptualizations resulting from thought experiments can enhance our understanding of the world. There is no need to recount the objections against Kuhn (see Sorensen 1992, pp. 111–131) or the possible ways in which to develop Lichtenberg’s and Kuhn’s conceptual constructivism (see Gendler 1998, pp. 415–420). We simply wish to make clear the possibil-

ity of reading Lichtenberg's experiments with thoughts and ideas as a meaningful and valuable contribution to the philosophical investigation of thought experiments that is ongoing.

One way we hope to have succeeded in this is by providing a consistent interpretation of Lichtenberg's work, one that is in some ways even an improvement on some more recent accounts, as it does not conclude by saying that thought experiments work by "a certain *constructive participation* on the part of the reader" (Gendler 1998, pp. 413–414). Rather, it promotes the use of specific mental tools—the controlled use of the subjunctive tense—to "unfreeze" concepts as a means of avoiding stagnation and encouraging scientific creativity.

As a result of the foregoing discussion, we may now ask why Lichtenberg's experiments with thoughts and ideas have not played the role they probably should have in current attempts to craft a comprehensive theory of thought experiments. This is most likely the result of several factors. First, aphorisms have come to "be regarded as a paradigm of literary decadence and aestheticism" (Stern 1963, p. 126). Second, few contributions in English to the ongoing discussion have been informed by the German literature on thought experiments. References to Lichtenberg can be found here and there, but there is no serious engagement. Third, almost no work has been done on the literary character of thought experiments. Thus there hasn't been much incentive to engage with Lichtenberg, a man of letters interested in grammar and speech, in a serious manner. The history of the current debate over thought experiments has made it difficult to get the literary component of thought experiments into focus. There are good reasons to believe that the current period of debate began in response to Brown's provocative Platonic account, in which the literary character plays no role. Nor does it matter for Brown's main opponent, John Norton, who even explicitly states that the particulars invoked by thought experiments are irrelevant for what they actually do (see Norton 1996, p. 336).

As the factors that might have caused Lichtenberg's contribution to disappear are incidental or accidental, we think that a close examination of Lichtenberg's work could prove promising for the study of thought experiments. Especially as such a study may provide additional arguments against the accounts of Brown and Norton. Lichtenberg has rationalist tendencies without Platonic elements, but would also clearly oppose the idea of reducing thought experiments to arguments while still promoting a kind of empiricism. And again, Lichtenberg could prove helpful in making clear the relationship between thought experiments and literary fiction, as well as answering the historical question of why there is so little research being done on the relationship between thought experiments and

literary fiction. This question is even more pressing, since it appears to us that there is a great amount of evidence to support the idea that the first thinkers to consider thought experiments philosophically worked at the intersection of science and literary fiction.

Finally, we think that a closer look at the work of Lichtenberg seems inevitable in order to understand the development of the notion of experiment in this period that, as it were, presages the epoch in which the term thought experiment emerges.

II Novalis's Poems of Productive Imagination

Like Lichtenberg, Novalis was a great humanist-scientist who came to lasting fame as a writer of outstanding literary and philosophical merit (see Beiser 2002, p. 408). His work brings us to German idealism, which lasted from approximately 1770 until 1840.

There is an emerging consensus “that early German romanticism was not only a literary but also a philosophical movement” (Beiser 2003, p. 1). This fact alone makes Novalis's attempt to unify science and art a most promising source for the study of thought experiments. This is true especially in light of recent claims that the cognitive efficacy of thought experiments is better understood given a close analysis of their aesthetic character (see Davies 2007, Macho and Wunschel 2004).

While there is no obvious relationship between Novalis and Lichtenberg, the latter's skepticism, resulting from his attachment to the Enlightenment, exemplifies very well the reason why German idealism emerged: it grew out of the crisis of the Enlightenment. That is to say, the two fundamental principles of the Enlightenment (rational criticism and scientific naturalism) seemed to lead to skepticism and materialism, both of which were widely considered philosophically unacceptable at the time. “There were few *Aufklärer* in Germany ready to admit such disastrous consequences; but there were also few willing to limit the principles of criticism and naturalism” (Beiser 2005, p. 18). German idealism was a philosophical movement to reach a middle ground.

Part of German idealism is *early German romanticism* (1797–1802). Novalis shaped it significantly. He was the one “who first declared the radical romantic manifesto in the striking sentence: “The world must be romanticized”” (Beiser 2003, p. 20). What he meant by this is difficult to say, but we can begin with Beiser, who interprets the motto as an attempt to give back *meaning, magic and mystery* to the world (Beiser 2003, p. 20). Such a romanticization of the world had become necessary due to the pre-revolution industrialization of modern culture, including its instrumental approach to nature: “The sciences are accused of having ceased to be human sciences . . . Their results are called meaningless, at least in so far as

they are concerned with the construction of an objective world detached from man” (Gode-von Aesch 1941, p. 27).

In Novalis, we find an organic conception of nature that facilitated his program to romanticize the world, and overcome the dualism between the mental and the physical, which had dominated the dispute between idealism and realism in the eighteenth century. Even more, it is the basis of the romantic doctrine that Novalis shared. The “doctrine consists in three fundamental propositions. First, there is a single universal substance in nature, which is the absolute. Second, this absolute consists in living force, so that it is neither subjective nor objective, but the unity of them both. Third, through its organic structure nature conforms to a purpose, plan, or design, which is not created by God but inherent in matter itself” (Beiser 2005, pp. 33–34). Contrary to widespread belief, this philosophy did not result in mere metaphysical speculation and dogmatism, or anything that was inconsistent with the contemporary idea of a natural science (see Beiser 2003, p. 156).

II.1 The Notion of Thought Experiment: Novalis

The notion of thought experiments present in Novalis is best defined as *poems of productive imagination* (this definition is inspired by Daiber 2001, pp. 65–67). They are indispensable for reaching “consilience” (in the sense of Wilson 1998), or in other words, the unity of all the sciences: “Novalis was a systematic philosopher, whose goal was to show how all the sciences form a unity” (Beiser 2002, p. 410). The unity of scientific knowledge in turn reflects the unity of nature (see Daiber 2001, pp. 109–110). Poems of productive imagination connect *writing on trial*, on the one hand, and *trials with nature*, on the other (see Daiber 2001, pp. 22–25) to achieve a *full experiment* (see Daiber 2001, pp. 110–111). Novalis reminds us that: “We will be physicists only if we use *imaginary* products and forces as regulative measures for *natural* products and forces” (Novalis 1978, p. 690, our emphasis).

Poems of pure imagination are a method in Novalis’s philosophy of nature, and can count therefore as a scientific method, because philosophy of nature was the science of its day—“not a metaphysical perversion of, or derivation from, ‘normal’ empirical science” (Beiser 2003, p. 156). Philosophy of nature was not opposed to the method of experiment and observation in studying nature.

Poesie, in the context of early German romanticism includes “works in prose” (Beiser 2003, p. 8). In addition, for the early German romantics, poetry denotes the ideal for all creative activity, whatever the medium, and whether or not expressed in language (see Beiser 2003, p. 12). Thus, according to the early German romantics, “all of nature and science should

become art, and . . . art should become nature and science” (Beiser 2003, p. 19). Romantic poetry is holistic in that it recreates the unity of all the arts and sciences, and re-establishes the unity of art and life (see Beiser 2003, p. 22).

In our choice of an exemplary poem of productive imagination, we follow Daiber (2001, pp. 227–233) in emphasizing *Heinrich von Aferdingen*, one of two novel fragments that Novalis left behind. Three of the characters of this novel are Fable, Eros, and Freya. We are interested in their appearance at the end of the first of the two parts of the novel fragment, namely within the so-called *Klingsobr's Fairy Tale* (see Novalis 1964, pp. 120–148). The fairy tale “demands an allegorical reading” (O'Brien 1995, p. 306). It takes place in the frozen realm of King Arcturus, and its “plot—insofar as the fairy tale has a plot—revolves around the kingdom's renewal, which is finally brought about by the marriage of the princess Freya and Eros” (O'Brien 1995, p. 305). The allegory of the fairy tale extends to a “universal renewal, including that of the sciences and nature” (O'Brien 1995, p. 306). At the moment of renewal, “the real and the fictitious, the true and the illusory, mysteriously interact” (O'Brien 1995, p. 310).

Of interest to us is the way in which Novalis has Freya come to life from a deep sleep that afflicted her. An

old hero received [Fable and Eros] at the gates of the palace. “Venerable Sir,” Fable said, “Eros needs your sword. *Gold has given him a chain, one end of which reaches down to the sea and the other is wound around his breast. Take hold of the chain with me and lead us into the hall where the princess is resting.*” Eros took the sword out of the hero's hand, set the hilt on his breast, and pointed the sword forward. The folding doors of the hall flew open, and Eros ecstatically approached the slumbering Freya. Suddenly there was a *loud discharge. A brilliant spark jumped from the princess to the sword; the sword and the chain grew luminous*; the hero held little Fable, who had nearly collapsed. The plume on Eros' helmet waved up. “*Throw the sword away,*” Fable cried, “*and awaken your beloved.*” Eros let the sword fall, flew to the princess, and fervently kissed her sweet lips. She opened her large dark eyes and recognized her beloved. A long kiss sealed the eternal union (Novalis 1964, p. 146, our emphasis).

Daiber (2001, pp. 227–233) reads this passage as a thought experiment, and we believe rightly so. The scientific background of the rise of Freya is *galvanism*. Novalis was familiar with speculations among scientists whether or not a return from death is possible by means of a galvanic arc. At the turn of the nineteenth century, they resulted in the perfor-

mance of “extraordinary experiments on human and animal corpses, attempting to animate and reanimate them under the influence of electricity from the voltaic pile” (Sleigh 1998, p. 220). Novalis modified the usual design for these experiments in his novel. Instead of using only one corpse, parts of it, or a liquid and two different metals to create a galvanic arc, he raised the question of the possibility of “galvanism among 2–3 and more people by means of metals” (Novalis 1960, volume III, p. 577). The idea was to create a galvanic arc between organic bodies and not metals, although by means of them, producing the usual effects of a galvanic arc, namely sparks and what looked like a return to life, or if you want, *of* life. This is exactly what happens in the scene quoted above.

There is first an incomplete galvanic arc, resulting from the golden chain (a metal) that reaches at one of its ends down to the sea (a liquid). Then Eros points the sword at Freya, who is electrostatically charged (as we know from an earlier passage of the fairy tale, when a number of girls busily rub Freya’s tender limbs [Novalis 1964, p. 121]). This explains the brilliant spark that jumps to Eros’ sword. The revival of Freya, however, does not happen before the physical bodies of Eros and Freya touch each other and complete the galvanic arc. Novalis remarks in his studies for the fairy tale: “To revive the princess in daylight—through a galvanic arc. [. . .] A kiss revives her” (Novalis 1960, volume III, p. 645). And: “It appears that in galvanism bodies first have to feel each other, before they express themselves to each other” (Novalis 1960, volume III, p. 609)—we need, so to speak, an “arc of lust” (Novalis 1960, volume I, p. 106).

Resisting the anachronistic urge to apply modern definitions of thought experiments to this pre-definitional case, the narrative of Freya’s revival is a thought experiment exploring the possibility of a revival from death by means of an arc of lust. Perhaps, Novalis speculated that any act of galvanism requires an element of lust to connect life and electricity. Erotic love is the deeply human and poetic side of this connection.

II.2 The Nature of Thought Experiments: Novalis

To unpack the noteworthy elements of a theory of thought experiments present in the writings of Novalis we want to look first at the notion of *romantic poetry*, and then to locate the method of experimentation in Novalis’s philosophy of nature as it results from the program of romantic poetry.

The notion of romantic poetry is central to the early German romantics’ understanding of science and its methods, because they deemed it not “possible or even desirable to distinguish between science and poetry” (Gode-von Aesch 1941, p. 30). Elaborating the notion of romantic poetry will help to clarify the nature of thought experimentation in Novalis.

This is because thought experiments derive their cognitive efficacy from the continuity he (and others) perceived between *nature* and *art*, as they applied the “term *Poesie* not only to literary creativity but to all artistic creativity, and indeed to the creativity of nature itself . . . The creativity of the artist was simply the highest organization, manifestation, and development of the same fundamental organic power active throughout all of nature” (Beiser 2003, p. 21). Accordingly, Novalis and others thought the sciences ought to be engaged in a disinterested exploration of nature (instead of an instrumentalist approach), which makes it possible to link them in co-operation with art (see Gode-von Aesch 1941, p. 30). Such co-operation we find in the poems of productive imagination.

In the metaphysics of Novalis, our minds are so closely tied to the world that it makes little sense to ask how the mind could reveal truths about nature without the support of mind-independent experimentation. “If nature is an organism, then it follows that there is no distinction in kind but only one of degree between the mental and the physical, the subjective and objective, the ideal and the real” (Beiser 2003, p. 168). This organic conception of nature is the answer young romantics gave to the problems caused by Kant’s dualisms, especially the dualism of an active, purely intellectual faculty and a passive, purely empirical faculty. Kant “had so radically divided them that any interchange between them seemed impossible” (Beiser 2003, p. 166). This is the context of Novalis’s philosophy of nature and his organic conception of nature. Only under the assumption that nature *is* an organism “is it possible to explain the *actual interaction* between the subjective and objective, the ideal and the real, the noumenal and phenomenal” (Beiser 2003, p. 167). In other words, drawing on the third fundamental proposition of the romantic doctrine stated above: reason can grasp nature only insofar as nature is reasonable—and accords with reason (see Daiber 2001, p. 63).

In thought experimentation, a creative activity is at play through which nature can be conformed to our will, an idea that does not seem absurd if we recall the doctrine of German idealism “that what we perceive depends on our own creative activity” (Beiser 2002, p. 423). The limits of such a creative activity are those that are “imposed by the physical world” (Beiser 2002, p. 426). There is no contradiction here. In order to see this, it is important to distinguish two concepts of idealism (for what follows, see Beiser 2002, p. 6): according to one reading, the ideal can be opposed to the physical. In this reading we contrast the word “idealism” with “realism” in order to express the view that everything there is depends on the mind. But there is another reading which captures better the thrust of German idealism. The “ideal” is the “ectypical,” that is, we are only cognizant of the impressions we have, and so everything we know in the world

must be knowable *for us*. These concepts of idealism are independent, in that you can be an idealist in the latter sense without being committed to idealism in the former sense. Our point is that the dependence of perceived objects on the perceiver, and the limitation on perception by physical constraints originating in what is perceived, are reconcilable in a framework that is idealistic in the second sense.

The conception of nature as organic can be expressed as follows: The soul externalizes itself into things of nature and nature internalizes itself in the mind. “By analyzing all nature into living force, [Novalis] has a unified means of understanding both the mental and the physical, which are simply different degrees of organization and development of living power or force. Matter is inchoate and nascent force, whereas mind is organized and developed force. Depending on one’s perspective, matter could be seen as a primitive form of mind, or mind as a developed form of matter” (Beiser 2002, p. 428). This metaphysics translates into an aesthetics according to which poetry is inherent to nature in its primitive state. Nature speaks, then, in the language of the poets (see Daiber 2001, p. 206).

The relationship between poetry and the world is thus not accurately described in terms of *mimesis*, *imitatio* or *copia*; poetry is participation in the process of discovering and understanding the world. Such a participatory view of poetry relates directly to a theory of thought experiments in that *thought* and *word* are *simultaneous*. It is not that words capture thoughts that precede the linguistic expression. Thus, writing can be an important way of discovering thought through the exploration of language (see Daiber 2001, p. 27). The epistemological consequence of Novalis’s metaphysics is that knowledge cannot be acquired unless the subject makes the object its own, and at the same time makes itself into the object. This is to say that the subject together with its creative activity are the “highest manifestation and realization of the powers of nature” (Beiser 2002, p. 4). Knowledge results from appropriation and self-alienation. In other words, mental life is the place where subject and object create each other, since

each is conceived as self-sufficient, the only relation between them is conceived to be one of external causality. Either the subject is the cause of the object (idealism) or the object is the cause of the subject (realism). But since these entities are so self-sufficient and heterogenous, even such a causal interaction becomes impossible. To get beyond this *aporia*, it is necessary to conceive the relation between subject and object in more organic terms, such that each becomes what it is only through the other. (Beiser 2002, p. 433)

1 LINE SHORT
REGULAR

With these metaphysical and epistemological premises in place, we can now turn to the role played by the poems of productive imagination in the program of Novalis's romantic poetry. Thought experiments combine with physical experiments in order to establish "complete experiments" (Novalis 1960, volume III, p. 408) within a "general theory of observation and experiments" (Novalis 1960, volume III, p. 437).

The range of their application is quite wide, including heterogeneous disciplines like physics and theology. For example, Novalis refers to the context of their use as "experimental physics of the mind" (Novalis 1960, volume III, p. 387) or "experiments in God" (Novalis 1960, volume III, p. 443).

In Novalis's philosophy of science, physical experiments are still the model of experimentation, and he models thought experiments on them (see Daiber 2001, pp. 115–175). But physical experiments are only a special case of experimentation. Generally speaking, "to experiment" is not an exercise for scientists only; it is available to everyone (Daiber 2001, p. 122). And it is never purely empirical or purely mental. While Novalis himself conducted and appreciated physical experiments (see Daiber 2001, pp. 88–98), he deemed them insufficient to realize the core objective of German idealism, namely to find the identity of identity and non-identity (see Beiser 2002, p. 14), as well as the unity of organic and inorganic nature from within an organic conception of nature (see Daiber 2001, p. 109). What we need, according to Novalis, is an art of invention without data, an absolute art of invention to approximate the unity of nature (see Daiber 2001, p. 145). What is required is a "true art of experimentation" (Novalis 1960, volume III, p. 445), which is only realized if there is a correspondence of physical and thought experiments. In other words, true knowledge of nature is not possible unless the subject and its inner world are considered in every experiment (see Daiber 2001, p. 110). Novalis says: "A good physical experiment can be used as a model of an inner experiment and is *itself* a good *inner*, subjective experiment as well" (Novalis 1960, volume III, p. 386). The true art of experimentation supervenes on the forces of nature and the forces of imagination, especially the force of productive imagination, which is also the source of poetry (see Daiber 2001, p. 111).

Obviously, in Novalis we see how in the period of the forerun "experiments" begin to play a role outside of "experimental philosophy"—the predecessor of today's science, as it were. There is no reason to think of Novalis's poetic departure from the primacy of reason in providing a foundation for our knowledge of nature as a departure from the trajectory to modern science if it is true that the "birth of modern experimental science was not attended with a new awareness of the powers and capacities of hu-

man reason, but rather the opposite—a consciousness of the manifold deficiencies of the intellect, of the misery of the human condition, and of the limited scope of scientific achievement” (Harrison 2009, p. 258).

II.3 Discussion: Novalis

As in Lichtenberg, we find in Novalis the seeds of a theory of thought experiments that is *unified* across disciplines. Poems of productive imagination are meant to play an important role wherever they are found, whether in the humanities or sciences. This fact might be taken as an additional reason to pursue a unified account of thought experiments as recommended by Cooper (2005).

Unlike Lichtenberg, Novalis’s notion of “complete experiments” as a synthesis of physical experiments and poems of productive imagination brings thought experiments much closer to physical experiments. This is similar to what Marco Buzzoni has in mind with respect to his Kantian account of thought experiments. Paraphrasing a famous Kantian dictum, Buzzoni puts his central claim this way: “The (empirical) thought experiment without a real world experiment is empty, the real world experiment without a thought experiment is blind” (Buzzoni 2011, p. 102, our translation).

In another departure from Lichtenberg, Novalis develops in a constructive manner a metaphysical framework that informs his theory of these poems of productive imagination. At the core of this theory, we find the project of romantic poetry. For this reason, our discussion of Novalis will focus on the observation that “in a thought experiment literature and natural science are almost forced to unite with each other . . . The thought experiment radicalizes a transgression between *fact* and *fiction*, natural science, literature, art and philosophy, as has often been observed” (Macho and Wunschel 2004, pp. 11–12, our translation). The question “of how writing can relate to the world has been a concern of literary criticism from Aristotle onward” (Oatley 1999, p. 103).

To begin with, we are used to a certain way in which thought experiments are presented to us. What is provided normally is a *scenario* in conjunction with a *framework* that guides the reader in manipulating that scenario. But such a presentation style is not necessary. For example, Galileo’s thought experiment about falling conjoined cannon balls of different weights was presented in a dialogue (see Galileo 1914[1638], pp. 62–63). The questions as to whether the dialogue form makes Galileo’s piece of scientific reasoning into genuine literature and what this would mean for the relationship between literature and science merits its own discussion.³

3. Seeskin makes the following remarks concerning the dialogical form sometimes

We cannot pursue this question here. Our point is that, like Galileo, neither Lichtenberg nor Novalis followed today's norms in presenting thought experiments. Their way of presenting thought experiments might make identification and access to them more difficult. But it certainly cannot count against the claim that we are dealing with genuine instances of thought experiments. The only way to bring this into question would be to use a theory of thought experiments. But such a theory would have to be informed by the different ways thought experiments have been presented, especially in the work of those who are among the first to think about them. The literary character of Galileo's, Lichtenberg's, and Novalis's thought experiments provides a very good reason to revisit them, because it seems that it is the *literary* features that are crucial for a proper assessment of their nature.

It seems useful to distinguish between two ways in which we can thematize the relationship between literature and thought experiments. One way is to highlight the frequent use of *narrative* to set up the *story* of a thought experiment. Accordingly, it has been argued that to account for thought experiments we need to aim for a proper understanding "of the way that we engage with fictional texts" (Ichikawa and Jarvis 2009, p. 223). For example, it has been suggested that it is the fictional character of a thought experiment that helps the thought experimenter in "*picking out* and *thinking about* propositions that are key" for the execution of a thought experiment (Ichikawa and Jarvis 2009, p. 229).

Another way to thematize the relationship between thought experiments and literature is to look for reasons that thought experiments might need to be embedded within the larger context of a fictional work such as a novel. For example, one could take the stance that such embedding is necessary for the same reason that some philosophical topics cannot be dealt with adequately but in a novel. For exactly this reason, in the face of philosophical discussions concerning free will and perennial conundrums in the philosophy of mind, the German analytic philosopher Peter Bieri, for instance, has begun to publish novels under the pseudonym Pascal Mercier.

Support for such a radical move comes from those who claim that

found in philosophy, but it could easily be applied to dialogues found in science: "Philosophy does not become literature merely because it is written in dialogue form. We could take the latest issue of the *Journal of Philosophy*, invent dummy characters, think of leading questions, and come up with a 'dialogue' which would have no literary significance whatever. Even if it were to be rewritten by a gifted stylist, it would not become literature unless the dialogue form were an integral part of the author's conception of philosophy" (Seeskin 1984, p. 181).

“fiction can be twice as true as fact” (Oatley 1999, p. 103). That is to say that a “narrative is a mode of thinking in which human agents with goals conceive plans that meet vicissitudes” (Oatley 1999, p. 103), and it is the fictional feature of a narrative that “provides context to understand the elliptical. It offers the context of a character’s goals and plans. It gives a sense of how actions lead to vicissitudes. It allows, too, the reader to experience something of emotions that can arise.” (Oatley 1999, p. 108). All this can serve to establish “truth as coherence within complex structures” (Oatley 1999, p. 103), and “it can serve as a personal truth and give rise to insight” (Oatley 1999, p. 109), especially if it is true that insights “of a personal kind when reading fiction are more likely to occur when the reader is moved emotionally by what he or she is reading and when the accompanying context helps the understanding of the resulting emotions” (Oatley 1999, pp. 114–115).

What we have said about Novalis’s poems of productive imagination and their nature leads to several interesting conclusions. First, the relationship between literature and thought experiments is important in both of the ways stated above. Second, this relationship illuminates why, according to Novalis, thought experimentation is not only about coherence in complex systems or the personal truths and insight they may provide. Rather, the relationship helps us to realize that the poems of pure imagination are about empirical truths as well, not just subjective fictions.

In reaction to the Enlightenment, Novalis developed a metaphysics according to which the mental and the physical represent or symbolize one another. This results in an epistemology of scientific experimentation that is described by Novalis himself as a “*science of active empiricism*” (Novalis 1960, volume III, p. 445). Within such an epistemology, any discovery of thought is equal to but no substitute for discoveries achieved by physical experimentation. Writing on trial is one way to make such discoveries. This is for two reasons. First, thought and language are simultaneous, and thoughts can therefore be explored by means of fictional literature. Second, “fiction does not really imitate” the world (Oatley 1999, pp. 107–108), and therefore it allows for the “inventions without data” that Novalis had in mind. Literary creativity is a function of human creativity, and this in turn is a function of the creativity of nature itself. Thought experimentation clearly articulates the organic power active throughout all of nature, which is conceived as an organism to depict metaphysically the actual interaction between the subjective and the objective. Thought experiments as a way to grasp nature by reason can be a method of the sciences because nature, the subject of scientific inquiry, itself accords with reason.

In conclusion, we can use the terminology and classification of Davies

(2007, p. 43) in order to classify Novalis's position on thought experiments as an "extreme inflationism." Like Brown and also Koyré (1968), Novalis views thought experimentation as a means to acquire knowledge of nature "without either (a) relying on new evidential input or (b) inferring that knowledge inductively or deductively from what is already known" (Davies 2007, p. 41). While such a position might not be attractive to everyone, Novalis should be taken seriously for his appreciation of the role of literature in thought experiments.

Davies (2007, pp. 31–33) and Elgin (this volume) do not see why thought experiments should not be counted as fictions. Of particular interest is Davies's claim that some works of fiction can be "properly viewed as much more elaborated" thought experiments (Davies 2007, p. 33). This could be problematic insofar as it is exactly *brevity* that makes a (good) thought experiment.⁴ Second, any exploration of thought experiments as fictions will be faced with the problem of "truth in fiction." It is unclear what exactly this might be, and relevant details are "still much in dispute" (Davies 2007, p. 34; see also Lewis 1978, Lihoreau 2011, and Suarez 2009). Third, there is a clear parallel between positions taken on the issue of whether or not learning from fiction without empirical testing is possible, and the epistemological puzzle that thought experiments pose (see Davies 2007, p. 43). But perhaps what is most interesting is that Novalis presents an account of thought experiments that directly attempts to unite the subjective and objective points of view in a way that dissolves the epistemological puzzle of thought experiments, by dissolving the distinction between mind and nature. As we will see, this idea finds another and perhaps more philosophically appealing expression in Kant.

III Kant's Experiments of Pure Reason

Kant's philosophy influenced the period of the *forerun* greatly, and both Lichtenberg and Novalis were strongly affected by his writings. But what makes Kant really stand out historically in the context of this paper is the fact that the Danish Kantian philosopher-scientist Hans-Christian Ørsted came up with the technical term "thought experiment" (*Tankeexperiment*) while engaging with Kant's *Metaphysical Foundations of Natural Science* ([1786] 2004). He introduces this term in his 1811 paper (Ørsted [1811] 1920), whose English title is the *First Introduction to General Physics* (Ørsted [1811] 1998). Ørsted himself classified this text as "an introduc-

4. We are indebted to Roy A. Sorensen who voiced this concern at a workshop on thought experiments in the summer of 2010 in Halifax. He had articulated this concern in writing in Sorensen 1992, pp. 222–224.

tion to experimental natural philosophy . . . a view through the corporeal world to the mental world.”⁵

Ørsted ([1811] 1998, p. 296) speaks of a “creative method” that is used “far more often than might be believed” in order to “satisfy the striving for insight of a vital and forceful mind.” This method does more than just show *that* “something is a certain way.” It helps to demonstrate “*why* it really is” that way. It allows us “to see every truth at its birth.” It is an essential method in physics, because it lets “the development of our thoughts follow that of the object’s” in order to demonstrate how “reason for being” and “our certainty about it” coincide. Such demonstrations are “thought experiments.” The most beautiful examples of these, claims Ørsted, are given to us by Kant in his *Metaphysical Foundations* “without, however, drawing attention to [them].”

Unfortunately, Ørsted does not identify any of the examples he refers to, nor does he tell us in which respects Kant fails to draw attention to them. Still, his remarks supply one of four reasons to revisit Kant in the context of a history of the inquiry into thought experiments. The second is that Kant greatly influenced Lichtenberg and Novalis in their stance on thought experiments. The third is that Kant himself speaks of “experiments of pure reason” (see Kant [1781] 1998, p. 112[BXXI]). This is the name Kant gives his method of testing transcendental principles, i.e., conditions for the possibility of human knowledge, such as “every event has a cause.” In the literature it has been noted already that these experiments make up a special class of thought experiments (see Kalin 1972). As for the fourth reason, Kant’s epistemology is promising with respect to finding a compromise between Brown’s Platonism and Norton’s eliminative empiricism (see Fehige 2012 and Forthcoming). In a certain sense the ongoing debate over thought experiments is a struggle for such a compromise. Kant defends, as it were, against Norton, *synthetic a priori* knowledge, but, contrary to Brown, he does not rely on Platonic entities or an intellectual perception with the mind’s eye.

III.1 The Notion of Thought Experiments: Kant

The main difficulty in identifying Kant’s notion of thought experiments is that he speaks of “experiments of pure reason” only in his *Critique of Pure Reason*, and not in his *Metaphysical Foundations*, although we find a number of thought experiments in the latter.

This amounts to a problem when dealing with his notion of thought

5. Ørsted in a letter to John Herschel (October 12, 1849). Taken in English from Kühne (2005, p. 129).

experiments because Kant does not pursue the same aim in both works. Thus, it is unclear how to relate the “experiments of pure reason” with the thought experiments of the *Metaphysical Foundations*. The aim of the *Critique of Pure Reason* is to identify the categories and principles of understanding in order to develop a theory as to what it is that we can know in principle. But in the *Metaphysical Foundations* those identified “categories and principles of understanding are taken simply as given, as premises for the further derivation of principles of pure natural science from them” (Friedman 2006, p. 323). What follows is that Kant obviously performed thought experiments *with two different purposes*. Looking at a few examples will bring these differences to light.

Kalin (1972, pp. 321–328) argues that the only way Kant was able to support his transcendental principles was by means of a kind of thought experiment. Transcendental principles are the conditions for the possibility of empirical knowledge. Therefore, Kant could not appeal to experience in order to present an inductive argument to support them. He could not appeal either to the kind of metaphysical reasoning that he meant to overcome, namely a mere analysis of concepts. Such an analysis leads, according to Kant, only to analytic propositions. However, transcendental principles are synthetic propositions. Hence, neither conceptual analysis nor inductive argument can establish them. But a thought experiment could. Here is an example to show how Kant advances the transcendental principle that “all alterations occur in accordance with the law of the connection of cause and effect” (Kant 1998, p. 304), i.e., that all events have causes. He begins:

I see a ship driven downstream. My perception of its position downstream follows the perception of its position upstream, and it is impossible that in the apprehension of this appearance the ship should first be perceived downstream and afterwards upstream. The order in the sequence of the perceptions in apprehension is therefore here determined, and the apprehension is bound to it. In the previous example of a house my perceptions could have begun at its rooftop and ended at the ground, but could also have begun below and ended above; likewise I could have apprehended the manifold of empirical intuition from the right or from the left. In the series of these perceptions there was therefore no determinate order that made it necessary when I had to begin in the apprehension in order to combine the manifold empirically. But this rule is always to be found in the perception of that which happens, and it makes the order of perceptions that follow one another (in the apprehension of

this appearance) **necessary** (Kant [1781] 1998, p. 307[A193/B238]).

The thought experiment proves for Kant that “all alterations occur in accordance with the law of the connection of cause and effect” (Kant [1781] 1998, p. 304), because if this was not true then “the subjective flow of mental contents would be indistinguishable from the objective sequence of actual events” (Kalin 1972, p. 322). This is certainly something like an “experiment in thought.”⁶ However, let us compare this to some examples from the *Metaphysical Foundations*, which was a mature expression of Kant’s thoughts on the nature of matter, given his critical philosophy and the natural philosophy (physics) of his day.

In the chapter on dynamics, Kant argues that the attractive and repulsive forces in matter are both fundamental. But the repulsive force is what gives matter its appearance of solidity, and this is the aspect of matter that we directly experience, not its attractive force. So why is there this difference, if these are equally fundamental properties of matter? He answers:

[E]ven if we had such a capacity [to sense attraction], it is still easy to see that our understanding would nonetheless choose the filling of space in order to designate substance in space . . . Attraction, even if we sensed it equally well, would still never disclose to us a matter of determinate *volume* and *figure*, but only the striving of our organ to approach a point outside us (the center of the attracting body). For the attractive force of all parts of the earth can affect us no more, and in no other way, than as if it were wholly united in the earth’s center, and this alone influenced our sense, and the same holds for the attraction of a mountain, or any stone, etc. But we thereby obtain no determinate concept of any object in space, since neither figure, nor quantity, nor even the place where it would be found can strike our senses. (Kant [1786] 2004, p. 47[510])

This thought experiment is used to establish a positive theoretical result, namely that our inability to sense the force of attraction is irrelevant as evidence concerning its status as a fundamental force. He shows this by considering what it would be like if we *could* sense this force, and points out that in such a scenario, our sense would only inform us concerning the location of the centers of mass of different objects. This would not give us any idea about the matter located around those inaccessible centers, so we can be confident that even though we cannot sense attraction, this does

6. For another thought experiment in the *Critique* see (A20/B 35), and Shabel’s discussion in Guyer 2010, pp. 95–96.

not mean it is not fundamental. This thought experiment does not have a transcendental principle as its goal; just as the thought experiments in the *Critique* do not aim to establish physical/theoretical results.

This use of thought experiments as evidence for physical claims in the *Metaphysical Foundations* is very interesting, so we will provide a few more examples. Again in the chapter on dynamics, Kant attempts to show that fluidity is another basic property that matter can exhibit. Kant argues that if fluidity were not a basic property but a derivative one, there would be imperfect fluids that could be made to experience friction in circumstances where there should be none. But this is impossible. Here is his thought experiment:

Consider, for example, a bent tube with two arms, one of which may be arbitrarily wide, and the other arbitrarily narrow . . . if one imagines both arms several hundred feet high, then, according to the laws of hydrostatics, the fluid matter in the narrow arm would stand precisely as high as in the wide one. But since the pressure on the bases of the tubes, and hence also on the part that joins them in common, can be thought as increasing to infinity in proportion to the heights, it follows that if the least amount of friction occurred between the parts of the fluid, a height for the tubes could be found, at which a small quantity of water, poured into the narrower tube, did not disturb that in the wider one from its place. So the water column in the former would come to stand higher than that in the latter, because the lower parts, at such great pressure against one another, could no longer be displaced by so small a moving force as that of the added weight of water. But this is contrary to experience, and even to the concept of a fluid. (Kant [1786] 2004, pp. 67–68[529])

Again, a thought experiment is used to establish a result about nature. In this case, a little experience with gravity and water are enough to show that fluidity must be a basic property of matter. One interesting feature of this thought experiment is that it includes constructing and manipulating devices that perhaps have never existed, and may never exist. To invoke glass tubes of arbitrary diameter hundreds of feet high (or higher), capable of creating infinite water pressure is indeed an interesting use of a thought experiment. While Lichtenberg suggested impossible situations, he used them only to falsify. And though Novalis illustrated his hypotheses with what seem fantastic narratives, Kant imagines a potentially impossible situation to make a point about the real world.

There are other cases in which Kant uses the result of a thought experiment as evidence for a claim. For example, Kant “represents to himself”

digging a hole through the earth and dropping a stone into it to see if its course would deviate in any direction. This is an imaginary experiment that demonstrates the possibility of testing the nature of the Earth's rotation, without using an external (non-moving) frame of reference. He goes on to imagine himself raising the stone higher and higher above the hole to see if at a certain height the rock would no longer enter the hole (Kant [1786] 2004, pp. 100–101[562]).

Besides the above examples, there is another way that Kant uses thought experiments. He occasionally invokes an imaginary scenario to make a difficult argument easier to understand. This sort of illustrative or mediative thought experiment is still definitely an exercise in thought experimentation, although it does not establish anything in addition to its accompanying argument. We include an example here to demonstrate the breadth of Kant's use of thought experiments, and perhaps to justify Ørsted in his conviction that Kant was a master thought experimenter.

Early in the *Metaphysical Foundations*, Kant reminds us that “Every motion, as object of a possible experience, can be viewed arbitrarily as motion of the body in a space at rest, or else as rest of the body, and, instead, as motion of the space in the opposite direction with the same speed” (Kant [1786] 2004, p. 23[488]). This is simply Galilean invariance, but Kant has a new use in mind. He wants to show that space “belongs merely to the subjective form of our sensible intuition of things or relations, which must remain completely unknown to us as to what they may be in themselves” (Kant [1786] 2004, p. 19[484]). He asks us to consider a motion within a space, “as when I see a ball moving on the table in the cabin of a ship.” Then, we are asked to expand to a larger space that includes the previous one: from “the bank of the river” we may see that the cabin on the ship is moving (Kant [1786] 2004, p. 23[488]). From this new perspective we may see the ball at rest. He goes on:

Now because it is completely impossible to determine for an empirically given space, no matter how enlarged it may be, whether it may or may not be moved in turn, in relation to an inclusive space of still greater extent, it must then be completely the same for all experience, and every consequence of experience, whether I wish to view a body as moved, or as at rest, but the space as moved in the opposite direction with the same speed. Further, since absolute space is nothing for all possible experience, the concepts are also the same whether I say that a body moves in relation to this given space, in such and such direction with such and such speed, or I wish to think the body as at rest, and to ascribe all this, but in the opposite direction, to the space. For any concept is entirely the

same as a concept whose differences from it have no possible example at all, being only different with respect to the connection we wish to give it in the understanding (Kant [1786] 2004, p. 23[488]).

This argument is meant to show that space is not an entity *or* an objective relation that obtains between objects, but rather a category that we apply to experience. This is because for space to be an entity, it must be able to move relative to something else. The space that encapsulates everything but doesn't move relative to anything is absolute space. However, such a concept is contradictory and couldn't be instantiated in the world. Space also cannot be a relation objectively instantiated, since the relation between objects depends on the choice of perspective. Rather, space must be a regulative ideal that we use to order our understanding. The role of the thought experiment in establishing this conclusion is merely illustrative (or mediative), but it is still very helpful, especially if one continues on from the ball to the cabin and river bank, earth, solar system, galaxy, universe, etc., until one reaches absolute space, which is a contradiction if it is assumed to physically exist.

III.2 The Nature of Thought Experiments: Kant

The experiments of pure reason and the thought experiments in the *Metaphysical Foundations* alike confirm Kant's confidence in the power of the human mind to gain knowledge by reasoning alone. According to Kant, the human mind, unaided by physical experiment, has the power to come up with "answers to questions that so far remained open or unanswered. Such answers may even constitute a revolution in science" (Witt-Hansen 1976, p. 51). Such a view of the potential in Kant's epistemology is not far-fetched.

On the one hand, Kant claims that real world experiments are the safest way to obtain knowledge of the natural world; he likens real world experiments to judges that rule in favor or against theories. These experiments, however, have their origin in the human mind, and are the condition of the possibility of the meaningful experiences that constitute real world experiments. Kant opposes this model to the claim that in real world experimentation, scientists are like passive students taught by nature. Rather, for Kant, there is a significant and indispensable contribution of the human mind to each real world experiment. This is true for natural science in general.

According to Kant, each science must have what he calls a "pure part," which is cognized a priori ([1786] 2004, p. 5[469]). Without this component, a system of inquiry cannot be a *proper* science. The science for

which Kant in fact aims to provide a metaphysical foundation “is Newtonian science: in particular, the science of Newton’s *Principia* (1687)” (Friedman 1992, p. 136). Much “of Kant’s philosophical development can be understood . . . as a continuous attempt . . . to construct . . . a genuine metaphysical foundation for Newtonian natural philosophy” (Friedman 1992, p. 4). In Kant we find the “life-long attempt to grapple philosophically with the exact sciences to the end” (Friedman 1992, p. 52). His own corresponding research program was already effectively outlined in his *Physical Monadology* (published in 1756). It occupied him for some thirty years, culminating “in his main work on natural philosophy, the *Metaphysical Foundations*; and he remained interested in the issues beyond that work until his death in 1804” (Pollok 2002, p. 62).

Contrary to Newton (according to a widespread reading, at least),⁷ Kant believed that physics stands in need of metaphysical principles, and he intended to provide them. Kant’s pure part of natural science is a “special metaphysics of nature” (Pollok 2002, 77), and as such it is closely related to Kant’s general metaphysics which he developed in the *Critique of Pure Reason*. But, unlike the latter, in the *Metaphysical Foundations* “the categories and principles of understanding are taken simply as given, as premises for the further derivation of principles of pure natural science from them” (Friedman 2006, p. 323). Still, in each case we deal with knowledge that is *synthetic* and *a priori*. An example of such knowledge is the following statement about the two forces of attraction and repulsion which are so central to Kant’s dynamical theory of matter, because Kant claimed that matter “fills the space it occupies by a continuous ‘balancing’ of the two fundamental forces of attraction and repulsion” (Friedman 2004, p. xvi). He defined these forces as follows:

Attractive force is that moving force by which a matter can be the cause of the approach of others to it (or, what is the same, by which it resists the removal of others from it). *Repulsive* force is that by which a matter can be the cause of others removing themselves from it (or, what is the same, by which it resists the approach of others to it). The latter force will also sometimes be called *driving* force, the former *drawing* force . . . Only these two moving forces of matter can be thought. For all motion that one matter can impress on another, since in this regard each of them is considered only as a point, must always be viewed as imparted in the straight line between the two points. But in this straight line there are only two possible motions: the one through which the two points *remove*

7. For a more balanced reading of the role of metaphysical principles in Newton see Janiak 2008.

themselves from one another, the second through which they *ap-proach* one another (Kant [1786] 2004, p. 35[498]).

The fact that Kant seems to attribute a key role to thought experiments in finding the *synthetic a priori* knowledge needed for the pure part of natural science is perhaps the strongest reason to look at Kant's views on thought experiments, since it is here that we may find a way to tread the middle ground between Platonism and default empiricism. Unlike Platonism, a Kantian theory of thought experiments won't posit Platonic entities, and therefore will not need an epistemological account of our access to them. And unlike empiricism, a Kantian approach makes possible knowledge of the natural world that is neither *analytic* nor *a posteriori* but *synthetic a priori*, and thus takes seriously the idea that thought experiments can reveal information about the world that goes beyond pure empirical input.

We don't want to tread too near the abyss of Kant-interpretation, but in order to address Kant's *synthetic a priori*, we should say at least this much: the idea, roughly speaking, is that "the mind so shapes or structures experience as to make the synthetic *a priori* propositions in question invariably come out true *within the experiential realm*" (BonJour 1998, p. 23). We can know, for instance, the proposition that every event has a cause *a priori* in spite of its *synthetic* character, because the mind so operates in structuring or "synthesizing" experience as to make this proposition invariably true within the experiential realm. Accordingly, the objective of Kant's proper natural science is to demonstrate that the laws of experience are in fact true with apodeictic certainty.

The fact that some of Kant's *synthetic a priori* statements in the pure part of natural science have been falsified, like the one about the two forces quoted above, shouldn't be seen as an insurmountable problem for a theory of thought experiments in a Kantian spirit. This is for three reasons. First, "Kant's own contributions to a dynamical theory of matter had a significant impact on the development of natural science itself, quite apart from the original more metaphysical setting within which it was first articulated" (Friedman 2004, p. x). This is to say that Kant's mistake was theoretically highly fertile. Second, contemporary rationalists claim that rationalism is compatible with fallibilism (see Bealer 2000, p. 9; and BonJour 1998, pp. 110–115, Brown 1986, p. 10), so some false *a priori* propositions should not worry us. Third, it depends on the particular reading of the transcendental nature of the human mind to assess the effects emerging from the empirical falsification of some of Kant's *synthetic a priori* statements for the feasibility of a Kantian theory of thought experiments. In fact, it is the latter task that will be at the heart of any Kantian account of thought experiments. This brings us straight to the discussion of Kant's

views on the power of the human mind as they pertain to thought experiments.

III.3 Discussion: Kant

Kant's theory of knowledge seems to allow us both to preserve the rationalistic character of thought experiments, and to concede to the empiricist the primacy of real world experiments for the scientific investigation of the world. That is to say, once more, that there is the potential in Kant to find a middle ground between Platonism and empiricism regarding thought experiments.

III.3.1 Buzzoni's Kantian Account of Thought Experiments Marco Buzzoni has done exactly this. The thrust of his proposed Kantian theory of thought experiments can be captured in a Kantian sounding manner in Buzzoni's own words as follows: "The (empirical) thought experiment without a real world experiment is empty, the real world experiment without a thought experiment is blind" (Buzzoni 2011, p. 102). This is to cohere with Kant's famous dictum that "thoughts" (*Gedanken*) without "intuitions" (*Anschauungen*) are empty and "intuitions" (*Anschauungen*) without "concepts" (*Begriffe*) are blind (see Kant 1998, pp. 193–194[B75]). Kant's claim is that both intuitions and concepts are constitutive elements for knowledge. They "must be united in any instance of knowledge" (Friedman 1992, p. 96), otherwise no human cognition can arise. Without sense perception, intuitions are empty, and "objects for any concept whatsoever can only be found in *empirical* intuition" (Friedman 1992, p. 101). But we would not have any meaningful thoughts about perceived objects without the contributions of the mind.

Buzzoni favors a *functional* reading of Kant's a priori over a *material* reading. For present purposes, this translates into the claim that thought experiments are the condition of the possibility of physical experiments. Despite his commitment to the principle of empiricism, Buzzoni's reading of the Kantian a priori enables him to argue for the indispensability of thought experiments as a part of scientific practice. Yet no thought experiment is complete unless it correlates to a physical experiment in the sense that the thought experiment is "a hypothetical experimental situation" (Buzzoni 2008, p. 93) that, "in principle," could be realized outside of the imagination. It is interesting, however, that this hypothetical experimental situation may not be realized outside of the imagination as a physical experiment. It might, and very often does, remain a mere thought experiment. Such thought experiments maintain cognitive efficacy in science, because they "anticipate the results of real experiments and in this way they inductively extend our knowledge" (Buzzoni 2008, p. 96). However

in cases of doubt, Buzzoni wants the principle of empiricism to be enforced. That is, “we must turn to real experiments, which remain the ultimate criterion for all empirical thought experiments” (Buzzoni 2008, p. 96). Skeptical scenarios will obtain if scientists cannot accept “as well-confirmed the facts and laws that they utilize” (Buzzoni 2008, p. 96).

Buzzoni models the relationship between thought experiments and real world experiments on the more general relationship between theory and physical experiment. He understands their relationship within a Kantian framework in response to “the empiricists’ failure to understand transcendental analysis: the *a priori*, as the condition of experience, is always already in experience, and all experience is always already synthesis *a priori*” (Buzzoni 2008, p. 58). Along the lines of a transcendental analysis so conceived, Buzzoni makes the methodological choice of taking *acting human beings* as the most fundamental starting point for his account of thought experiments. Through their bodies, human agents find themselves always already in a situation in which conceptualizations and evaluations of reality are entangled with specific operations (see Buzzoni 2008, p. 22). In scientific experimentation, we build on this situation:

The experimental natural sciences extend the intentionality of empirical knowledge that consists in translating the theoretical-conceptual contents of propositions into operations that are in principle testable by means of their real execution. In this way, the experimental natural sciences undertake to exemplify each proposition by a technical apparatus, the functioning of which exhibits the empirical truth of the corresponding propositional content in a way that is now independent of our will (Buzzoni 2008, p. 24).

In thought experimentation, we develop operations in order to test new scientific claims. The propositional content of these claims is represented by a technical apparatus. Unless it is realized outside of the imagination, this apparatus confirms the claim in question and this is done independently of the mind. This is acceptable scientific practice, according to Buzzoni, as long as the employed scientific facts and laws of nature are not called into question. Then a realization outside of the mind is mandatory. The principle of empiricism demands such testability outside of the imagination. But even then, Buzzoni argues, thought experimentation is inevitable, because it is the condition of the possibility of real-world experiments.

Despite their limited epistemic power, thought experiments are very important in science because they enable real world experiments, and unless challenged, can support new scientific claims. At the most fundamental epistemological level, thought experimentation exemplifies what

Buzzoni calls an inductive-experimental use of reason (see Buzzoni 2008, p. 65). Induction is defined by him as the methodological-discursive ability of human intelligence to grasp general and reproducible structures immanent in real events and processes (see Buzzoni 2008, p. 68). An inductive use of reason is experimental for Buzzoni in three respects. First, as Mach (1897) claimed, it employs the *method of variation*, whereby some variables are systematically modified to establish which relation of dependence, if any, holds between them (see Buzzoni 2008, p. 21). Second, as Kant explained, it *puts a question to nature* and *anticipates* nature's answer (see Buzzoni 2008, p. 32). Third, it specifies technical operations that are *in principle* testable by means of their real execution (Buzzoni 2008, p. 39). The condition of the possibility of an inductive-experimental use of reason, in turn, is the mind's ability to conceive of every real entity as merely possible (see Buzzoni 2008, p. 109).

Buzzoni obviously distinguishes between two levels of the mind, namely the *reflexive-transcendental* where the actual is turned into the mere possible, and the *positive level* where hypothetical speculation takes place (see Buzzoni 2008, p. 102). While we welcome Buzzoni's Kantian account, we feel that he has not exhausted the resources we find in Kant available to develop a theory of thought experiments. To support our claim we thus turn to a somewhat tentative discussion of the prospects of a Kantian account of thought experiments in light of today's cognitive science.

III.3.2 Kant and Cognitive Science There has been a great deal of work in the last 30 years on Kant's theory of the mind. This started with Karl Ameriks's book in 1983, and was followed by Allison (1983), Guyer (1987), Martindale (1987), Aquila (1989), Kitcher (1990), Powel (1990), and Waxman (1991). According to Andrew Brook, each generation tries to apply Kant's ideas to the issues of its time (1994, p. 3), so it's natural that we should examine the relationship between Kant and cognitive science, which is used by some to model the mechanisms underlying thought experimentation (e.g., Nersessian 1992, 1993, 2007; Mišćević 1992, 2007).

There are at least two ways to see the relationship between Kant and cognitive science. One is of ancestry. For example, "Kant has virtually been adopted as an intellectual grandfather by cognitive science" (Brook 1994, p. 12). This is not ancestry for the sake of rhetoric and tradition: there are *living* elements of Kant in cognitive science. For one, transcendental arguments have "become a major, perhaps the major, method of cognitive science" (1994, p. 12). This is because in many cases we know what the mind is capable of, so the job of cognitive science is to figure out

what mechanisms and abilities it *must* possess in order for it to have those capacities. Furthermore, Kant is the one from whom we inherited modern functionalism, the idea that the best way to model the mind is to model what it can and does do. This is a widely accepted view of the mind in modern cognitive science (1994, p. 12). Finally, Kant played a role in the development of the “representational view of the mind,” which claims that “the function of the mind is to shape and transform representations” (1994, p. 12).

The other way to relate Kant to cognitive science is to use his work to drive novel, positive contributions. In this respect, “Patricia Kitcher’s work is the leading example” (Brook 1994, p. 3). She claims that by ignoring Kant’s work on the mind, “an opportunity is being lost” (Kitcher 1990, p. 205). It should not be surprising, says Kitcher, that “a great theoretician who devoted enormous intellectual efforts to determining what cognition requires of our mental faculties might have had some ideas that are useful to cognitive science” (1990, p. 230). To prove her point, Kitcher uses Kant to suggest an answer to an ongoing problem in cognitive science: the problem of the nature of concepts. Do concepts consist of necessary and sufficient conditions, prototype-matching algorithms, or what? (For more on this problem, see Margolis and Lawrence 1999). The answer Kitcher provides on Kant’s behalf is noteworthy. This makes us wonder: what could a cognitive science-based Kantian account of the mind teach us about thought experiments?

For one, Kant claims that for our mind to generate and manipulate concepts, it needs percepts, or sense-input. But sense-input alone is useless; it needs to be synthesized by the imagination. Cognitive scientists now argue that this seems to indeed be the case, at least in humans. High-level concepts (like *cause*) can tie together representations created by different faculties. That is, sights, sounds, smells, emotions, etc., which may have been recorded in memory at different places and times, are combined in cognition by the brain (*Critique* [1771] 1998, p. 238[A120]; see Prinz 2002; Kitcher 1990, p. 152). The main thing that has been added to Kant’s account is that in place of the traditional understanding of representations, we now understand these to be something like “patterns of activation in populations of neurons” (Thagard 2010, p. 78). Still, Kant is right that given our sense-input we apply processes that yield mental representations, which are then meaningfully combined with other representations (Thagard and Stewart 2011).

We can transition to thinking about thought experiments by considering again Lichtenberg’s melting and freezing of concepts. Perhaps a cognitive scientific Kantian account should understand thought experiments as tools that allow us to break the ties that are created between concepts and

their representations. But this is not how Kant would have understood them. Besides not being a falsificationist, Kant would never have allowed that concepts could be “frozen,” since he claimed that they are always changing as new experiences force us to update their content (Kitcher 1990, p. 213). In fact, Kitcher claims that given Kant’s arguments for the malleability of concepts, they should not be thought of as stable *objects* of cognitive scientific study, but rather “something like conceptual worms that continually evolve through time” (Kitcher 1990, p. 213).

Since the Kantian relationship between concepts and thought experiments is not as simple as in Lichtenberg and Kuhn, Kant would give a more complex answer to the question of the role of thought experiments. Recall that Kant used thought experiments to justify beliefs, to gain a priori knowledge, to explore possibilities, and to illuminate difficult ideas. Thus, it might be possible to form a cognitive science-based Kantian account of thought experiments in each of these directions. To justify beliefs, a Kantian account could assume that concepts are “something like conceptual worms,” and argue that thought experiments are sometimes used to cultivate and expand these ever-changing concepts. One philosopher who takes such a view is Imre Lakatos. Lakatos’s view is not well-known in the thought experiments literature (but see Glas 1999 and Buzzoni 2011), and perhaps this is because it is quite atypical. For Lakatos, thought experiments are the means by which our concepts find increasingly accurate expression in increasingly advanced theoretical languages. Their purpose is to yield and modify concepts that accurately track the perceived “facts” about scientific and mathematical objects (Lakatos 1976, pp. 90–92). According to Lakatos, mathematics progresses by informal reasoning, not by valid formal proofs. Proofs only represent completed reasoning once fully worked out. A typical example of mathematicians at work involves the presentation of some theorem, definition, or proposed axiom, which is then deconstructed and analyzed by means of examples and counterexamples. These, for Lakatos are thought experiments, and they form the heart of mathematical practice.

However, a complete Kantian account of the thought experiments meant to justify theoretical knowledge cannot stop here. For Kant, thought experiments can also generate synthetic a priori knowledge, and this is perhaps their most exciting Kantian function—especially since here as elsewhere there are hints in cognitive science that Kant’s ideas may be born out. For example, the kind of conceptual combination and manipulation that takes place in thought experimentation can yield entirely new concepts, without any need for additional empirical input. Recently, Thagard and Stewart (2011) presented a cognitive scientific mechanism for this function. These developments can be understood as a priori in the

Kantian sense, since for Kant, a priori knowledge must be *acquired* a priori, not just *known* a priori, and this process of conceptual combination discussed by Thagard and Stewart does not rely for its justification on any empirical input.⁸ It would be fascinating if cognitive science made possible a Kantian account of thought experiments in which synthetic *a priori* knowledge was gained through thought experimentation. We now explore what such an account might look like.

III.3.3 A Sketch of a Modern Kantian Account of Thought Experiments A modern Kantian account of thought experiments along the lines sketched above would portray thought experiments as mental processes that provide access to transcendental truths. It would delimit clearly the set of those transcendental truths which are thus accessible (e.g., laws of nature, essential dispositions of objects, etc.), and clarify our epistemological connection to those facts.

The idea that there is a modal element to the laws of nature is common to most realist accounts of laws. If there is such a modal feature, then laws of nature could be a viable target for Kantian thought experiments, and the account under consideration should explain our knowledge of them. We must be clear: such thought experiments would not aim to generate knowledge of metaphysical *de re* necessities, e.g., that the mass of an electron is $9.10938291 \times 10^{-31}$ kilograms. This is because, according to Kant, nothing can be known about the world merely by thinking about it. To determine whether something is true, for Kant, we must go out into the world and experience it. But Kantian thought experiments could provide access to epistemological *de dicto* necessities, e.g., tautologies, conceptual truths, etc. Access to this latter brand of necessity would not be objectionable for naturalists or empiricists, as it merely helps to display features of pre-constructed *facts*, not metaphysical *objects*. That is, it is not objectionable to claim that we may be led to the necessary truth of some fact, without having direct access to its truthmakers. This is because Kantian transcendental arguments are meant to show what must be true, given what is known to be true. And very often in science we are given some theoretically interpreted phenomenon which we then proceed to consider thought-experimentally. This process can yield facts about our theories, our instruments, our concepts, or about the relationships be-

8. This method is not *a priori* in the sense that we necessarily gain knowledge of apodictic truths. However it would be *a priori* in the sense that we have here an action that provides the concepts needed to justify propositions in general. We merely suggest a *prima facie* consilience with some of Kant's ideas, and do not wish to take a stand on the correct interpretation of the *a priori* in Kant. For some discussion of this issue, see, e.g., Greenberg 2001, Kitcher 2006, and Longuenesse 2006.

tween objects in the world. We think it is clear that scientists perform this type of activity quite often. For example, according to one interpretation of the Einstein-Podolsky-Rosen thought experiment, we learn what must be the case concerning “entangled” states of quantum particles, given features of the quantum mechanical formalism (Einstein et al. 1935). And there are recent accounts of the laws of nature, such as Alexander Bird’s, which claim that laws are entailed by the dispositions of objects that are essential to their nature (Bird 2005, p. 356). Given scientifically grounded knowledge of the dispositions of an object, a Kantian could certainly claim to know by a transcendental thought experiment some of the laws of nature in which it must figure. And going one step further, it is equally likely that given some simple facts about our experience, a Kantian could perform transcendental thought experiments to discover the dispositions of some *mental* objects, which might lead to knowledge of something like mental laws. These laws could then figure into more thought experiments that eventually lead to transcendental truths about the mind and its epistemological powers. To claim that this rudimentary sketch in fact captures the intended thrust of Kant’s actual system in the *Critique* would be premature, anachronistic, and naïve. However, we think it does accord at least *prima facie* with Kant’s own remarks about his method.

Another interesting consequence of a Kantian account concerns the role of emotion in thought experiments. Cognitive scientists claim that emotion is necessarily linked to cognition (e.g., Thagard 2010). According to Maria Borges, Kant allows for a full continuum of emotions. She claims, “Kant’s account of emotion includes both physiological aspects and cognitive contents, mainly evaluative beliefs” (2004, p. 140). Specifically, there is a hierarchy of emotions that could be arranged according to their level of abstraction, and their importance to the reasoning process. She concludes that “Kantian moral theory contributes an outstanding theory of emotions to contemporary debates, one which acknowledges both the physiological as well as cognitive aspects of emotions (Borges 2004, p. 140). One of the many interesting aspects of Kant’s account of emotion, according to Borges, is that emotions can carry cognitive content, and they can therefore be awakened “through imagination or reason” (Borges 2004, p. 155). This is precisely what we find in cognitive scientific studies on emotion, and these findings should be related to the view that thought experiments involve some kind of strong feeling of certainty regarding their results’ intuitive plausibility (see Brendel 2004, p. 96). This is one of the reasons we find a Kantian account of thought experiments so interesting: it does not merely offer an explanation of the epistemological status of the output of thought experiments; it also enables us to unify seemingly disparate features of this method, such as the role of emotion. This, we

think, is in part due to the explanatory power of Kant's theory of the mind in general as interpreted by current cognitive science.

A question that has not been asked until now concerns the epistemological nature of such emotional content. What epistemological merits or dangers are heralded by the role of emotion? What would it mean for emotion to be part of the scientific context of justification and not merely the context of discovery? We have now come full circle, as some answers were already suggested in the above discussion. According to Novalis, narrative and other literary devices are the only means to certain conclusions. Perhaps this is because of the role that literary devices (narrative, metaphor, etc.) play in appealing to the emotions necessary for the functioning of certain cognitive processes. If thought and word and world are one, experimenting with thoughts is but another way of experimenting with the world, and for romantic poets, this involves appeal to human emotion.

Conclusion

We hope it has become clear why Lichtenberg, Novalis, and Kant are good choices to characterize the forerun period in the history of the investigation into thought experiments. The German Enlightenment focused on modern experimental science, the mind's relation to nature, and how we might learn about the world using literature. It seems that it was therefore natural for intellectuals to begin to think about thought experiments, and how they might work.

We saw in Lichtenberg a master of experimenting with thoughts and ideas to melt down concepts in order to learn about our world. Novalis likewise used his poems of productive imagination to probe nature. An organic conception of nature enabled Novalis to establish a correspondence between trials in nature and writing on trial. In Novalis and Lichtenberg, the literary component of thought experiments is crucial to understanding their function, which raises questions concerning the relation between literature and the world, and between thought experiments and literature. Novalis and Kant shared the belief that thought experiments and physical experiments are very closely tied to one another, perhaps due to their idealism. And in all three, we find the notion that whatever thought experiments are, they will be equally applicable in any domain of enquiry.

However it is with Kant that we begin to see instances of all the kinds of thought experiments tabulated by Brown in recent times ([2010] 1991), as well as an appreciation of the mind's capacity to reach truths about the world by this method. Kant respected it as one of the means, or perhaps the only means, by which to reach the transcendental principles, which themselves tell us about the way the world must be. But he also ap-

preciated that thought experiments can reveal physical truths directly when they are used in physical contexts. And finally, due to Kant's indefatigable research on the mind, many of his insights are *still* applicable and we think offer new directions for research concerning thought experiments.

The interaction of Kant studies and cognitive science may be the best hope we have to find a middle ground between Brown and Norton. One place to begin might be a closer study of Lichtenberg's and Novalis's extensions of Kant's ideas, which could help us to understand the epistemic contribution of the literary feature of thought experiments. As the intellectual descendants of the forerun, it is little wonder that Ørsted and Mach, having inherited these ideas, would open the door to serious work on the philosophy of thought experiments.

References

- Allison, Henry. 1983. *Kant's Transcendental Idealism: An Interpretation and Defense*. New Haven: Yale University Press.
- Ameriks, Karl. (1983) 2000. *Kant's Theory of the Mind*. New York: Oxford University Press.
- Aquila, Richard. 1989. *Matter in Mind*. Bloomington: Indiana University Press.
- Bacon, Francis. 1901. *The Works*. Edited by J. Spedding. London: R. L. Ellis, and D. D. Heath.
- Bealer, George, 2000. "A Theory of the A Priori." *Pacific Philosophical Quarterly* 81: 1–30.
- Beiser, Frederick C. 2005. "The Enlightenment and Idealism." Pp. 18–36 in *The Cambridge Companion to German Idealism*. Edited by Karl Ameriks. Cambridge: Cambridge University Press.
- Beiser, Frederick C. 2003. *The Romantic Imperative: The Concept of Early Romanticism*. Cambridge: Harvard University.
- Beiser, Fredrick C. 2002. *German Idealism: The Struggle against Subjectivism, 1781–1801*. Cambridge: Harvard University Press.
- Bird, Alexander. 2005. "The Dispositionalist Conception of Laws." *Foundations of Science* 10: 353–370.
- Bokulich, Alisa. 2001. "Rethinking Thought Experiments." *Perspectives on Science* 9: 285–307.
- Borges, Maria. 2004. "What Can Kant Teach Us about Emotions?" *The Journal of Philosophy* 101: 140–158
- Brook, Andrew. 1994. *Kant and the Mind*. New York: Cambridge University Press.
- Brook, Andrew. 2007. "Kant and Cognitive Science." Pp. 117–36 in *The*

- Prehistory of Cognitive Science*. Edited by Andrew Brook. Basingstoke: Palgrave Macmillan.
- Brown, James R. 1986. "Thought Experiments since the Scientific Revolution." *International Studies in the Philosophy of Science* 1: 1–15.
- Brown, James R. (1999) 2010. *The Laboratory of the Mind* 2nd Edition. New York; London: Routledge.
- Brown, James R. 2011. "Über das Leben im Labor des Geistes." *Deutsche Zeitschrift für Philosophie* 59: 65–73.
- Buzzoni, Marco. 2008. *Thought Experiment in the Natural Sciences*. Würzburg: Königshausen & Neumann.
- Buzzoni, Marco. 2011. "Kant und das Gedankenexperimentieren." *Deutsche Zeitschrift für Philosophie* 59: 93–107.
- Buzzoni, Marco. 2011. "Mathematical Thought Experiments." *Epistemologia* XXXIV: 61–88.
- Cooper, Rachel. 2005. "Thought Experiments." *Metaphilosophy* 36: 328–47.
- Daiber, Jürgen. 2001. *Experimentalphysik des Geistes: Novalis und das romantische Experiment*. Göttingen: Vadenhoeck & Ruprecht.
- Davies, David. 2007. "Thought Experiments and Fictional Narratives." *Croatian Journal of Philosophy* VII: 29–45.
- Einstein, Albert, Boris Podolsky, and Nathan Rosen. 1935. "Can Quantum-Mechanical Description of Physical Reality be Considered Complete?" *Physical Review* 47: 777–80.
- Fehige, Yiftach J. H. Forthcoming. "The Relativized A Priori and the Laboratory of the Mind: Towards a Neo-Kantian Account of Thought Experiments in Science." *Epistemologia: Italian Journal for Philosophy of Science* 36: 55–73.
- Fehige, Yiftach J. H. 2012. "'Experiments of Pure Reason.' Kantianism and Thought Experiments in Science." *Epistemologia: Italian Journal for Philosophy of Science* 35: 141–160
- Friedman, Michael. 2007. "Kant, Naturphilosophie, Electromagnetism." Pp. 135–158 in *Hans Christian Ørsted and the Romantic Legacy*. Edited by R. M. Brain et al. Dordrecht: Springer, 135–158.
- Friedman, Michael. 2006. "Philosophy of Natural Science." Pp. 303–41 in *The Cambridge Companion to Kant and Modern Philosophy*. Edited by Paul Guyer. Cambridge: Cambridge University Press.
- Friedman, Michael. 2004. "Introduction." Pp. vii–xxx in *Metaphysical Foundations of Natural Science*. Immanuel Kant. Translated by Michael Friedman. Cambridge: Cambridge University Press.
- Friedman, Michael. 2002. "Kant, Kuhn, and the Rationality of Science." *Philosophy of Science* 69: 171–190.

- Friedman, Michael. 1992. *Kant and the Exact Sciences*. Cambridge: Harvard University Press.
- Gendler, Tamar, S. 1998. "Galileo and the Indispensability of Scientific Thought Experiment." *The British Journal for the Philosophy of Science* 49: 397–424.
- Glas, Eduard. 1999. "Thought Experimentation and Mathematical Innovation." *Studies in the History and Philosophy of Science* 1: 1–19.
- Gode-von Aesch, Alexander. 1941. *Natural Science in German Romanticism*. New York: Columbia University Press.
- Greenberg, Robert. 2001. *Kant's Theory of A Priori Knowledge*. University Park, PA: Pennsylvania State University Press.
- Guyer, Paul. 1987. *Kant and the Claims of Knowledge*. Cambridge: Cambridge University Press.
- Guyer, Paul (ed.) 2010. *The Cambridge Companion to Kant's Critique of Pure Reason*. Cambridge: Cambridge University Press.
- Harrison, Peter. 2009. *The Fall of Man and the Foundations of Modern Science*. Cambridge: Cambridge University Press.
- Hermann, Armin. 1974. "Das Wissenschaftliche Weltbild Lichtenbergs." Pp. 44–59 in *Aufklärung über Lichtenberg*. Göttingen: Vadenhoeck & Ruprecht.
- Hollindgale, Reginald J. 1990. *The Waste Books: Georg Christoph Lichtenberg*. New York: New York Review of Books.
- Ichikawa, Jonathan, and Jarvis, Benjamin. 2009. "Thought-experiment Intuition and Truth in Fiction." *Philosophical Studies* 142: 221–46.
- Ierodiakonou, Katerina, and Roux, Sophie. 2011. *Thought Experiments in Methodological and Historical Contexts*. Leiden: Brill.
- Janiak, Andrew. 2008. *Newton As Philosopher*. Cambridge: Cambridge University Press.
- Kalin, Martin G. 1972. "Kant's Transcendental Arguments as *Gedankenexperimente*." *Kant Studien* 63: 315–28.
- Kant, Immanuel. (1781) 1998. *Critique of Pure Reason*. Cambridge: Cambridge University Press.
- Kant, Immanuel. (1786) 2004. *Metaphysical Foundations of Natural Science*. Edited and Translated by Michael Friedman. Cambridge: Cambridge University Press.
- King, Peter. 1991. "Mediaeval Thought-Experiments: The Metamethodology of Mediaeval Science." Pp. 43–64 in *Thought Experiments in Science and Philosophy*. Edited by Tamara Horowitz and Gerald Massey. Lanham: Rowman & Littlefield.
- Kitcher, Patricia. 1990. *Kant's Transcendental Psychology*. New York: Oxford University Press.
- Kitcher, Phillip. 2006. "A Priori." Pp. 28–59 in *Kant. The Cambridge Com-*

- panion to Kant and Modern Philosophy*. Edited by P. Guyer. Cambridge: Cambridge University Press.
- Koyré, Alexandre. 1968. *Metaphysics and Measurement*. London: Chapman and Hall.
- Kuhn, Thomas S. (1964) 1977. "A Function for Thought Experiments." Pp. 240–265 in *The Essential Tension*. Chicago: University of Chicago Press.
- Kuhn, Thomas S. (1962) 1996. *The Structure of Scientific Revolutions* 3rd Edn. Chicago; London: University of Chicago Press.
- Kühne, Ulrich. 2005. *Die Methode des Gedankenexperiments*. Frankfurt: Suhrkamp.
- Lakatos, Imre. 1976. *Proofs and Refutations: The Logic of Mathematical Discovery*. Cambridge: Cambridge University Press.
- Lewis, David. 1978. "Truth in Fiction." *American Philosophical Quarterly* 15: 37–46.
- Lichtenberg, Georg Ch. 1806. *Physikalische und Mathematische Schriften*. Edited by Ludwig Christian Lichtenberg and Friedrich Kries, vol. 4. Göttingen: Dietrich.
- Lichtenberg, Georg Ch. 1968. *Schriften und Briefe: Erster Band*. Edited by Wolfgang Promies. München: Carl Hanser.
- Lichtenberg, Georg Ch. 1971. *Schriften und Briefe: Zweiter Band*. Edited by Wolfgang Promies. München: Carl Hanser.
- Lihoreau, Franck (ed.) 2011. *Truth in Fiction*. Piscataway, NJ: Transaction Books.
- Longuenesse, Béatrice. 2006. "Kant on A Priori Concepts: The Metaphysical Deduction of the Categories." Pp. 129–68 in *Kant. The Cambridge Companion to Kant and Modern Philosophy*. Edited by Paul Guyer. Cambridge: Cambridge University Press.
- Macho, Thomas and Annette Wunschel (eds.) 2004. *Science and Fiction: Über Gedankenexperimente in Wissenschaft, Philosophie und Literatur*. Frankfurt am Main: Fischer Verlag.
- Margolis, Eric and Lawrence, Stephen (eds.) 1999. *Concepts: Core Readings*. Cambridge: MIT Press.
- Martindale, Colin. 1987. "Can We Construct Kantian Mental Machines?" *Journal of Mind and Behaviour* 8: 261–268.
- Mautner, Franz and Henry Hatfield. 1959. *The Lichtenberg Reader*. Beacon Hill; Boston: Beacon Press.
- Mautner, Franz and Franklin Miller, Jr. 1952. "Remarks on G. C. Lichtenberg, Humanist-Scientist." *Isis* 43:223–231.
- Miščević, N. 1992. "Mental Models and Thought Experiments." *International Studies in the Philosophy of Science* 6: 215–26.

- Miščević, N. 2007. "Modelling Intuitions and Thought Experiments." *Croatian Journal of Philosophy* VII: 181–214.
- Moue, Aspasia S. et al. 2006. "Tracing the Development of Thought Experiments in the Philosophy of the Natural Sciences." *Journal for General Philosophy of Science* 37: 61–75.
- Nersessian, N. 1992. "How Do Scientists Think? Capturing the Dynamics of Conceptual Change in Science." Pp. 3–44 in *Cognitive Models of Science*. Edited by Ron Giere. Minneapolis: University of Minnesota Press.
- Nersessian, N. 1993. "In the Theoretician's Laboratory: Thought Experimenting as Mental Modeling." *Proceedings of the Philosophy of Science Association* 2: 291–301.
- Nersessian, N. 2007. "Thought Experiments as Mental Modelling: Empiricism without Logic." *Croatian Journal of Philosophy* VII: 125–161.
- Norton, John D. 1996. "Are Thought Experiments Just What You Thought?" *Canadian Journal of Philosophy* 26:333–366.
- Novalis. 1978. *Werke, Tagebücher und Briefe Friedrich von Hardenbergs*, three volumes, edited by Hans-Joachim Mähl. München: Carl Hanser Verlag.
- Novalis. 1964. *Henry von Ofterdingen: A Novel*, translated by Palmer Hilty. Prospect Heights, IL: Waveland Press.
- Novalis. 1960. *Novalis Schriften*, edited by Paul Kluckhorn and Richard Samuel. Stuttgart: W. Kohlhammer.
- Oatley, Keith. 1999. "Why Fiction may Be Twice as True as Fact: Fiction as Cognitive and Emotional Simulation." *Review of General Psychology* 3: 101–117.
- O'Brien, Arctander Wm. 1995. *Novalis: Signs of Revolution*. Durham; London: Duke University Press.
- Ørsted, Hans Christian. (1811) 1920. "Første Indledning til den almindelige Naturlaere." Pp. 151–190 in: *Hans Christian Ørsted: Naturvidenskabelige Skrifter: Samlet Udgave med to Afhandlinger om Hans Virke*, volume III, edited by Kristine Meier, Kopenhagen: Andr. Fred. Høst & Søn.
- Ørsted, Hans Christian. (1811) 1998. "First Introduction to General Physics: A Prospectus of Lectures in this Science." Pp. 282–309 in *Selected Scientific Works of Hans Christian Ørsted*. Edited by Karen Jelved, Andrews D. Jackson, and Ole Knudsen. Princeton: Princeton University Press.
- Pollok, Konstantin. 2002. "'Fabricating a World in Accordance with Mere Fantasy . . . ?' The Origins of Kant's Critical Theory of Matter." *The Review of Metaphysics* 56: 61–97.
- Pollok, Konstantin. 2001. *Kants "Metaphysische Anfangsgründe der*

- Naturwissenschaft*": *Ein Kritischer Kommentar*, Hamburg: Felix Meiner Verlag.
- Popper, Karl. 1959. "On the Use and Misuse of Imaginary Experiments, Especially in Quantum Theory." Pp. 442–56 in *The Logic of Scientific Discovery*. London: Hutchinson.
- Powell, Thomas C. 1990. *Kant's Theory of Self-Consciousness*. New York: Oxford University Press.
- Prinz, Jesse. 2002. *Furnishing the Mind: Concepts and Their Perceptual Basis*. Cambridge: MIT Press.
- Schildknecht, Christiane. 1990. *Philosophische Masken: Literarische Formen der Philosophie bei Platon, Descartes, Wolff und Lichtenberg*. Stuttgart: Metzler.
- Schöne, Albrecht. 1982. *Aufklärung aus dem Geist der Experimentalphysik: Lichtenbergsche Konjunktive*. München: Verlag C.H. Beck.
- Seekins, Kenneth. 1984. "Socratic Philosophy and the Dialogue Form." *Philosophy and Literature* 8: 181–94.
- Sleigh, Charlotte. 1998. "Life, Death, and Galvinism". *Studies in the History and Philosophy of Science. Part C* 29: 219–48.
- Sorensen, Roy A. 1992. *Thought Experiments*. Oxford: Oxford University Press.
- Stern, Joseph P. 1963. *Lichtenberg: A Doctrine of Scattered Occasions*. London: Thames and Hudson.
- Suarez, Mauricio (ed.) 2009. *Fictions in Science: Philosophical Essays in Modeling and Idealization*. New York: Routledge.
- Sutton, C. 1994. "'Nullius in Verba' and 'Nihil in Verbis'": Public Understanding of the Role of Language in Science." *The British Journal for the History of Science* 27: 55–6.
- Thagard, Paul. 2010. *The Brain and the Meaning of Life*. Princeton: Princeton University Press.
- Thagard, Paul and Terence C. Stewart. 2011. "The Aha! Experience: Creativity through Emergent Binding in Neural Networks." *Cognitive Science*: 1–33.
- Waxman, Wayne. 1991. *Kant's Model of the Mind*. New York: Oxford University Press.
- Willaschek, Marcus. 2003. *Der Mentale Zugang zur Welt*. Frankfurt am Main: Klostermann.
- Wilson, Edward O. 1998. *Consilience: The Unity of Knowledge*. New York: Vintage Books.