LSE CPNSS

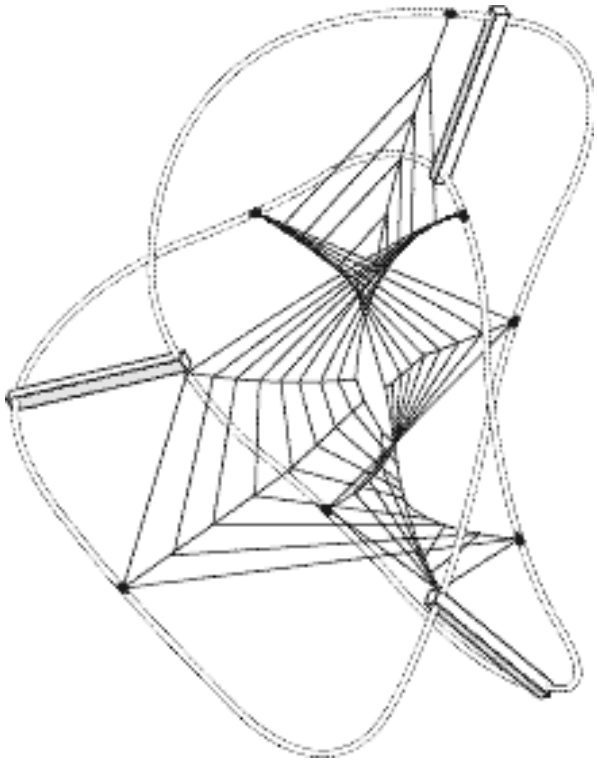
Centre for Philosophy of Natural and Social Science

Causality: Metaphysics and Methods

Technical Report 03/03

Causal Inference in the Abstract or Seven Myths About Thought Experiments

Julian Reiss



Editor: Julian Reiss

Causal Inference in the Abstract or Seven Myths About Thought Experiments*

Julian Reiss
Centre for Philosophy of Natural and Social Science
London School of Economics
Houghton St
London WC2A 2AE

October, 2002

*Thanks to Nancy Cartwright and Roman Frigg for helpful comments and suggestions.

Abstract

I analyse and criticise the following seven commonly held, but to my mind, mistaken beliefs about thought experiments: (1) The history of science is full of significant thought experiments; (2) A good thought experiment provides evidence in its own right; (3) We learn from thought experiments in essentially the same way as we learn from concrete experiments; (4) It is puzzling that thought experiments allow us to learn about the world without providing new empirical data; (5) Thought experiments make acceptance of their result(s) compelling; (6) Mental experiencing is essential to thought experimentation; (7) Thought experimentation involves intervention. After clearing the ground in this way, I sketch a *positive* theory of the thought experiment. The basic idea of the new theory is to integrate thought experiments into a broader inductive scientific methodology. Within such a broader methodology, thought experiments can assume a number of functions, four of which I briefly discuss: (a) concept formation, (b) establishing a causal hypothesis, (c) nomological refutation and (d) suggestion of "new works".

1 Introduction

The 1990s experienced a sudden outburst in philosophical interest in thought experiments that still has not waned. A number of features thought experiments are said to have are often used to justify such philosophical curiosity: thought experiments have played a significant role in the history of science; they give us new knowledge about nature; they are curious devices because they give us new knowledge without new data; they "force us" to accept their results. I am sceptical about each of these claims. If thought experiments have really played such a great role in the history of science, why is it the case that in the philosophical literature only a handful of examples is analysed *ad nauseam*? While I do believe that thought experiments can play a role in the process of producing and refining our knowledge about nature, that role is relatively modest and always contingent upon real experiments and observations. But given there are such experiments and observations, it comes to no surprise that thinking and mental organisation should help to exploit experimental and observational data. In so doing, no result is "forced upon us". All reasoning about experiments and observations is tentative, hypothetical, fallible or whatever one may want to call it, and thought experiments are no different.

Because of the overwhelming tendency in the literature to marvel at the "wonder of thought experimentation", I think it is time for a critical re-appraisal. I have organised my re-appraisal into two main parts, a pessimistic and an optimistic part. In the pessimistic part, I review seven "myths", commonly held but, in my view, mistaken beliefs about thought experiments and try to demonstrate what is wrong with each of them. In the (much shorter) optimistic part, I sketch a number functions thought experiments can fulfil within the scope of a wider scientific methodology. I hope in this way to give an account of both limits and potentials of thought experimentation.

2 Seven Myths Uncovered

Of all claims that have been made about thought experiments, I think the following seven play a special role in the philosophical analysis: (1) The history of science is full of significant thought experiments; (2) A good thought experiment provides evidence in its own right; (3) We learn from thought

experiments in essentially the same way as we learn from concrete experiments; (4) It is puzzling that thought experiments allow us to learn about the world without providing new empirical data; (5) Thought experiments make acceptance of their result(s) compelling; (6) Mental experiencing is essential to thought experimentation; (7) Thought experimentation involves intervention. Far be it from me to claim that all commentators share these beliefs or even a majority of them. But they are widely held and not often criticised. In this section, I will investigate each belief in turn and attempt to show why it is false.

Myth 1 The history of science is full of significant thought experiments.

> It frequently occurs in science that a claim is established or discredited by a thought experiment (McAllister 1996, p. 233)
>
> Thought experiments have more than once played a critical important role in the development of physical science (Kuhn 1981 [1964], p. 6)
>
> We need only list a few well-known thought experiments to be reminded of their enormous influence and importance in the sciences... (Brown 1995, p. 135)
>
> Despite their centrality and importance to both science and philosophy, relatively little has been written about thought experiments (Horowitz and Massey 1991, p. 1)
>
> Historically, thought experiments have played a central role in the articulation and evaluation of scientific theories (Bokulich 2001, p. 285)

Let us start lightly. The first myth about scientific thought experiments I want to discuss concerns their alleged significance for the growth of science. What I mean by significance here is evidential significance, that is, the importance thought experimentation has as a practice to provide evidence for a theory or a causal or nomological claim.

A great number of commentators subscribe to this doctrine in one or another form. James Brown writes about thought experiments as if it was the dominant practice in science. Roy Sorensen and James McAllister treat thought experiments as on par with concrete experiments.[1] Others, such as Thomas Kuhn ascribe a significant role in scientific change to them.

I am sceptical about this enthusiasm for thought experimentation. One reason for scepticism is that it does not seem to me that the history of science underwrites the belief that thought experiments frequently provide evidence for theoretical advance. Thought experiments come nowhere near concrete experiments in playing this role. But even in absolute numbers, thought experiments seem to be a rare species in scientific history.

The claim just made is difficult to substantiate in both its absolute (the significance of thought experiments in science) and its relative (their significance vis-à-vis concrete experiments) form without an extended survey of the history of science. Not only am I not competent to conduct such a survey; it is also well beyond the scope of this paper. But one might attempt to take a short-cut through the woods of philosophical commentary on the topic. Coupled with the belief that philosophers of science do their work thoroughly, we may be able to argue that (a) if there is only a small number of thought experiments analysed in the philosophical literature, then the number of noteworthy thought experiments in science is small and (b) if the number of analysed thought experiments compared to the number of analysed concrete experiments is small, then thought experiments are relatively unimportant.

Surveying the literature on thought experiments quickly reveals that the philosophical contributions centre around only a handful of standard case studies. The above quote from Brown 1995 continues: "Newton's bucket, Maxwell's demon, Einstein's elevator, Heisenberg's gamma-ray microscope,

[1] With James McAllister, I call the instantiated partner of thought experiments *concrete* experiments. Though not perfect, I think "concrete" is least misleading predicate among the alternatives such as *real*, *instantiated*, *actual*, *physical* or experiment *simpliciter*. I want to allow that a thought experiment is a species of experiment, and that it can be just as real or actual as any other. A thought experiment, when conducted, is certainly instantiated but not all concrete experiments are physical. By contrast, with Roy Sorensen I believe that thought experiments are fictionally incomplete, which makes them abstract (*i.e.* the objects that figure in a thought experiment lack many characteristics such as a definite size, colour, the sex of Maxwell's demon and so on). Concrete experiment, on the other hand, are not incomplete in this sense.

Schrödinger's cat". We may also mention further Einsteinian examples, Galileo's free fall and the inclined plane, Stevin's prism, Leibniz's account of *vis viva*, and Huygen's boat.[2] The reader can add his or her own favourite to the list. I do not to claim that the list is exhaustive. But it covers the overwhelming majority of philosophical discussion.[3]

Compare this with the wealth and variety of case studies analysed in the literature on concrete experiments. Already the list of philosophical stock examples that form part of the curriculum of every undergraduate in the philosophy of science is longer: Michelson-Morley, Millikan's oil drops, Kettlewell's moths, Fresnel's white spot, Newton's prism, Fisher's randomised experiments, Röntgen's X-rays, Hertz's cathode rays and Kelvin's rejoinder, Descartes's rainbow, Stern-Gerlach's slits. Opening any classic experimentalist philosophy (such as Hacking's 1983, Latour and Woolgar 1986, Gooding, Pinch and Schaffer 1989, Franklin 1986 and 1990 and Galison's 1987, 1997 and forthcoming) will easily multiply the number of entries. An experimentalist philosopher is spoilt for choice. His challenge consists in picking one or a small set of significant case studies from a million with which he can make his point effectively. The thought experimenter has no choice. Her challenge consists in shedding a new light on an old case.

It is not an artefact of my choice of experimentation as a contrast to thought experimentation. What is true of concrete experiments is just as true of many other scientific practices: observation, measurement, simulation, modelling, calculating (in Hacking's 1983 sense) or theorising. There is abundance of all of these activities in science. Not so of thought experimentation. Thought experiments are rare. They are insignificant, both absolutely and relatively to other practices. This is not to say, yet, that thought experiments are not philosophically interesting in their own right. They may well be perplexing and they are surely able to attract philosophical curiosity. But they lag behind many other scientific activities in number and significance.

[2]I did not mention Plato's Meno and Poincaré's and Reichenbach's thought experiments on Euclidean geometry because they are mathematical rather than scientific thought experiments; that is, they tell us about mathematical rather than natural (or social) objects. I acknowledge, however, that there is no sharp boundary between the two spheres and that one might as well include these mathematical cases.

[3]This claim is made on the basis of c. 80 philosophical analyses of *scientific* thought experiments that I have looked at, including the three monographs Brown 1991b, Sorensen 1992 and Gendler 2000.

One of the few commentators who notices this for economics is Margaret Schabas. She notes in an essay on thought experiments in Hume:[4]

> Are there thought experiments in contemporary economics? No doubt there are, though my impression is that they are rare. [...] Both models and thought experiments are thus present in eighteenth-century political economy, but as theorists drifted more and more to the use of abstraction and idealization, it is fair to say that models became the more common form of reasoning.

Myth 2 A good thought experiment provides evidence in its own right.

As far as I can tell, this thesis is not *explicitly* argued for but it is implicit in most accounts of thought experiments. The claim is that whether a thought experiment provides evidence in favour or against a scientific claim depends only on features the thought experiment has itself rather than any extraneous characteristic. James Brown, for example, notes: "After the thought experiment has been performed and the new theory adopted, the degree of rational belief in Galileo's theory is r', where $0 < r < r' < 1$ [r is the degree of belief in Aristotle's theory prior to the thought experiment]. That is, I make the historical claim that the degree of rational belief in Galileo's theory was higher just after the thought experiment than it was in Aristotle's just before" (Brown 1994, p. 115). No mention of any evidence apart from the Galileo's thought experiment or of background knowledge is made. The view is implicit also in John Norton's and Roy Sorensen's accounts.[5] Both regard thought experiments as species of arguments. But arguments are sound or unsound, valid or invalid, reliable or unreliable in their own right.

It is a commonplace today that an experiment requires background knowledge in order to play the role of evidence for a scientific claim.[6] In this respect thought experiments and concrete experiments *are* alike. Just like a concrete experiment, a thought experiment requires a substantial number of

[4]Schabas 2002

[5]For the former, see Norton 1991 and 1996, for the latter, Sorensen 1992a and b and 1995a and b.

[6]See for example Longino 1990, ch. 3.

background assumptions in order to play any role on the path from evidence to hypothesis.

In the case of concrete experimentation, there are two analytical steps to get from experimental results to theory. First, the experimenter has to make sure that his or her results are genuine phenomena rather than artefacts of the experimental or measurement apparatus. In clinical trials, for example, randomisation is often held to guarantee that estimates of the causal effect of a treatment are unbiased.[7] But the fact that the causal effect of a treatment can be measured without bias does not by itself make the estimate evidence for anything. Hence, in a second step, the phenomenon must be related to a theoretical claim.[8] The methods that extract genuine phenomena from experimental results and those that relate phenomena to theory are often not the same. For example, in order to establish whether the background radiation in the universe is really 3K, we need to make sure that our radio telescopes function properly, that all other known sources of radio signals have been controlled for and so on. But in order to have reason for believing that the background radiation is evidence for the big bang, we need auxiliary hypotheses about, *e.g.* the age of the universe, the initial concentration of energy and so on.

The point is that for both steps we need a lot of background knowledge if an experimental result should play the role of evidence for a theory. This background knowledge can be methodological ("randomisation is or is not a reliable technique to estimate causal effects of treatments"), conceptual ("to measure inflation *means* to compare averages of prices") causal ("osmium fixation tends to produce artefacts in cellular microscopy") or theoretical ("hadron showers are more likely to occur near the vicinity of the bubble chamber wall").[9] Thought experiments require background knowledge in a similar way.

Consider Galileo's famous thought experiment about the free fall. Galileo asks us to imagine that two balls are dropped from a height, a heavy and

[7]This claim has, in my view correctly, been criticised by Bayesians and others. For a recent discussion, see Worrall 2002.

[8]We may, of course, stop here and not worry about theory. But my point is that *given* an experimental result is supposed to be evidence for a theory, there are two analytical steps.

[9]The examples are from Worrall 2002, Jevons 1884 (essay II, ch. I), Hudson 1999 and Bogen and Woodward 1988, respectively.

a light one, which are connected by a string.[10] Galileo tells us that the Aristotelian theory leads to a contradiction in this scenario. According to the Aristotelian theory, falling objects move with their *natural motion*, which is proportional to their weight. The standard story than goes that in the envisaged scenario, we are led to believe that, on the one hand, the ensemble is faster than the heavy ball alone since the ensemble is heavier. On the other hand, however, it falls less fast because the slower falling light ball drags up the heavier one and thus slows down the ensemble. Since both cannot be true at the same time, but both can be derived from the Aristotelian theory, the theory must be false. In addition, Galileo's own theory, *viz.* that barring confounders such as a heavy wind blow bodies fall independently of their weight is shown to be true.

My point is that if Galileo's story should be anywhere near convincing, he better makes sure that all other causes of the rate of fall except gravity have been controlled for and that observers are competent. No one knows what will happen to his two balls if the experiment is conducted near a big magnetic field or in a hurricane or if the observers are lunatics. He also must make sure that the phenomenon, as he envisages it, falls under the Aristotelian theory. It would be easy for the Aristotelian to deny the relevance of the thought experiment because two objects joined in the way Galileo requires is "unnatural" or because the law is silent about what happens in a vacuum because a vacuum is a physical impossibility.

I think, therefore, that a *negative* lecture can be transferred from concrete to thought experiments: they both can play the role of evidence for theoretical claims only in the light of substantial background knowledge. In the next subsection I want to argue that thought experiments *differ* from concrete experiments in that they are epistemically derivative in a way that concrete experiments are not.

Myth 3 We learn from thought experiments in essentially the same way as we learn from concrete experiments.

> ... thought experiment is experiment (albeit a limiting case of it), so that the lessons learned about experimentation carry over to thought experiment, and vice versa (Sorensen 1992, p. 3)

[10]This is the version of Galileo's thought experiment that is discussed in the relevant literature.

> [T]hought experiments are experiments, albeit an extreme form of them. On [the experimentalist view], a thought experiment, like a concrete experiment, provides evidence about the world; and a thought experiment establishes or discredits a scientific claim in the way a concrete experiment does, in the light of the evidence about the world that it provides (McAllister 1996, p. 233)

There has been a movement within the literature on thought experiments, which draws analogies with concrete experiments. According to this movement a thought experiment just *is* a species of experiment. Lessons learned from the analysis of concrete experiments can be exploited (with caveats, to be sure) to understand thought experiments and vice versa (with equal caveats). I do not wish to deny that many a lesson can be learned mutually, nor that the two types of activity share some, maybe even crucial, aspects, nor that they are fundamentally the same kind of thing. In fact, below I will argue just the latter point: that thought experiments belong to the same fundamental category as concrete experiments.

I also do not want to argue against the view that there is a kind of continuity between thought and concrete experiments. Surely, some important thought experiments could be implemented as concrete experiments but they just have not been implemented for whatever reason: it is believed that an implementation does not improve evidential weight of the experiment, it is too costly, it is unethical or what have you. There may even be some real borderline cases where it is impossible for us to decide whether the activity at hand is a concrete (psychological) experiment or a thought experiment. But a continuum of cases does not render a distinction invalid. Because there may be a continuum of cases between the two sexes, the distinction between male and female humans is not ineffective. Nor does it entail that there are no important differences.

But there is an important difference between thought and concrete experiments in their respective relation to theoretical claims. There are two aspects of the difference. First, thought experiments are tied much tighter to theories than concrete experiments. While it is the case that many concrete experiments are designed to test theories, others are conducted more or less independently of and prior to any theorising. Thought experiments, by contrast, never precede theories and cannot be conducted independently of them. And second, the validity of a thought experiment always depends

on the results of similar concrete experiments. But the validity of a concrete experiment is not derivative in this sense.

Let us discuss experimental autonomy first. Ian Hacking famously claimed that experimentation has a "life of its own".[11] He argued against a background of theory-dominated philosophy of science at the time. Experiments were not a significant issue in philosophical discussions. More importantly, experiments were treated only as means to test or establish scientific *theories*. Against this, Hacking pointed out that experimentation is a largely autonomous activity within science. Many experiments are conducted without aiming at testing or establishing theories but to find out about nature in a particular way. Scientists achieve great experimental results despite holding completely false theoretical beliefs. And a lot of experimental knowledge can be produced before any significant theorising has taken place. Often, experimentally created phenomena await theoretical treatment for many years.

If Hacking is not altogether off track, there is an important difference between concrete and thought experiments in this matter. No thought experiment I know of, and certainly no one that is discussed in this paper would be possible without a theory playing an essential role it.[12] Just think of the two most frequently discussed thought experimenters, Galileo and Einstein. All Galilean thought experiments function by assuming the truth of an aspect of Aristotelian physics and deriving a contradiction from it. In Einstein's, relativity or quantum mechanics or both played an indispensable role—either as initial ingredient or as result. Most thought experiments in economics depend on a theory of rationality. If they do not, as Hume's thought experiment which will be discussed below, they are still designed to show the inadequacy of a theory, in Hume's case of a mercantilist theory of the interest rate. Schrodinger's cat tests quantum mechanics while Leibniz's *vis viva* thought experiment tests Cartesian mechanics. Unlike concrete experimentation, thought experimentation cannot be conducted in a theoretical vacuum.

[11] Hacking 1983, ch. 11

[12] Bokulich 2001 argues that "thought experiments are no more bound to any one particular theory than ordinary physical experiments..." (p. 293). But while I think she convincingly demonstrates that thought experiments do not have to be tied to a *particular* theory, her paper is a striking confirmation of the fact that thought experiments are tied to *some* theory. Thanks to Nancy Cartwright for providing me with the reference to Bokulich's work.

Thought experimentats do not have a life of their own. Not surprisingly, Hacking says just this in a comment on accounts of thought experiments by Nancy Nersessian, David Gooding and James Brown. He says,[13]

> Once the thought experiment is written out in perfection it is an icon. Icons, to reiterate, do not have a life of their own.

The second aspect of the difference between concrete and thought experiments is, I believe, one reason for the fact that thought experiments do not have a life of their own. This reason is that their epistemic worth is derivative from other, concrete experiments. This is because of the following dilemma. We either have experience with a situation that is relevantly similar to the thought experimental situation or we don't. If we do, then our thought experiment just tells us what we know from concrete experiments (or observations). But if don't, then the thought experimental result amounts to mere guesswork.

Let me expand on this. The main point is that we cannot learn from thought experiments in the same way we can learn from concrete experiments. In a concrete experiment we need a lot of background knowledge in order to achieve a sound experimental design, but once we have that, nature herself answers our questions. This is not what happens in a thought experiment. Here we need the same kind of background knowledge for a good design but a good design does not assure sound inference. Not nature answers our questions but we answer them on the basis of intuitions, be they derived from past observations or somehow "hardwired" in our brains. There is no reason why our intuitions should give the same answer as nature.

For example, one does not have to be particularly imaginative to come up with responses on behalf of the Aristotelians in the light of the Galileian "refutation". Maybe the string rips because of the speed differential. If it is assumed that it cannot rip, there is good reason to treat the ensemble as one heavier object, so it travels faster. Or we say that the ensemble is an unnatural object, so it falls outside the scope of the Aristotelian theory.

In order to demonstrate that people's intuitions about what happens in the Galilean scenario differ and that we are in no way forced into finding

[13]Hacking 1993, p. 307. Hacking comments on Nersessian 1993, Gooding 1993 and Brown 1993.

the contradiction that allegedly follows from the Aristotelian theory, I conducted a little "experiment"[14] in which philosophy students were presented with Galileo's thought experiment and then asked to answer a number of questions. The students were from an undergraduate course in mathematical logic.[15] Most of them were enrolled as philosophy students but there were also students from other departments taking logic as an outside option. The physics background varied from "none" to a completed undergraduate degree. 65 students returned the questionnaire.

To introduce the experiment, I read the following statement to the students:

> According to an Aristotelian theory, a falling body falls at the rate of its so-called *natural motion*, which is proportional to its weight. This implies that heavy bodies fall faster than light ones. Galileo constructed a thought experiment, which, he claimed, bears on the Aristotelian theory. He asked his readers to imagine two balls of different weights to be dropped simultaneously from a tower. The Aristotelian predicts that the heavier one falls faster and hits the ground first. "But now", Galileo said, "suppose that the two balls were joined to each other by a string or rope and then dropped". What is the outcome in this scenario?

Among others, the students were then asked to answer the following questions:

1. Have you heard of this thought experiment before? (Yes | No)
2. What is your physics background? (none | GCSE | A level | BSc | MSc | PhD | other—please specify)

[14] I put "experiment" in quotes to express that there is no claim made here that I have conducted a well-designed experiment. All I want to demonstrate is that there is reason to believe that people's intuitions differ about what happens in a thought experimental scenario. Unless LSE students can be shown to be particularly inapt to conduct thought experiments, the methodology employed here seems good enough to show that this is indeed the case.

[15] My thanks go to Talal Debs who provided me with the opportunity to conduct the experiment in his logic class.

3. What is your philosophy background (none | 1st year | 2nd year | 3rd year | Master's | PhD | other—please specify)

4. What do you think will happen in the scenario as described?

5. Does that have a bearing on the Aristotelian theory of the free fall? (Yes | No) If Yes, in what way?

To me, the results came with no surprise. Only a single student wrote in his or her answer that the Aristotelian theory is falsified because a contradiction can be derived from it. Less than twenty per cent knew the thought experiment already—which, interestingly, did not have a great influence on the result. One student who knew the story answered even that the lighter ball will hit the ground first. Just over half thought that Galileo's story had a bearing on the theory but a much smaller number noted that it actually disconfirms it (some did not answer the question and almost two fifths said the thought experiment has no bearing on the theory). A little over ten per cent said that the rate of fall is *independent* of the object's weight. An additional good twelve per cent claimed the rates of fall are identical, but did not state what rate that is. A relative majority of about a quarter maintained that the amalgam falls at the rate of the heavier body because it pulls the light one down such that the latter is sped up (one—if one can tell from the handwriting, female—student said, "Heavier ball will be on the bottom; lighter one on top. Heavier ball will speed things up; lighter ball will try to slow things down. Lighter ball will be unsuccessful. Isn't that the way it *always* goes!" [emphasis original]). Some fourteen percent each said that the joined object falls at a slower and a faster rate than the heavy object, respectively. Another fifth gave "other" replies such as that the balls fall at different rates, the heavier one hits the ground first, the lighter one hits the ground first, it depends on surface area or material, there will be a "togging and spinning", they fall at the sum of the individual rates, or they gave inconsistent answers without pointing out that there is an inconsistency.

Again to no surprise, a background in physics matters. I divided the group into a "low physics" (GCSE/equivalent or less) and a "high physics" half. Among the low physics students, less than ten per cent said the rate of fall was independent of weight, while this number was twenty percent exactly for high physics students. More than two fifths of the low physics group thought the amalgam falls at the rate of the faster ball as it drags

the other one down while less than seven per cent of the high physics group thought this.

The point of this "experiment" was to demonstrate that it is far from true that our intuitions are a reliable guide to physical truth, even if the scenario regards an everyday situation. Any answer is possible, and one will always find someone who actually maintains it.[16]

I expect a number of objections to the design of my "experiment". One is that we need competent experimenters rather than untrained people. Just as we cannot expect from a mathematically untrained person to find the correct result of a complex derivation or to find any result "compelling", we cannot expect the analogous qualities from untrained thought experimenters. To be a trained thought experimenter may either imply that one is trained in the subject area or in thought experimentation or both. To me it seems that the students are certainly competent enough as regards their knowledge of the subject area. Everyone has a great deal of experience with falling objects. A strong background in physics rather *impedes* the ability to conduct this thought experiment because one thinks one knows the "right" answer already and as a consequence does not actually perform the mental experiment. As regards skill as a thought experimenter, I think that little is required to complete the Galilean experiment successfully. The students were told the relevant aspects of the Aristotelian theory. I do not think that practice with thought experiments helps much to notice that that theory is ambiguous about what happens in the Galilean scenario. But once that is clear, the contradiction arises, and logic students should be able to figure out that there is something wrong with a theory that implies a contradiction.

A second potential objection is that the questions were too general. As a mathematical result may appear compelling only once one sees its demonstration, a thought experiment will appear compelling only when one is guided through it explicitly. This can happen in two ways. Either we can use a Socratic midwife strategy and tickle the result out of students by asking the right questions; or we can present the result and ask, "*Given* you know result, do you think this is what must happen?". I deliberately formulated general questions because the focus of this paper is on epistemology rather than, say, on pedagogical or rhetorical uses of thought experiments. If we are

[16] Although no student answered that the string would rip—which is an actual physical possibility if the conditions are right.

to be able to learn from thought experiments, we must get the results from contemplating the scenario itself rather than from hearing them from others. I thus do not think that presenting the result would improve the informativeness of my "experiment". One could, however, ask more directed questions, such as, "Does the scenario falsify the Aristotelian theory?" or "What does the Aristotelian theory imply if we treat the amalgam as one object or as two loosely connected objects, respectively?". But it seems to me that such questions bias the answers too much in the Galilean direction. I wanted it to be a real possibility that the scenario has no evidential bearing on the Aristotelian theory at all, and that many different outcomes can obtain. The structure of the case plus the students' intuitions and background knowledge are what is supposed to do the job of finding the right thought experimental outcome. So I do think that the presentation did not skew the "experiment" in a way that favours scepticism about thought experiments. The "right" solution was a real possibility but so were other solutions.

To adjudicate between the different intuitions of our thought experimenters, we can conduct concrete experiments. And this is just what Galileo appears to have done.[17] To be sure, in practice it is not so easy to control for all possible confounders. At Galileo's time, air resistance was particularly difficult to control for. Air pumps were not too effective and travel outside the Earth's atmosphere was unthought of. Galileo thus receded to the next best: a series of experiments of dropping balls in liquids of decreasing density. Extrapolating the results from these experiments vindicated (or motivated) his solution to the thought experiment: that the rate of fall is independent of the body's weight once extraneous factors have been controlled for. But this just means that the the thought experimental result piggy-backs on the results of a range of concrete experiments.[18]

My analysis is of course heavily critical of major philosophical accounts of thought experiments, in particular James Brown's and Roy Sorensen's. James Brown's defends a rationalist account of scientific epistemology in general and of thought experiments in particular. For him, scientific knowl-

[17]See, for example, Losee 1993, ch. 7 and Laymon 2000.

[18]The physicist David Atkinson seems do endorse a very similar view. For example, he writes that the conclusion of Galileo's thought experiment "is in the case of free fall approximately correct, however, is not an accident: it is the result of real experiments performed with steel balls and inclined planes" (Atkinson 2002, p. 11). Thanks to Jan-Willem Romeyn for providing me with this reference.

edge is codified as knowledge about laws of nature. Laws of nature, in turn, are relations between universals of a certain kind (a relation of *nomic necessitation*). In fortuitous circumstances, thought experiments allow us to "see" these relations directly. Thought experiments are windows to Plato's heaven.[19]

Brown's argument in favour of the rationalist epistemology is essentially a *tu quoque* answer. To the challenge that his account involves a mysterious route to knowledge, he replies that ordinary perception (an understanding of which he seems to accept as a necessary condition for a coherent empiricism) is understood just as poorly. But Platonism demystifies the thought experiment because it shows us how we can learn about nature from thinking alone and has a number of other advantages. Hence we might as well be Platonists.

I disagree with Brown about the relevance of having a deep understanding of the process of ordinary perception for being an empiricist. From a non-foundationist point of view (and Brown is not a foundationist), all we need to know is how observation and experiment can be *reliable* guides to knowledge. There is a wealth of techniques that are supposed to help us with making sure that the experimental procedure is reliable (such as control for confounding factors, empirical investigation of laboratory equipment, simulations, robustness tests, reliable process reasoning, minimalist overdetermination *etc.*). Such techniques are analysed at length in the philosophical literature. The rationalist has no such techniques at hand; maybe they exist, but Brown and other rationalists have so far failed to reveal them. Therefore, I think, Brown's rationalism is untenable.

Pseudo-empiricist positions such as Roy Sorensen's do not fare better.[20] Sorensen accepts that, fundamentally, we learn about nature only through observing her and experimenting with her. The reason he gives for the fact that thought experiments often give us right results is that we have evolutionarily developed an intuition about physical matters of fact. For example, we do not have to make the experience with the perils of the dark *ourselves* in order to fear it. This is because our ancestors have made such experiences (many will have perished because of that) and we inherited that kind of knowledge from them. So knowledge is based on experience though not our own. But there are two problems with such an account. First, as Sorensen

[19] Brown 1986, 1988, 1991a, 1991b, 1998

[20] Sorensen 1992a and b, 1995a and b

himself acknowledges, his theory covers only thought experiments that concern everyday situations:[21]

> The laws of nature have led us to develop rough and ready intuitions of physical possibility which are then exploited by thought experimenters to reveal some of the very laws responsible for those intuitions. The good news is that natural selection ensures a degree of reliability for the intuitions. The bad news is that the evolutionary account seems to limit the range of reliable thought experiment to highly practical and concrete contexts.

But second, the good news turns out to be bad news too. Sorensen does not give us guidance as to what kinds of thought experiments to accept as reliable and what to reject. We neither know what contexts exactly are "highly practical and concrete" nor what we should do when facing more abstract situations. He is explicitly agnostic about, *e.g.*, the Newton-Mach dispute (see below for a discussion). Worse, as my experiment showed, our intuitions can fail even in highly practical and concrete situations. In my view, an empiricism that accepts observation and experiment as epistemic tools without providing us with criteria to decide when these tools are *reliable* producers of knowledge is as blind an empiricism as that of Bacon's ants[22]—it is a version of pseudo-empiricism. A proper empiricism tells us what kinds of cases of observation and experiment are reliable. If no such guide to reliability exists for thought experiments, we must accept that thought experiments tell us little more than what we know from concrete experiments. We learn from thought experiments in a different way because in thought experiments, we experiment on *us*, not on nature.

Myth 4 It is *puzzling* that thought experiments allow us to learn about the world without providing new empirical data.

This claim has nothing to do with thought experiments themselves but with the reception that they get in the literature. I think it underlies almost all of the recent work in the analysis of thought experiments but particularly clear statements are the following:

[21] Sorensen 1992b, p. 15

[22] See Bacon's *Novum Organum*, book I, aphorism 95

> Granting that every successful thought experiment embodies in its design some prior information about the world, that information is not itself at issue in the experiment. On the contrary, if we have to do with a real thought experiment, the empirical data upon which it rests must have been both well known and generally accepted before the experiment was even conceived. How, then, relying exclusively upon familiar data, can a thought experiment lead to new knowledge or to new understanding of nature? (Kuhn 1981 [1964], p. 7)

> Scientific thought experiments are typically factive; they are attempts to elicit physical intuitions about what would happen under certain conditions. Such thought experiments are puzzling because they seem to describe cases where we learn something new about the physical world, even though we have no new empirical information about the world (Gendler 2000, p. 150)

The puzzle to me is rather why so many intelligent people can find it puzzling that we should learn by thinking. Surely, we do not learn about nature by thinking alone.[23] With the possible exception of James Brown, none of the authors makes a claim to this effect. To the contrary, all accept that the raw material of a thought experiment is experience. The puzzle, then, is how we can learn about the world by re-arranging past experience in a way that enables causal or nomological inference (inductively speaking; we may add: ... or in a way that enables refutation of theories, deductively speaking). But why should that be puzzling? Do we not have to organise our experiences first before inferences can be made? Does the repeated observation of a black raven lead us automatically to believe in the proposition "All ravens are black"? It does, if we follow the staunch Humean line which makes us believe that repeated observations of some event-type forces us qua psychological law to make the corresponding inference to a regularity. But according to anything like an epistemology à la Bacon, Kant, Popper, Goodman, Kuhn or others no such automatism is possible. Bacon taught that

[23]However, James Brown possibly thinks that we do: "Thought experiments provide us with scientific understanding and theoretical advances which are sometimes quite significant, yet they do this without new empirical input, and possibly without any empirical input at all" (Brown 1993, p. 271).

a vital ingredient of the scientific endeavour is concept formation and the organisation of experience; Kant is famous for his dictum "Percepts without concepts are blind"; Popper used the "searchlight" metaphor to express that the mind is active; Goodman pointed out that the description of evidence will be relevant for the inductive inferences we make; finally, Kuhn highlighted that we learn about concepts and the laws into which they enter at the same time. If one role of a scientific thought experiment is to organise our stock of past experience, it should cause little amazement that we learn from it. Further, a thought experiment is not alone in being a technique that aids in nomological or causal inference (or theoretical refutation) without adding to the stock of raw experience. Other techniques abound, especially in sciences where the "data", *i.e.* results of observational procedures, come in a form which is unusable for the theoretician. All disciplines that use statistics for inference and refutation provide examples. Running a regression in econometrics is such a technique. Imagine, an econometrician finds a new method with which to reliably estimate the natural rate of unemployment. So far, estimates have been x, say, and now they are y. This gives rise to some new theory. This example, in fact, satisfies James Brown's characteristic of *a priori* science:[24]

> First, there have been no new empirical data. [...] Second, [the] new theory is not logically deduced from old data. Nor is it any kind of logical truth. [...] Third, the transition from [the old] to [the new] theory is not just a case of making the simplest overall adjustment to the old theory.

Econometricians, by and large, do not gather but manipulate data. If we further assume that the econometrician in question used the same data as have been used in the procedure that yielded x, the first point is surely satisfied. Nothing in the data suggests how a regression model should be set up. Thus the second point is fulfilled. The third point is fulfilled by stipulation. But it does not matter for the argument whether it is fulfilled or not. What matters is that we have a technique which extracts new, genuine knowledge from old data. Historiography is another paradigmatic example for a supplementary science whose objective it is to provide tools for generating knowledge from existing data, in this case historical records. By no

[24] Brown 1994, pp. 114f.

means are these techniques germane only to social science. Astronomy, cosmology, particle physics, medicine, epidemiology and many other sciences use similar techniques.

The upshot is that we do not have to be Kantians, idealists or rationalist in order to accept that without a contribution of the mind, observations are not very informative. What exactly that contributions is, is of course a matter of controversy. But if thought experiments are helping us to organise experience, there is no puzzle that they help us to learn about nature.

Myth 5 Thought experiments make acceptance of their result(s) compelling.

> A thought experiment is usually so compelling that even in those cases where it is possible to carry it out, the reader feels no need to do so (Nersessian 1993, p. 296)

> Implications [of a thought experiment] are compelling: however surprising the resolution may be, the intellectual experience engendered by a thought experiment is compelling; whether the experiment could be performed in the world seems as irrelevant as whether the story-line of a good joke is actually true (Gooding 2000)

Contemplating a thought experiment the experimenter is forced to accept its solution. There is no way out, no arbitration. Thought experiments initiate scientific revolutions, they refute theories that have been believed for millennia in a single stroke, they coerce their audience to accept some ground-breaking new theory, they convince of the implausible. So goes the myth. I want to argue in a series of three steps that this belief is mistaken. First, thought experiments are fallible. This is the second *negative* lesson that can be learned from concrete experiments. Like them, thought experiments can fail because they overlook a significant factor, a disturbance confounds the result, the design is poor for this or that reason, the experimenter is incompetent and so on. Second, in some cases, they may fall short of giving an unequivocal result. But if they are fallible and/or give ambiguous results, they cannot be rationally compelling. They may still be subjectively compelling. That is, the experimenter may feel that she is forced to accept the solution without having conclusive evidence. One cannot, of course, refute

the claim that some experimenters feel compelled; a number of proponents of thought experimentalism obviously do feel so. But to get some evidence that not everyone feels compelled I included a question in my "experiment" to see whether the students were sure of their results. As one might expect, only few students were so sure.

But first things first. Einstein is arguably the most famous thought experimenter in recent history of physical science. As ingenious as his other thought experiments may have been, one was refuted by Niels Bohr. The story is told in detail by Michael Bishop. His presentation of the original thought experiment by Einstein reads:[25]

> During the 1930 Solvay Conference on magnetism, Einstein presented Niels Bohr with the clock-in-the-box counterexample to the uncertainty principle. Suppose we have a box full of photons that has, on one of its walls, a shutter that is controlled by a clock. Weigh the box. Now set up the shutter mechanism so that it opens for a brief interval at which time a single photon escapes. Weigh the box again. The change in the weight of the box gives us the weight of the photon, which gives us its mass. And using Einstein's famous equation $E = mc^2$, we can determine the photon's energy. In principle, therefore, we can measure the photon's energy and its time of passage to any arbitrary degree of accuracy.

Unfortunately for Einstein, after a sleepless night, Bohr came up with a counter-argument. When the photon flees, it passes on momentum of unknown magnitude to the box. This will induce the box to move in the gravitational field. That, in turn, will affect the rate of the clock. Hence, as the initial magnitude of induced momentum is uncertain, time can be measured only to a degree of uncertainty too. Heisenberg's principle stands unshattered. Apparently, Bohr could convince Einstein of this reading of the thought experiment.[26]

Any thought experiment can fail on a number of different counts. Allen Janis[27] distinguishes three *kinds* of potential failure: (a) the thought experimenter fails to reach a conclusion altogether, (b) the thought experiment

[25] Bishop 1999, p. 536

[26] See *ibid.*, p. 538 and Sorensen 1992, p. 183.

[27] Janis 1991

produces incorrect results and (c) it produces results that are not answers to the question that initially motivated the thought experiment. The Einstein-Bohr debate clearly fits in with the second category. Janis gives examples of the other types of failure.

That thought experiments can fail should be uncontroversial. A more interesting case is the following. With his famous bucket (thought) experiment, Newton sought to establish the reality of absolute space.[28] Newton's narrative, simplified, is this. Suppose there is a bucket full of water suspended by a twisted cord in otherwise empty space. Initially, the bucket is at rest (stage 1). When the cord unwinds, the bucket gains angular momentum. The bucket will move relative to the water (stage 2). Slowly friction will cause the water to move with the bucket such that the surface of the water acquires a parabolic shape. At some point there is no relative motion between bucket and water (stage 3). There is no difference in *relative* motion between stages 1 and 3. And yet, the water is flat in stage 1 but parabolic in stage 3. Newton argues that the difference is due to absolute space (in stage 1 the bucket is at rest relative to absolute space while it is moving in stage 3).[29]

Ernst Mach criticised this specimen of Newtonian reasoning.[30] Essentially, he argued that the centrifugal effects must be caused by a force that depends on the acceleration and mass of the interacting bodies. If there are no bodies to interact with, there cannot be any centrifugal effects. Observations of such effects are due to, *e.g.*, the fixed stars. Regarding Newton's thought experiment, Mach seems to think that our intuitions founder in cases of such a degree of unrealisticness. He writes, "No one is competent to predicate things about absolute space and absolute motion: they are pure things of thought, pure mental constructs, that cannot be produced in experience".[31]

We could interpret this episode as a type-a failure according to Janis's scheme. Mach simply was not able or willing to bring Newton's thought experiment to an end. But I think it is more illuminating to point out that different people have different intuitions about what happens in hypothetical situations, and wanting an arbiter between them, we best conclude with a

[28]Newton did actually conduct it but presented it also as a thought experiment.

[29]For a detailed description of the experiment, see e.g. Sorensen 1992, pp. 144ff. and Brown 1991, pp. 7ff.

[30]Mach 1942

[31]Mach 1942, p. 284

draw.[32]

Thought experiments can be either accessible or closed to empirical investigation.[33] If they are accessible, our results piggy-back on the results of concrete experiments (see above).[34] If they are closed (such as Newton's bucket), results depend on our intuitions and these can differ—even between competent experimenters, which makes them ambiguous. But ambiguous results cannot be rationally compelling.

They can still be subjectively compelling. I am sure that many thought experimenters do feel that the results of their mental exercises are not only true but that they also must be true. There is no principled argument one can invoke to refute that claim. But I think we can shed doubt on it by finding evidence that people do disagree about the results of a thought experiment to a large degree and that they do not claim to feel "compelled" by its results.

That people differ in their opinions about what happens in a thought experimental scenario, I hope to have demonstrated with my "experiment" described above. In order to see whether they feel subjectively compelled by the thought experiment, I also asked them the following question:

> Do you feel compelled by your "solution" to the thought experiment? That is, are you absolutely sure of it? (Yes, I am absolutely sure | I am quite sure | No, I am not so sure)

Interestingly, only a quarter of students was absolutely sure about their result. And this was largely but not completely independent of their actual result: about half of the "absolute sure" group answered that rate of fall is independent of weight but the remainder divides between rate of the heavy ball, faster than the heavy ball, "equal rate" and "other" replies. Very much to the general point of my paper, one student said he was absolutely sure about his result because he had conducted a (concrete) *experiment*. A good third was quite sure and the remaining forty per cent were no so sure. There is nothing subjectively compelling about thought experiments. Of course, some people will always be sure of whatever they say. But many other people will not be so sure.

[32]This is also confirmation that intuitions of even experts can and do differ.

[33]Pierre Duhem used a similar dilemma to make a point about thought (or as he said, fictitious) experiments. See Duhem 1991 [1914], p. 202.

[34]Whether or not the results of a concrete experiment are compelling is a different matter. I of course do not believe so, but that's another paper.

I think these three arguments show that thought experiments are not compelling unless one seriously waters down the meaning of "compelling". Because they are fallible and sometimes their results ambiguous, they cannot be rationally compelling. They will be subjectively compelling to some thought experimenters. But I hope to have demonstrated at least that not everybody feels compelled in the same way.

Myth 6 Mental experiencing is essential to thought experimentation.

> Sight is perhaps our most important sense and we have undoubtedly let this condition our thought experiments as well. I have made being 'visualizable' or 'picturable' a hallmark of any thought experiment. Perhaps 'sensory' would be a more accurate term. After all, there is no reason why a thought experiment couldn't be about imagined sounds, tastes, or smells. What *is* important is that it be experiencable in some way or other (Brown 1991, pp. 16f.)

The last two "myths" are much less important and consequently my discussion will be rather brief. I discuss this claim because I want a theory of thought experiments that is at least in general to be applicable to other sciences than physics, for example, the social sciences. Thought experiments in the social sciences, however, rarely have the quality of being "visualisable" or "mentally experienceable".

James Brown and others claim that thought experiments are always mentally experienceable, and that that is an important characteristic. Brown, of course, has an agenda. He wants to defend a Platonist picture of science. A fundamental difficulty historical Platonisms have had was the so-called "problem of access". The Platonist claims that real knowledge is knowledge about abstract entities and their relations—entities and relations that reside in a realm outside space and time. If these abstract entities exist outside space and time, we cannot be in causal contact with them. But if, as many people believe, in order to know something, we have to be somehow causally connected with it, we cannot know abstract entities. So how is knowledge possible?

Brown likens seeing abstract entities with ordinary seeing. Just as we perceive concrete objects with our bodily eyes, we can perceive abstract

objects with the "mind's eye". Thought experiments, according to Brown, provide a tool that enables us to see abstract things with the mind's eye.

I suspect that the metaphor about "seeing things with the mind's eye" has led Brown to believe that it is an essential characteristic of thought experiments to be visualisable or, in general, mentally experienceable. But I neither believe that visualisation is a characteristic that all (or even most) thought experiments have nor that it is essential to how they work.[35] That thought experiments are visualisable and that visualisation is an essential characteristic actually boils down to the same thing if one realises that everything is visualisable in a trivial sense. If you want to visualise a concept x, just draw a symbol and dub it 'x'. For many concepts there will be standard symbols, for others one can invent them. Thus the real question is whether visualisability helps in the derivation of the thought experimental results.

A very straightforward counterexample to this claim is Hilary Putnam's Twin-Earth. With it, Putnam wants to demonstrate that a concept's meaning is not exhausted by what is in their users' mind. To make this claim, he asks us to imagine a planet, Twin Earth, which is identical in all respects to our own planet except that what we call water, the chemical structure H_2O, on Twin Earth has the structure XYZ. All phenomenal qualities of H_2O and XYZ are identical. Therefore, all beliefs and desires of the users are the same. But on earth, 'water' designates H_2O while on Twin Earth it designates XYZ. Therefore, the semantics cannot be reduced to psychology.

This is a nice counterexample to the visualisability thesis because the gist of Putnam's argument is that H_2O and XYZ are phenomenally indistinguishable. Therefore, the difference cannot be visualisable except in the abovementioned trivial sense (say, by writing down 'H_2O' and 'XYZ' and drawing arrows pointing to the oceans on each planet, respectively). Hence visualisation cannot be essential to Putnam's story.

Unfortunately, Putnam's thought experiment is a philosophical rather than a scientific specimen. But I believe the same can be said about thought experiments in science. Brown's own beloved EPR thought experiment is a good *scientific* counterexample. This is due to the fact that one crucial step in the argument is that "measurements which are done simultaneously at a large distance from one another do not interfere with one another. (Accord-

[35]For brevity, I concentrate on the sense of sight. This does not effect the results of my discussion though.

ing to special relativity, two events outside each other's light cones cannot be causally connected.)" But if this is so, how are we supposed to see the measurements "with the mind's eye"? Again, we can surely visualise this assumption in a trivial sense (by drawing the two apparatuses and something that symbolises "far" in between). But the thought experiment goes through only if we have the right conceptual knowledge, not if we can produce a mental image of the situation.

No thought experiment I can think of in economics is visualisable in but a trivial sense. One example due to David Hume and discussed by Margaret Schabas is the following:[36]

> For suppose, that, by miracle, every man in GREAT BRITAIN should have five pounds slipt into his pocket in one night; this would much more than double the whole money that is at present in the kingdom; yet there would not next day, nor for some time, be any more lenders, nor any variation in the interest [rate]

Again, we could attempt to trivially visualise this thought experiment. But it will succeed only if we understand that the doubling of the money stock is indeed simultaneous, that people with more money in their hands will take some time to change their behaviour, that institutions are rigid and all that.

The upshot is that a thought experiment's success depends, as argued above, on the right conceptual (and other) background knowledge and in no way on whether the situation that it envisages is visualisable in any but a trivial sense.

Myth 7 Thought experimentation involves intervention.

> Intervention is important but unproblematic: the reader/observer is involved, but is perfectly competent... (Gooding 2000)

> As well as being sensory, thought experiments are like concrete experiments in that something gets often manipulated: the balls are joined together, the links are extended and joined under the inclined plane, the observer runs to catch up to the front of the

[36] Hume 1985 [1777] , 299, quoted from Schabas 2002, p. 1

> light beam. As thought experimenters, we are not so much passive observers as we are active interveners in our own imaginings. We are doubly active; active in the sense of imaginative (but this is obvious), and active in the sense of imagining ourselves to be actively manipulating (rather than passively observing) our imaginary situation. (Brown 1991, p. 17)

This last claim is, I believe, a relict from the analogy between thought and concrete experiments. I discuss it because it is a widely shared conviction with respect to both thought and concrete experiments. It does more harm in the context of concrete experiments but it is worthwhile being discussed in this context, too.

So here is a third negative spill over from thinking about concrete experiments to thinking about thought experiments. I do not believe that intervention is essential to experimentation. Nor do I believe (*a fortiori*, one might say) that it is essential to thought experimentation. My argument is similar to the argument about visualisation above. This is because, *in a trivial sense*, all thought experimentation involves "interventions". If by that is meant that we deliberately have to think up the thought experimental scenario, the claim is trivially true (or "obvious", as Brown says). We surely cannot passively contemplate a thought experiment happening in our head.[37] But if the claim is that the thought experimental scenario involves human agency, it is wrong.

The claim is wrong because it is inessential to the success of a thought experiment whether the "change" or "difference" which is crucial to the production of the result is brought about by human agency or by "natural" causes. It does not matter whether we envisage, in Galileo's free fall experiment, a gunner drops the cannon and the musket ball, whether it is a musketeer or whether two apples fall from a tree as windfall. It may be *unlikely* that two apples of different weights grown together such that they are conjoined should fall from a tree but the thought experimental result does not depend on that. It does not matter whether a Georgian King Midas slips the five pounds into every Briton's pocket (as long it is gold, of course), whether it is due to an influx from Australia and California or whether it happens, as Hume says, "by miracle". Important is the sudden *change*. Lastly, whether

[37] I am not sure about a dreamt thought experiment. But if we believe Freud, our subconsciousness will probably count as the engine of intervention.

Newton's bucket is twirled by you, by me or by God is irrelevant.[38] Important is the *difference* between the stages with the flat water surface and with the parabolic surface.

In each case just discussed, some factor is "wiggled" and its effects on some other feature of the situation is recorded. Who does the wiggling is inessential. What is important is the arrangement of the factors in the envisaged situation. Only if the experiment is designed in particular ways, can it be informative about the issue at hand (the Aristotelian theory of the free fall, the reality of absolute space or the time needed for a change in the money stock to channel through, say). Whether or not the factors are arranged in a fortuitous way we may know if we have a detailed and accurate enough background knowledge. But if we have that knowledge, whether one brings about the change by a voluntary intervention or whether it obtains naturally is a side-issue.

3 Functions for Thought Experiments

In the last section I have been deliberately negative in my discussion because I think thought experiments are largely overrated in the literature. From an empiricist point of view, I also think it is obligatory to combat some rationalist tendencies that have become popular in recent years. The wide acceptance of the epistemic significance of "thinking without looking" that we find in thought experiments is one aspect of these tendencies.

But I do not think that there is anything wrong with thought experimentation as long as one is clear about what it can do for you. Accordingly, in this section I want to sketch a number of functions that I believe a thought experiment could fulfil within a wider scientific methodology.

My preferred model for a scientific methodology is taken from Francis Bacon's philosophical works. The reason for this choice in the context of this paper is that Bacon's methodological ideas are particularly amenable to thought experiments. Three features of his philosophy of science point towards this amenability. First, it is thoroughly pluralist. Although Bacon defends a general three-stage inductive scheme, he allows a broad range of in-

[38] One should, however, strictly speaking, abstract from the mass of the intervener in order not to confound the result. So maybe God is a good choice. Is masslessness a perfection?

dividual techniques (called "Prerogatives of Instances") to aid the inductive scheme. Thought experiments are easy to integrate into the broader methodology as yet another supportive technique. Just like the religious pluralism of the ancient Greeks allowed to incorporate new elements whenever new land with its people and their religious beliefs was conquered, a methodological pluralism welcomes new techniques whenever they are helpful. Second, *concept formation* is an explicit component of the general inductive scheme. Bacon emphasised time and again the importance of sound concepts if induction should be successful. Since concept formation is one important function of thought experimentation (which I will discuss momentarily), a Baconian methodology is very welcome to thought experiments. Third, another important Baconian topic is the collaboration of scientific faculties. In the critical part of his *Novum Organum*, Bacon bemoaned not that there is a lack of experimentation in the science of his day; nor that there is too little speculation. He criticised that these two functions are not exercised in a collaborative effort. One famous aphorism puts it in this way:[39]

> Those who have handled the sciences have been either Empiricists or Rationalists. Empiricists, like ants, merely collect things and use them. The Rationalists, like spiders, spin webs out of themselves. The middle way is that of the bee, which gathers its material from the flowers of the garden and field, but then transforms and digests it by a power of its own. And the true business of philosophy is much the same, for it does not rely only or chiefly on the powers of the mind, nor does it store the material supplied by natural history and practical experiments untouched in its memory, but lays it up in the understanding changed and refined. Thus from a close and purer alliance of the two faculties—the experimental and the rational, such as has never yet been made—we have good reason for hope.

Bacon thus anticipated Kant's "Percepts without concepts are blind; concepts without percepts are empty". There is a clear role for the mind in Bacon's scientific philosophy. The mind imposes order on the chaos of phenomena by way of classifying them according to a conceptual scheme; the

[39]Book I, aphorism 95

mind hypothesises laws and causal explanations; the mind devises experiments and other techniques to investigate nature; the mind stores, compares and transforms previous experience such that it is put to fruitful use. The "wonder of armchair enquiry" (Sorensen) consists in how we can learn about nature without making new experiences. Within Bacon's scientific methodology, this feature follows naturally. We need the mind to make sense of any experience. That thought experience can help us in so doing is no mystery.

Let me come back to the first point, Bacon's methodological pluralism. It is well known that Bacon's inductive methodology consisted of a three-stage process of (1) observing or experimentally creating, (2) classifying and (3) causally explaining phenomena. Causal laws should in turn be tested by new experiments; they lead to "new works" as Bacon said. This general process is aided by what Bacon called "Prerogatives of Instances", that is, techniques and methodological principles that support the general inductive process. In this, Bacon was pluralist—anything goes, as long as the technique supports the overall aim of finding causal explanations of phenomena. Bacon discusses twenty-seven Prerogatives of Instances, with no hint that this list should be understood as being exhaustive. Among the examples Bacon mentions are different kinds of measurements, the use of instruments such as the microscope and the telescope, the "crucial" experiment[40] and the use of analogical reasoning.

I think thought experiments can be added to the list as a twenty-eighth Prerogative. Thought experiments can help us in various ways on the road from bare observations to established causal explanations. In what follows, I want to discuss four such possible functions. Self speaking, these are not the only functions thought experiments can assume within a scientific methodology, but I think they are important and frequent. They are (a) concept formation, (b) establishing a causal hypothesis, (c) nomological refutation and (d) suggestion of "new works". Let us look at each of them in turn.

Thomas Kuhn re-introduced the importance of concept formation for scientific change to the philosophical discussion. It is no surprise, then, that the function he ascribes to thought experiments in his account is exactly that:

[40]I put "crucial" in quotes because there is good reason to believe that Bacon did not think that such experiments are crucial in the sense discussed in 20th century philosophy of science. See Hacking 1983, ch. 15, for a critique of 20th century views (in particular Imré Lakatos's). A better translation of the Latin *experimentum crucis* is "experiment of the crossroads".

they help in conceptual change.

The example Kuhn examines in his famous essay on thought experimentation is Galileo's inclined plane.[41] Galileo asks us to consider two planes, one vertical, the other inclined. Along the two planes, two bodies are imagined to slide without friction from the top. If we ask, Galileo reasons, which of the two bodies acquires the greater velocity at bottom, we are led into a contradiction. This is because, on the one hand, we know that both bodies must have the same speed—the speed necessary to carry them back the the height from which they started. On the other hand, traditional criteria that tell us which of two bodies is faster (*e.g.*, "that which arrives at the goal first") imply that the body moving along the perpendicular is the faster one.

The matter is complicated by the fact that the two planes have a different length. The faster of two motions is usually defined as "that which covers the same distance in less time". But then, traditional criteria can imply that either body is faster or that they travel at the same speed, depending on whether we compare the length of the perpendicular plane with the top or bottom segment or a segment in the middle of the inclined plane.

At any rate, we are led into a contradiction. The contradiction arises, Kuhn interprets, because the traditional concept of speed is incoherent. It does work for many situations, but not the one Galileo envisages in the thought experiment. The way out of the paradox is to separate out two senses of speed: "instantaneous speed" and "average speed". Kuhn writes:[42]

> 'Faster' and 'speed' must not be used in the traditional way. One may say that at a particular instant one body has a faster instantaneous speed than another body has at that same time or at another specified instant. Or one may say that a particular body traverses a particular distance more quickly than another traverses the same or some other distance. But the two sorts of statements do not describe the same characteristics of motion. 'Faster' means something different when applied, on the one hand, to the comparison of instantaneous rates of motion at particular instants, and, on the other, to the comparison of the times required for the completion of the whole of the two speci-

[41] Kuhn, *op. cit.*
[42] *ibid.*, p. 15

> fied motions. A body may be 'faster' in one sense and not in the other.
>
> That conceptual reform is what Galileo's thought experiment helped to teach... The concepts that Aristotle applied to the study of motion were, in some part, self-contradictory, and the contradiction was not entirely eliminated during the Middle Ages. Galileo's thought experiment brought the difficulty to the fore by confronting readers with the paradox implicit in their mode of thought. As a result, it helped them to modify their conceptual apparatus.

I agree with most of Kuhn's analysis but want to emphasise his use of the verb "help": thought experiments *help* us with conceptual reform, they do little on their own. On the basis of observations or experiments that make plausible that the thought experimental scenario is possible and that the idealisations involved are harmless, we see that the traditional concept is incoherent. The thought experiment also points us towards new concepts that avoid the contradiction. But there is no guarantee that the new concepts of "instantaneous" and "average" speed are helpful in any way but the scenario envisaged here. They still await the formulation of laws or causal explanations that use these concepts and the experimental confirmation of such laws or causal explanations. The thought experiment does not allow us to peep into Plato's heaven; it gives new concepts that may help us in formulating fruitful laws and causal explanations.

One might object that my use of Galileo's second thought experiment shoots back at me like a boomerang. Have I not just pointed out how uncertain the results of his first thought experiment are, and does that not apply to his second one as well. The answer is that it does not matter. Even if Galileo were the sole person in the world who achieves the result, he still has a hypothesis about how to classify phenomena. But the Baconian treatment of the thought experiment implies that the job is not finished until the paper work is done, which means in this case until the laws that are formulated on the basis of the new concepts are experimentally confirmed *etc.* Using a thought experiment as a creative tool underlines that we need further experiments; it emphasises that no result is compelling.

A second function of thought experimentation is to establish a causal hypothesis. In a famous *economic* thought experiment, George Akerlof analyses

the phenomenon of a large differential between the price of new cars and cars that have "just left the showroom". He asks us to:[43]

> Suppose... that there are just four kinds of cars. There are new cars and used cars. There are good cars and bad cars (which in America are known as "lemons"). A new car may be a good car or a lemon, and of course the same is true of used cars.
>
> The individuals in this market buy a new automobile without knowing whether the car they buy will be good or a lemon. But they do know that with probability q it is a good car and with probability $(1-q)$ it is a lemon; by assumption, q is the proportion of good cars produced and $(1-q)$ is the proportion of lemons.
>
> After owning a specific car, however, for a length of time, the car owner can form a good idea of the quality of this machine; *i.e.*, the owner assigns a new probability to the event that his car is a lemon. This estimate is more accurate than the original estimate. An asymmetry in available information has developed: for the sellers now have more knowledge about the quality of a car than the buyers. But good cars and bad cars must still sell at the same price-since it is impossible for a buyer to tell the difference between a good car and a bad car. It is apparent that a used car cannot have the same valuation, it would clearly be advantageous to trade a lemon at the price of a new car, and buy another new car, at a higher probability q of being good... Thus the owner of a good machine must be locked in. Not only is it true that he cannot receive the true value of his car, but he cannot even obtain the expected value of a new car.

In my reading, Akerlof provides us with a hypothetical causal factor (asymmetric information) that *could* explain the phenomenon of the price differential. He does that by constructing a hypothetical world in which that factor *does* cause the phenomenon. But by this we have reason to suppose that, in real situations, asymmetric information *could* cause the price differential. The inquiry is not finished, however, unless the causal hypothesis has been tested by concrete experiments. The thought experiment does not

[43] Akerlof 1970, p. 489

establish that there is even a single real-world situation in which asymmetric information does the job Akerlof wants it to do; it gives us a hypothesis.[44] But in conjunction with a series of concrete experiments, we can learn from Akerlof's thought experiment, just as we can learn from Galileo's free fall experiment.

As an aside, in this context I may mention that there is an important difference between the mathematical models we frequently find in economics and economic thought experiments of the kind Akerlof employs.[45] A mathematical model differs from a thought experiment in its level of explicitness and logical stringency. A model is a fully articulated system in the sense that all claims it makes follow deductively from a set of assumptions. A thought experiment, by contrast, leaves many details unspecified. For example, in Akerlof's scenario, there are four types of cars, good and bad, new and old. But nothing is said about any other properties they might have. In a mathematical model, cars have nothing but these properties. This is made sure by making quality and nothing else an argument in agents' utility function and by defining "new" as "agents know about the quality" (*i.e.* the car's quality rather than an expected value is an argument in their decision function) and "old" as "agent do not know about the quality".

As a consequence, the way we derive results from models and from thought experiments differs. A model *proves* a conditional of the form: $A \rightarrow R$, *i.e.*, for systems in which the model's assumptions A hold true, the results R must be true. The inference proceeds much more informally in thought experiments. Thought experiments make a result plausible in the light of our intuitions about an economic situation. But reverse results are always possible. Nothing forces us to "lock in the owners of good cars". Owners may reveal the quality of their cars, and buyers may believe them. People may give away the cars because they suddenly prefer biking. Or owners of bad cars may not notice the arbitrage opportunity that arises when they learn about the inferior quality of their cars. The result follows on account of our intuitions about what economic behaviour consists in. But outside the more formal mathematical model, nothing necessitates it.

Another way of putting the issue is as follows. Modelling proceeds hypoth-

[44] For a very similar reading of Akerlof's thought experiment, see Sugden 2000.

[45] To be sure, Akerlof presents a mathematical model after discussing the thought experiment.

etico-deductively while thought experiments reason in a way which closely resembles Mill's method *a priori*[46]. For Mill, truths in political economy cannot be found by induction from specific phenomena as there is no *experimentum crucis*. Rather, we need to establish the fundamental tendency laws on the basis of introspection. In my view, thought experiments can help us to establish such fundamental laws. They answer questions of the form, "Had I been in situation S and acted on the economic motive alone, what would I have done?" The answer to that is obviously hypothetical but it is not arbitrary: it is based on my experience with situations similar to S and with being greedy. The result of the mathematical model is more and less certain at the same time. It is more certain in that the conclusion must follow from the premisses, given of course we make no mistakes in the calculation. But it is less certain in that the initial premisses are more arbitrary, more purely hypothetical (I take, say, the assignment of a definite utility function to an agent to be more arbitrary than the ascription of an economic motive). Two final remarks. First, the difference between thought experiment and mathematical model is one of degree, not one of kind. The assumptions of a model may well be very plausible, and in many cases even consist of well-established laws. The other way round, we may design thought experiments so tightly such that our results follow as if by deduction. Either way, both techniques deliver only hypotheses. This is because a chain of reasoning is only as strong as its weakest link. In mathematical models, the weakest link often consists of the initial premisses; in thought experiments, it is the inference itself which is not always reliable. Second, although I have focused on economic models and thought experiments in this context, I take the results to be applicable similarly outside economics, especially with respect to physical models and thought experiments.

Let us go back to the functions for thought experiments. A third function is nomological refutation. One can devise thought experiments in order to demonstrate that some laws, conjoined with certain assumptions, cannot all be true. EPR is of course one important example, in which the thought experimenters show us that it cannot be the case that both relativity and quantum mechanics are true under the assumptions that both theories are complete and under a specific "criterion of reality".[47] Another example, also

[46]Mill 1844

[47]For a discussion in the context of thought experimentation, see Brown 1991b, ch. 6.

due to Einstein, is his failed attempt at showing that Heisenberg's uncertainty principle cannot be right, as has been discussed above.[48] Importantly, even this kind of thought experiments does not function independently. The reason is that, as before, the thought experiment presents a *possible world* in which two or more laws cannot be true at the same time. But unless we are fundamentalists and believe that laws must be, if true at all, *universally* true, it gives us only reason to believe that they *might* conflict in real situations, too.

A fourth function is to suggest "new works" as Bacon called them. In a recent contribution, Alisa Bokulich analyses a physical thought experiment that reveals interesting implications of special relativity.[49] In the thought experiment, we are asked to imagine two identically constructed rockets, one 100 metres behind the other and initially at rest with respect to inertial frame S. They are connected by a thin piece of thread that exactly spans the distance between the two rockets. We are further asked to imagine that the two rockets start their engines simultaneously and accelerate at identical rates up to a terminal velocity of 80% of the speed of light, at which point the engines stop and the rockets travel with uniform velocity. According to Bokulich, careful analysis of the scenario within special relativity reveals that the thread must break, *i.e.*, that Lorentz contraction can cause measurable stress on moving bodies. Bokulich takes this case to demonstrate that one function of thought experiments is to draw out the physical implications of our theories. But in so doing, they are in fact doing more: they suggest effects or "phenomena" whose existence is yet to be proved. Therefore they suggest new experiments to test whether these predicted phenomena are actually real.

This list of functions is surely not exhaustive but I think it does cover a number of important uses thought experiments can have within a broader scientific methodology. As last item, I want to discuss how the theory of thought experiment that I endorse bears on the claims that I attempted to refute in section two. I will concentrate on the more significant myths. The most important point is, I think, that thought experiments do not function

[48] See Bishop, *op. cit.*, for a discussion of the laws that are involved in the Einstein-Bohr debate. For an analysis of this and other Einsteinian thought experiments of this kind, see also Norton 1991.

[49] Bokunich 2001. This is one of the rare examples of a contribution that analyses a thought experiment that is not part of the canon I have described above.

in their own right but in conjunction with other methods and within a much broader enterprise. This follows directly from Bacon's methodology because no single method or "Prerogative" does anything in its own right. Bacon is famous for being an inductivist. But he often stressed the fact that claims established by inductive methods must be tested by deductive ones: "Such a road [of scientific investigation] is not level, but rises and falls; first ascending to axioms, then descending to works". Any claim, no matter how established, needs independent confirmation by other methods. And Bacon's "First Vintage"—the result of running the three-stage process once—is only a *first* vintage after all; it surely demands further investigation. As no result is compelling in any way, it follows naturally that thought experiments should not be regarded as compelling.

That the epistemic worth of thought experiments piggy-backs on observations and experiments also follows nicely from Baconian ideas. Only if the rational and the empirical faculty co-operate fruitfully, will we gain knowledge about nature. So we need both: thinking, inventing, hypothesising and observing, measuring, experimenting. We need one in order to have the other. In order to fill a thought experiment with empirical content, we therefore have to have made the requisite experiences. But the reverse is not true. Experiment and observation require background knowledge to make sense. But they do not require thought experiments in particular. Thought experiments are elegant means for concept formation, establishing hypotheses, finding "new works" and fiddling with our framework of laws. But there are many alternative means (mathematical modelling has been mentioned as an example).

Finally, as mentioned above, the Baconian account also makes thought experiments seem less puzzling. As a significant contribution of the mind is an integral part of the methodology, it should appear as no surprise that a method which does not use novel empirical input has a distinctive role within a broader scientific enterprise.

4 Conclusion

I tried to fulfil two aims with this paper. This first was to take the spell out of thought experiments. I believe that thought experiments are neither particularly important nor special nor mysterious in any kind of way. The

second aim was to defend an empiricist view of thought experiments. I hope to have shown that, like many other scientific practices, thought experiments do have a number of functions. They help us in forming new concepts; they invent causal or nomological hypotheses; they give us ideas about hitherto unobserved phenomena; they test accepted laws for compatibility. But in none of the functions I identified, a thought experiment does any work all by itself. In each case, if epistemic advance is made, that advance is derivative on other, concrete experiments (or observations). With this, I hope to have honoured the empiricist belief that only nature can teach us about nature. Whatever method we devise to steal some secrets from her, it is her not us who has the last say.

References

Akerlof, George 1970, "The Market for "Lemons": Quality Uncertainty and the Market Mechanism", *Quarterly Journal of Economics* **84**:3, 488-500

Atkinson, David 2002, "Experiments and Thought Experiments in Natural Science", manuscript, University of Groningen, Netherlands

Bacon, Francis 1620, *Novum Organum*, translated and edited by Peter Urbach and John Gibson 1994, Chicago and LaSalle, IL: Open Court

Bishop, Michael 1998, "An Epistemological Role for Thought Experiments", in Shanks, Niall (ed.): *Idealization IX: Idealization in Contemporary Physics, Poznan Studies in the Philosophy of the Sciences and Humanities*, Amsterdam: Rodopi, 19-33

— 1999, "Why Thought Experiments Are Not Arguments", *Philosophy of Science* **66**, 534-41

Bogen, James and James Woodward 1988, "Saving the Phenomena", *Philosophical Review* **97**, 302-52

Bokulich, Alisa 2001, "Rethinking Thought Experiments", *Perspectives on Science* **9**:2, 285-307

Brown, James 1986, "Thought Experiments Since the Scientific Revolution", *International Studies in the Philosophy of Science* **1**, 1-15

— 1988, "Platonic Explanation: Or, What Abstract Entities Can Do For You", *International Studies in the Philosophy of Science* **3**, 51-67

— 1991a, "Thought Experiments: A Platonic Account", in Horowitz and Massey 1991b, 119-28

— 1991b, *The Laboratory of the Mind: Thought Experiments in the Natural Sciences*, London: Routledge

— 1993, "Why Empiricism Won't Work", *Proceedings of the Biennial Meetings of the Philosophy of Science 1992*, 271-9

— 1995, "Critical Notice of Roy Sorensen 'Thought Experiments'"

— 1998, "Einstein's Principle Theories", *Protosociology* **12**, 144-57

Duhem, Pierre 1991 [1914], *The Aim and Structure of Physical Theory*, Princeton: Princeton University Press

Franklin, Allen 1986, *The Neglect of Experiment*, Cambridge: CUP

— 1990), *Experiment, Right or Wrong*, Cambridge: CUP

Galison, Peter 1987, *How Experiments End*, Chicago, IL: University of Chicago Press

— 1990, *Image and Logic*, Chicago, IL: University of Chicago Press

— forthcoming, *Theory Machines*, Chicago, IL: University of Chicago Press

Gendler, Tamar Szabó 2000, *Thought Experiment: On the Powers and Limits of Imaginary Cases*, New Work/London: Garland

Gooding, David 1993, "What is Experimental' about Thought Experiments?", *Proceedings of the Biennial Meetings of the Philosophy of Science 1992*, 280-90

—, Trevor Pinch and Simon Schaffer (1989), *The Uses of Experiment*, Cambridge: CUP

Hacking, Ian 1983, *Representing and Intervening*, Cambridge: CUP

— 1993, "Do Thought Experiments have a Life of Their Own? Comments on James Brown *et al.*", *Proceedings of the Biennial Meetings of the Philosophy of Science 1992*, 302-8

Horowitz, Tamara and Garald Massey 1991a, "Introduction" in Horowitz and Massey 1991b, 1-28

— 1991b, *Thought Experiments in Science and Philosophy*, Lanham: Rowman and Littlefield

Hudson, Robert 1999, "Mesosomes: A Study in the Nature of Experimental Reasoning", *Philosophy of Science* **66**, 289-309

Hume, David 1985 [1777], "Of Interest", in *Essays: Moral, Political and Literary*, Indianapolis: Liberty Fund

Janis, Allen 1991, "Can Thought Experiments Fail?", in Horowitz and Massey 1991b, 113-8

Jevons, William Stanley, *Investigations in Currency and Finance*, London: Macmillan

Kuhn, Thomas 1981 [1964], "A Function For Thought Experiments", in Hacking, Ian 1981, *Scientific Revolutions*, 6-27

Latour, Bruno and Woolgar, Steve 1986, *Laboratory Life*, Princeton, NJ: Princeton University Press

Laymon, Ronald 2000, "Idealizations", in *Routledge Encyclopedia of Philosophy*, Online Edition

Longino 1990, *Science As Social Knowledge: Values and Objectivity in Scientific Inquiry*, Princeton: Princeton University Press

Losee, Brian 1993, *A Historical Introduction to the Philosophy of Science*, Oxford: OPUS/OUP

McAllister, James 1996, "The Evidential Significance of Thought Experiment in Science", *Studies in History and Philosophy of Science* **27**:2, 233-250

Mach, Ernst 1942, *The Science of Mechanics*, 9th ed., London: Open Court

Mill, John Stuart 1844, "On the Definition of Political Economy; and on the Method of Investigation proper to it", in *Essays on Some Unsettled Questions of Political Economy*, London: Parker, 120-164

Nersessian, Nancy 1993, "In the Theoretician's Laboratory: Thought Experimenting as Mental Models", *Proceedings of the Biennial Meetings of the Philosophy of Science 1992*, 291-301

Norton, John 1991, "Thought Experiments in Einstein's Work" in Horowitz and Massey 1991b, 129-48

— 1995, "Are Thought Experiments Just What You Thought?", *Canadian Journal of Philosophy* **26**:3, 333-66

Schabas, Margaret forthcoming, "Hume on Thought Experiments", unpublished manuscript, University of British Columbia, Vancouver

Sorensen, Roy 1992a, *Thought Experiments*, New York, NY: OUP

— 1992b, "Thought Experiments and the Epistemology of Laws", *Canadian Journal of Philosophy* **22**:1, 15-44

— 1995a, "Precis of 'Thought Experiments'", *Informal Logic* **17**:3, 385-7

— 1995b, "Sorensen's Reply to Bunzl and Feldman, *Informal Logic* **17**:3, 399-405

Sugden, Robert 2000, "Credible Worlds: The Status of Theoretical Models in Economics", *Journal of Economic Methodology* **7**:1, 1-33

Worrall, John 2002, "*What* Evidence in Evidence-Based Medicine?", *Causality: Metaphysics and Methods Technical Reports CTR 01/02*, CPNSS, LSE