Sharvy's Lucy and Benjamin puzzle

Thomas Forster

July 12, 2010

Abstract

Sharvy's puzzle concerns a situation in which common knowledge of two parties is obtained by repeated observation each of the other, no fixed point being reached in finite time. Can a fixed point be reached?

Keywords: Bourbaki-Witt fixed point theorem; Incorrigible knowledge; Supertask; Common knowledge; Liar paradox.

et l'on revient toujours
à ses premiers amours.

Charles-Guillaume Etienne

Many years ago Richard Sharvy showed me a wonderful puzzle which he had written up for a brief article in Philosophia: [4]. I wrote a discussion of it which was accepted by Philosophia but I withdrew it at the last minute: something in me recognised that the offering was not completely baked. It has now been on the back burner for 20 years and is slightly more presentable as a result.

I was spurred to look again at my notes on it by making the discovery—on a visit to *Logic and Philosophy of Science* in UC Irvine—of Sharvy's copy of the fascicule of the *Bulletin of the American Mathematical Society* 1963 with Saunders MacLane's famous seminal article on Category theory in it. I should have known that Sharvy had worked at UCI, but I didn't. The copy has annotations in Sharvy's unmistakeable hand. I was rather taken aback by this: I hadn't realised that Sharvy knew that much mathematics. I am hoping to exploit this as an excuse for the amount of mathematical gadgetry I am about to employ in what follows below!

Sharvy died on July 1st 1988, and it seems appropriate to commemorate the 20th anniversary of his passing by having another look at the puzzle he bequeathed us. Here it is.[1]

[1] I am endebted to Peter Johnstone for finding a mistake in the published version. I claimed that the common fixed point whose existence is proved on page 5 is the least fixed point: it might not be. No use was made of the claim that it was the least fixed point.

The Lucy and Benjamin Puzzle

Two people—Lucy and Benjamin[2]—are placed (each in ignorance of the other's presence there) on a bush-clad island where there is a blackboard and a large supply of chalk. There then ensues a (potentially) infinite sequence of events. At stage i (i even) Lucy writes on the blackboard, below what is already there "Lucy was here" and signs it "Lucy", and is observed—unbeknownst to her—by Benjamin, who is hiding in the shrubbery. Dually for i odd: Benjamin writes on the blackboard, below what is already there "Benjamin was here" and signs it "Benjamin", and is observed—unbeknownst to him—by Lucy, who is hiding in the shrubbery. Now let H_0 be

Lucy and Benjamin are both on the island (H_0)

and thereafter, for $i \geq 1$

Benjamin knows H_{2i} (H_{2i+1})

and

Lucy knows H_{2i-1} (H_{2i})

That is to say, we have a sequence of propositions which begins

Lucy and Benjamin are both on the island (H_0)

Benjamin knows (H_0) (H_1)

[. . . because he sees her writing for her first time. . .]

Lucy knows (H_1) (H_2)

[. . . because she sees him writing for his first time after reading what she had written . . .]

Benjamin knows (H_2) (H_3)

[. . . because he sees her writing for her second time after reading what he had written. . .]

Lucy knows (H_3) (H_4)

[2] I'm not sure why the protagonists bear the names they do. Sharvy had a son called 'Benjamin'; there was a Lucy in his life too, who was his partner when he came to Auckland once. She was a budding architect who worked on the project of saving the Old Custom House on Quay Street (an undistinguished Victorian building of the kind that passes for *colonial heritage* in Auckland).

[. . . because she sees him writing for his second time after reading what she had written. . .]

etc.

Evidently H_i becomes true at stage i and not before. Sharvy says: what happens should their eyes meet? It seems that they would come to know some very simple H that implies all the H_i but is not equivalent to the conjunction of any finite number of them. Sharvy challenges us to formulate H.[3]

Discussion

When I first started thinking about this problem many years ago I assumed that the proposition H that they both come to know would be not only something that they both know but would be such that each knows that the other knows it, and knows that the other knows that . . . and so on. In other words, it would be a case of what in the Artificial Intelligence/Computer Science literature is called *Common Knowledge*:

Lucy and Benjamin are both on the island and both know H_∞. (H_∞)

This possibility is raised also by Harman in [5] p 150. Philosophers *love* self-reference, but unfortunately a satisfactory treatment of it demands more mathematics than is to most tastes. Although there is nothing in Sharvy's challenge that actually *requires* self-reference it turns out that seeking a self-referential solution does not unduly further complicate a situation that is already complicated. But let us ignore self-reference for the moment: Sharvy is merely after something that they both know, which could be a much weaker assertion than H_∞. So let us consider the infinite conjunction:

$$\bigwedge_{i\in\mathbb{N}} H_i \qquad\qquad H_\omega$$

Might H_ω be the H that Sharvy is after? How might they both come to know H_ω? This would seem to be an example of what the literature nowadays calls a *supertask*. See [3] for a treatment of some supertask arguments and a good bibliography. (See also [6].) Their eyes meet for—say—one second. In the first half-second Benjamin realises that Lucy is in the forest so at the end of that half-second H_1 becomes true; in the next 250 milliseconds Lucy realises that Benjamin knows she's in the forest and at that point H_2 becomes true. And so on.

The typical finding with supertasks is that there comes a point in the sequence of subtasks after which all subsequent subtasks are physically infeasible. In some of the supertasks in the literature that is not a crippling problem since the point for which the supertask was being invoked is a point about *logical*

[3] I have doctored Sharvy's original version—in which they were merely in a forest—to isolate them on an island so that neither can escape. If they both know there is no escape from the island then certain distracting complications can be ignored.

possibility: there one thinks for example of the supertasks in [3] that concern the logical possibility of there being physical machines that can solve the halting problem. However it most definitely is a problem here.

There is also the question of whether or not Lucy's (or Benjamin's) knowledge of all the H_i is the same as knowledge of the infinite conjunction H_ω. Does knowledge distribute over conjunction? Let's flag this as an assumption:

Knowledge of a conjunction is simply knowledge of all its conjuncts. (K)

(K) looks all right, but notice that its dual—the assertion that knowledge of a disjunction is knowledge of at least one of its disjuncts—is definitely not all right with people who believe the law of excluded middle: if they are right then we all (presumably) know $A \vee \neg A$ without knowing either A or $\neg A$.

Is their coming to know H_ω really a supertask? The way the story is told it looks as though H_{2i+1} cannot become true (that is to say, Lucy cannot know H_{2i}) until H_{2i} becomes true. However all that the story establishes is that there is a uniform way in which, for each i, H_j can be true for all $j < i$ without H_i being true. Might there be no lower bound on the time it takes—after H_i becomes true—for H_{i+1} to become true? Can they become true simultaneously?

There is a consideration that suggests that they cannot all become true simultaneously and indeed even suggests that there is a lower bound on the time it takes for H_{i+1} to become true once H_i is true. This is the idea that because of the nature of human nervous systems there is a certain minimum period for which H_i has to be true while the penny drops so that H_{i+1} becomes true. If there is such a minimum period then the task of making H_ω true really is a supertask, and an impossible one at that. If that is the case then H_ω can never be true.

So for us the key question at this stage is whether or not H_i and H_{i+1} can become true at the same time. What kind of proposition p has the feature that Lucy comes to know p simultaneously with p becoming true? Or even that there is no lower bound on the time it takes Lucy to realise that p? The only propositions about which anyone has ever plausibly made claims like this are propositions involving Lucy's internal states. "I believe p" (claimed by Lucy); "*Red-here-now-for-Lucy*"[4] that sort of thing. If there are any propositions that Lucy can come to know simultaneously with their becoming true, or even with an arbitrarily short delay that are *not* of this special self-regarding kind then they are not well-documented, and propositions concerning Benjamin's internal state would seem to be well down the list of plausible candidates. The conclusion seems to be that H_ω cannot be true. Since Sharvy's H must imply every H_i it implies their conjunction, which is H_ω. Since Sharvy's H must become true when they meet and it implies something that can never be true then there can be no such H.

Interestingly, there is a refinement of the Bourbaki-Witt theorem that implies that there really is a coherent proposition H_∞ whose content is that Lucy and Benjamin are both on the island and both believe H_∞. Clearly this proposition

[4]It probably sounds better in Viennese.

will imply H_ω. If I am correct in believing that H_ω cannot be true, then H_∞ cannot be true either. However the fact that such a proposition can be successfully formulated seems worth noting, and the proof is worth spelling out in detail.

A Refinement of the Bourbaki-Witt Theorem

The Bourbaki-Witt theorem ([1], [2]) says that every inflationary function from a chain-complete poset with a bottom element into itself has a fixed point. ($f : P \to P$ is inflationary iff $(\forall p \in P)(p \leq f(p))$. Here we need a slight refinement.

Theorem 1 *Suppose $\langle P, \leq \rangle$ is a chain complete poset*[5] *with a least element $\perp$, and $f : P \to P$ and $g : P \to P$ are inflationary functions with $\perp < f(\perp)$, $\perp < g(\perp)$ satisfying the extra condition:*

$$(\forall x)(g(x) \leq g(f(x))) \tag{1}$$

and

Then f and g have a common fixed point.

Proof:

We observe that $g \cdot f$ is an inflationary function $P \to P$ and—by the usual Bourbaki-Witt result—will have a fixed point, c. Further

$$c \;=\; g \cdot f(c) \;\geq^{(a)}\; g(c) \;\geq^{(b)}\; c$$

so $c = g(c)$. ((a) holds because of 1; (b) holds because g is inflationary.) Also

$$c \;=\; g \cdot f(c) \;\geq^{(a)}\; f(c) \;\geq^{(b)}\; c$$

so $c = f(c)$. (a) holds because g is inflationary; (b) holds because f is inflationary.

Thus c is a fixed point for both f and g. ■

Now we want to apply this to the problem in hand, namely that of establishing that there really is a proposition H_∞ with the desired self-referential properties. It cannot be emphasised too strongly that the mere fact that one can write down things like

Lucy and Benjamin are both on the island and both know H_∞. (H_∞)

[5] I write '$\perp$' here for the bottom element of the poset because this is standard notation in the literature on posets. No identification of the bottom element with the `false` is intended.

does not assure us that what we write down expresses a proposition: the liar paradox is a constant reminder of this, there being no proposition which is true iff it is false.[6] There is thus no *prima facie* guarantee that there is a proposition whose content is that Lucy and Benjamin are both on the island and both know that proposition: each case like this needs to be argued for individually. I propose to use the above modification of the Bourbaki-Witt theorem to show that there really is such a proposition H_∞. Whether or not this proposition is *true* is another matter altogether!

The poset P is going to be the Lindenbaum algebra of all propositions that imply that Lucy and Benjamin are both on the island, $p \leq q$ is going to be q implies p (so the Lindenbaum algebra is upside-down), f is going to be 'Lucy knows that ...' and g is going to be 'Benjamin knows that ...' and $\perp$ is going to be 'Lucy and Benjamin are both on the island'. Both f and g clearly satisfy the special condition (1): if Lucy knows that Benjamin knows that p then she clearly knows p herself.[7] The gadgetry will be of no relevance to us unless the poset of propositions is chain-complete. Is it the case that a nested conjunction of propositions is a proposition? It sounds plausible enough, but when things go wrong later on we will find life easier if this possible error was earlier flagged for our subsequent attention. For the moment let us assume that

Any nested conjunction of propositions is another proposition: (C)

At any rate, given assumption (C) then there really is a proposition H_∞ as above. Let us note parenthetically that on our earlier assumption K (that knowledge distributes over conjunction, so that knowledge of every conjunct of a conjunction guarantees knowledge of the conjunction) then H_∞ is in fact just H_ω. It might be worth spelling this out.

Assume H_ω. If H_ω is true then Lucy knows all the H_i. But then—by (C)—she knows their conjunction, which is H_ω. So H_ω is a proposition p whose content is that Lucy and Benjamin are both on the island and both know p. So it is either H_∞ (which is well-defined by theorem 1) or something possibly even stronger. We reason about Benjamin analogously.

For the other direction observe that H_∞ implies all the H_i, so it implies their conjunction, which is H_ω

Summary

- H_ω is coherent (easy) but false (arguably);
- H_∞ is coherent (hard) and implies H_ω;
- If knowledge distributes over conjunction then H_∞ and H_ω are the same proposition.

[6] It has to be admitted that there are people who contest this.

[7] Notice that neither f nor g satisfy 1 if f is merely "Lucy *believes* that ..." and g is "Benjamin *believes* that ...". This will matter later.

Finally we can minute an observation made by my Canterbury colleague Douglas Campbell. Is there a proposition D_∞ as below?

Lucy and Benjamin are both on the island and both *believe* D_∞: (D_∞)

The significance of this speculation is that—as Campbell points out—D_∞ might very well be true: Lucy and Benjamin might have a simultaneous moment of madness and both decide to believe D_∞—at which point it becomes true! This illustrates clearly the point that there are no *logical* difficulties with H_∞—because any logical difficulties would arise also with D_∞. The difficulties with H_∞ are *epistemic*.

However, as was emphasised above, the Liar paradox is an ever-present reminder that semantics does not routinely meekly obey syntax, and there is no guarantee that formulations like D_∞ bear the meaning they purport to bear, so it is not clear that there is such a proposition. In the case of H_∞ we could use a spiced-up version of Bourbaki-Witt as above. Sadly the same argument cannot be used in this case: although 'Lucy knows p' follows from 'Lucy knows that Benjamin knows that p' the anologous implication from 'Lucy believes that Benjamin believes p' to 'Lucy believes p' is not good. This is the condition (1)-and-(2) above, it is essential to the proof that H_∞ is well defined, and essential use is made of it in the proof that H_∞ exists.

References

[1] N. Bourbaki, Sur le theorème de Zorn, Arch. Math. **2** (1950), 434–437.

[2] E. Witt, Beweisstudien zum Satz von M. Zorn, Math. Nachr. **4** (1951), 434–438.

[3] B. J. Copeland. Accelerating Turing Machines. Minds and Machines **12** (20020 pp 281–301.

[4] Richard Sharvy, An Epistemic Puzzle. Philosophia **4** pp 553–4.

[5] Philosophia **6** fascicule 1, March 1976.

[6] J.M. Thompson. Tasks and Super-tasks. *Analysis* **15** (1954) pp 1–13.

Department of Pure Mathematics and Mathematical Statistics, University of Cambridge, *and*

Department of Philosophy and Religious Studies, University of Canterbury.