

### 3. *Feldman on Sorensen's Thought Experiments*

RICHARD FELDMAN *University of Rochester*

Roy Sorensen's *Thought Experiments*<sup>1</sup> is an original, provocative, and funny discussion of thought experiments. According to Sorensen, "An *experiment* is a procedure for answering or raising a question about the relationship between variables by varying one (or more) of them and tracking any response by the other or others."(p. 186) A *thought experiment* is "an experiment that purports to achieve its aim without the benefit of execution."(p. 205) Through the use of fascinating examples from science Sorensen makes clear that thought experiments play a significant role in science as well as in philosophy, and that their role in science is similar to their role in philosophy.

One of Sorensen's main themes is that all thought experiments are "reducible to two highly specific kinds of paradox."(p. 5) A paradox, he says, "is a small set of individually plausible yet jointly inconsistent propositions."(p. 5) Some thought experiments are "necessity refuters." These are thought experiments that contain a *modal source statement*, such as a definition. The experiments identify a consequence of this source statement and claim that the consequence is absurd. In the successful necessity refuter thought experiment, the modal source statement is rejected as a result.

Gettier's famous thought experiments designed to refute the definition of knowledge as justified true belief illustrates the pattern.<sup>2</sup> One example concerns Smith who justifiably believes (a) that Jones is the man who will get a certain job and that Jones has ten coins in his pocket. He deduces from this (b) that the man who will get the job has ten coins in his pocket. Although (a) is false, Smith is luckily right about (b) because the job-getter coincidentally also has ten coins in his pocket. Sorensen puts this into his preferred format as follows (p. 137):

1. *Modal source statement*: knowledge is justified true belief.
2. *Modal extractor*: if knowledge is justified true belief, then necessarily, if a person has a justified true belief that p, then the person knows that p.
3. *Counterfactual*: if all justified true believers that p have knowledge that p and Smith's justified belief that (b) is only true by luck, then Smith knows that (b) even though it is only true by luck.
4. *Absurdity*: it is impossible for someone to know something when it is only true by luck.
5. *Content possibility*: it is possible for Smith's belief to be justified and only true by luck.

Since these five statements are individually plausible yet inconsistent, they constitute a paradox. Most philosophers, following Gettier, conclude that the

source statement is false. Critics have to reject a different statement, arguing that the definition doesn't have the stated implication, that the content possibility relied upon—that a justified belief could be true by luck or coincidence—isn't really possible, or that the consequence isn't actually absurd. Gettier's examples are so successful because these other responses are so implausible.

The other kind of thought experiment takes the form of a possibility refuter. These are quite similar to necessity refuters, except that the consequence extracted from the modal source statement is a possibility statement.

I agree with many of Sorensen's central theses. I think he's right to reject the general reasons for thinking that thought experiments cannot yield knowledge. I agree that philosophical thought experiments are typically used to refute statements by denying a modal consequence. And I was delighted to see the similarity of scientific and philosophical thought experiments. However, I do have a few questions about some of the details of Sorensen's story. I will take up three of them here.

### **I. On the connection between philosophical thought experiments and paradoxes.**

I believe that philosophers use thought experiments in ways that don't conform to Sorensen's paradoxical structure. For example, the instructor of an epistemology class might ask her students to consider the proposition that knowledge is true belief. The students generally reject the idea, fumbling ineptly with the idea of lucky guesses as counterexamples. With the instructor's help, they eventually formulate one. They've done a thought experiment and we could construct a set of inconsistent statements here, using the proposition that knowledge is true belief as the modal source statement.

Still, this does not amount to one Sorensen's paradoxes. What's missing is the initial plausibility of all the statements required to form the paradox. In the situation described, the statement that knowledge was true belief did not have initial plausibility. Yet, we used a thought experiment to refute the statement. In this case, I think, the thought experiment is used to support a premise in a simple argument such as:

1. If knowledge is true belief, then all lucky guessers know the propositions they luckily guess.
2. Not all lucky guessers ... (based on example)
3. Knowledge isn't true belief.

So, thought experiments don't always involve jointly inconsistent but individually plausible statements. Sometimes they help to make explicit what is wrong with statements that are not initially plausible.

Sometimes thought experiments are used just to establish a possibility. For example, I wonder whether there is ever a case in which the best available action (the one with the greatest overall value) is at least mildly harmful to everyone affected. I eventually come up with a case in which doing the lesser of two (or more) evils is best. By establishing this possibility I do refute the proposition that

it is necessary that there is no such case. So, the case could be reconstructed as a necessity refuter. But doing so complicates matters unnecessarily. The thought experiment shows that a certain sort of situation is possible. No statement implying the denial of that possibility is taken to be initially credible. Again, we don't have a paradox in Sorensen's sense.

Thus, I think that Sorensen's over-emphasizes the paradoxical nature of the cases involving thought experiments.

## **II. Executing Experiments**

I believe that there may be a difference in the value of actually executing philosophical and scientific thought experiments. Many philosophical thought experiments involve situations that we could set up. We could provide you with evidence justifying the belief that Jones owns a Ford and, without telling you, send Brown to Barcelona. We've thereby set up the conditions required for one of Gettier's proposed counterexamples to the justified true belief analysis of knowledge.<sup>3</sup> We also could arrange to have two people stranded on an island and have one make a promise to the other just before the second dies, thereby setting up the deathbed promise case that is supposed to be a problem for utilitarians. And so on, for many more examples. While these ideas may provide the basis for applications for travel grants, executing these experiments would be of absolutely no philosophical value. We would know nothing more about whether the case is a case of knowledge or whether keeping the promise was right. I conjecture that it is never the case that "executing" a philosophical thought experiment would be useful. In contrast, performing scientific thought experiments often, perhaps always, would be of some value, even if many of them could not in practice be executed.

## **III. Thought Experiments and Fallacies**

In Chapter 10 Sorensen discusses "fallacious thought experiments." These are thought experiments in which the experimenter commits some characteristic error in reasoning. Sorensen identifies the following kinds of fallacious thought experiments: missupposition, including oversupposing and undersupposing; perspectival illusions; framing effects, biases of various kinds; jumping the is/ought gap; overweighing negative thought experiments; the additive fallacy; and the blindspot fallacy.

Elsewhere, Sorensen says, "The adequacy of a classification system is more a question of efficiency and suggestiveness. A good scheme consolidates knowledge in a way that minimizes the demand on your memory and expedites the acquisition of new knowledge by raising helpful leading questions."(p. 132) I agree entirely. However, I doubt that Sorensen's classification of fallacies associated with thought experiments is particularly efficient or suggestive.

I probably should put my cards on the table here. I'm no fan of the informal fallacies in the teaching of reasoning. I find the classification systems confusing, the things identified as fallacies an assorted lot of bad arguments, good

arguments, non-arguments, false beliefs, unjustified beliefs, true beliefs, jokes, analogies, and debating tactics. I fear that teaching this material to students inclines them to play mindless games of "Name That Fallacy" rather than to think seriously about the arguments they encounter. I routinely find myself amazed by the answers textbook writers provide in their teacher's manuals for their exercises on fallacy spotting. I can't imagine what students make of all of it. (Well, I can: not much, in my experience.) So, I think that, as a practical matter, these classification systems are not efficient or suggestive.

I don't know that Sorensen would disagree with much of this. Sorensen says that fallacies are *inference rules*.(p. 254) Sticking to this eliminates as fallacies many of the things routinely so classified. He also includes a section on anti-fallacies, thereby providing an antidote to the excessively negative and uncharitable attitude fostered by focusing on the so-called fallacies. Still, I found his discussion of the fallacies associated with thought experiments unhelpful and confusing. I have the space here to look at only a couple of examples.

One is guilty of the fallacy of oversupposing when one "assumes too much."(p. 257) Sorensen says that Anscombe committed this fallacy when, in the discussion of an example in which a person is given a choice between punishing an innocent person and bringing on a nuclear war, she said, "Why not fake the punishment."(p. 258) Anscombe is apparently guilty of oversupposing because she imports something extra into the example, giving the person an option not originally intended to be there. Perhaps she's made a mistake, but I don't see that she's used any bad inference rule in this case, so I don't see what the *fallacy* is supposed to be.

A few pages later Sorensen explains the fallacies associated with perspectival illusions. In an effort to reject mechanistic theories of mind, Leibniz says that if a machine could produce thinking, then we can imagine an enlarged version of the machine, large enough for us to enter inside. But, he claims, all we would observe are "parts which push and move each other, and never anything that could explain perception."<sup>4</sup> Sorensen comments that the "counterfactual fuelling this thought experiment is 'If there were thinking machines and they were enlarged for inspection, then the inspectors would observe mental things responsible for the machine's thoughts'. However, this counterfactual illicitly assumes that properties of the whole must be possessed by the parts."(pp. 260-1) Again, I'm unsure of the fallacious inference Leibniz is supposed to have made.

Moreover, I don't see why Leibniz is guilty of a perspectival illusion rather than of oversupposing. Didn't he "suppose too much" about what we'd find in an enlarged thinking machine? I suppose that he could be guilty of both fallacies, but then I'm less clear on the value of the classifications.

In the section on biases, Sorensen describes biases in favor of continuity. These are a subcategory of literary biases. He notes that:

The audience of Star Trek stories believes that the characters travel by teletransportation. Captain Kirk enters the chamber, is deconstituted, and is then reconstituted in some distant place. Most philosophers working on personal identity think this is overly optimistic. They say Kirk is destroyed when deconstituted and then a replica is created at the "destination" point. The

philosophers deny that any travel takes place because no one survives such a disruptive process. (p. 264)

The audience, he suggests, is misled by a bias in favor of continuity, whereby they assume character continuity in a story. Yet again, I don't know what the *fallacy* is supposed to be here.

Moreover, and more importantly, I don't see what is gained by this name calling. Some people who think about teletransportation cases have the intuition that a person survives teletransportation. Others don't. Reflecting carefully on the case may help clarify matters. Maybe there's a "bias in favor of continuity" that leads some to think, mistakenly, that we would survive in such cases. Or, maybe there is a "bias against disruptive survival" that leads some philosophers to think, mistakenly, that we cannot survive a disruptive process like teletransportation. The merits of the case are what matter, and the names are of little value.

As I said at the beginning of this section, in agreement with Sorensen, the value of a classification is a function of its usefulness. I don't see that this sort of classification of thought-experimental fallacies has a great deal of value. The individual examples may well be instructive. But I don't see much value in the system.

In general, I believe that Sorensen's book is an excellent discussion of thought experiments. I have raised here only a few minor points of disagreement, all perhaps stemming from Sorensen's effort to regiment and systematize thought experiments more than the subject matter will bear.

## Notes

<sup>1</sup>Oxford University Press, Oxford, 1992.

<sup>2</sup>"Is Justified True Belief Knowledge?" *Analysis* 23 (1963): 121-123.

<sup>3</sup>See "Is Justified True Belief Knowledge?"

<sup>4</sup>Leibniz, *Monadology*, trans. Schrecker, Bobbs-Merrill, New York, 1965. See sec. 17.

*RICHARD FELDMAN*  
*UNIVERSITY OF ROCHESTER*  
*ROCHESTER, NY 14627*

