

## 4. Sorensen's Reply to Bunzl and Feldman

ROY A. SORENSEN *New York University*

Both Martin Bunzl and Richard Feldman voice reservations about the logical schemas presented in the chapter entitled "The Logical Structure of Thought Experiments". Both suspect that the schema oversophisticates some thought experiments. Martin Bunzl thinks that some of the modal operators in the necessity refuter scheme are gratuitous. Richard Feldman suggests that some thought experiments are simply counterexamples that support the negative premise of a *modus tollens*.

Since I presented the classification system in a pragmatic spirit, I need not deny that thought experiments can be formulated in other ways. Indeed, from logic we know that one argument can instantiate infinitely many forms. Most of these are invalid forms. When a valid argument instantiates more than one valid form, we generally rest content with the simplest one. "If it doesn't itch, don't scratch." This is a good motto but its application will vary with different interpretive goals. A taxonomist of thought experiments will seek a uniform translation of all thought experiments. Such global hermeneutics will often lead him to attribute more structure to a thought experiment than your local hermeneuticist with only a passing interest in the same thought experiment. Wide-ranging classificatory ambition may also lead the global interpreter to assign less structure than an historian of science doing an intensive case study. Our second motto should be "Different strokes for different folks".

So my classification scheme is compatible with the fact that some thought experiments can be formulated in other ways. My taxonomy is only threatened by thought experiments that do not instantiate either of the necessity refuter schema or the possibility refuter schema. Given my pragmatic attitude towards classification, the global nature of my interpretive enterprise, plus the well known vagaries of interpretation, most thought experiments should fit even if they occasionally pinch at the heel. After all, I admit that the scheme is a *regimentation* of thought experiments, that it puts them in an artificial state like butterflies propped on pins. A study of *gedankenexperimente* in this posture yields copious benefits: the exercise explains the persuasiveness of thought experiments as inconsistency reduction, it accounts for some of the psychological aspects of thought experiments in terms of dissonance, it lays out the possible moves in response to a thought experiment, it implicitly constitutes a special theory of thought experiment fallacies, and it integrates well with the conflict vagueness analysis of Kuhnian thought experiments.

So I am submitting the classification system in the same spirit as the proponents of nomenclature reforms. A nomenclature system generally carries

factual commitments plus a world view such as Linneaus' commitment to the existence certain organisms plus his vision of life as organized in a hierarchy: individual, species, genus, phylum, kingdom. However, a system of nomenclature is mainly designed to have virtues as a convention. Thus the system is unlikely to be adopted by those who are skeptical of the implicit factual and metaphysical claims and by those who view the system as a clumsy convention. I think my taxonomy does well on the conventional side. It purports to be a neat and tidy way of summarizing a multitude of facts about thought experiments and projects a modally robust outlook. Under my view, even modest descriptive claims have rich modal implications. For example, 'This page is a piece of paper' implies counterfactuals such as 'If the temperature of this page were to rise to 400 degrees Centigrade, it would burn'. This counterfactual has to be analyzed modally, perhaps even in terms of possible worlds.

So Richard Feldman is right that the source statement must be a modal claim in the sense that it must have modal consequences. However, it need not be a modal claim in the sense being *about* necessity. The conjunction of a contingency and a necessity is always a contingency. My list of source statements is heterogeneous: "semantic theses (definitions, synonymy claims, entailment theses), testability theses (unverifiability, unfalsifiability, undetectability), feasibility claims, law statements, disposition and intention attributions, validity verdicts, and clusters of these—theories." (135) However, the source statement does not itself need to be a necessary truth. After all, reportive definitions register contingent fact about usage.

The source statement can also be quite particular such as Feldman's hypothesis that his students will enjoy and benefit from a discussion of thought experiments. Feldman agrees that I might be able to force this thought experiment into one of my favored schemas but thinks it would be misleading to do so. I agree that there are several ways in which that exercise would be misleading. For instance, the formalization would falsely suggest a larger lesson to be learned from the modest thought experiment. However, insofar as it is a thought experiment, then the formalization seems innocent. Feldman's case is underdescribed. For example, there is no mention of the extraction of an absurdity. As it stands, the alleged thought experiment only appears to be a prediction supported by mental imagery. Or it might be an application of pedagogical *verstehen* in which the teacher hypothetically reverses role with a student. In *Thought Experiments* I distinguish thought experiment from lots of close cousins such as imagistic reasoning (209-210), thought simulation, models, reenactments, (225-8) internal psychological experiment (208-209), imaginary experiments, fictional experiments, and other look-alikes.

Feldman may be right about me overdoing the paradoxical nature of thought experiments. I have a weakness for this model. About ten years ago I gave a job talk entitled "Philosophy as Paradoxology". The thesis was that paradoxes were the subject matter of philosophy. During discussion I was presented with paradoxes of home insulation, automotive repair, etc. The result was an embarrassingly complete consensus that philosophy is not paradoxology. (But I

still got the job!) Feldman's counterexample is directed against my thesis that all thought experiments are special types of paradoxes. In particular, he thinks that the initial plausibility condition is not required. For sometimes we construct hypothetical counterexamples to principles that we find implausible such as 'Knowledge is true belief'.

I think Feldman has a successful counterexample to the thesis that the members of the paradox must be plausible to the thought experimenter. So instead of relativizing 'plausible' in this natural direction, I must fish for fancier relations. To keep the revision principled, I'll draw an analogy with other phenomena that are partially defined in terms of a propositional attitude. A surprise party is a party that *purports* to be unexpected by the honoree. The celebration can still be a surprise party if he actually expects it and even if each person at the party knows the secret was broken. However, if it becomes common knowledge (so that everybody knows that everybody knows), then the party throwers cannot be *intending* the event as a surprise party. In the case of thought experiment, I now say that each member of the paradox must be intended as being plausible to someone who *would* find the source statement plausible. So when the thought experimenter himself does not find the source statement plausible he must direct the paradox to others who do—perhaps even a *hypothetical* proponent of the source statement.

Generally, thought experimenters construct their scenarios for their own edification. But after dwelling on Feldman's case, one can identify others. To my chagrin, some even appear in the very chapter laying out the classification scheme. For example, I illustrate the possibility refuter scheme with a thought experiment that has a source statement which is a "dead issue" in our culture (153-4):

- I. Polytheism: more than one god exists.
- II. If polytheism is correct then two omnipotent beings can co-exist (because gods are perfect and perfection implies unlimited power).
- III. However, if there were two omnipotent beings and they were to have a shoving match, an irresistible force would meet an immovable object. (For one would have the power to move anything and the other would have the power to resist any movement.)
- IV. But it is impossible for an irresistible force to meet an immovable object: either can exist but it is contradictory to say they co-exist.
- V. If it is possible for two omnipotent beings to exist then it is possible for them to have a shoving match.

Although we atheists and monotheists are not tempted by polytheism, we want the rational backing that comes from refuting the best polytheist we can imagine. And we get it.

This gives me a quick response to Martin Bunzl's skepticism about possibility refuters. The thought experiment shows how polytheism entails a contradiction. Contradictions don't hold in any possible world. That answers Professor Bunzl's query about "how a thought experiment can rule out a state of affairs in every possible world." It should also be noted that sometimes the

possibility refuter is only concerned with a certain class of possible worlds—like those in which Newton's physics holds.

Feldman suggests that scientific thought experiments differ from philosophical ones in that the execution of the experimental design provides significant new evidence for scientific thought experiments but not for philosophical ones. But consider hypothetical counterexamples to the principle of sufficient reason. They refute the principle that there must be a reason for each event by showing how it implies absurd equilibria. For example, suppose that a perfectly uniform hair is pulled at both ends by equal forces. Since there is no more reason for the hair to break in one place rather than the other, the principle of sufficient reason has the counterintuitive consequence that the hair remains intact regardless of how strongly it is pulled. But let us suppose that an empirical minded barber finds a promising hair and executes the experiment. To everyone's astonishment, the hair does remain intact! This would undermine the absurdity claim of the original thought experiment. (For an example of an execution sensitive thought experiment outside this genre, see the nurse example on page 198). In contrast, actually executing Einstein's lightning bolt and train thought experiment (discussed on 178) would add nothing to this demonstration of the relativity of 'simultaneous'. So the relevance of execution is not an asymmetry between scientific and philosophical thought experiments.

I share some of Professor Feldman's worries about fallacies. His textbook on the analysis of arguments illustrates how informal logic can be taught by concentrating solely on what makes an argument good. When I first started teaching critical thinking I was dismayed to find that my performance at Name that Fallacy was not much better than my students. After a while, I became test-wise and improved my hit rate by concentrating on what the textbook author would count as a fallacy. This is an unhappy situation but not so desperate as to warrant total defenestration of fallacies. The fallacy categories in textbooks are useful in the way that folk categories for diseases are useful. Physicians are dismayed by the crudity of *cold*, *whiplash*, *bug*, *fits*, but acknowledge that this terminology is a useful first step in securing an accurate diagnosis. Physicians also complain about the unsystematic character of professional nosology (disease taxonomy). But none have reacted by writing medical textbooks that concentrate solely on well-functioning bodies. Philosophy of medicine shows that 'disease' is about as conceptually problematic as the logician's 'fallacy'. So the physician's greater confidence in negative thinking is best explained in terms of the greater reliability and validity of extant classification systems. There are no principled obstacles to a future system of fallacies achieving a level of success comparable to current nosologies.

But for now, honesty is the best policy. Students should be told that the theory of fallacies is in an ill-developed state and that there is unclarity as to which aspects of the subject belong to logic and which to psychology. The sophomoric tendency to become overly dismissive of other's reasoning can be countered with the concept of an anti-fallacy (a good argument that looks like a bad one) plus the lore of argument interpretation. Psychologists of reasoning have

cast doubt on the topic neutrality of fallacies. Instead of being governed by a few overarching principles, our reasoning appears to proceed via thousands of low level heuristics. Formal logicians naturally gravitate towards macro-structural theories of fallacies because they are familiar with Euclid-style logical systems that economize on inference rules.

A psychologically realistic theory of fallacy should be micro-structural and domain specific. That's why I modeled my account of thought experiment fallacies on textbooks devoted to occupational diseases. Although no fallacy is unique to thought experiment (just as no disease is unique to miners), they will occur in unusual frequencies or with unusual effect.

Actually, my book has two accounts of fallacy. The first, logically oriented account is a spin off of my classification system. The second, more psychological version is reserved for the final chapter. That's the one Feldman finds reminiscent of the disorganized fallacy chapters in critical thinking texts. I don't agree with Feldman's particular cases. Peter Unger's accusation of literary bias was directed against Robert Nozick's closest continuer account of personal identity in *Philosophical Explanations*. Nozick appeals to our comprehension of teletransportation in science fiction as grounds for believing it is logically possible. As for the Leibniz case, I think over-supposition can be ruled out because Leibniz does not fall to the temptation "to annihilate obstacles through a sheer act of will." (257) It is more plausible to ascribe a perspectival illusion because he is, after all, re-scaling a specimen in a way that sets him up for the fallacy of composition. Nevertheless, Feldman may be right on inductive grounds. Just about everyone who takes a stab at fallacies creates a bloody mess, therefore I may well have made a bloody mess of it!

Martin Bunzl sharply distinguishes between the experiment and the use to which it is put. He appears to agree that an experiment can be put to many of the uses associated with thought experiment but he denies that this builds much of an analogy. After all, a screwdriver has many of the uses that we associate with a knife but this does not show that screwdrivers are knives—or even limiting cases of knives. However, functional analogy is stronger when the "objects" under comparison are not objects with their own autonomous properties. Experiments and thought experiments are abstract, hence they are largely what they do. Experiments provide fresh information but they also do things that we associate with thought experiment. Compare with 'Checks are not money; they only have many of the uses of money'.

Martin Bunzl is skeptical about "philosophical appeals to evolutionary arguments . . . they are just too easy!". This has the same over-dismissiveness of those who complain that anything can be proved by statistics. However, Bunzl does think that my biological account has an explanatory gap. In particular, he alleges that I cannot account for the contrast between the straightforwardness of the Gettier counterexample as opposed to the instability and indecisiveness of personal identity transfer cases. He thinks that "imaginative" thought experiments tend to spin out of control in a way that "inferential" ones do not. Professor Bunzl's forthcoming work on thought experiment may make that distinction

specific enough to afford a precise reply. But for now, I'll launch two general remarks. The first is that some executed experiments also spin out of control. This is especially noticeable when a "crucial experiment" peters out into baroque stalemate. For example, the question of whether light is composed of particles or waves initially seemed simply resolvable. The basic experimental design is to pour light on a very mobile object. If light is made of particles, then one should be able to move the object by the pressure exerted by light particles. This hypothesis led to long series of experiments in which the particle theory and the wave theory see-sawed for ascendancy. Early experiments were undermined by the discovery of nuisance variables (such as air currents formed from the heat created by the light beam) and then by refinements of the particle theory and the wave theory that undermined the apparent difference in what they predicted. Experiments that spin out of control are frequently labeled pseudo-scientific (like the notorious polywater and N-ray studies) but John Worrall defends the pressure of light experiment as an edifying invitations to theoretical precisification.<sup>1</sup>

My second general point is directed against the idea that the thought experiments in the personal identity literature are more likely to spin out of control than the ones in epistemology. In reply, first note that these are both semantic thought experiments, one for the meaning of '*A* knows that *p*' and the other for '*x* is the same person as *y*'. Semantic thought experiments are supported by our linguistic ability. My biological model harmonizes Noam Chomsky's thesis that human beings have a species-specific language organ. Linguistic intuitions will be affected by biological limitations of attention and memory. But we should not expect any special problems with personal identity cases. And indeed there are personal identity thought experiments that are just as decisive as Gettier's. An example is Thomas Reid's thought experiment targeting John Locke's principle that personal identity requires direct memory links with one's own past. Suppose there was a boy who was flogged for robbing an orchard. He grew to be a brave officer who took a standard from an enemy in his first campaign, and finally became a general late in life.<sup>2</sup> As a brave officer, he remembers being flogged. As an old general, he loses the memory but still remembers taking the standard. When we apply Locke's principle we get a failure of transitivity: the old general is the same person as the brave officer and the brave officer is the same person as the boy, but the old general is not the same person as the boy! Even those sympathetic to Locke's analysis agree that Reid's thought experiment requires at least a revision of the memory principle.

Martin Bunzl thinks that thought experiments have a far more limited role than I do. Bunzl illustrates this deflationism with some doubts about the role of thought experiments in Einstein's work. I'm no Albert Einstein scholar. But my account finds the popular picture of Einstein congenial. Einstein spoke highly of thought experiments and it is that aspect of his work that struck him and his fellow physicists as amongst the most interesting. Einstein cooperated with the gestalt psychologist Max Wertheimer in a study of his thought patterns. Einstein was intrigued by imagistic thinking and characterized discursive thinking as secondary, almost as an imposition needed to disseminate his thinking. Although

a fine mathematician, Einstein was not as skilled at mathematical physics as many of his colleagues. Einstein did not take particular pride in mathematical aspects of his work. His emphasis on geometric imagery is typical. The originator of our conception of electric and magnetic fields, Michael Faraday, had no mathematical training. He visualized lines of force as narrow tubes curving through space. The mathematically brilliant James Clerk Maxwell set down the equations for Faraday's lines of force but only after elaborating with imaginary models of sheets and fluids. Similar stories have been related for Ernest Lawrence's conception of the cyclotron, James Watson and Francis Crick's discovery of the DNA double helix, and appropriately enough, Roger Shepard's seminal studies of mental rotation.

Although Bunzl and the logical positivists use the reverential term "real work" to cover the math, Einstein and I think most top physicists did not think of it that way. My impression is that their attitude is similar to the philosopher's attitude towards formalization of intuitive arguments. Although philosophers are impressed by the development of new logical methods (Saul Kripke on modal logic, Gerald Massey on multi-grade predicates, Donald Davidson on adverbs), they do not regard corroborating proofs as the most interesting or valuable part of their work. Indeed, philosophers typically relegate formalizations to an appendix. Thomas Edison may have been right about invention being 1% inspiration and 99% perspiration. But we still cherish the inspiration. Anyone can perspire.

We must be wary of real chauvinism—the question-begging marking of loyalties with assurer words such as 'genuine', 'authentic', and 'real'. The 'real' in 'real work' or 'real experiment' is just a device of emphasis. Proof: Every real  $F$  is an  $F$ , Every  $F$  is a real  $F$ , therefore, 'real' is semantically redundant (though it is not pragmatically redundant).<sup>3</sup> Talk of real science is just as tricky as talk of pseudo-science!

## Notes

<sup>1</sup>John Worrall "The Pressure of Light: The Strange Case of the Vacillating 'Crucial Experiment'" *Studies in the History and Philosophy of Science* 13 (1982): 133-71. I draw on his account on pages 232-33 of *Thought Experiments*.

<sup>2</sup>Reid presents the counterexample in the third essay of *Essays on the Intellectual Powers of Man* 1785.

<sup>3</sup>I defend a redundancy theory of 'real' in the second chapter of *Pseudo-Problems* (London: Routledge, 1993).

ROY SORENSEN  
DEPARTMENT OF PHILOSOPHY  
NEW YORK UNIVERSITY  
100 WASHINGTON SQUARE EAST  
NEW YORK, NY 10003-6688

