Surprise Exam Paradox

PARADOX AND INFINITY
Benjamin Brast-McKie
May 6, 2024

The Exam

Setup: A single surprise exam is announced for next week (9am m, w, or f).

- Let 'E_i' read 'The exam occurs on i' where $i \in \{m, w, f\}$.
- Let '$\mathcal{B}_i(A)$' read 'The students believe A at 9am on i'.
- E_i is a *surprise iff* $E_i \wedge \neg\mathcal{B}_i(E_i)$.
- Let $S = (E_m \wedge \neg\mathcal{B}_m(E_m)) \vee (E_w \wedge \neg\mathcal{B}_w(E_w)) \vee (E_f \wedge \neg\mathcal{B}_f(E_f))$.
- The students believe this announcement S throughout the week.

Closure: If $\mathcal{B}_i(A)$ for all $A \in \Gamma$ and $\Gamma \vdash B$, then $\mathcal{B}_i(B)$.

- We only need limited instances of *Closure* to hold.

Informed: The students learn each day if there is an exam, forming a true belief.

Memory: The students maintain their beliefs from the previous days.

Friday: On Monday 8am, the students reason as follows:

- If E_f, then $\neg E_m$ and $\neg E_w$, so $\mathcal{B}_m(\neg E_m)$ and $\mathcal{B}_w(\neg E_w)$ by *Informed*.
- So $\mathcal{B}_f(\neg E_m)$ and $\mathcal{B}_f(\neg E_w)$ by *Memory*, where $\mathcal{B}_f(S)$ is a premise.
- But $S, \neg E_m, \neg E_w \vdash E_f$, and so $\mathcal{B}_f(E_f)$ by *Closure*.
- Thus E_f is not a surprise, i.e., $\neg\neg\mathcal{B}_f(E_f)$, and so $\neg(E_f \wedge \neg\mathcal{B}_f(E_f))$.
- In this way, the students come to $\mathcal{B}_m(\neg(E_f \wedge \neg\mathcal{B}_f(E_f)))$.
- However, $S, \neg(E_f \wedge \neg\mathcal{B}_f(E_f)) \vdash (E_m \wedge \neg\mathcal{B}_m(E_m)) \vee (E_w \wedge \neg\mathcal{B}_w(E_w))$.
- By *Closure*, $\mathcal{B}_m(S')$ where $S' = (E_m \wedge \neg\mathcal{B}_m(E_m)) \vee (E_w \wedge \neg\mathcal{B}_w(E_w))$.

Wednesday: The students (on Monday 8:05am) turn to reason about Wednesday:

- If E_w, then $\neg E_m$, so $\mathcal{B}_m(\neg E_m)$ by *Informed* and $\mathcal{B}_w(\neg E_m)$ by *Memory*.
- However, $\mathcal{B}_m(S')$ by *Friday*, and so $\mathcal{B}_w(S')$ by *Memory*.
- Since $S', \neg E_m \vdash E_w$, it follows by *Closure* that $\mathcal{B}_w(E_w)$.
- Thus if E_w, then E_w is not a surprise, and so $\neg(E_w \wedge \neg\mathcal{B}_w(E_w))$.
- In this way, the students come to $\mathcal{B}_m(\neg(E_w \wedge \neg\mathcal{B}_w(E_w)))$.
- However, $S', \neg(E_w \wedge \neg\mathcal{B}_w(E_w)) \vdash E_m \wedge \neg\mathcal{B}_m(E_m)$.
- By *Closure*, $\mathcal{B}_m(S'')$ where $S'' = E_m \wedge \neg\mathcal{B}_m(E_m)$.

Monday The students now turn to consider Monday (still on Monday 8:10am):

- $\mathcal{B}_m(S'')$ entails $\mathcal{B}_m(E_m)$ and $\mathcal{B}_m(\neg\mathcal{B}_m(E_m))$.
- $\mathcal{B}_m(E_m) \vdash \mathcal{B}_m(\mathcal{B}_m(E_m))$ leads to believing a contradiction.
- So the students would seem to have reason to reject $\mathcal{B}_m(S'')$.

Moore's Problem

Rain: It is raining (R) but I do not believe that it is raining $\neg\mathcal{B}(R)$.

- Can be true, but can't be asserted (normally).
- Can *assert* either R or $\neg\mathcal{B}(R)$, but not both.
- OK to assert: It is raining but *you* do not believe that it is raining.

Belief Norm: Don't assert what you don't yourself believe (in normal circumstances).

- One could appeal to this norm to infer $\mathcal{B}(R)$ from an assertion of R.
- Similarly, $\mathcal{B}(\neg\mathcal{B}(R))$ can be inferred from an assertion of $\neg\mathcal{B}(R)$.
- Can *believe* R or $\neg\mathcal{B}(R)$, but not both?

Introspection: The following introspection principles have many true instances.

(*Positive*) $\mathcal{B}(A) \vdash \mathcal{B}(\mathcal{B}(A))$. (*Negative*) $\neg\mathcal{B}(A) \vdash \mathcal{B}(\neg\mathcal{B}(A))$.

- Nothing seems to block introspection for $A = E_m$.
- As above, $\mathcal{B}_m(S)$ entails $\mathcal{B}_m(E_m)$ and $\mathcal{B}_m(\neg\mathcal{B}_m(E_m))$.
- So $\mathcal{B}_m(\mathcal{B}_m(E_m))$ follows by *Positive Introspection*.
- Moreover $\mathcal{B}_m(E_m), \neg\mathcal{B}_m(E_m) \vdash \mathcal{B}_m(E_m) \wedge \neg\mathcal{B}_m(E_m)$.
- Hence $\mathcal{B}_m(\mathcal{B}_m(E_m) \wedge \neg\mathcal{B}_m(E_m))$ follows by *Closure*.
- But $\mathcal{B}_m(E_m) \wedge \neg\mathcal{B}_m(E_m)$ is a contradiction.

Contradiction: Don't believe contradictions (revise your beliefs accordingly).

- Since $\mathcal{B}_m(S) \vdash \mathcal{B}_m(\mathcal{B}_m(E_m) \wedge \neg\mathcal{B}_m(E_m))$, we get $\neg\mathcal{B}_m(S)$.
- But the students are able to believe that there will be a surprise exam.

Blindspot: Is the paradox solved by claiming that it is impossible for $\mathcal{B}_m(S)$?

- Is it still possible for S to be true?