

SECTION 3

THOUGHT EXPERIMENTS
IN THE SCIENCES

CAN THOUGHT EXPERIMENTS FAIL?

Allen I. Janis

SINCE not everyone agrees on exactly what a thought experiment is, let me begin by saying what I mean by the phrase. In the first place, I shall restrict myself to the context of physics. Perhaps the ideas presented here apply to other fields, but I shall leave that judgment to those that know those fields better than I do. There are two characteristics of a thought experiment that will be relevant for the following discussion and which I take to be the defining characteristics. First, a thought experiment is a description of an experimental procedure as well as its outcome or possible outcomes. (In referring to an experimental procedure, I do not mean to imply that the experiment could actually be carried out. The reasons could range from technological difficulties that could, in principle, be overcome to difficulties of principle; for example, it may be desirable in a thought experiment to suppose that the universe were quite different from the way it actually is.) Second, the outcome or possible outcomes must be deduced by reasoning consistent with a given theoretical framework.

As an example, let me cite the well known Bohm modification of the Einstein-Podolsky-Rosen (EPR) thought experiment.¹ One imagines a source that emits two particles, each of spin one-half (in the standard units of Planck's constant divided by 2π), in a state whose total angular momentum and orbital angular momentum are both zero. The particles travel in opposite directions, so that simply by waiting long enough the particles can be as widely separated as one wishes. One then imagines that the component of spin of each of the particles in a given direction, say the z -direction, is measured. According to the laws of quantum mechanics, these measurements will show that the z -component of spin of one of the particles will be $+1/2$ and that of the other will be $-1/2$, regardless of which direction is chosen to be designated as the z -direction.

I shall comment later on the significance of the EPR experiment. My purpose here is only to illustrate the requirements that I wish to impose on thought experiments. This example does meet the two requirements that I listed. First, it is a description of an experiment and its possible outcomes. And second, the possible outcomes are deduced by using the theoretical framework of quantum mechanics, the exploration of which is indeed the purpose of the thought experiment.

It is well known that real experiments sometimes fail. The question that I wish to explore here is whether thought experiments can also fail. As a

preliminary, let me distinguish three different senses in which real experiments are said to fail.

First, an experiment can fail because it cannot be carried to completion. For example, there might be a breakdown of the equipment. Or perhaps some outside influence interferes with the proper functioning of the experiment (or of the experimenter). In this sense of failure, the experiment (or attempted experiment) simply fails to produce any result whatsoever. Examples of this sort of failure abound, but not normally in the published literature!

Second, an experiment can fail by giving a result that is incorrect. This also might be due to equipment breakdowns (probably undetected ones, in this case). Or perhaps the experiment was poorly designed, so that the results obtained are tainted by unintended influences. Although some definite result is obtained from the experiment, in this sense of failure, the result is not a correct representation of what the experiment was intended to disclose. A well known example of this sort of failure is provided by Kaufmann's measurements, in the early 1900's, of the dependence of the electron's mass on velocity. Kaufmann's results were inconsistent with the results of special relativity, or at least they seemed to be. Planck, however, studied the experiments and concluded that the values Kaufmann assumed for the electric field in his apparatus were wrong because of ionization produced in the gas residue of the evacuated apparatus. Planck's reanalysis of the data supported the predictions of relativity.²

Third, an experiment can fail by producing results that, although correct, fail to provide answers to the questions that motivated the experiment in the first place. An example of this sort of failure is provided by experiments that have attempted to determine whether or not the proton is stable. These experiments give no firm evidence for proton decay but indicate that, if the proton decays at all, its lifetime is probably at least 10^{32} years.³ At the time these experiments were begun, however, theories then in vogue indicated a shorter lifetime that would have been detectable by the experiments. Although it is generally believed that these experiments are giving correct results, they have failed to establish the expected instability of the proton.

One might argue that this third sense of failure is not really failure at all, since correct results are, in fact, obtained. If, in my example, the experimenters' motivation had been to set a lower bound on the lifetime of the proton, then the experiment would not even have fit my definition of failure in the third sense. The establishment of this lower bound, however, does not shed any light on the question of whether or not the proton actually decays, and so at least some interested parties would undoubtedly consider there to be a sense in which the experiment had failed. Although I grant that one could reasonably consider experiments of this type to be successes

rather than failures, I shall (at least for the purposes of the present analysis) continue to refer to them as failures in the third sense.

Let me refine my definition of failure in the third sense by contrasting the search for proton decay with Millikan's experimental examination of Einstein's photoelectric equation.⁴ Millikan undertook this work believing that Einstein was wrong in his views on the nature of light, and expecting that the experiments would refute Einstein's analysis; in fact, they were in excellent agreement with Einstein's predictions. So certainly these experiments did not accomplish the aims that originally motivated them. The distinction I would make between Millikan's experiments and those that attempt to measure proton decay is that Millikan's results established the correctness of Einstein's photoelectric equation, whereas the only conclusion that can be drawn from the proton experiments is that *if* the proton decays its lifetime must be greater than a certain value. Thus if an experiment sets out to test certain ideas and it establishes (at least to the extent that experiments ever really establish anything) that those ideas are either true or false, then regardless of what the experimenter's expectations or motivations may have been, I would classify the experiment as a success. On the other hand, if an experiment sets out to establish the existence of a certain effect and ends up only setting limits (beyond the range available to the experiment) on the size of the effect, then I would say the experiment has failed in the third sense.

It is my contention that thought experiments can also fail in each of these three ways. Let me consider them in turn.

A thought experiment can fail in the first sense if the thought experimenter (to whom I shall henceforth refer as "the TE") lacks the knowledge or ability to carry the analysis through to a conclusion. As in the case of real experiments, this may be due to poor equipment (i.e., inadequate brain power or training) or to outside influences (e.g., a disabling accident or illness, or the influence of alcohol or drugs). In this type of failure, there is simply no result that emerges from the process of thinking. Nevertheless, in analogy to the case of a real experiment, it can make sense to say that there has been a thought experiment that failed rather than that there has been no thought experiment at all. The thought experiment is conceived when the TE thinks that conceptual analysis along certain lines will lead to certain clarifications or conclusions, it is launched when the TE starts to carry out this analysis, and it fails when the TE is unable to bring it to a conclusion. As in the case of this sense of failure of real experiments, I suspect that there is an abundance of failed thought experiments in this sense but that one rarely, if ever, hears about them.

A thought experiment can fail in the second sense if the TE reaches an incorrect conclusion. A famous example of this sense of failed thought experiment is one proposed by Einstein at the 1930 Solvay Conference. He described a device which, in principle, seemed to allow measurements that

would contradict the form of the uncertainty principle that involves energy and time. Bohr, however, found that Einstein's analysis was incorrect. Ironically, it was a certain aspect of general relativity that Einstein had overlooked; when general relativity was properly taken into account, the analysis was in fact consistent with the uncertainty principle.⁵

A thought experiment can fail in the third sense when its analysis leads to a correct conclusion, but not one that accomplishes the purpose that provided its motivation. I shall argue that the EPR thought experiment is an example of this sort of failure, but to do so it is necessary first to make some additional comments about the experiment.

In view of what I have already explained about the EPR experiment, it is clear that if a measurement is made of the z -component of spin of one of the widely separated particles, then the z -component of spin of the other particle can be predicted with certainty. Similarly, if the x -component of spin of the first particle were measured instead of the z -component, then the x -component of the second particle's spin could be predicted with certainty. The EPR argument⁶ was that the measurement of the first particle's spin does not disturb the second particle in any way, and that if a physical attribute of a system can be predicted with certainty without disturbing the system, then that attribute corresponds to an element of physical reality. EPR further argued that, since either the x -component or the z -component of the second particle's spin can be predicted with certainty, depending on which measurement one chooses to perform on the first particle, both components of spin of the second particle correspond to elements of physical reality. Since according to the conventional ("Copenhagen") interpretation of quantum mechanics these spin components cannot simultaneously have well defined values, EPR concluded that the description of reality given by quantum mechanics is incomplete.

At this point, I should like to draw a line between the thought experiment itself and the conclusions that one might draw from the results of the thought experiment. It seems to me that, in this example, the line should come just after one deduces what the results of the spin-component measurements would be; the conclusions about the nature of quantum mechanics are on the other side of the line. Then there would seem to be no doubt that the results of the thought experiment are correct. But there has certainly been no general agreement that EPR's conclusions about the nature of quantum mechanics are correct.⁷ Thus the thought experiment failed to provide a clear basis for concluding that quantum mechanics is incomplete. Since this goal was the motivation for the thought experiment, however, the thought experiment failed in the third sense.

Let me explain why I consider the EPR thought experiment to be more analogous to the proton-decay experiments, which I cited as an example of failure of real experiments in the third sense, than to Millikan's photoelectric experiments, which I considered to be successful real experiments. To

be analogous to the Millikan experiment, a thought experiment would have to lead to the conclusion that certain ideas are correct, regardless of the experimenter's expectations. The EPR experiment, however, does not show that quantum mechanics is complete. It only fails to establish that quantum mechanics is incomplete. Like the proton-decay experiments, the EPR experiment shows neither the truth nor the falsity of the ideas that provided its motivation.

It seems to me that the third sense of failure, for both real and thought experiments, is the most interesting one. As I have already mentioned, failures in the first sense are rarely made public. Although failures in the second sense sometimes are published, it is unusual for the mistakes that were made to be of sufficient interest that the experiments are worth discussing once the mistakes have been discovered. The third sense of failure, however, often leads to a deeper understanding of the situation that prompted the experiment. In the example I gave for the third sense of failure of a real experiment, the theoretical arguments leading to the prediction of proton decay were subsequently refined to give a clearer understanding of the processes involved (and, of course, to give a decay time consistent with the experimental limits); whether what emerged is the correct understanding is not yet clear, however, for proton decay has still not been established experimentally. And in my example of the third sense in which a thought experiment can fail, the arguments the EPR experiment fueled about the nature of quantum mechanics are still raging. I suspect that many more failed thought experiments in the third sense may be necessary before real progress in understanding the nature of quantum mechanics is achieved.

NOTES

1. A more complete discussion of this thought experiment can be found in Bohm (1951).
2. The Kaufmann experiments and Planck's analysis of them are described, for example, in Miller (1981).
3. A review of these experiments is given in LoSecco (1985).
4. An account of Millikan's work on the photoelectric effect and his beliefs about the nature of light can be found, for example, in Kargon (1982).
5. An account of this thought experiment and Bohr's response to it is given in Bohr (1951).
6. In addition to the discussion in Bohm (1951), cited above, see also the original paper of Einstein, Podolsky, and Rosen (1935). Although this paper refers to a different form of the thought experiment, the ideas discussed by EPR in connection with the experiment are applicable to both versions.
7. Bohr (1935a) almost immediately wrote a short note to *Nature* in which he gave the basic idea of his response to the EPR analysis and very quickly followed it with a more detailed analysis in *The Physical Review* (Bohr 1935b).

REFERENCES

- Bohm, D. (1951): *Quantum Theory* (New York: Prentice-Hall, Inc.), pp. 611-22.
- Bohr, N. (1935a): "Quantum Mechanics and Physical Reality," *Nature*, vol. 136, p. 65.
- Bohr, N. (1935b): "Can Quantum-Mechanical Description of Physical Reality be Considered Complete?," *Physical Review*, vol. 48, pp. 696-702.
- Bohr, N. (1951): "Discussion with Einstein on Epistemological Problems in Atomic Physics," in P. A. Schilpp (ed.), *Albert Einstein: Philosopher-Scientist* (New York: Tudor Publishing Co., 2nd edition), pp. 224-28.
- Einstein, A., Podolsky, B., and Rosen, N. (1935): "Can Quantum-Mechanical Description of Physical Reality be Considered Complete?," *Physical Review*, vol. 47, pp. 777-80.
- Kargon, R. H. (1982): *The Rise of Robert Millikan* (Ithaca, NY: Cornell University Press), pp. 66-74.
- LoSecco, J. M. (1985): "Status Report on the Search for Proton Decay," *Comments on Nuclear and Particle Physics*, vol. 15, pp. 23-34.
- Miller, A. I. (1981): *Albert Einstein's Special Theory of Relativity* (Reading, MA: Addison-Wesley Publishing Co., Inc.), pp. 47-54, 61-67, 228-35, 340-41.