DO THOUGHT EXPERIMENTS TRANSCEND EMPIRICISM?

In a nutshell, empiricism is the view that all of our knowledge of the world ultimately derives from sense experience. There is, arguably, no other philosophical principle of comparable scope and historic significance that is so widely held by philosophers today. James Robert Brown argues that *thought experiments* provide a counterexample to the principle of empiricism. Thought experiments, which are widely used throughout the sciences, seem to give us insight into the nature of the world; yet at least some thought experiments do not rely, even implicitly, upon knowledge gained through observation. In defending this view, Brown criticizes a position held by John Norton, to the effect that thought experiments are colorful presentations of ordinary arguments from empirical premises. In his response to Brown, Norton challenges Brown's critique of empiricism, and calls for a more thorough investigation into the epistemology of thought experiments. In addition to its implications for broader debates about empiricism, the subject of the role of thought experiments within science is of interest in its own right.

CHAPTER ONE

Why Thought Experiments Transcend Empiricism

James Robert Brown

1.1 Galileo's Rationalism

A recent issue of *Physics World* lists the ten "most beautiful experiments" of all time, the result of a poll of contemporary physicists (Crease, 2002; see also Rogers, 2002). "Beauty" was left undefined, but judging from the list of winners it seems to involve a combination of simplicity in design and execution with deep and far reaching consequences. The examples include the double-slit experiment with single electrons, Millikan's oil-drop experiment to measure the charge on an electron, Newton's decomposition of sunlight through a prism, and so on.[1] They are wonderful examples.

Galileo's experiment on falling bodies is ranked number two, but there are reasons for thinking that this might not be appropriate. Galileo is supposed to have climbed the Leaning Tower of Pisa and dropped objects of different weights. His result, that all bodies fall at the same rate regardless of their weight, is taken to refute the prevailing Aristotelian view, a theory that said that heavy objects fall faster than light ones and that their rates of fall vary in proportion to their weights. There are several problems with proposing that this is one of the most beautiful experiments ever. For one thing, dropping things from a tower isn't really that innovative. As novel ideas go, it's not in the same league as the double-slit idea. Moreover, it is unlikely that Galileo actually performed such an experiment. More striking yet, when a variety of different objects are dropped, the do *not* all fall at the same rate. In particular, cannon balls will hit the ground before musket balls when released at the same time. True,

1 According to the poll, the top ten are: (1) Young's double-slit experiment applied to the interference of single electrons; (2) Galileo's experiment on falling bodies; (3) Millikan's oil-drop experiment; (4) Newton's decomposition of sunlight with a prism; (5) Young's light-interference experiment; (6) Cavendish's torsion-bar experiment; (7) Eratosthenes's measurement of the earth's circumference; (8) Galileo's experiments with rolling balls down inclined planes; (9) Rutherford's discovery of the nucleus; (10) Foucault's pendulum.

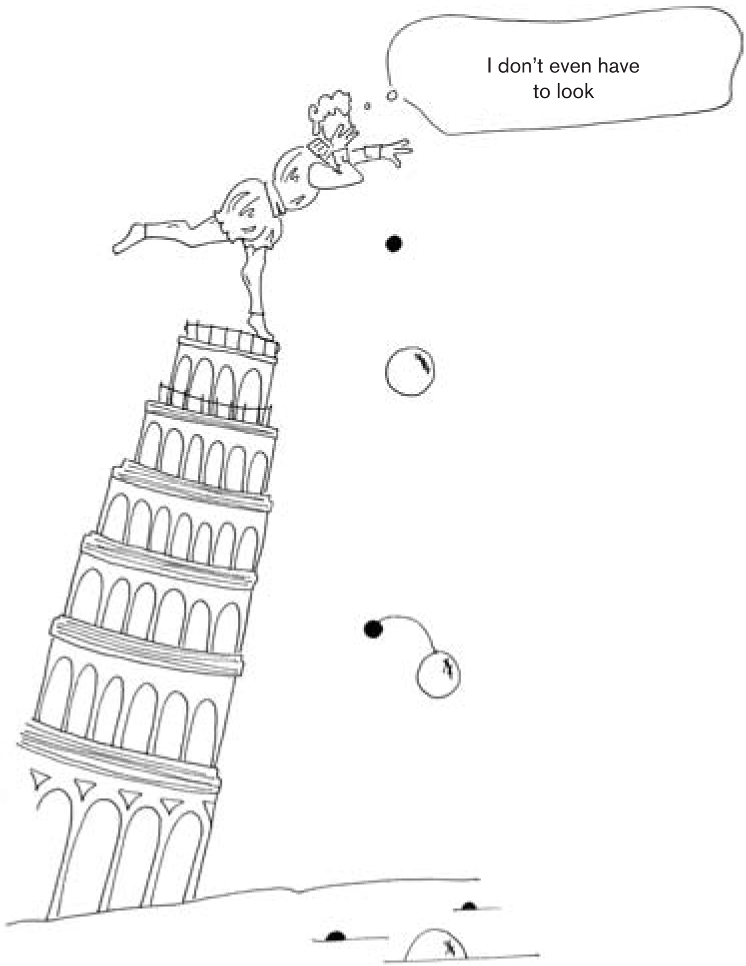

Figure 1.1 Galileo and the determination of rates of fall.

they do not fall in the way Aristotle claimed, but they do not fall as Galileo contended either.

At this point in a discussion of falling objects one often hears the interjection, "But in a vacuum . . . " This caveat must be quickly set aside – Galileo did not perform any experiment in a vacuum, and Aristotle had arguments against the very possibility. It is certainly true that all objects fall at the same rate in a vacuum, but this was not an experimental discovery made by Galileo.

Instead of performing a real experiment on falling bodies, Galileo did something much more clever and profound. He established his result by means of a thought experiment (*Discoursi*, p. 66f). It was, arguably, the most beautiful thought experiment ever devised. It would certainly get my vote as the "most beautiful" ever. It's brilliantly original and as simple as it is profound.

Aristotle and common sense hold that a heavy body will fall faster than a light one. (We can symbolize this as H > L). But consider figure 1.1, where a heavy cannon ball is attached to a light musket ball (H + L). It must fall faster than the cannon ball alone (H + L > H). Yet the compound object must also fall slower (H + L < H), since the lighter part will act as a drag on the heavier part. Now we have a contradiction (H + L > H & H + L < H). That's the end of Aristotle's theory. But we can go further. The right account of free fall is now perfectly obvious: they all move at the same speed (H = L = H + L).

Galileo once remarked that "Without experiment, I am sure that the effect will happen as I tell you, because it must happen that way" (*Diologo*, p. 145). What he had in mind in this surprising pronouncement is that often a thought experiment will yield the result. An actual experiment is not needed and may even be impossible. The great French historian of science, Alexandre Koyré, once remarked "Good physics is made *a priori*" (1968, p. 88), and he claimed for Galileo "the glory and the merit of having known how to dispense with [real] experiments" (1968, p. 75). This remarkable assessment of Galileo is right. Some thought experiments do transcend empirical sensory experience.

1.2 Some Examples

It's difficult to say precisely what a thought experiment is. However, it is not important. We know them when we see them, and that's enough to make talking about them possible. A few features are rather obvious. Thought experiments are carried out in the mind and they involve something akin to experience; that is, we typically "see" something happening in a thought experiment. Often, there is more than mere observation. As in a real experiment, there might be some calculating, some application of theory, and some guesswork and conjecture.

Thought experiments are often taken to be idealizations. This is sometimes true, but idealization is neither necessary nor sufficient. Thinking about how something might move along a frictionless plane needn't be a thought experiment at all, since it might involve nothing more than a calculation. And no idealizations are involved in Schrödinger's cat. Some people also want to include the claim that a genuine thought experiment *cannot* be performed in the actual world. I doubt this very much, since what seems to matter is whether we can get the result by thinking, not whether we can get it in some other way, as well. For now I'd prefer simply to leave these considerations out of the definition, allowing that maybe these features should or maybe they should not be part of the definition. These are things to be argued, debated, and, with luck, resolved at the end of inquiry, not fixed by stipulation at the outset. The best way to get a grip on what thought experiments are is to simply look at lots of examples. For the sake of illustration, I'll briefly give a few.

One of the most beautiful early examples is from Lucretius's *De Rerum Natura* (although it has an earlier history). It attempts to show that space is infinite. If there is a boundary to the universe, we can toss a spear at it. If the spear flies through, then it isn't a boundary after all. And if the spear bounces back, then there must be something beyond the supposed edge of space, a cosmic brick wall that stopped the spear, a wall that is itself in space. Either way, there is no edge of the universe; so, space is infinite.

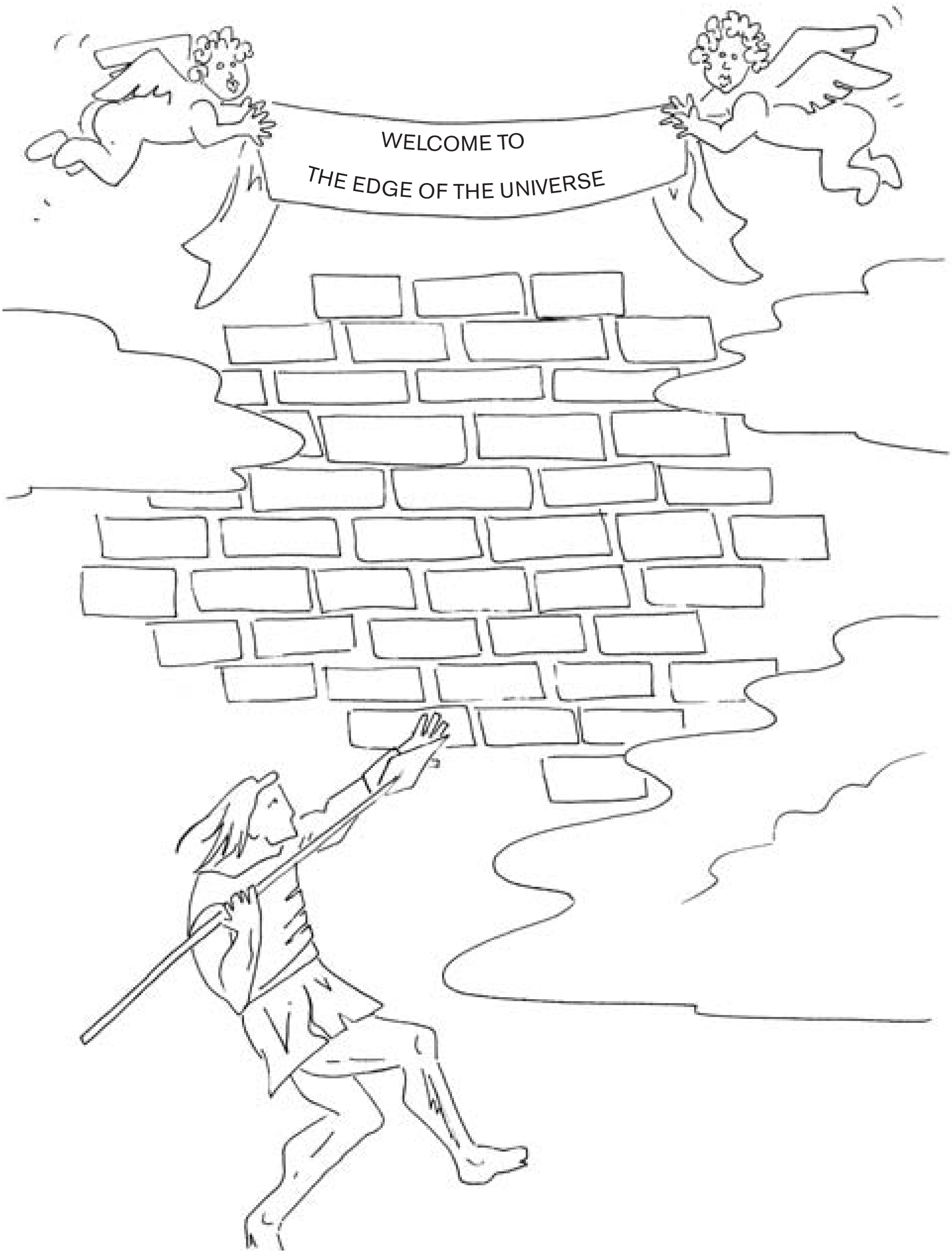

Figure 1.2 Throwing a spear at the edge of space.

This example nicely illustrates many of the common features of thought experiments: We visualize some situation; we carry out an operation; we see what happens. Although we use empirical concepts, we often can't carry out an empirical test. It also illustrates their fallibility. In this case we've learned how to conceptualize space so

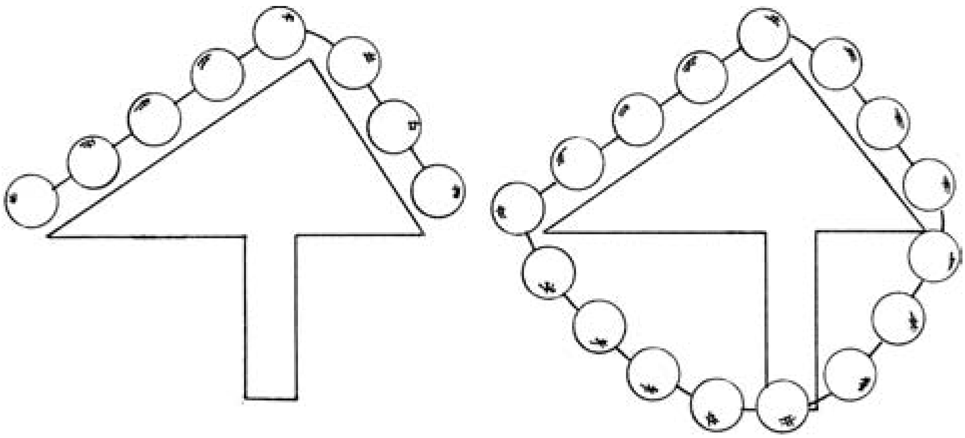

Figure 1.3 A chain draped over frictionless planes.

that it is both finite and unbounded. It also illustrates the fact that a thought experiment is a scheme that could be implemented in different ways. Here, a spear is thrown. John Norton (see chapter 2, this volume) gives a slightly different version, stemming from an older source. It's still reasonable to call them the same thought experiment.

Ernst Mach (who seems to have coined the much used expression *Gedankenexperiment*) developed an interesting empiricist view in his classic, *The Science of Mechanics*. He claims we have lots of "instinctive knowledge" picked up from experience. This needn't be articulated at all, but we become conscious of it when we consider certain situations, such as in one of his favorite examples, due to Simon Stevin. A chain is draped over a double frictionless plane (figure 1.2, left). How will it move? We could plausibly imagine it sliding one way or the other or staying fixed. But we don't know which. Add some links (as in figure 1.2, right) and the correct answer becomes obvious. It will remain fixed. So, the initial setup must have been in static equilibrium. Otherwise, we would have a perpetual motion machine. Our experience-based "instinctive knowledge," says Mach (1960, p. 34ff), tells us that this is impossible.

This beautiful example illustrates another feature of thought experiments: they often use background information, just as real experiments do – in this case, that there is no perpetual motion.

According to Maxwell's theory of electrodynamics, light is an oscillation in the electromagnetic field. The theory says that a *changing* electric field, E, gives rise to a magnetic field, M, and a *changing* magnetic field gives rise to an electric field. If I were to jiggle a charge, it would change the electric field around it, which in turn would create a magnetic field which in turn would create an electric field, and so on. This wave motion spreading through the electromagnetic field with velocity *c* is none other than light itself.

When he was only 16, Einstein wondered what it would be like to run so fast as to be able to catch up to the front of a beam of light. Perhaps it would be like running toward the shore from the end of a pier stretched out into the ocean, with a wave coming in alongside. There would be a hump in the water that would be stationary

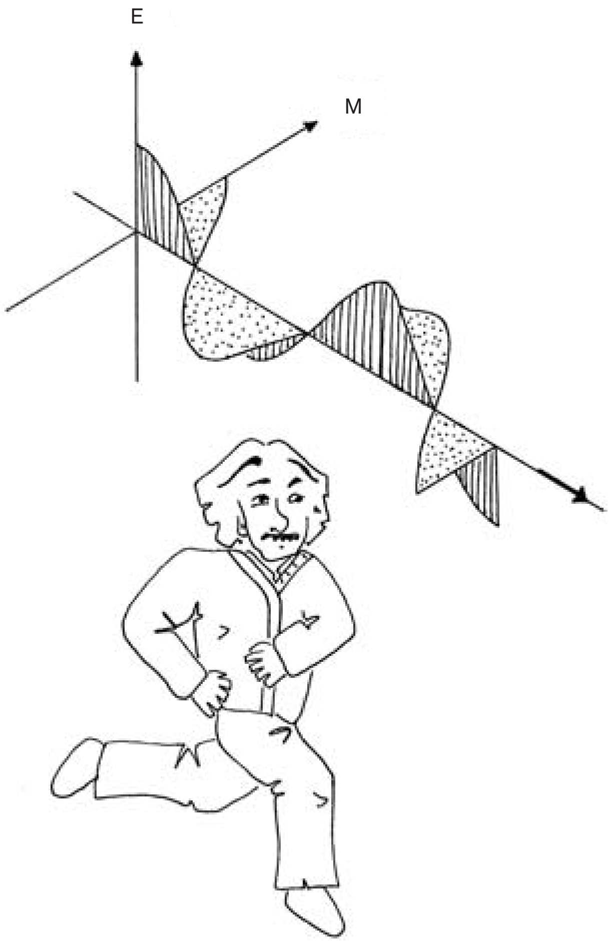

Figure 1.4 Einstein chases a light beam.

with respect to the runner. Would it be the same for a light wave? No, since change is essential; if either the electric or the magnetic field is static, it will not give rise to the other and hence there will be no electromagnetic wave. "If I pursue a beam of light with the velocity *c* (velocity of light in a vacuum)," said Einstein, "I should observe such a beam of light as a spatially oscillatory electromagnetic field at rest. However, there seems to be no such thing, whether on the basis of experience or according to Maxwell's equations" (1949, p. 53).

Newton's bucket is one of the most famous thought experiments ever. It's also perfectly doable as a real experiment, as I'm sure just about everyone knows from personal experience. The two-spheres example, which is also described in the famous *scholium* to definition 8 of the *Principia*, is not actually doable. So let's consider it. We image the universe to be completely empty except for two spheres connected by a cord. The spheres are of a material such that they neither attract nor repel one another. There is a tension in the cord joining them, but the spheres are not moving

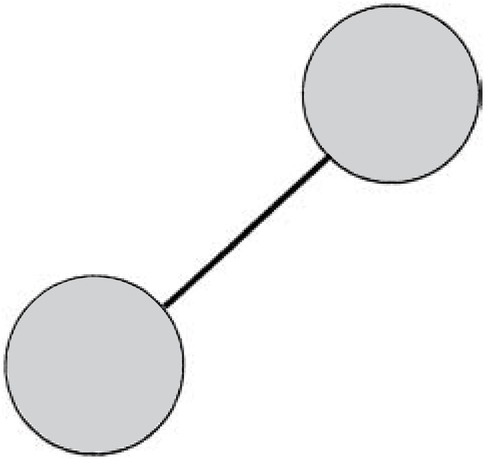

Figure 1.5 Newton's spheres in otherwise empty space.

toward one another under the force in the cord. Why not? Newton offers an explanation: they are rotating with respect to space itself; their inertial motion keeps them apart. And so, Newton concludes, absolute space must exist.

1.3 Different Uses of Thought Experiments

Even with only a few examples such as these, we can begin to draw a few conclusions. For one thing, thought experiments can mislead us, not unlike real experiments. The Lucretius example leads us to infinite space, but we have since learned how to conceive of an unbounded but still finite space. (A circle, for example, is a finite, unbounded one-dimensional space.) Some of these thought experiments couldn't be actually performed. Einstein can't run that fast, so catching up to a light beam is something done in thought alone. Needless to say, Newton's two-spheres thought experiment is impossible to perform, since we can't get rid of the entire universe to try it out. On the other hand, the thought experiments of Galileo and Stevin come close to being doable – not exactly, perhaps, but we can approximate them rather well. This, I think, shows that being able or not being able to perform a thought experiment is rather unimportant to understanding them. The so-called counterfactual nature of thought experiments is overstressed. And, as I mentioned earlier, so is idealization.

More interesting is the different roles that they play (for a taxonomy, see Brown, 1991, ch. 2). Some play a negative or refuting role. The Einstein example amounts to a *reductio ad absurdum* of Maxwell's electrodynamics, in conjunction with what were then common-sense assumptions about motion (i.e., Galilean relativity). Stevin, by contrast, teaches us something positive, something new about static equilibrium. Newton's spheres do not so much give us a new result but, rather, give us a remarkable phenomenon, something that needs to be explained. The thought experiment establishes a phenomenon; the explanation comes later. And the best explanation, according to Newton, is the existence of absolute space. This way of looking at it is confirmed, it seems to me, by the way in which Berkeley and Mach reacted to the thought experiment. They didn't deny that rotation with respect to absolute space is the best explanation for the tension in the cord. Instead, they denied there would be any tension in the first place, if the spheres are not moving together. Or the two spheres would move together, if there were any tension. That is, they didn't bother to challenge the explanation of the phenomena that Newton posited; they challenged the alleged phenomenon itself.

1.4 *A priori* Knowledge of Nature

The most interesting example is surely Galileo's. This seems to play a negative role – it refutes Aristotle by means of a *reductio ad absurdum* – then, in a positive vein, it establishes a new theory of motion. There are lots of wonderful thought experiments, but only a small number work in this way. Elsewhere I have called them Platonic (Brown, 1991). I think they are rather remarkable – they provide us with *a priori* knowledge of nature.

My reasons for calling this thought experiment *a priori* – it transcends experience – are rather simple and straightforward. First, although it is true that empirical knowledge is present in this example, there are no new empirical data being used when we move from Aristotle's to Galileo's theory of free fall. And, second, it is not a logical truth. After all, objects could fall on the basis of their color, say, with red things falling faster than blue things. If this were the case, then Galileo's thought experiment wouldn't be effective in refuting it. After all, sticking two red things together doesn't make one thing that is even redder.

This last point is worth a moment's reflection. Properties such as weight or mass are commonly called "extensive," while other properties such as color or heat are known as "intensive." Extensive properties can be added in a way that suggests that they have a structure similar to adding real numbers; the intensive properties don't. Thus, physically adding two bodies at 50 kg each yields one body at 100 kg. But physically adding two bodies at 50 degrees will not yield one body at 100 degrees. Galileo's thought experiment leads to a result that is very much more powerful than is initially apparent, even though the initial result is already very powerful. What Galileo has in effect shown (though it is not manifest), is that for any extensive property, any theory along the lines of Aristotle will succumb to the same problem that the particular case based on weight did. Suppose that bodies fall due to X (where X is an extensive property; for example, weight, height, and so on), and that their (supposed) rates of fall differ due to differing quantities of X. We could now follow the Galileo pattern: we could join two bodies with different amounts of X and ask how fast this new body will fall. The inevitable answer will be: "On the one hand, faster because . . . , and on the other hand, slower because . . . " We would have generated a contradiction.

As an exercise, readers are invited to try it for themselves with various extensive properties such as height, volume, and so on. They should also try Galileo's style of argument on the so-called "fifth force," which was postulated a few years ago as a slight modification to gravity. Objects would fall, according to this theory, at slightly different rates, due to their differing chemical composition. Would the Galileo style of thought experiment work in this case? Is the property *being made of aluminum* intensive or extensive?

In considering an arbitrary X, I have remained focused on X being the cause of *falling*. But I'm sure this could be generalized too, although I'm not sure to what extent and in what directions this generalization might go. In any case, the importance of Galileo's discovery goes well beyond the case of objects falling *due to their heaviness*. This alone is enough to say that we have *a priori* knowledge of nature. The additional power of the generalized Galileo pattern of thinking is rather clearly not derivable from empirical premises; it goes too far beyond anything we have experienced. Yet it is clear as day. This, it seems to me, is another case of *a priori* intuition at work.

Let me call the argument so far *the prima facie case for a priori knowledge of nature*. It is, I admit, not yet a completely decisive argument. Two things are needed to strengthen it. One is an explanation of how we are able to acquire this *a priori* knowledge via thought experiments. The second is to cast doubt on rival views.

1.5 Mathematical Intuitions and Laws of Nature

Even if one thought the Galileo example impressive, one might still resist the idea of *a priori* knowledge of nature, since it seems so contrary to the currently accepted principles of epistemology, where empiricism reigns supreme. How is it even possible to have experience that transcends knowledge? I can't address all concerns, but I will try to develop an analogy that will give an account of how thought experiments might work. I'll start with mathematical Platonism. Kurt Gödel famously remarked,

> Classes and concepts may, however, also be conceived as real objects . . . existing independently of our definitions and constructions. It seems to me that the assumption of such objects is quite as legitimate as the assumption of physical bodies and there is quite as much reason to believe in their existence. They are in the same sense necessary to obtain a satisfactory system of mathematics as physical bodies are necessary for a satisfactory theory of our sense perceptions . . . (1944, p. 456f)

> . . . despite their remoteness from sense experience, we do have something like a perception also of the objects of set theory, as is seen from the fact that the axioms force themselves upon us as being true. I don't see any reason why we should have any less confidence in this kind of perception, i.e., in mathematical intuition, than in sense perception, which induces us to build up physical theories and to expect that future sense perceptions will agree with them and, moreover, to believe that a question not decidable now has meaning and may be decided in the future. The set-theoretical paradoxes are hardly more troublesome for mathematics than deceptions of the senses are for physics . . . [N]ew mathematical intuitions leading to a decision of such problems as Cantor's continuum hypothesis are perfectly possible . . . (1947/1964, p. 484)

I take these passages to assert a number of important things, including the following: mathematical objects exist independently from us; we can perceive or intuit them;

our perceptions or intuitions are fallible (similar to our fallible sense perception of physical objects); we conjecture mathematical theories or adopt axioms on the basis of intuitions (as physical theories are conjectured on the basis of sense perception); these theories typically go well beyond the intuitions themselves, but are tested by them (just as physical theories go beyond empirical observations but are tested by them); and in the future we might have striking new intuitions that could lead to new axioms that would settle some of today's outstanding questions. These are the typical ingredients of Platonism. The only one I want to focus on is the perception of abstract entities, commonly called intuition, or seeing with the mind's eye. Gödel took intuitions to be the counterparts of ordinary sense perception. Just as we can see some physical objects – trees, dogs, rocks, the moon – so we can intuit some mathematical entities. And just as we can see that grass is green and the moon is full, so we can intuit that some mathematical propositions are true.

I cannot argue here for mathematical Platonism, since that would take us too far off course. I present it dogmatically, with only Gödel's authority to strengthen its plausibility (for more detail, see Brown, 1999). The key thing to take from this is that we can have a kind of perception, an intuition, of abstract entities. This, of course, is *a priori* in the sense that ordinary sense perception is not involved. Seeing with the mind's eye transcends experience.

In addition to mathematical Platonism, we need a second ingredient, this time involving laws of nature, in order to have the full analogy that I'm trying to develop.

What is a law of nature? What is it about a law that makes it a law? Let's take "Ravens are black" to be a law. It's not a very good example of a law, but it's simple and will easily illustrate the issue. There are several accounts of laws – the one favored most by modern empiricists is some version of Hume's account (for further discussion, see the contributions to this volume by John Roberts and Harold Kincaid, chapters 7 and 8). According to Hume, a law is just a regularity that we have noticed and come to expect in the future. We have seen a great many ravens and they have all been black; and we expect future ravens also to be black. "All events seem entirely loose and separate." says Hume: "One event follows another, but we never can observe any tie between them." (*Enquiry*, p. 74) ". . . after a repetition of similar instances, the mind is carried by habit, upon the appearance of one event, to expect its usual attendant, and to believe that it will exist" (*Enquiry*, p. 75). Causality and the laws of nature are each nothing more than regularities that are expected. To say that fire causes heat, or that it is a law of nature that fire is hot, is to say nothing more than that fire is constantly conjoined with heat. Hume defined cause as "an object, followed by another, and where all the objects similar to the first are followed by objects similar to the second" (*Enquiry*, p. 76). We can't see a "connection" between fire and heat such that if we knew of the one we could know that the other must also occur. All we know is that whenever in the past we have experienced one we have also experienced the other – hence the "regularity" or "constant conjunction" view of causality and laws of nature.

The appeal to empiricists is evident. All that exists are the regular events themselves; there are no mysterious connections between events – no metaphysics to cope with. The general form of a law is simply a universal statement. "It is a law that As are Bs" has the form $(\forall x)\ (\mathrm{A}x \supset \mathrm{B}x)$.

There are all sorts of problems with this view. For one thing, laws are subjective. If there were no conscious beings with expectations, there would be no laws of nature. This is very counterintuitive. Secondly, there is a problem with so-called accidental generalizations. Consider: "All the books in Bob's office are in English." Suppose this is true and, moreover, that in the entire history of the universe a non-English book never makes its way into Bob's office. When I pick any book from the shelf I fully expect it to be in English, and I am never wrong. This has the form of a law of nature, on Hume's view. But this too seems absurd.

Problems such as these and other difficulties have cast serious doubt on the hope that a simple empiricist account might work. A new, rival account of laws has been proposed (independently) by David Armstrong, Fred Dretske, and Michael Tooley. Each claims that laws of nature are relations among universals; that is, among abstract entities which exist independently of physical objects, independently of us, and outside of space and time. At least this is so in Tooley's version (which I prefer); it's a species of Platonism.

The "basic suggestion," according to Tooley, "is that the fact that universals stand in certain relationships may logically necessitate some corresponding generalization about particulars, and that when this is the case, the generalization in question expresses a law" (1977, p. 672).

Thus, a law is not a regularity; rather, it is a link between properties. When we have a law that Fs are Gs, we have the existence of universals, F-ness and G-ness, and a relation of necessitation between them (Armstrong symbolizes this as N(F,G)). A regularity between Fs and Gs is said to hold in virtue of the relation between the universals F and G. "[T]he phrase 'in virtue of universals F and G' is supposed to indicate," says Armstrong, that "what is involved is a real, irreducible, relation, a particular species of the necessitation relation, holding between the universals F and G . . . " (1983, p. 97).

The law entails the corresponding regularity, but is not entailed by it. Thus we have

$$N(\mathrm{F},\mathrm{G}) \rightarrow (\forall x)(\mathrm{F}x \supset \mathrm{G}x)$$

and yet

$$(\forall x)(\mathrm{F}x \supset \mathrm{G}x) \not\rightarrow N(\mathrm{F},\mathrm{G})$$

The relation N of nomic necessity is understood to be a primitive notion. It is a theoretical entity posited for explanatory reasons. N is also understood to be contingent. At first sight, this seems to be a contradiction. How can a relation of necessitation be contingent? The answer is simple: In this possible world, Fs are required to be Gs, but in other possible worlds Fs may be required to be something else. The law N(F,G) is posited only for this world; in other worlds perhaps the law N(F,G′) holds.

Some of the advantages of a realist view of laws are immediately apparent. To start with, this account distinguishes – objectively – between genuine laws of nature and accidental generalizations. Moreover, laws are independent of us – they existed before we did and there is not a whiff of subjectivity or relativism about them. Thus, they can be used to *explain* and not merely summarize events. As with mathematical

Platonism, I cannot attempt to defend this account of laws of nature beyond what I have briefly mentioned. I can only hope to have said enough about it that readers can glimpse its advantage and see why it might be preferred over a Humean account.

The reason for this digression into mathematical Platonism and the realist account of laws of nature is to try to develop some idea of how it is that a thought experiment could transcend experience; that is, how it could yield *a priori* knowledge of nature. I'll now try to explain with a simple argument. According to Platonism, we can intuit some mathematical objects, and mathematical objects are abstract entities. Thus, we can (at least in principle) intuit abstract entities. According to the realist account of laws of nature, laws are also abstract entities. Thus, we might be able (at least in principle) to intuit laws of nature as well. There is one situation in which we seem to have a special access to the facts of nature, namely in thought experiments. Thus, it is possible that thought experiments (at least in some cases) allow us to intuit laws of nature. Intuitions, remember, are nonsensory perceptions of abstract entities. Because they do not involve the senses, they transcend experience and give us *a priori* knowledge of the laws of nature.

Let me head off one concern right away. *A priori* knowledge is not certain knowledge. Intuitions are open to mistakes, just as ordinary sense perceptions are. Mathematics often presents itself as a body of certain truths. This is not so. The history of mathematics is filled with false "theorems." Just think of Russell's paradox, for example. But certainty, as Gödel remarked in the passage quoted above, is not a part of contemporary Platonism. And certainty need not be part of any thought experiment.

1.6 Norton's Arguments and Empiricism

Let me review the argument so far. I described the Galileo thought experiment concerning falling objects, then gave what I called the *prima facie* case for *a priori* knowledge of nature. This was followed by an account of how such *a priori* knowledge could come about, drawing on mathematical Platonism and the realist view of laws of nature. Now I move to the final ingredient in my overall argument, which is to cast doubt on rivals. Ideally, I would refute them all. That's impossible. What I can do is look at one alternative, John Norton's, which I take to be my most serious rival, and try to undermine his account. Norton himself surveys a number of rival accounts of thought experiments in chapter 2 of this volume (see his section 2.5). His descriptions and criticisms of these various views are admittedly brief, but (aside from his criticism of my own view, of course) I'm inclined to agree with what he says about these other views.

In arguing that thought experiments do not transcend experience, Norton is making two claims that give him this conclusion:

1 (Argument thesis): A thought experiment is an argument. It may be disguised, but on reconstruction it begins with premises and the conclusion is derived by means of the rules of deductive logic or inductive inference. (These rules are

understood quite liberally.) Unlike a real experiment, no observation or observation-like process occurs.

2 (Empiricism): All knowledge stems from sensory experience.

These two assumptions lead to the following conclusion:

> A thought experiment is a good one (or "reliable," as Norton puts it) insofar as the premises are empirically justified and the conclusion follows by good rules of inference.

From this, he can now derive the crucial conclusion:

> Thought experiments do not transcend experience.

This conclusion does indeed follow from his two key assumptions, so if I am to reject his conclusion I must attack his premises. One way is to refute empiricism. The other way is to refute the thesis that thought experiments are really just arguments. Before going into this, a few remarks are in order.

A way of seeing the difference between Norton and me is to consider, first, real experiments. We would agree (as would most people) that a real experiment carries us from a perception (and some possible background propositions) to a proposition (a statement of the result). The so-called experimental result may be the culmination of a great deal of theorizing and calculating, but somewhere along the way the experimenter has had to look at something; for example, a thermometer, a streak in a cloud chamber, or a piece of litmus paper. I hold that a thought experiment has a similar structure. The only difference is that the perception is not a sense perception but, rather, is an intuition, an instance of seeing with the mind's eye. In other words, a thought experiment (or at least some of them) carries us from a *(nonsensory) perception* (and some possible background propositions) to a proposition, just as a real experiment does. Norton would deny this similarity and instead claim that a thought experiment carries us from a proposition (and some possible background propositions) to a proposition. For him, it is argument and inference all the way; there is no perception of any kind involved during the thought experiment. Let me quickly add a warning about a possible misunderstanding. In a thought experiment one sees *in the imagination* falling bodies, just as in a real thought experiment. Norton and I would both agree that one "sees" something in this sense. But I would claim – and he would deny – that one also "sees" something else, a law of nature, and that this is a nonsensory experience that is different from any inference that might be made in the thought experiment.

Norton has two important considerations in favor of his deflationary view. One is that his account involves only unproblematic ingredients; that is, empirical observations and empirically acceptable inference patterns. The naturalist-minded take a view such as mine, involving intuitions of abstract entities, to be highly problematic and they regard the idea of seeing with the mind's eye as hopeless. Secondly, Norton has so far managed to reconstruct every thought experiment that he has examined into the argument pattern that he champions. There seem to be no counterexamples. These two considerations make Norton's view very plausible. He goes so far as to claim that unless his view is defeated, it should be accepted as the "default" account. I think he's right about this. The current philosophical climate favors empiricism and opposes abstract entities. Since my view flies in the face of this, the burden of proof is on me.

A second aside, this time historical: the debate between Norton and myself is somewhat reminiscent of an older debate over how to understand Descartes's *cogito, ergo sum*. Is it an argument or an immediate intuition, or something else? The term "*ergo*" suggests an argument, with "I think" as the premise and "I exist" as the conclusion. As an argument, however, it's grossly invalid as it stands. The alternative that I'm inclined to favor is that "I exist" is an immediate intuition, evident on noticing that I happen to be thinking. I will not pursue this point, but I think the parallel with interpreting thought experiments is interesting and possibly instructive (for a discussion, see Hintikka, 1962).

In trying to make my case, let me raise a small point, first. Even if every thought experiment could be reconstructed in Norton's argument form, this would not guarantee that this is what thought experiments essentially are. Norton calls it the "elimination thesis." It says: "Any conclusion reached by a (successful) scientific thought experiment will also be demonstrable by a non-thought-experimental argument" (1991, p. 131). This claim – bold though it is – may be true without thought experiments actually being arguments, disguised or not. For instance, most of us would say that we make judgments of the relative size of other people based on our perception of their geometric shape. Suppose that Norton claimed this is not so, and that what we really do is count molecules. He then shows us that every time we judge that person A is bigger than person B, it turns out that on his (laborious) reconstruction, A, has more molecules than B. Even a success rate of 100 percent in reconstructing judgments of size does not refute the common-sense claim that we judge size by means of visual shape.

There is no denying that such reconstructions would be significant. But for Norton to make his case for arguments, mere reconstruction is not enough. He must also make the case that a thought experiment gives some sort of clue to the (hidden) argument; perhaps it's there in a sketchy or embryonic form and can be readily grasped. Norton acknowledges this point when he remarks that ". . . we should expect the schema of this logic not to be very complicated, so that they can be used tacitly . . . " (this volume, chapter 2, p. **[13]** – I'll return to this point below). If this were not the case, then there must be something else going on of great epistemic significance. This, of course, is what I think, even if a Norton-type reconstruction is possible in every case.

My strategy will be to find examples of thought experiments that fail to fit the Norton pattern of being an argument with empirically justified premises. It is the argument ingredient that I will focus upon, but the empiricism ingredient will also be challenged as a byproduct. To do this I will use a mathematical example, so the condition requiring the premises to be empirically justified will be set aside for now. My aim is to undermine Norton's claim that a thought experiment is an argument. I'll begin with an easy, but impressive, warm-up example.

Consider the following simple picture proof of a theorem from number theory.

THEOREM. $1 + 2 + 3 + \ldots + n = n^2/2 + n/2$.

Proof:

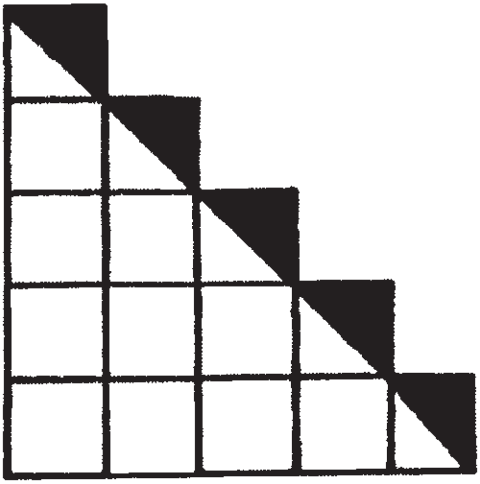

Figure 1.6 A picture proof.

This is a kind of thought experiment, since the evidence is visual. It's quite different than a normal mathematical proof, which is a verbal/symbolic entity. Norton would grant that the picture works as a proof, but only because there is a traditional proof that we can give of this theorem. That is, we can reconstruct such pictorial evidence with proper proofs, in this case a proof by mathematical induction. Such a proof would, of course, be an argument in Norton's sense.

It is certainly true that a traditional proof (by mathematical induction) exists for this theorem. But I think we can still ask: Does the picture proof suggest the traditional proof? When given a "proof sketch" we can fill in the details, but is something like that going on here? Speaking for myself, I don't "see" induction being suggested in the proof at all. Perhaps others do. For Norton to be right, the picture must be a crude form of induction (or a crude form of some other correct argument form); otherwise, he has not made his case. The mere fact that an argument *also* exists does not refute the claim that the thought experiment works evidentially in a nonargument way. Norton, as I mentioned above, recognizes this point. In describing his generalized logic (his allowable rules of inference in a argument), he says (I repeat the quote): ". . . we should expect the schema of this logic not to be very complicated, so that they can be used tacitly by those who have the knack of using thought experiments reliably."

An example such as this one is unlikely to be decisive. I would claim that the picture is conclusive evidence in its own right for the theorem. Norton would deny this, claiming that the real evidence is the proof by mathematical induction. I might then claim that the inductive proof is not "tacitly" contained in the picture. He might disagree. We're at a standoff. At this point, readers might try to decide for themselves which account is more plausible and persuasive.[2]

2 I have often presented this example to audiences of mathematicians and asked if any "see" induction in the picture. About half of any audience says "Yes," while the other half assert just as strenuously that they see no induction at all.

Because this example is not decisive, I'll try another, one that is more complex and certainly bound to be more controversial. It is, however, quite interesting in its own right, and worth the effort of working through the details, though it will not likely be decisive either.

Christopher Freiling (1986) produced a remarkable result that has gone largely unnoticed by philosophers. He refuted the continuum hypothesis. Did he *really* refute the continuum hypothesis? That's hard to say. The standards for success in such a venture are not normal mathematical standards, since the continuum hypothesis (CH) is independent of the rest of set theory. So any "proof" or "refutation" will be based on considerations outside current mathematics. Because of this, Freiling calls his argument "philosophical."

First of all, we need some background. The cardinality of the set of natural numbers is the first infinite cardinal number. If $\mathrm{N} = \{0, 1, 2, \ldots\}$, then $|\mathrm{N}| = \aleph_0$. Any set this size or finite is called *countable* because its members can be paired with the counting numbers. But not all sets are countable. The infinite cardinal numbers increase without bound: $\aleph_0, \aleph_1, \aleph_2, \ldots$ It is known that the cardinality of the real numbers is uncountable; it is greater than that of the natural numbers, $|\mathrm{R}| = 2^{\aleph_0} > \aleph_0$. The interesting question is which one of the cardinal numbers this is. Does $|\mathrm{R}|$ equal $\aleph_1$ or $\aleph_{27}$, or which? The continuum hypothesis is the conjecture that $|\mathrm{R}| = \aleph_1$. Is it true? It has been shown that CH is independent of the rest of set theory, which means that it cannot be proven or refuted on the basis of the existing standard axioms. In short, it cannot be proven or refuted in the normal sense of those terms.

The second thing to mention is that we shall take ZFC for granted. ZF is Zermelo–Frankel set theory, which is standard. The C refers to the axiom of choice, which we also assume. An important consequence of ZFC is the so-called *well ordering principle*. It says that any set can be well ordered; that is, can be ordered in such a way that every subset has a first element. The usual ordering, <, on the natural numbers is also a well ordering. Pick any subset, say $\{14, 6, 82\}$; it has a first element, namely 6. But the usual ordering on the real numbers is not a well ordering. The subset $(0, 1) = \{x\colon 0 < x < 1\}$, for instance, does not have a first element. Nevertheless, ZFC guarantees that the real numbers can be well ordered by some relation, <, even though no one has yet found such a well ordering. Now we can turn to CH.

Imagine throwing darts at the real line, specifically at the interval [0,1]. Two darts are thrown and they are independent of one another. The point is to select two random numbers. We assume ZFC and further assume that CH is true. Thus, the points on the line can be well ordered so that, for each $q \notin [0, 1]$, the set $\{p \notin [0, 1]\colon p < q\}$ is countable. (Note that < is the well ordering relation, not the usual *less than*, <.) The well ordering is guaranteed by ZFC; the fact that the set is countable stems from the nature of any ordering of any set that has cardinality $\aleph_1$. To get a feel for this, imagine the set of natural numbers. It is infinite, but if you pick a number in the ordering, there will be only finitely many numbers earlier – and infinitely many numbers later. Similarly, pick a number in an well ordered set that is $\aleph_1$ in size and you will get a set of earlier members that is at most $\aleph_0$ in size, and possibly even finite. In any case, it will be a countable set.

We shall call the set of elements that are earlier than the point q in the well ordering S_q. Suppose that the first throw hits point q and the second hits p. Either $p < q$

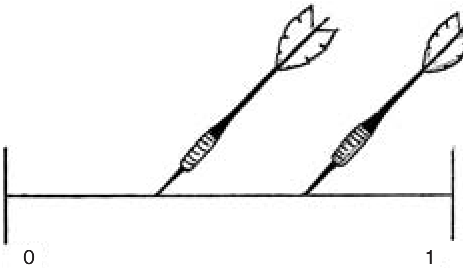

Figure 1.7 Darts picking out independent real numbers at random.

or vice versa; we'll assume the first. Thus, $p \in S_q$. Note that S_q is a *countable* subset of points on the line. Since the two throws were independent, we can say that the throw landing on q defines the set S_q "before" or "independently from" the throw that picks out p. The measure of any countable set is 0. So the probability of landing on a point in S_q is 0. (It is not important to understand the notion of *measure*, only the consequence for probability theory.) While logically possible, this sort of thing is almost never the case. Yet it will happen every time there is a pair of darts thrown at the real line. Consequently, we should abandon the initial assumption, CH, since it led to this absurdity. Thus, CH is refuted and so the number of points on the line is greater than $\aleph_1$.

Notice that if the cardinality of the continuum is $\aleph_2$ or greater, then there is no problem (at least as set out here), since the set of points S_q earlier in the well ordering need not be countable. Thus, it would not automatically lead to a zero probability of hitting a point in it.[3]

There is one aspect of this example on which I should elaborate. The darts give us a pair of real numbers picked at random. These are "real random variables" says Mumford (2000), whose version of the thought experiment I have followed. The concept of random variable at work here is not the mathematical concept found in measure theory (a defined concept inside set theory that will not yield ~CH). Moreover, the two real numbers are picked independently; either could be considered as having been chosen "first." This means that the example cannot be dismissed in the way we might dismiss someone who said of a license number, say, 915372, on a passing car: "Wow, there was only a one in a million chance of that!" We're only impressed if the number is fixed in advance. The independence and randomness of the darts guarantees the symmetry of the throws, so either could be considered the first throw that fixes the set of real numbers that are earlier in the well ordering.

An example such as this must be controversial. Only a minority of mathematicians has accepted it. I am going to assume that it works as a refutation of CH. But I realize that Norton could quite reasonably dismiss it. However, because the example is of the utmost intrinsic interest and is of the greatest importance to mathematics, and because it might prove decisive in the debate with Norton, I find it wholly irresistible.

3 Freiling (1986) actually goes on to show that there are infinitely many cardinal numbers between zero and the continuum.

Recall Norton's view: We start with established propositions and then, using deductive or inductive principles, we infer a conclusion. There is nothing in the process of getting to the conclusion that could count as a perception of any sort. Norton also requires the premises to be empirically acceptable, but we can ignore that here, since the example is mathematical. My view is that we perhaps start with some established propositions and rules of inference, but somewhere in the process we make some sort of empirical experience-transcending observation, we have an intuition, we see something with the mind's eye, and this is essential in reaching the conclusion. Argument alone from already established premises is not sufficient. Which pattern, Norton's or mine, does the CH example follow?

It cannot follow the Norton pattern. The conclusion cannot be derived from the initial assumptions, since CH is independent of the rest of set theory. That is, if we tried to formalize the Freiling example, we would violate existing principles of mathematics. So it cannot be a deductive argument from established premises. Is it perhaps inductive? This seems most unlikely too. Of course, we might be suspicious of the reasoning involved, but if it is correct, it feels far too tight to be called an inductive argument. If it's neither deductive nor inductive, then it's not an argument at all.

Of course, there are sub-arguments within it. In that respect, it is similar to a real experiment, which often includes calculation and theory application. But, as I stressed above, there is at least one point in an experiment at which an observation is made, perhaps as trivial as reading a thermometer or checking the color of some litmus paper. I say, analogously, that there is at least one point in the CH thought experiment at which an intuition occurs. Where? I don't know, but I will conjecture. I suspect that it has to do with the perception (with the mind's eye) of randomness and independence of the darts. I'm not at all confident of this, but ignorance does not upset the case for intuitions. We may be in the same situation as the famous chicken-sexers who can indeed distinguish male from female day-old chicks, and they do this through sense perception, but they have no idea what it is that they actually see. We could be having an appropriate intuition without knowing what it is or when it happened.

1.7 The Moral of the Story

If the analysis of the dart-throwing thought experiment is right, then the moral is fairly straightforward. Mathematical thought experiments are sometimes not arguments and they sometimes involve mathematical intuitions. It is now a small step by analogy to a similar moral for thought experiments in the natural sciences. If we have established that experience-transcending intuitions can happen in the case of a mathematical thought experiment, then they should be able to happen when the laws of nature are the issue.

Recall the Galileo case. There, I argued that we have good reason to think that intuitions, not sensory experience, are at work. Then I tried to explain how all of this might work, tying it to mathematical Platonism and a realist account of the laws of nature. Finally, we saw the dart-throwing example, which tends to reinforce mathematical Platonism, but in a way that is highly analogous to scientific thought experiments. These different strands seem to dovetail very nicely in support of my general contention that thought experiments are sometimes not arguments from empirically justified premises and they sometimes involve nonsensory intuitions of laws of nature. In other words, some thought experiments transcend experience.

Acknowledgments

I am very grateful to John Norton for numerous discussions over many years. We adopted our very different views at the same time and quite independently of one another, and when almost no one else was the least interested in the topic. That was in 1986. Since then, he has made me re-think and modify my views over and over again. I count myself very lucky to have him as a friendly rival. I am also very grateful to my daughter, Elizabeth, for preparing the diagrams.

Bibliography

Armstrong, D. 1983: *What is a Law of Nature?* Cambridge: Cambridge University Press.
Arthur, R. 1999: On thought experiments as *a priori* science. *International Studies in the Philosophy of Science*, 13, 215–29.
Benacerraf, P. and Putnam, H. (eds.) 1984: *Philosophy of Mathematics*, 2nd edn. Cambridge, UK: Cambridge University Press.
Brown, J. R. 1986: Thought experiments since the scientific revolution. *International Studies in the Philosophy of Science*, 1, 1–15.
—— 1991: *Laboratory of the Mind: Thought Experiments in the Natural Sciences*. London: Routledge.
—— 1993a: Seeing the laws of nature [author's response to Norton, 1993]. *Metascience* (new series), 3, 38–40.
—— 1993b: Why empiricism won't work. In Hull et al., op. cit., pp. 271–9.
—— 1999: *Philosophy of Mathematics: An Introduction to the World of Proofs and Pictures*. London: Routledge.
—— 2002: Peeking into Plato's heaven. Manuscript prepared for Philosophy of Science Association Biennial Meeting, Milwaukee, Wisconsin.
Crease, R. 2002: The most beautiful experiment. *Physics World*, September 2002.
Dretske, F. 1977: Laws of nature. *Philosophy of Science*, 44, 248–68.
Einstein, A. 1949: Autobiographical notes. In P. A. Schilpp (ed.), *Albert Einstein: Philosopher–Scientist*. La Salle, IL: Open Court, 1–95.
Freiling, C. 1986: Axioms of symmetry: throwing darts at the real number line. *Journal of Symbolic Logic*, 51(1), 190–200.
Galileo Galilei 1967: (*Dialogo*), *Dialogue Concerning the Two Chief World Systems*, trans. S. Drake. Berkeley, CA: University of California Press (original work published 1632).
—— 1974: (*Discoursi*), *Two New Sciences*, trans. S. Drake. Madison, WI: University of Wisconsin Press (original work published 1638).
Gendler, T. S. 1998: Galileo and the indispensability of scientific thought experiment. *The British Journal for the Philosophy of Science*, 49, 397–424.
—— 2002: Thought experiments rethought – and reperceived. Manuscript prepared for Philosophy of Science Association Biennial Meeting, Milwaukee, Wisconsin.
Genz, H. 1999: *Gedanken-experimente*, Weinheim: Wiley–VCH.

Gödel, K. 1944: Russell's mathematical logic. Reprinted in Benacerraf and Putnam, op.cit.
—— 1947/1964: What is Cantor's continuum problem? Reprinted in Benacerraf and Putnam, op. cit.
Gooding, D. 1993: What is experimental about thought experiments? In Hull et al., op. cit., pp. 280–90.
Hacking, I. 1993: Do thought experiments have a life of their own? In Hull et al., op. cit., pp. 302–8.
Häggqvist, S. 1996: *Thought Experiments in Philosophy*. Stockholm: Almqvist & Wiksell International.
Hintikka, J. 1962: *Cogito, ergo sum*: inference or performance? *Philosophical Review*, 71, 3–32.
Horowitz, T. and Massey, G. (eds.) 1991: *Thought Experiments in Science and Philosophy*. Savage, MD: Rowman and Littlefield.
Hull, D., Forbes, M., and Okruhlik, K. (eds.), *PSA 1992*, vol. 2. East Lansing, MI: Philosophy of Science Association.
Hume, D. 1975: *(Enquiry), Enquiry Concerning Human Understanding*. Oxford: Oxford University Press.
Humphreys, P. 1994: Seven theses on thought experiments. In J. Earman *et al.* (eds.), *Philosophical Problems of the Internal and External World*. Pittsburgh: University of Pittsburgh Press.
Koyré, A. 1968: Galileo's treatise *De Motu Gravium*: the use and abuse of imaginary experiment. In *Metaphysics and Measurement*. London: Chapman and Hall, 44–88.
Kuhn, T. S. 1964: A function for thought experiments. In *L'Aventure de la Science, Mélanges Alexandre Koyré*, vol. 2. Paris: Hermann, 307–34. Reprinted in *The Essential Tension: Selected Studies in Scientific Tradition and Change*. Chicago: University of Chicago Press, 1977, 240–65.
Mach, E. 1906: Über Gedankenexperimente. In *Erkenntnis und Irrtum. Skizzen zur Psychologie der Forschung*, 2nd edn. Leipzig: Verlag von Johann Ambrosius, 183–200. Translated from the 5th edn. as: On thought experiments. In *Knowledge and Error: Sketches on the Psychology of Enquiry*, trans. T. J. McCormack. Dordrecht: D. Reidel, 1976, 134–47.
—— 1960: *The Science of Mechanics*, trans. J. McCormack, 6th edn. La Salle, IL: Open Court.
Maddy, P. 1997: *Naturalism in Mathematics*. Oxford and New York: Oxford University Press.
McAllister, J. 1996: The evidential significance of thought experiments in science. *Studies in History and Philosophy of Science*, 27, 233–50.
—— 2002: Thought experiments and the belief in phenomena. Manuscript prepared for Philosophy of Science Association Biennial Meeting, Milwaukee, Wisconsin.
Miscevic, N. 1992: Mental models and thought experiments. *International Studies in the Philosophy of Science*, 6, 215–26.
Mumford, D. 2000: Dawning of the age of stochasticity. In V. Arnold et al. (eds.), *Mathematics: Frontiers and Perspectives*. Washington, DC: American Mathematical Society, 197–218.
Newton, I. 1999: (*Principia*), *Mathematical Principles of Natural Philosophy*, trans. I. B. Cohen. Berkeley, CA: University of California Press (original work published 1687).
Nersessian, N. 1993: In the theoretician's laboratory: thought experimenting as mental modeling. In Hull et al., op. cit., pp. 291–301.
Norton, J. D. 1991: Thought experiments in Einstein's work. In Horowitz and Massey, op. cit., pp. 129–48.
—— 1993: Seeing the laws of nature [review of Brown, 1991]. *Metascience*, 3 (new series), 33–8.
—— 1996: Are thought experiments just what you always thought? *Canadian Journal of Philosophy*, 26, 333–66.
—— 2002: On thought experiments: Is there more to the argument? Manuscript prepared for 2002 Philosophy of Science Association Biennial Meeting, Milwaukee, Wisconsin.

Rogers, P. 2002: Editorial: the double slit experiment. *Physics World*, September 2002.
Schabas, M. 2002: Hume on thought experiments. Unpublished manuscript, presented at the Canadian Society for History and Philosophy of Science, Toronto, June.
Sorensen, R. 1992: *Thought Experiments*. Oxford: Oxford University Press.
Tooley, M. 1977: The nature of laws. *Canadian Journal of Philosophy*, 7, 667–98.

CONTEMPORARY DEBATES IN PHILOSOPHY OF SCIENCE

Edited by

Christopher Hitchcock

Blackwell
Publishing

350 Main Street, Malden, MA 02148-5020, USA
108 Cowley Road, Oxford OX4 1JF, UK
550 Swanston Street, Carlton, Victoria 3053, Australia

First published 2004 by Blackwell Publishing Ltd

Library of Congress Cataloging-in-Publication Data

Contemporary debates in philosophy of science / edited by Christopher Hitchcock.
p. cm. – (Contemporary debates in philosophy ; 2)
Includes bibliographical references and index.
ISBN 1-4051-0151-2 (alk. paper) – ISBN 1-4051-0152-0 (pbk. : alk. paper)
1. Science–Philosophy. I. Hitchcock, Christopher, 1964– II. Series.

Q175.C6917 2004
501–dc22

2003016800

A catalogue record for this title is available from the British Library.

Set in 10/12½ pt Rotis serif
by SNP Best-set Typesetter Ltd., Hong Kong
Printed and bound in the United Kingdom
by MPG Books Ltd, Bodmin, Cornwall

For further information on
Blackwell Publishing, visit our website:
http://www.blackwellpublishing.com