Why Thought Experiments Are Not Arguments*

Michael A. Bishop†‡
Department of Philosophy, Iowa State University

Are thought experiments nothing but arguments? I argue that it is not possible to make sense of the historical trajectory of certain thought experiments if one takes them to be arguments. Einstein and Bohr disagreed about the outcome of the clock-in-the-box thought experiment, and so they reconstructed it using different arguments. This is to be expected whenever scientists disagree about a thought experiment's outcome. Since any such episode consists of two arguments but just one thought experiment, the thought experiment cannot be the arguments.

1. Introduction. A number of philosophers have argued that thought experiments are arguments. John Norton presents a sophisticated account and defense of this view.

> Thought experiments are arguments which: (i) posit hypothetical or counterfactual states of affairs, and (ii) invoke particulars irrelevant to the generality of the conclusion. . . . Thought experiments in physics provide or purport to provide us information about the physical world. Since they are *thought* experiments rather than *physical* experiments, this information does not come from the reporting of new empirical data. Thus there is only one non-controversial source from which this information can come: it is elicited from information we already have by an identifiable argument, although that argument might not be laid out in detail in the statement of the thought experiment. The alternative to this view is to

*Received April 1998; revised May 1999.

†Send requests for reprints to the author, Department of Philosophy, 402 Catt Hall, Iowa State University, Ames, IA 50011; e-mail: mikebish@iastate.edu.

‡I would like to thank John Norton, Niall Shanks, Jonathan Sutton, Loretta Torrago, David Rudge, and an anonymous referee for very helpful comments.

Philosophy of Science, 66 (December 1999) pp. 534–541. 0031-8248/1999/6604-0002$2.00

suppose that thought experiments provide some new and even mysterious route to knowledge of the physical world. (1991, 129)

Without any new information *from* the world, thought experiments can yield new information *about* the world. Norton claims that only arguments uncontroversially have this property, since the conclusion of an argument can make explicit information that was implicit in the argument's premises. Others who identify thought experiments with arguments are Nicholas Rescher (1991, 31), Andrew D. Irvine (1991, 150) and John Forge (1991, 210).

Norton (1996) reconstructs various thought experiments in standard premise-conclusion form. This invites the interpretation that when he identifies thought experiments with arguments, he means 'argument' to be understood conventionally—as consisting of premises and a conclusion. So let us call the kind of argument that is supposed to be identified with a thought experiment a *t-argument*: it is an argument whose premises posit hypothetical states of affairs and general hypotheses (perhaps implicit) about the workings of nature and whose conclusion is some claim about the world that is supposed to follow from the premises.[1] My aim in this paper is to argue that thought experiments are not t-arguments.[2] Such a view cannot make sense of the historical trajectory of certain thought experiments. In particular, it cannot account for episodes in which different thinkers disagree about a thought experiment's results. Many such episodes have occurred in the history of science, e.g., Galileo on falling bodies and Newton's bucket (see Norton 1996 for useful discussions of these examples). In this paper, I will focus on an example from twentieth century physics.

2. The Clock-in-the-Box Thought Experiment. In 1927, Werner Heisenberg introduced the uncertainty principle that bears his name. The uncertainty principle says that there is an irreducible limit to the accuracy

1. The primary alternative to the view considered in this paper is the view that thought experiments are some sort of mental entity. They are *thought (about) experiments*. For a defense of this view, see Lennox 1991, Nersessian 1993, and Bishop 1998.

2. Roy Sorensen (1992) brackets the question of whether thought experiments are arguments by proposing, as a methodological matter, "a parity thesis: thought experiments are arguments if and only if experiments are arguments" (214). As an opponent of the argument view, and as one who agrees with Sorensen that there are many similarities between thought experiments and real ones, I accept his parity thesis. But I am aware of no defender of the argument view who has claimed that real experiments are arguments. There is good reason for this. Real experiments consist of things like beakers and cloud chambers. Arguments do not. Sorensen's parity thesis might reasonably be the *conclusion* of an argument for why thought experiments are not t-arguments. But as a premise, I think most parties to the debate would see it as question-begging.

to which pairs of conjugate variables (variable pairs like position-momentum, and energy-time) can be measured. This relation is described by the following equation,

$$\Delta p \times \Delta q > h$$

where Δp and Δq stand for the uncertainty in the measurement of the respective variables and h is a constant (Planck's constant divided by 2π).

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. So on the basis of this thought experiment, Einstein concluded that Heisenberg's uncertainty principle is false.

L. Rosenfeld describes Bohr's reaction to Einstein's clock-in-the-box thought experiment.

> It was quite a shock for Bohr . . . he did not see the solution at once. During the whole evening he was extremely unhappy, going from one to the other and trying to persuade them that it couldn't be true, that it would be the end of physics if Einstein were right; but he couldn't produce any refutation. I shall never forget the vision of the two antagonists leaving the club: Einstein a tall majestic figure, walking quietly, with a somewhat ironical smile, and Bohr trotting near him, very excited. . . . The next morning came Bohr's triumph. (Rosenfeld, quoted in Pais 1982, 446–447)

Bohr's "triumph" consisted in focusing on the practical instruments and procedures that one would have to use in order to measure the photon's energy at a particular time. Note Bohr's realistic illustration of the thought experiment—in particular, note that weighing the box will require the clock-in-the-box apparatus to move in a gravitational field (see Figure 1). Bohr showed that there was a fundamental limit to the accuracy to which any clock-in-the-box apparatus could measure a photon's weight (and hence energy) at a particular time, given by Heisenberg's formula.

$$\Delta E \times \Delta T > h$$

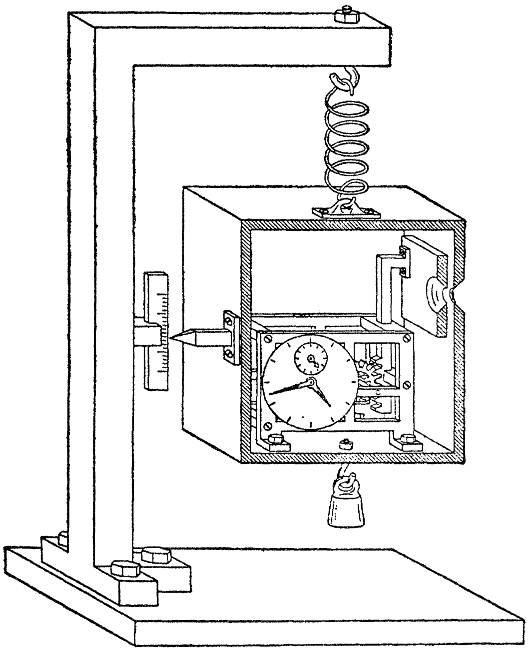

Figure 1. (Bohr 1949, 227; reprinted by permission of Open Court Publishing Company, a division of Carus Publishing Company, Peru, IL).

As a result of Bohr's argument, Einstein abandoned this particular attempt to undermine the uncertainty principle. It will be useful to recap Bohr's argument, though it is not necessary to understand it in order to see the point of this paper. (In other words, readers can now skip to Section 3 if they wish.)

Bohr's argument begins with a proposal for measuring the box's weight change after the photon's exit: add a weight to the bottom of the box in order to bring the pointer back to zero. Bohr's argument presupposes the uncertainty principle with respect to position and momentum. So while it is possible to bring the pointer's position back to zero to within any desired degree of accuracy, Δq, this quantity will imply a minimum latitude, Δp, in the uncertainty of the box's momentum.

$$\Delta p \approx h/\Delta q$$

The total impulse (change in momentum $\Delta(v \times m)$) of the box during the weighing procedure is this,

$$I = T \times g \times \Delta m$$

where T is the duration of the weighing procedure and g is the gravitational constant. Clearly, Δp (the uncertainty in the box's momentum) will be smaller than the box's total momentum change (impulse).

$$\Delta p \approx h/\Delta q < T \times g \times \Delta m \tag{1}$$

At this point, Bohr appeals to Einstein's own theory of general relativity. According to the red shift formula, a clock displaced by Δq in the direction of a gravitational force implies a change, Δt, in the clock's reading.

$$\Delta t = T \times g \times \Delta q/c^2 \tag{2}$$

Now come some very simple manipulations of formulas (1) and (2). We can isolate T in formula (2) on the left side of the equation.

$$T = (\Delta t \times c^2) / (g \times \Delta q) \tag{3}$$

Replace T in formula (1) with the right side of (3).

$$h/\Delta q < (\Delta t \times c^2 \times g \times \Delta m) / (g \times \Delta q)$$

After canceling, we get the following.

$$h < \Delta t \times \Delta m \times c^2$$

Given that $E = mc^2$ and hence that $\Delta E = \Delta m \times c^2$, Bohr derives Heisenberg's uncertainty principle.

$$h < \Delta t \times \Delta E$$

So in the clock-in-the-box thought experiment, there is an irreducible limit to the accuracy to which the conjugate variables *time* and *energy* can be measured. This is the denial of the result Einstein had derived from the thought experiment the previous evening. Bohr's argument convinced Einstein that the clock-in-the-box thought experiment failed. It did not show that a photon's energy at a particular time could be measured to within any arbitrary degree of accuracy.

3. Why the Clock-in-the-Box Thought Experiment Cannot Be a T-Argument. This episode raises what I believe is an insuperable problem for any attempt to identify the clock-in-the-box thought experiment with a t-argument. The problem is that Bohr and Einstein were analyzing one thought experiment, but they were proposing two different t-arguments. Therefore, the clock-in-the-box thought experiment cannot be a t-argument. To make this argument stick, let us begin with a prosaic claim: Both real and thought experiments can be, and sometimes are, repeated. For this to be so, it must be possible for there to be different tokens of one experiment-type. If thought experiments are t-arguments, then thought experiment-types are t-argument-types. The argument view of thought experiments is committed to the following thesis.

> (A) Two tokens of a thought experiment are tokens of the same thought experiment-type if and only if they are tokens of the same t-argument type.

The problem is that in the clock-in-the-box episode, A is false. The t-arguments proposed by Bohr and Einstein were not type-identical, but the thought experiments they proposed were type-identical. If this is correct, then the argument view of thought experiments is false. In order to elude this objection, the defender of the argument view must show that one and only one of the following claims is true:

1. The t-arguments Einstein and Bohr proposed were type-identical.
2. Einstein's thought experiment was not type-identical to Bohr's.

If both claims are false, then thought experiments cannot be t-arguments. Let us consider each claim in turn.

Were the t-arguments proposed by Einstein and Bohr type-identical? One reason to think they were not is that their conclusions are contradictory. But as Sorensen has noted, if the argument view is to account for the historical trajectory of thought experiments, the identity conditions on t-argument types must be somewhat "lenient" (1992, 163). So the proponent of the argument view might suggest that t-arguments can be type-identical even if their conclusions are contradictory. Indeed, it is always open to a proponent of the argument view of thought experiments to jigger the identity conditions on t-argument types so that thesis A (i.e., same thought experiment type iff same t-argument type) is saved. Such jiggering might bruise our intuitions about when arguments are type-identical. But if the view has epistemic benefits, they might be worth the cost of injured intuitions. The problem is that this version of the argument view brings serious epistemic costs. The assumption that the arguments Bohr and Einstein proposed were type-identical makes a muddle of the episode. Consider what happened. At first, Einstein proposed a t-argument that Bohr did not accept. This t-argument appeared to threaten the uncertainty principle. Later, Bohr proposed a t-argument that forced Einstein to disown his original t-argument. Here is what one must say if one believes that Einstein and Bohr were presenting tokens of the same t-argument-type: Bohr did not accept Einstein's t-argument, and so he presented Einstein with a token of that same argument; and then when faced with a token of his own t-argument, Einstein proceeded to disavow that very argument. The problem here is not that our intuitions about what counts as the "same argument" are offended. The problem is that when we put this view of thought experiments to work, its account of the clock-in-the-box episode seriously misrepresents what happened.

Let us now turn to the second claim. Were the thought experiments proposed by Einstein and Bohr tokens of the same type? Note that while they were very similar, they were also importantly different. The defender of the argument view might hold that these differences were sufficient to entail that the thought experiments were not tokens of the same type after all. Of course, this is only the faintest sketch of a line of defense. Even so, it will not work, or so I shall argue.

Thought experiments, like real experiments, can be repeated. And in order to repeat an experiment, it is not necessary (or even possible) to duplicate the original in all its details. In fact, if one thinks that an experiment has been botched, it would be folly to try to duplicate it, mistake and all. This was Bohr's predicament. He suspected that Einstein had botched the clock-in-the-box thought experiment, but he couldn't see how right away. Bohr's "triumph" came because he was able to show that *Einstein's thought experiment* did not have the result Einstein thought it had. This understanding of the clock-in-the-box episode makes sense only if we suppose that Bohr replicated Einstein's thought experiment, i.e., that they were tokens of the same thought experiment-type.

To see this more clearly, suppose that they really were different thought experiments (i.e., different types). If Bohr had not replicated the clock-in-the-box thought experiment, then he could not possibly have done that very thought experiment properly. And so Einstein could have rightly accused Bohr of changing the subject. Einstein would have had absolutely no reason to abandon the clock-in-the-box critique of the uncertainty principle. But of course, Bohr's thought experiment did convince Einstein, as well as the rest of the physics community, that the first time around, Einstein had erred. The only way to make sense of this is to suppose that Bohr had repeated—though not duplicated, warts and all—the original clock-in-the-box thought experiment.

In the clock-in-the-box episode, there are tokens of two t-argument-types but tokens of just one thought experiment-type. Thesis A is false. Since we have two different arguments but just one thought experiment, the thought experiment cannot be the arguments. Episodes like this one are not rare in the history of science. People often disagree about the results of thought experiments. Any attempt to identify thought experiments with t-arguments will lead to false and misleading characterizations of such episodes. Thought experiments are not arguments.

REFERENCES

Bishop, Michael (1998), "An Epistemological Role for Thought Experiments", in Niall Shanks (ed.), *Idealization IX: Idealization in Contemporary Physics.* Poznan Studies in the Philosophy of the Sciences and the Humanities, vol. 63. Amsterdam/Atlanta, GA: Rodopi, 19–33.

Bohr, Niels (1949), "Discussion with Einstein on Epistemological Problems in Atomic Physics", in P. Schilpp (ed.), *Albert Einstein: Philosopher-Scientist.* Evanston, IL: The Library of Living Philosophers, 199–242.

Forge, John (1991), "Thought Experiments in the Philosophy of Physical Science", in Horowitz and Massey 1991, 209–222.

Horowitz, Tamara and Gerald J. Massey (eds.) (1991), *Thought Experiments in Science and Philosophy.* Savage, MD: Rowman & Littlefield Publishers

Irvine, Andrew D. (1991), "On the Nature of Thought Experiments in Scientific Reasoning", in Horowitz and Massey 1991, 149–166.

Lennox, James (1991), "Darwinian Thought Experiments: A Function for Just-So Stories", in Horowitz and Massey 1991, 223–246.

Nersessian, Nancy (1993), "Thought Experimenting as Mental Modeling", in D. Hull and M. Forbes (eds.), *PSA 1992,* vol. 2. East Lansing, MI: Philosophy of Science Association, 291–301.

Norton, John (1991), "Thought Experiments in Einstein's Work", in Horowitz and Massey 1991, 129–148.

———. (1996), "Are Thought Experiments Just What You Thought?", *Canadian Journal of Philosophy* 26: 333–366.

Pais, Abraham (1982), *'Subtle is the Lord . . . '*. Oxford: Clarendon Press.

Rescher, Nicholas (1991), "Thought Experimentation in Presocratic Philosophy", in Horowitz and Massey 1991, 31–42.

Sorensen, Roy (1992), *Thought Experiments.* Oxford: Oxford University Press.