THOUGHT EXPERIMENTS IN SCIENTIFIC REASONING

Andrew D. Irvine

I. Introduction

CONSIDER the following common definition of a thought experiment:[1] a thought experiment is an instance of reasoning which attempts to draw a conclusion about how the world either is or could be by positing some hypothetical, and perhaps even counterfactual, state of affairs. In short, it is an instance of hypothetical reasoning whose antecedent assumptions may well be false but which leads us to conclusions about the nature of the world or about our surroundings.

As a first attempt, such a definition has its merits, but it is not difficult to conclude that it also has its shortcomings. After all, it is unlikely that we would want to claim, as this definition allows, that *any* instance of hypothetical reasoning which purports to result in information about the world or about our surroundings is a thought *experiment*. Surely some instances of hypothetical reasoning, whatever their accompanying intentions, are simply not scientific enough to be deemed thought experiments. After all, if thought experiments in fact result in genuinely scientific knowledge, it would seem plausible to postulate that something like the kind of rigorous and systematic methods of inquiry which characterize other scientific practices should also be applicable in the case of thought experiments. If so, some more detailed explanation of what distinguishes genuine thought experiments from other, less scientific, forms of hypothetical reasoning needs to be given.[2]

With this in mind, the purpose of what follows is to examine the nature and role of several forms of reasoning used in the sciences that are commonly grouped together under the heading "thought experiments." Specifically, the task shall be to distinguish thought experiments within science from other, equally common, forms of hypothetical reasoning. Just what is a thought experiment? What is its relationship—its similarities and its differences—to a physical experiment? How do we determine its outcome? What relationship do thought experiments have to scientific theories? And so on.

In the course of deliberating about such questions, the following thesis will be defended: thought experiments, in the first instance, are simply arguments concerning particular events or states of affairs of a hypothetical (and often counterfactual) nature which lead to conclusions about the nature *of the* world around us. After all, thought experiments do not purport to introduce any new empirical data and, since they do purport to yield some new information about the world, the only other source of obtaining such new information must result from the reconsideration of previous data by way of argument. This much, at least, should be uncontroversial.[3] However, in addition, like physical experiments, thought experiments must stand in a privileged relationship both to past empirical observations and to some reasonably well-developed background theory. It is this additional requirement that distinguishes thought experiments from other, less scientific and less experimental, forms of reasoning. The result is that, just as not every observation made of the world around us can fairly be said to constitute a physical experiment, not every instance of hypothetical or counterfactual reasoning can fairly be said to constitute a thought experiment.

This parallel between physical experiments and thought experiments is a strong one. For example, just as with physical experiments, many (although not necessarily all) of a thought experiment's assumptions must be supported by independently confirmed empirical observations; the thought experiment must take place within the context of a reasonably well-developed background theory; it must have some bearing on how we answer questions determined within this practical and theoretical context; there must be independently isolatable variables in order that we are able to determine correlations between variations of these variables and the experiment's outcome; and so on. In addition, it should also be noted that, just as a physical experiment often has repercussions for its background theory in terms of confirmation, falsification or the like, so too will a thought experiment. Of course, the parallel is not exact: thought experiments, unlike most physical experiments, do not include actual interventions within the physical environment. It may also be the case that certain other characteristics of physical experiments (such as practical concerns regarding time or the environment) may be only occasional factors in the case of thought experiments. Nevertheless, without many such shared characteristics, thought experiments could never be distinguished, even in principle, from other equally common forms of reasoning.

Recognizing this, it is worth considering one further point. Sometimes it is claimed that the reason that our initial definition is too wide is simply that it allows under the rubric of "thought experiment" instances of reasoning in which all the premisses may in fact turn out to be true. On this account, what distinguishes a thought experiment from other forms of hypothetical reasoning is that a thought experiment must necessarily have some counterfactual element. After all, typical examples of thought experiments are often based upon counterfactual assumptions. Thought experiments regularly posit things like perfectly frictionless surfaces, absolute vacuums, perfectly elastic bodies, or an ideal market economy.[4]

However, this stronger requirement of counterfactuality in the definition of a thought experiment can be seen to be uncalled for, especially when it is pointed out that (i) many thought experiments are meant to precede real experiments in which the original thought experiment's premisses are actually instantiated, and (ii) many thought experiments include only premisses which are either known to be true or about which we remain agnostic. In this latter case, such premisses may in fact turn out to be counterfactual but, as far as we know at the time, they may also turn out already to have been actualized.[5] Both cases indicate that a requirement of necessary counterfactuality should be abandoned.

II. Thought Experiments in Physics

Let us now consider whether the above thesis regarding the nature of thought experiments can be substantiated. We shall begin by characterizing a "typical" physical experiment, the kind of experiment used on a day-to-day basis in a discipline which has undeniably used both physical experiments and thought experiments to great effect throughout its history, the discipline of physics.

Typically,[6] a physical experiment can be defined as a designed intervention in nature or the laboratory whose observable consequences serve to test some hypothesis previously developed in light of both past observations and a particular theoretical context. Typically, too, the experiment is carried out under controlled, and hence reproducible, conditions. These conditions are associated with particular events or states of affairs from which more general conclusions regarding the nature of the world can then be generated. The observations made then often lead to a supplementing or revising of the original theoretical context.

In addition, it must be added that physical experiments typically allow for the identification and distinguishing of some well defined set of independent (or antecedent) variables. This is in order that observational correlations can be made between variations in the original conditions of the experiment and a further set of dependent variables. In other words, there is an assumption in every experiment that the conditions which in fact cause observed effects can be isolated into independently variable factors in such a way that by altering one or more of these factors, while at the same time holding others constant, different causal relationships can be determined. In sum, actual experiments typically include procedures for the testing of hypotheses in which the experiment's antecedent conditions are regulated or controlled in order to discover their relationship with the experiment's outcome.

Now, if there are indeed strong similarities between physical experiments on the one hand and thought experiments on the other, we should expect that many of the above same characteristics will be incorporated within a typical thought experiment. By way of a test case, let us consider the following thought experiment associated with the theory of special relativity:[7] Let us assume hypothetically that we are in a glass room which is moving at a high but uniform velocity towards an outside observer. In the exact center of the room is a light bulb which flashes periodically. At the exact moment that we and the room reach the outside observer the light bulb flashes. Is there any difference between what we see and what the observer sees?

Einstein's answer to this question was of course "yes." Since the light from the bulb travels away from the bulb at a constant velocity, and since all the walls of the room are equidistant from the light source, it follows that we will see the light strike both the front wall and rear wall at the same instant. However, for the outside observer there is a crucial difference. From his or her viewpoint outside the room, the front wall is moving forward, away from the approaching light. At the same time the rear wall is moving forward to meet it. Thus, from the point of view of the outside observer, the light will reach the rear wall before it reaches the front wall. What was simultaneous from our position inside the room will now be distinct. The conclusion we therefore draw is that the same event will appear differently depending upon the observer's frame of reference.

This thought experiment resembles a physical experiment in several essential aspects. Many of its premisses (such as that stating the constancy of the velocity of light) are assumptions made on the basis of past observations or as a result of a particular theoretical context. There is a hypothesis (in this case concerning the simultaneity of events) that is being tested and whose confirmation or falsification will have implications for the background theory. It is also possible to isolate particular antecedent variables (such as the velocity of the room relative to the observer) and to vary them in repeated instances of the experiment in order to determine their altered effect (if any) on the outcome.

With these points in mind, it also is possible to see that many of the parameters of the experiment have been determined solely by its theoretical context. For example, since the special theory of relativity is applied only to co-ordinate systems moving uniformly relative to each other, it is essential that the imaginary room be understood to be moving uniformly relative to the outside observer if the experiment is to have its intended application. Like a physical experiment, too, this thought experiment deals primarily with particular events in order to help substantiate a more general conclusion regarding the nature of the world around us. Past and present theories may have to be revised in light of the experiment's conclusion.

Where the thought experiment differs from a typical physical experiment is in the case of what might be called its *formal*, as opposed to its *scientific*, characteristics. For example, while a physical experiment typically depends upon an actual intervention in nature, the thought experiment takes the form of an argument based upon hypothetical premisses. While a physical experiment will usually be heavily reliant upon its instrumentation and methods of measurement, a thought experiment will, as a rule, emphasize differences primarily of a qualitative nature. However, what must be noted is that differences in these formal characteristics fail to affect the thought experiment's basic scientific role. Thought experiments, like physical experiments, still take place within a particular theoretical context, they still provide evidence concerning the nature of the world around us and they still are subject to the requirements inherent within some form of scientific method.

III. Thought Experiments in Antiquity

By way of contrast to the above observations, let us now consider an additional hypothesis, namely, that it was the Presocratics who introduced the use of thought experiments in their reasoning about nature and, in so doing, it was they who introduced a versatile, efficient instrument that would prove to be essential for the later development of the sciences.[8]

If our earlier thesis is correct and thought experiments, like physical experiments, can take place only within the context of a reasonably sophisticated scientific method, we should immediately have a certain uneasiness with such a contention. Put simply, this uneasiness will arise once we recall that, despite their importance in the history of western thought, the Presocratics represented only a first step in the development of science as we know it today. It is unlikely that the initiators of what was (at best) only a primitive science, or (what is more likely) a sophisticated pre-science, could have introduced a scientific instrument of the sophistication of a thought experiment.

Of course, if so, it must also be added that this would in no way diminish the intellectual contribution of these thinkers' many achievements. Such a conclusion would not negate the fact that the Presocratics were pioneers in the development of thought experiments, just as they were pioneers in the natural sciences and in the science of reasoning in general. They introduced into the West an intellectual curiosity about the natural world coupled with an awareness of the importance of both observation and the critical method that, today, we still do well to emulate. In one author's helpful phrase, the Presocratic Greeks "dared to look nature in the face,"[9] and it was this, as opposed to any particular theory or exact methodology, that constituted their greatest accomplishment.

However, just as it cannot be said that the initial inquiries into the nature of the world on the part of the Presocratics could constitute the fruition or realization of science, neither can it be said (at least in most cases) that

Presocratic hypothetical reasoning could constitute anything more than an initial step towards the development and implementation of thought experiments. Just as it was to take centuries for science and its methodology to develop into a mature, recognizable form, so it was too with thought experiments. In short, even though the Presocratics were adept in the use of hypothetical reasoning, *reductios*, and explanatory conjectures, it doesn't follow that the use of these forms of reasoning should, in itself, be considered sufficient to prove the existence of thought experiments. It is much more likely that the use of hypothetical reasoning in and of itself constituted only an initial but necessary first step in the development of what were later to become full-fledged thought experiments.

In defense of this conclusion, let us compare two examples of the use of hypothetical reasoning in early Greek thought. One is an example of Thales' concerning the natural world. The other is a much later example from Euclid concerning the infinity of primes.

Let us first consider the Thales example.[10] Thales introduces the hypothetical assumption that the earth is like a floating log in order to account for its constancy of position. His idea is that if we make the assumption that the earth is like a floating log, this assumption will in turn help to explain our belief in the stability of the earth's position. The conjecture will then itself gain support because of its explanatory role with respect to our original, independently confirmed belief.

This use by Thales and other Presocratics of arguments involving explanatory conjectures is important in several respects. For example, it indicates a fundamental interest in, and an awareness of, the natural world. In addition it also shows that, at least to some extent, the Presocratics may have anticipated the modern theory of abduction. Thus, without a doubt, Thales and his contemporaries were involved in a form of hypothetical reasoning about the natural world. Now, the question that must be asked is: do these two factors together mean that Thales' conjecture should be understood to be a thought experiment?

According to C. S. Peirce, abduction constitutes, along with deduction and induction, a third general form of reasoning. In Peirce's view, abduction concerns the selection and evaluation of hypotheses on the basis of what can be called their explanatory content. Consider some known fact F. According to Peirce, to the extent that some hypothesis H can be said to *explain* F (as opposed to simply entailing F, or providing inductive support for F), this fact can be said to more or less constitute evidence in favor of H. H is then said to be supported by abduction. In such cases, *explanation* must be more than simply having F follow deductively or inductively from H since, for example, even though F follows deductively from, say, $p \& \sim p$, this conjunction cannot in any way be said to *explain* F. Along similar lines, even though the sight of smoke might be a good reason for believing that there is a fire, the sight of smoke would not explain the presence of fire in the way that the presence of fire would explain the sighting of smoke.[11] In short, explanations are to be distinguished from good reasons.

Now, although it is clear that Thales' explanatory conjecture was concerned with the natural world and involved reasoning of a form resembling that outlined in Peirce's theory of abduction, the question still remains: Is this sufficient for an instance of reasoning to be judged a thought experiment? As it stands, this appears unlikely. Without some more detailed observational and theoretical context in which the hypothetical reasoning is to take place, it is implausible to categorize Thales' conjecture as an *experiment*, either a thought experiment or otherwise. Thales' conjecture, remains simply a case of reasoning by analogy. As it stands, it just does not have enough that is recognizably experimental about it to be considered a thought experiment.

What we are forced to conclude is that, although many of the types of reasoning used by the Presocratics represent an important first step towards the development and use of thought experiments in science, this is all they represent, namely, a first step. Just as this degree of sophistication in experimentation in general was not found in the first Presocratic attempts at science, neither does it appear in their initial attempts at reasoning about the natural world.

By way of comparison, let us next consider Euclid's famous proof of the existence of an infinity of primes.[12] The proof is a *reductio*. Consider first the series of known primes arranged in ascending order, namely,

2, 3, 5, 7, 11, 13, 17, 19, 23,....

To prove that this series never comes to an end, let us first suppose that it does. If so, there will have to be some prime P, such that P is the largest prime and that

2, 3, 5, 7, 11, 13, 17, 19, 23,...P

is the completed series.

Next, let us consider the number Q as defined by the formula

$$Q = (2 \cdot 3 \cdot 5 \cdot 7 \cdot 11 \cdot 13 \cdot 17 \cdot 19 \cdot 23 \ldots\ldots P) + 1.$$

Q cannot itself be a prime since by definition it is larger than any prime in the series ending with P. But if it is not a prime, it must be divisible by some factor other than 1 and itself and hence divisible by some prime. Clearly Q is not divisible by any of the primes within the completed series, since in each case there will be a remainder of 1. Thus, contrary to our original assumption, there must be some additional prime greater than P. Since P was arbitrary, there must exist an infinity of primes.

Now, what are we to say of this piece of hypothetical reasoning? In contrast to the case of Thales' conjecture, this argument is much more like a thought experiment in many respects. It uses hypothetical reasoning to test a particular hypothesis (regarding the number of primes) that arose within a particular, quite detailed theoretical context. The assumptions made can be seen to arise out of, and to be justified within, that same context. There appear several independently isolatable variables and the result of the reasoning will have ramifications, not only for the theory of mathematics, but also for many associated theoretical beliefs concerning the finite and the infinite. In many more respects than in the case of Thales' example, this instance of hypothetical reasoning appears to be more like that of a thought experiment.

However, it must also be added that the above example of hypothetical reasoning appears, *prima facie*, to have very little to do with the physical world. If this instance of reasoning is in fact to be classed as a thought experiment, it will clearly be a thought experiment in mathematics rather than in the natural sciences. Unlike thought experiments in the natural sciences, this thought experiment has comparatively little to do with past observational data. It is also not quite clear (at least not without some quite detailed metaphysical account[13]) whether such a thought experiment could ever be tested by way of related physical experiments and, if so, what would constitute such tests.

What these latter concerns indicate is that the defining characteristics of thought experiments will turn out, in part, to be subject-dependent. Thus, it may turn out that within some bodies of organized knowledge several of the characteristics that we would generally attribute to thought experiments in other disciplines will not be present. For example, in some cases there may exist a satisfactory theoretical context as well as a well-defined set of antecedent variables. However, it may also turn out that some scientific characteristics, such as the normally required high degree of reliance upon past observational data, will vary. In such cases this variability will best be understood as a function of the kind of restrictions inherent in the experimental method of the discipline itself, rather than as a result of any intrinsic feature of thought experiments in general. Nevertheless, in such cases, the essential characteristic of the thought experiment would still be that it remain closely tied to the concerns, practices and methodology of the particular branch of science in question.

IV. Thought Experiments in Astronomy

As an example of the variability of particular scientific characteristics within disciplines, consider the case of experiments in astronomy. Physical experiments in astronomy differ from other ordinary physical experiments in a number of respects: they do not take place within the controlled conditions of a laboratory, they do not usually involve interventions in nature, and quite often their observational consequences are not reproducible (at least not within what might be deemed a reasonable time period). Despite this, absence in astronomy of many of the "typical" characteristics associated with physical experiments in, say, physics or chemistry, important physical experiments and thought experiments still take place within the discipline.

By way of example, mention may be made of the following 18th century thought experiment which gave rise to the now famous Olbers' paradox.[14] Let us assume hypothetically that all light-emitting objects or radio sources such as galaxies are distributed uniformly throughout space.[15] We also initially assume that the universe is infinite in size. For simplicity, we might even assume it to be Euclidean. Under such assumptions, we then attempt to compute the amount of total background radiation present in the night sky.

Now, given that the intensity of light to reach us from any galaxy at a distance r from the earth will decrease as a factor of $1/r^2$, it may appear that only the closest of galaxies will make any significant contribution to the light of the night sky. However, light sources at a great distance, although they may be too faint to be detected individually, may still play some role in aggregate. Since we have assumed a uniform distribution of galaxies, this turns out to be just the case. Under these conditions the number of galaxies at distance r will increase as a factor of r^2. In other words, if we double the distance only a quarter of the light will reach us from any particular galaxy but, since the number of galaxies at that new distance will be proportional to the square of the new distance, the two factors cancel one another out.[16] Thus the contribution of light received from galaxies at distance r will be independent of r itself. The decrease in intensity from any one galaxy will be exactly offset by the increase in radiation due to the greater number of sources at that distance. Since we have also assumed that r can be taken to infinity, the inevitable conclusion is that, as we extend r, the increase in radiation will also increase proportionately until the whole sky should appear as if covered by one great sun.

As before, this thought experiment contains many of the scientific characteristics that are required of any experiment, thought experiment or physical experiment. Importantly, too, it has ramifications for the background theory in which the paradox was originally developed. For example, during the nineteenth century the most commonly proposed solution to the resulting paradox was to assume that the universe was not uniformly populated and that galaxies were therefore localized in some areas of the universe and not in others. Today, however, this attempted solution has been abandoned. Observations made out to distances of hundreds of megaparsecs[17] have given no indication whatsoever of the kind of localization that would be required in order to avoid the paradox in this manner.

In contrast to the nineteenth century proposal, the most satisfactory contemporary resolution of the paradox emphasizes two separate factors. Together these two factors account for the required lessening of the contribution made by distant galaxies to the observable extragalactic background light. The first of these two factors is that the universe, including all very distant sources of radiation, has existed for only a finite period of time. Because the time required for emissions from very distant galaxies to reach the earth is in many cases greater than the age of these galaxies, it turns out that the finite life-span of the universe becomes a crucial factor in any calculation of background radiation.[18]

The second factor is that it is now generally accepted that the universe is expanding. This expansion in turn causes the further weakening of radiation received from distant galaxies. Since all galaxies are now thought to be moving away from one another with speeds proportional to their distance apart, it turns out that the further away a galaxy is from the earth, the greater the weakening of the radiation that we receive from it. One estimate is that under these conditions no light could ever reach us from distances of greater than about 2,000 megaparsecs.[19]

Today the theory of an expanding universe of finite age receives confirmation from two independent sources. First, confirmation comes from observations of the so-called red shift, observations of the shift of spectral lines toward longer wave lengths. The shift results from a change in frequency caused by the relative motion between the observer and the source, in short, by the expansion of the universe. Second, confirmation comes as a result of calculations of the age of the universe based on the rate of radioactive decay (for example, of uranium isotopes U-235 and -238) which yield figures comparable to those determined by the red shift.

V. On The Value of Thought Experiments for Science

Given the above considerations, what now can be said in general concerning thought experimentation within the physical sciences? Our initial conclusions have been twofold. First, the intuition that there should be a distinction drawn between hypothetical reasoning in general and thought experiments in particular can be defended. In disciplines as diverse as astronomy and physics a case can be made for claiming that thought experiments can and should be distinguished from instances of simple, hypothetical reasoning. Second, if our earlier analysis is accurate we have gone some way towards indicating why it is that not every instance of hypothetical or counterfactual reasoning will constitute a thought experiment. Thought experiments are to be understood as arguments concerning particular hypothetical events or states of affairs which stand in a privileged relationship both to previous empirical observations and to a particular background theory. This privileged relationship between a thought experiment and its operative observational/theoretical context requires that the thought experiment exhibit at least several, if not all, of those scientific characteristics which typify experiments in general. As a representative (but probably not exhaustive) list of such characteristics, the following have been mentioned:

- A thought experiment must be relevant to the testing of some hypothesis (or to the answering of some set of questions) which has arisen within a particular observational/theoretical context.
- Many, although not all, of the assumptions within the thought experiment must be supported by independently confirmed empirical observation. In short, at least some features of the thought experiment must be grounded in the observable world if it is to have any relevancy to general scientific inquiry.
- The thought experiment needs to be set out in enough detail (in a controlled enough environment, so to speak) that it is capable of being repeated. Good thought experiments, like most good physical experiments, are repeatable.
- It must be possible to identify a number of independent (or antecedent) variables within the thought experiment in order to determine correlations between variations of these variables and a further set of dependent variables used to characterize the experiment's outcome.
- The outcome of the thought experiment should have repercussions for the original background theory. The reasoning involved concerning the particular hypothetical event or state of affairs in question should provide evidence either for or against some general conclusion regarding the world or our environment. This will in turn lead to either a supplementing or a revising of the thought experiment's original theoretical context.

Thus, just as not every observation or manipulation of the world around us can lay claim to constituting a physical experiment, not every instance of hypothetical reasoning concerning the natural world will be considered a genuine thought experiment. Only instances of hypothetical and counterfactual reasoning which have their parameters in large measure determined by a corpus of relevant observational and theoretical concerns will count as genuine thought experiments. The existence and utility of any experiment (thought experiment or physical experiment) will appear only within the context of a well-developed scientific method and, although they will differ from physical experiments with respect to their formal characteristics, genuine thought experiments must share with physical experiments at least some of those scientific characteristics which help define particular disciplines and the scientific enterprise in general. If this thesis is not accepted, any distinction that we may wish to make between hypothetical reasoning in general and thought experimentation in particular will be in danger of collapse.

Given these conclusions, something further can now be said by way of explanation of the value of thought experiments within the natural sciences. In particular, something can be said concerning both the advantages and the limitations of using thought experiments. In a sense, the advantages of thought experiments are clear: thought experiments require no particular apparatus and, as a result, they are free from the practical limitations of the laboratory. They also, in many circumstances and under quite varied theoretical conditions, have been proven to be a powerful tool for the advancement of science.

In addition, thought experiments have the advantage of being able to reach conclusions regarding hypothetical ideal environments (frictionless surfaces, perfectly elastic bodies, situations of absolute zero, and the like), thereby helping to identify and distinguish between nonaccidental variations between variables and parameter sets. By omitting from consideration particular variables (such as the friction of a surface) the dependence or independence of other variables may become more clear. Once this occurs, the underlying laws can then be applied profitably to more complex, real-world situations.

Our earlier analysis also helps to indicate just how it is that such hypothetical (or even counterfactual) reasoning concerning ideal limits can be informative about the physical laws of the real world. Our insistence that thought experiments stand in a privileged relationship to past observational data guarantees that they be grounded in the real world. In such cases, the postulating of an ideal limit (or some other counterfactual premiss) is akin to ignoring what, rightly or wrongly, are believed to be extraneous factors within an experiment, factors which are deemed irrelevant to the particular purpose or issue at hand. On this account thought experiments can easily be viewed as simple hypothetical or counterfactual predictions, as predictions of what would be the case should certain (hypothetical or counterfactual) conditions ever obtain. In more mundane, practical environments, however, they will be characterized more profitably as theoretical heuristics, as guiding cases which to a greater or lesser extent (depending upon the type of assumption made) will approximate real world applications.

In contrast to such advantages, the limitations of thought experiments may be less readily apparent. The most important of these limitations is that a thought experiment, by definition, cannot introduce any new empirical data. As a result, we run the risk of being mislead by a thought experiment whenever the conclusion of a thought experiment goes beyond the information given in its premisses or whenever such information is itself unjustified. Thought experiments are limited to whatever extent we are capable of drawing unwarranted conclusions from the evidence at hand.[20] It must be remembered that the reason we believe thought experiments to be useful is because we believe that *if* they were carried out, the results would be as predicted. This is the case even with regard to counterfactual claims such as those concerning frictionless surfaces and the like. The claim in such cases is that *if* it were possible to construct a frictionless surface, then such-and-such would result. In the end, there is no test like a real test.

The reason that thought experiments so often can be misleading is that behind almost every thought experiment will lie a large number of unquestioned auxiliary assumptions, assumptions which are assumed to be true but which, if false, would overturn the result in question. Thought experiments, despite their advantages, can never replace observation and actual experiment. For proof of this it need only be pointed out that, should a thought experiment and an actual experiment conflict, it will almost always be the thought experiment, not the physical experiment, which will have to be revised.

As an example of the kinds of auxiliary assumptions present within a thought experiment, consider Galileo's famous thought experiment concerning falling bodies, presented as part of the dialogue during the first day of his *Discorsi*. The then contemporary Aristotelian theory held that the "natural speeds" of falling bodies were a function of their weight. Galileo's thought experiment vitiates this view, paving the way for a theory in which a body's "natural" acceleration is independent of its mass. The dialogue is between Salviati and Simplicio:[21]

> SALVIATI But, even without further experiment, it is possible to prove clearly, by means of a short and conclusive argument, that a heavier body does not move more rapidly than a lighter one.... But tell me, Simplicio, whether you admit that each falling body acquires a definite speed fixed by nature, a velocity which cannot be increased or diminished except by the use of force [*violenza*] or resistance.
>
> SIMPLICIO There can be no doubt but that one and the same body moving in a single medium has a fixed velocity which is determined by nature and which cannot be increased except by the addition of momentum [*impeto*] or diminished except by some resistance which retards it.
>
> SALVIATI If then we take two bodies whose natural speeds are different, it is clear that on uniting the two, the more rapid one will be partly retarded by the slower, and the slower will be somewhat hastened by the swifter. Do you not agree with me in this opinion?
>
> SIMPLICIO You are unquestionably right.
>
> SALVIATI But if this is true, and if a large stone moves with a speed of, say, eight while a smaller moves with a speed of four, then when they are united, the system will move with a speed of less than eight; but the two stones when tied together make a stone larger than that which before moved with a speed of eight. Hence the heavier body moves with less speed than the lighter; an effect which is contrary to your supposition. Thus you see how, from your assumption that the heavier body moves more rapidly than the lighter one, I infer that the heavier body moves more slowly.[22]

Here the unstated auxiliary assumptions include ones such as the assumption that falling bodies in fact can be successfully united during a fall in the manner that Salviati describes. They also include the assumption that two stones, when joined together, make a single stone and thus have a natural speed which is comparable in principle to that of other stones. If we alter such assumptions even slightly, a different result occurs.

As an example, let us postulate that upon further consideration Simplicio is unwilling to grant the empirical premiss that two bodies will remain joined together throughout a fall. He is not at all confident that in the case where two objects are joined together they may not become detached at some point early in their fall. He may even postulate that this is as a result of the fact that the rapid one is being retarded by the slower, and that the slower is being hastened by the swifter. If knots in rope and the like regularly do come undone during a fall, one account of this might just be that the pre-Galilean theory was correct. The quite detailed discussion between Salviati and Simplicio regarding similar considerations shows that Galileo was sensitive to just this fact.[23] The strength of the thought experiment depends crucially upon the relevancy of its unstated empirical assumptions.

The lesson, of course, is that thought experiments, despite their power and versatility, are simply fallible. To the extent that they are employed for anything other than exploring the consequences of statements made within a particular observational and theoretical context, we run the risk of being misled.[24]

NOTES

1. A thought experiment is also sometimes called a "*Gedankenexperiment.*" The term appears to have originated with Ernst Mach. See Mach (1883), in which he discusses several examples, *e.g.*, pp. 32-41, 159-62, and Mach (1897).

2. Discussions concerning the distinction between science and non-science have played a central role within much of twentieth century philosophy, first as a result of the logical positivist movement and later through the writings of Popper and his contemporaries. For the purpose at hand, the debate over *exactly* which characteristics distinguish science from non-science need not be entered into. For present purposes, it need only be assumed that scientific and non-scientific enterprises are distinct, and that the major distinguishing characteristics of the scientific enterprise, whatever they may be in detail, can be determined in reasonably broad outline. For a recent contribution to the historical debate which helps indicate why the distinction between science and non-science has become blurred in recent times, see Stove (1982).

3. Alas, although (for the reasons given) this much *should* be uncontroversial, it is not. For an opposing view, in which thought experiments are held to be *a priori* windows to the unobserved, see Brown (1986). Also see Koyré (1968), which Brown cites approvingly but not uncritically. The difficulty with all such views, of course, is in giving a satisfactory naturalistic epistemological account of the new factual knowledge which allegedly arises in such cases.

4. As an example of someone who claims that thought experiments are characterized (at least in part) by their typically counterfactual nature, see Brown (1986). Brown holds that thought experiments must be distinguished from *merely imagined* experiments since the former, but not the latter, are either "technologically, physically, or conceptually impossible" (p. 3). By way of an *ad hominem* against Brown, it need only be pointed out that Brown himself is unable to maintain this requirement for he, too, recognizes Galileo's famous thought experiment concerning free falling objects (which was neither technologically, physically nor conceptually

impossible, even in the seventeenth century) as a paradigm *thought* experiment (pp. 9f).

5. For example, consider the case of a hypothetical physicist who has not yet heard of the invention of the modern cyclotron. Working in extreme isolation, he might reason as follows: "Today we understand quite well the natural process of radioactive decay, at least in theory, as exhibited in such elements as radium-226 which decays eventually to lead-206. Now, if we could just introduce some mechanism capable of accelerating certain sub-atomic particles to very high speeds, we would be able artificially to induce the same result. By bombarding an element with high speed neutrons, we would destabilize its nucleus and thereby begin a process through which one of the incident neutrons is converted into a proton through beta decay. Without a doubt, the result would be the creation of another, slightly heavier element." Although he doesn't realize it, his thought experiment has already been actualized. Surely this fact alone should not determine whether his instance of reasoning is to be counted a thought experiment.

6. Of course, not every physical experiment has every one of the following characteristics. (For example, observations related to experiments in astronomy may not be repeatable. See section IV below.) Thus, in the case of both thought experiments and physical experiments, it is unlikely that there exists a single set of specific characteristics which are together both necessary and sufficient for a definition. Nevertheless, it should be plain that in both cases there are *typical* or *paradigm* examples which may be characterized.

7. For Einstein's more familiar example concerning trains, see Einstein (1917), pp. 25-27.

8. Just such a thesis is defended by Rescher. See *this volume*, pp. **-**.

9. Kline (1980), p. 10.

10. Rescher cites this conjecture of Thales' as an instance of a Presocratic thought experiment. See this volume, pp. 6-7.

11. The example is due to Armstrong who in effect uses abduction in his arguments against a regularity theory of laws of nature. See Armstrong (1983), p. 40.

12. What follows is a paraphrase of the proof of Proposition 20 of Book IX of Euclid's *Elements*. See Heath (1956), volume 2, pp. 412f.

13. As it happens, I believe that such an account can be given and that, as a result, mathematics is best understood to be more like the natural sciences than is traditionally assumed. Under such an account mathematics is seen to be distinct from physics, chemistry and the like, only as a matter of degree, not as a matter of kind. However, within the context of the current argument, nothing much hinges on whether such an account turns out to be successful.

14. The problem was first proposed by Edmund Halley in 1720 and restated more formally by the Swiss astronomer Chéseaux in 1743. However, it was not until the 1820s that it received its most famous formulation in the ruminations of Heinrich Wilhelm Matthäus Olbers (1758-1840). For a more detailed scientific discussion of the problem, see Sciama (1971), Chapter 8, pp. 98-127 (especially pp. 122ff); Hoyle (1955), Chapter 18, pp. 304-21; and Wesson, Vale & Stabell (1987).

15. More specifically, we assume that the universe is both homogeneous and isotropic. Therefore the average density of galaxies, and hence the average luminosity, can be assumed to be the same everywhere.

16. That is, on the one hand, the intensity of the radiation received from any particular source is inversely proportional to the square of the distance; but on the other hand, the number of radiation sources is directly proportional to the square of the distance. Thus, the number of galaxies increases as a function of r in a way

that exactly compensates for the weakening of radiation due to the increase in distance.

17. One parsec is equal to 3.2615 light years, or the apparent displacement of an object whose annual shift in the sky due to the earth's motion around the sun (or parallax) is one second of arc. A megaparsec is equal to one million parsecs.

18. For details, see Wesson, Valle & Stabell (1987).

19. Hoyle (1955), p. 314.

20. Even Koyré (1960), who claims that "Good physics is made *a priori*" (p. 88), admits that even Galileo's thought experiments sometimes lead to "assertions that reality persistently refutes" (p. 82).

21. Salviati, Galileo's mouthpiece, appears to have been named after a friend of Galileo whose memory he wished to perpetuate. It is also tempting to think that Galileo named Simplicio after the Aristotelian commentator Simplicius, but see Koyré (1960), p. 46, n. 1, for a discussion of the issue.

22. Galileo (1638), pp. 62f.

23. Galileo (1638), pp. 63ff.

24. Work on this paper began while I was a visiting fellow at the Center for Philosophy of Science, University of Pittsburgh, and continued later at the University of Toronto. During this time, financial support was gratefully received from the Social Sciences and Humanities Research Council of Canada. I am indebted to John Norton, Nicholas Rescher and Jim Brown for their stimulating discussions on this topic, and especially to John Forge for his thoughtful comments on an earlier draft of this paper. Finally, I would like to express my thanks to Gerald Massey and Tamara Horowitz for their invitation, first, to participate in their 1986 workshop on thought experiments and, later, to contribute to the present volume.

REFERENCES

Armstrong, D. M. (1983), *What is a Law of Nature?*, Cambridge: Cambridge University Press.

Brown, James Robert (1986), "Thought Experiments since the Scientific Revolution," *International Studies in the Philosophy of Science*, vol. 1, pp. 1-15.

Einstein, Albert (1917), *Relativity* (15th edition, translated by Robert W. Lawson), London: Methuen, 1979.

Galileo [Galileo Galilei] (1638), *Dialogues Concerning Two New Sciences* (translated by Henry Crew & Alfonzo de Salvio), New York: Macmillan, 1914.

Heath, Thomas L. (1956), *The Thirteen Books of Euclid's Elements* (3 volumes), New York: Dover Publications.

Hoyle, Fred (1955), *Frontiers of Astronomy*, New York: Harper & Brothers.

Kline, Morris (1980), *Mathematics*, New York: Oxford University Press.

Koyré, Alexandre (1960), "Galileo's Treatise *De Motu Gravium*: the Use and Abuse of Imaginary Experiment," in *Metaphysics and Measurement*, London: Chapman and Hall, 1968, pp. 44-88.

Kuhn, Thomas S. (1964), "A Function for Thought Experiments," in *The Essential Tension*, Chicago: University of Chicago Press, 1977, pp. 240-65.

Mach, Ernst (1883), *The Science of Mechanics* (6th edition, translated by Thomas J. McCormack), LaSalle, Illinois: Open Court, 1960.

Mach, Ernst (1897), "On Thought Experiments," in *Knowledge and Error* (translated by Thomas J. McCormack and Paul Foulkes), Dordrecht, Holland: Reidel, 1976, pp. 134-47.

Rescher, Nicholas (1987), "Thought Experimentation in Presocratic Philosophy," this volume, pp. 5-15.

Sciama, D. W. (1971), *Modern Cosmology*, Cambridge: Cambridge University Press.

Stove, D. C. (1982), *Popper and After*, Oxford: Pergamon Press.

Wesson, Paul S., K. Valle & R. Stabell (1987), "The Extragalactic Background Light and a Definite Resolution of Olber's Paradox," *The Astrophysical Journal*, vol. 317, pp. 601-6.