Puzzles

Eric Pacuit

Department of Philosophy
University of Maryland

pacuit.org/esslli2019/puzzles

August 11, 2019

The Surprise Exam Paradox

Variants of our first paradox, the surprise exam prediction (also called the prediction paradox), were first discussed by logicians and philosophers in the 1940s.

The Surprise Exam Paradox

Variants of our first paradox, the surprise exam prediction (also called the prediction paradox), were first discussed by logicians and philosophers in the 1940s.

Reportedly Tarski passed it along from Berkeley to Quine in the early 40s, and Gödel presented it at Princeton in ’47.

The Surprise Exam Paradox

A teacher announces to her student, a clever logician, that she will give him a **surprise exam** in a term of $n \geq 2$ days.

The Surprise Exam Paradox

A teacher announces to her student, a clever logician, that she will give him a **surprise exam** in a term of $n \geq 2$ days. He replies:

- you can't wait until day n to give the exam, because then I'd know on the morning of n that the exam must be that day;

The Surprise Exam Paradox

A teacher announces to her student, a clever logician, that she will give him a **surprise exam** in a term of $n \geq 2$ days. He replies:

- you can't wait until day n to give the exam, because then I'd know on the morning of n that the exam must be that day;
- you also can't wait until day $n - 1$ to give the exam, because then I'd know on the morning of $n - 1$ that it must be that day, having ruled out day n by the previous reasoning.

The Surprise Exam Paradox

A teacher announces to her student, a clever logician, that she will give him a **surprise exam** in a term of $n \geq 2$ days. He replies:

- you can't wait until day n to give the exam, because then I'd know on the morning of n that the exam must be that day;
- you also can't wait until day $n - 1$ to give the exam, because then I'd know on the morning of $n - 1$ that it must be that day, having ruled out day n by the previous reasoning.
- you also can't wait until day $n - 2$ to give the exam, etc.

The Surprise Exam Paradox

A teacher announces to her student, a clever logician, that she will give him a **surprise exam** in a term of $n \geq 2$ days. He replies:

- you can't wait until day n to give the exam, because then I'd know on the morning of n that the exam must be that day;
- you also can't wait until day $n - 1$ to give the exam, because then I'd know on the morning of $n - 1$ that it must be that day, having ruled out day n by the previous reasoning.
- you also can't wait until day $n - 2$ to give the exam, etc.

He concludes that the teacher cannot give him a surprise exam.

The Surprise Exam Paradox

A teacher announces to her student, a clever logician, that she will give him a **surprise exam** in a term of $n \geq 2$ days. He replies:

- you can't wait until day n to give the exam, because then I'd know on the morning of n that the exam must be that day;
- you also can't wait until day $n - 1$ to give the exam, because then I'd know on the morning of $n - 1$ that it must be that day, having ruled out day n by the previous reasoning.
- you also can't wait until day $n - 2$ to give the exam, etc.

He concludes that the teacher cannot give him a surprise exam. But then he is surprised to receive an exam on, say, day $n - 1$.

The Surprise Exam Paradox

A teacher announces to her student, a clever logician, that she will give him a **surprise exam** in a term of $n \geq 2$ days. He replies:

- you can't wait until day n to give the exam, because then I'd know on the morning of n that the exam must be that day;
- you also can't wait until day $n - 1$ to give the exam, because then I'd know on the morning of $n - 1$ that it must be that day, having ruled out day n by the previous reasoning.
- you also can't wait until day $n - 2$ to give the exam, etc.

He concludes that the teacher cannot give him a surprise exam. But then he is surprised to receive an exam on, say, day $n - 1$.

QUESTION: what went wrong in the student's reasoning?

The Designated Student Paradox

Here is a version of Sorensen's *designated student paradox*:

The Designated Student Paradox

Here is a version of Sorensen's *designated student paradox*:

A teacher shows her class of $n \geq 2$ clever logicians one gold star and $n - 1$ silver stars. After lining the students up, single file, she walks behind each student and sticks one of the stars on his back. No student can see his own back, but each can see the backs of all students in front of him. The teacher announces that the student with the gold star will be **surprised** to learn that he has it.

The Designated Student Paradox

Here is a version of Sorensen's *designated student paradox*:

A teacher shows her class of $n \geq 2$ clever logicians one gold star and $n - 1$ silver stars. After lining the students up, single file, she walks behind each student and sticks one of the stars on his back. No student can see his own back, but each can see the backs of all students in front of him. The teacher announces that the student with the gold star will be **surprised** to learn that he has it.

(This is clearly analogous to the surprise exam setup. Is there a difference?)

The Designated Student Paradox

Student 1, at the front of the line, replies:

The Designated Student Paradox

Student 1, at the front of the line, replies:

- you can't give the gold star to student *n*, because then he'd see all silver stars and therefore know he has the gold star;

The Designated Student Paradox

Student 1, at the front of the line, replies:

- you can't give the gold star to student *n*, because then he'd see all silver stars and therefore know he has the gold star;
- you also can't give the gold star to student $n - 1$, because then he'd see all silver stars and therefore know he has the gold star, having ruled out the possibility that student *n* has the gold star by the previous reasoning.

The Designated Student Paradox

Student 1, at the front of the line, replies:

- you can't give the gold star to student *n*, because then he'd see all silver stars and therefore know he has the gold star;
- you also can't give the gold star to student $n - 1$, because then he'd see all silver stars and therefore know he has the gold star, having ruled out the possibility that student *n* has the gold star by the previous reasoning.
- you also can't give the gold star to student $n - 2$, etc.

He concludes that the teacher's claim about a surprise is false.

The Designated Student Paradox

Student 1, at the front of the line, replies:

- you can't give the gold star to student n, because then he'd see all silver stars and therefore know he has the gold star;
- you also can't give the gold star to student $n-1$, because then he'd see all silver stars and therefore know he has the gold star, having ruled out the possibility that student n has the gold star by the previous reasoning.
- you also can't give the gold star to student $n-2$, etc.

He concludes that the teacher's claim about a surprise is false.

But then the students pull the stars off their backs and it is, say, student $n-1$ who has the gold star, and he is surprised.

The Designated Student Paradox

Student 1, at the front of the line, replies:

- you can't give the gold star to student n, because then he'd see all silver stars and therefore know he has the gold star;
- you also can't give the gold star to student $n - 1$, because then he'd see all silver stars and therefore know he has the gold star, having ruled out the possibility that student n has the gold star by the previous reasoning.
- you also can't give the gold star to student $n - 2$, etc.

He concludes that the teacher's claim about a surprise is false.

But then the students pull the stars off their backs and it is, say, student $n - 1$ who has the gold star, and he is surprised.

QUESTION: what went wrong in the student's reasoning?

The Surprise Exam & Designated Student Paradoxes

Many “solutions” of the surprise exam paradox and its variations have been given by philosophers and logicians in the last 60+ years.

The Surprise Exam & Designated Student Paradoxes

Many “solutions” of the surprise exam paradox and its variations have been given by philosophers and logicians in the last 60+ years.

We won’t try to survey the solutions that have been given or argue for a particular solution here. Instead, we’ll just try to get a clearer understanding of the paradox by formalizing it in epistemic logic.

The Surprise Exam & Designated Student Paradoxes

Many “solutions” of the surprise exam paradox and its variations have been given by philosophers and logicians in the last 60+ years.

We won’t try to survey the solutions that have been given or argue for a particular solution here. Instead, we’ll just try to get a clearer understanding of the paradox by formalizing it in epistemic logic.

One thing that formalization forces us to do is to make explicit a number of suppressed assumptions behind the clever student’s reasoning, without which a paradox can’t be generated.

We will follow in the tradition of those who have formalized the prediction paradox in static epistemic/doxastic logic:

R. Binkley. *The Surprise Examination in Modal Logic*. Journal of Philosophy, 1968.

C. Harrison. 1969.. *The Unanticipated Examination in View of Kripke's Semantics for Modal Logic*. Philosophical Logic..

J. McLelland and C. Chihara. *The Surprise Examination Paradox*. Journal of Philosophical Logic, 1975.

R. Sorensen. *Blindspots*. Oxford University Press, 1988.

Our brief discussion here is based on a more detailed analysis in:

W. Holliday. *Simplifying the Surprise Exam*. 2016.

Step 1: Choosing the Formalism (language)

To formalize the paradoxes, we use the epistemic language

$$\varphi ::= p_i \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid K_i\varphi$$

where $i \in \mathbb{N}$.

Step 1: Choosing the Formalism (language)

To formalize the paradoxes, we use the epistemic language

$$\varphi ::= p_i \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid K_i\varphi$$

where $i \in \mathbb{N}$. For the surprise exam paradox, we read

$K_i\varphi$ as "the student knows on the *morning* of day i that φ";

p_i as "there is an exam on the *afternoon* of day i".

Step 1: Choosing the Formalism (language)

To formalize the paradoxes, we use the epistemic language

$$\varphi ::= p_i \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid K_i\varphi$$

where $i \in \mathbb{N}$. For the surprise exam paradox, we read

$K_i\varphi$ as "the student knows on the *morning* of day i that φ";

p_i as "there is an exam on the *afternoon* of day i".

For the designated student paradox, we read

$K_i\varphi$ as "the i-th student in line knows that φ";

p_i as "there is a gold star on the back of the i-th student".

Step 1: Choosing the Formalism (reasoning system)

$$\frac{(\varphi_1 \wedge \cdots \wedge \varphi_n) \rightarrow \psi}{(K_i\varphi_1 \wedge \cdots \wedge K_i\varphi_n) \rightarrow K_i\psi}$$

Step 2: Formalizing the Assumptions ($n = 2$)

Starting with the $n = 2$ case, consider the following assumptions:

Step 2: Formalizing the Assumptions ($n = 2$)

Starting with the $n = 2$ case, consider the following assumptions:

(A) $K_1((p_1 \wedge \neg K_1 p_1) \vee (p_2 \wedge \neg K_2 p_2))$;

(B) $K_1(p_2 \to K_2 \neg p_1)$;

(C) $K_1 K_2(p_1 \vee p_2)$.

Step 2: Formalizing the Assumptions ($n = 2$)

Starting with the $n = 2$ case, consider the following assumptions:

(A) $K_1((p_1 \wedge \neg K_1 p_1) \vee (p_2 \wedge \neg K_2 p_2))$;

(B) $K_1(p_2 \rightarrow K_2 \neg p_1)$;

(C) $K_1 K_2(p_1 \vee p_2)$.

For the surprise exam, (A) states that the student knows on the morning of day 1 that the teacher's announcement is true.

Step 2: Formalizing the Assumptions ($n = 2$)

Starting with the $n = 2$ case, consider the following assumptions:

(A) $K_1((p_1 \wedge \neg K_1 p_1) \vee (p_2 \wedge \neg K_2 p_2))$;

(B) $K_1(p_2 \rightarrow K_2 \neg p_1)$;

(C) $K_1 K_2(p_1 \vee p_2)$.

For the surprise exam, (A) states that the student knows on the morning of day 1 that the teacher's announcement is true. (B) states that the student knows on the morning of day 1 that if the exam is on the afternoon of day 2, then the student will know on the morning of day 2 that it was not on day 1 (on the basis of memory).

Step 2: Formalizing the Assumptions ($n = 2$)

Starting with the $n = 2$ case, consider the following assumptions:

(A) $K_1((p_1 \wedge \neg K_1 p_1) \vee (p_2 \wedge \neg K_2 p_2))$;

(B) $K_1(p_2 \rightarrow K_2 \neg p_1)$;

(C) $K_1 K_2 (p_1 \vee p_2)$.

For the surprise exam, (A) states that the student knows on the morning of day 1 that the teacher's announcement is true. (B) states that the student knows on the morning of day 1 that if the exam is on the afternoon of day 2, then the student will know on the morning of day 2 that it was not on day 1 (on the basis of memory). Finally, (C) states that the student knows on the morning of day 1 that she will know on the morning of day 2 the part of the teacher's announcement about an *exam*.

Step 2: Formalizing the Assumptions ($n = 2$)

Starting with the $n = 2$ case, consider the following assumptions:

(A) $K_1((p_1 \wedge \neg K_1 p_1) \vee (p_2 \wedge \neg K_2 p_2))$;

(B) $K_1(p_2 \rightarrow K_2 \neg p_1)$;

(C) $K_1 K_2(p_1 \vee p_2)$.

For the designated student, (A) states that student 1 knows that the teacher's announcement is true.

Step 2: Formalizing the Assumptions ($n = 2$)

Starting with the $n = 2$ case, consider the following assumptions:

(A) $K_1((p_1 \wedge \neg K_1 p_1) \vee (p_2 \wedge \neg K_2 p_2))$;

(B) $K_1(p_2 \rightarrow K_2 \neg p_1)$;

(C) $K_1 K_2(p_1 \vee p_2)$.

For the designated student, (A) states that student 1 knows that the teacher's announcement is true. (B) states that student 1 knows that if student 2 has the gold star, then student 2 knows that student 1 does not have the gold star (on the basis of seeing the silver star on student 1's back).

Step 2: Formalizing the Assumptions ($n = 2$)

Starting with the $n = 2$ case, consider the following assumptions:

(A) $K_1((p_1 \wedge \neg K_1 p_1) \vee (p_2 \wedge \neg K_2 p_2))$;

(B) $K_1(p_2 \rightarrow K_2 \neg p_1)$;

(C) $K_1 K_2(p_1 \vee p_2)$.

For the designated student, (A) states that student 1 knows that the teacher's announcement is true. (B) states that student 1 knows that if student 2 has the gold star, then student 2 knows that student 1 does not have the gold star (on the basis of seeing the silver star on student 1's back). (C) states that student 1 knows that student 2 knows that one of them has the gold star.

Step 3: Showing Inconsistency with a Proof ($n = 2$)

Let us first show: $\{(A), (B), (C)\} \vdash_{\mathbf{K}} K_1(p_1 \wedge \neg K_1 p_1)$

Step 3: Showing Inconsistency with a Proof ($n = 2$)

Let us first show: $\{(A), (B), (C)\} \vdash_{\mathbf{K}} K_1(p_1 \wedge \neg K_1 p_1)$

(A) $K_1((p_1 \wedge \neg K_1 p_1) \vee (p_2 \wedge \neg K_2 p_2))$ premise

(B) $K_1(p_2 \rightarrow K_2 \neg p_1)$ premise

(C) $K_1 K_2(p_1 \vee p_2)$ premise

Step 3: Showing Inconsistency with a Proof ($n = 2$)

Let us first show: $\{(A), (B), (C)\} \vdash_{\mathbf{K}} K_1(p_1 \wedge \neg K_1 p_1)$

(A) $K_1((p_1 \wedge \neg K_1 p_1) \vee (p_2 \wedge \neg K_2 p_2))$ premise

(B) $K_1(p_2 \rightarrow K_2 \neg p_1)$ premise

(C) $K_1 K_2(p_1 \vee p_2)$ premise

(1.1) $((p_1 \vee p_2) \wedge \neg p_1) \rightarrow p_2)$ propositional tautology

Step 3: Showing Inconsistency with a Proof ($n = 2$)

Let us first show: $\{(A), (B), (C)\} \vdash_{\mathbf{K}} K_1(p_1 \wedge \neg K_1 p_1)$

(A) $K_1((p_1 \wedge \neg K_1 p_1) \vee (p_2 \wedge \neg K_2 p_2))$ premise

(B) $K_1(p_2 \rightarrow K_2 \neg p_1)$ premise

(C) $K_1 K_2(p_1 \vee p_2)$ premise

(1.1) $((p_1 \vee p_2) \wedge \neg p_1) \rightarrow p_2)$ propositional tautology

(1.2) $(K_2(p_1 \vee p_2) \wedge K_2 \neg p_1) \rightarrow K_2 p_2$ from (1.1) by RK_2

Step 3: Showing Inconsistency with a Proof ($n = 2$)

Let us first show: $\{(A), (B), (C)\} \vdash_{\mathbf{K}} K_1(p_1 \wedge \neg K_1 p_1)$

(A) $K_1((p_1 \wedge \neg K_1 p_1) \vee (p_2 \wedge \neg K_2 p_2))$ premise

(B) $K_1(p_2 \rightarrow K_2 \neg p_1)$ premise

(C) $K_1 K_2(p_1 \vee p_2)$ premise

(1) $(K_2(p_1 \vee p_2) \wedge K_2 \neg p_1) \rightarrow K_2 p_2$ using PL and RK_2

Step 3: Showing Inconsistency with a Proof ($n = 2$)

Let us first show: $\{(A), (B), (C)\} \vdash_{\mathbf{K}} K_1(p_1 \wedge \neg K_1 p_1)$

(A) $K_1((p_1 \wedge \neg K_1 p_1) \vee (p_2 \wedge \neg K_2 p_2))$ premise

(B) $K_1(p_2 \rightarrow K_2 \neg p_1)$ premise

(C) $K_1 K_2(p_1 \vee p_2)$ premise

(1) $(K_2(p_1 \vee p_2) \wedge K_2 \neg p_1) \rightarrow K_2 p_2$ using PL and RK_2

(2) $K_1((K_2(p_1 \vee p_2) \wedge K_2 \neg p_1) \rightarrow K_2 p_2)$ from (1) by Nec_1

Step 3: Showing Inconsistency with a Proof ($n = 2$)

Let us first show: $\{(A), (B), (C)\} \vdash_{\mathbf{K}} K_1(p_1 \wedge \neg K_1 p_1)$

(A) $K_1((p_1 \wedge \neg K_1 p_1) \vee (p_2 \wedge \neg K_2 p_2))$ premise

(B) $K_1(p_2 \rightarrow K_2 \neg p_1)$ premise

(C) $K_1 K_2(p_1 \vee p_2)$ premise

(1) $(K_2(p_1 \vee p_2) \wedge K_2 \neg p_1) \rightarrow K_2 p_2$ using PL and RK_2

(2) $K_1((K_2(p_1 \vee p_2) \wedge K_2 \neg p_1) \rightarrow K_2 p_2)$ from (1) by Nec_1

(3) $K_1(K_2 \neg p_1 \rightarrow K_2 p_2)$ from (C) and (2) using PL and RK_1

Step 3: Showing Inconsistency with a Proof ($n = 2$)

Let us first show: $\{(A), (B), (C)\} \vdash_{\mathbf{K}} K_1(p_1 \wedge \neg K_1 p_1)$

(A) $K_1((p_1 \wedge \neg K_1 p_1) \vee (p_2 \wedge \neg K_2 p_2))$ premise

(B) $K_1(p_2 \to K_2 \neg p_1)$ premise

(C) $K_1 K_2(p_1 \vee p_2)$ premise

(1) $(K_2(p_1 \vee p_2) \wedge K_2 \neg p_1) \to K_2 p_2$ using PL and RK_2

(2) $K_1((K_2(p_1 \vee p_2) \wedge K_2 \neg p_1) \to K_2 p_2)$ from (1) by Nec_1

(3) $K_1(K_2 \neg p_1 \to K_2 p_2)$ from (C) and (2) using PL and RK_1

(4) $K_1 \neg(p_2 \wedge \neg K_2 p_2)$ from (B) and (3) using PL and RK_1

Step 3: Showing Inconsistency with a Proof ($n = 2$)

Let us first show: $\{(A), (B), (C)\} \vdash_{\mathbf{K}} K_1(p_1 \wedge \neg K_1 p_1)$

(A) $K_1((p_1 \wedge \neg K_1 p_1) \vee (p_2 \wedge \neg K_2 p_2))$ premise

(B) $K_1(p_2 \rightarrow K_2 \neg p_1)$ premise

(C) $K_1 K_2(p_1 \vee p_2)$ premise

(1) $(K_2(p_1 \vee p_2) \wedge K_2 \neg p_1) \rightarrow K_2 p_2$ using PL and RK_2

(2) $K_1((K_2(p_1 \vee p_2) \wedge K_2 \neg p_1) \rightarrow K_2 p_2)$ from (1) by Nec_1

(3) $K_1(K_2 \neg p_1 \rightarrow K_2 p_2)$ from (C) and (2) using PL and RK_1

(4) $K_1 \neg(p_2 \wedge \neg K_2 p_2)$ from (B) and (3) using PL and RK_1

(5) $K_1(p_1 \wedge \neg K_1 p_1)$ from (A) and (4) using PL and RK_1

Step 3: Showing Inconsistency with a Proof ($n = 2$)

Given $\{(A), (B), (C)\} \vdash_{\mathbf{K}} K_1(p_1 \wedge \neg K_1 p_1)$, although we haven't yet derived a contradiction, we have derived something paradoxical.

Step 3: Showing Inconsistency with a Proof ($n = 2$)

Given $\{(A), (B), (C)\} \vdash_{\mathbf{K}} K_1(p_1 \wedge \neg K_1 p_1)$, although we haven't yet derived a contradiction, we have derived something paradoxical.

If we just add the "factivity" axiom T_1, $K_1\varphi \rightarrow \varphi$, or the "weak factivity" axiom J_1, $K_1\neg K_1\varphi \rightarrow \neg K_1\varphi$ (e.g., reading K as belief instead of knowledge), then we can derive a contradiction:

$$\{(A), (B), (C)\} \vdash_{\mathbf{KT_1}} \bot \text{ and } \{(A), (B), (C)\} \vdash_{\mathbf{KJ_1}} \bot.$$

Step 3: Showing Inconsistency with a Proof ($n = 2$)

Given $\{(A), (B), (C)\} \vdash_{\mathbf{K}} K_1(p_1 \wedge \neg K_1 p_1)$, although we haven't yet derived a contradiction, we have derived something paradoxical.

If we just add the "factivity" axiom T_1, $K_1\varphi \rightarrow \varphi$, or the "weak factivity" axiom J_1, $K_1\neg K_1\varphi \rightarrow \neg K_1\varphi$ (e.g., reading K as belief instead of knowledge), then we can derive a contradiction:

$$\{(A), (B), (C)\} \vdash_{\mathbf{KT_1}} \bot \text{ and } \{(A), (B), (C)\} \vdash_{\mathbf{KJ_1}} \bot.$$

Thus, we must reject either (A), (B), (C), or the rule RK_i...

"One of you don't know it, but you have a gold star on your back"

vs.

"One of you has the gold star on your back, but will not know it until she pulls the star off her back."

"One of you don't know it, but you have a gold star on your back"

vs.

"One of you has the gold star on your back, but will not know it until she pulls the star off her back."

Using the previous argument the second announcement reduces to:

"Student 1 has a gold star on her back but won't know it until she pulls the star off her back."

- Self-refuting: “I am not speaking now”

- Self-refuting: “I am not speaking now”
- Anti-peformatory: “You don’t know it, but my birthday is in April”

> *If you know that I am well informed and if I address the words . . . to you, these words have a curious effect which may perhaps be called anti-performatory. You may come to know that what I say was true, but saying it in so many words has the effect of making what is being said false.* (Hintikka, 1962)

- Self-refuting: “I am not speaking now”
- Anti-peformatory: “You don’t know it, but my birthday is in April”

  > *If you know that I am well informed and if I address the words . . . to you, these words have a curious effect which may perhaps be called anti-performatory. You may come to know that what I say was true, but saying it in so many words has the effect of making what is being said false. (Hintikka, 1962)*

- Unassimilable: “You won’t know it by the end of this party, but my birthday is in April”

Summary

- $\{(A^2), (B^2), (C^2)\} \vdash_{\mathbf{K}} K_1(p_1 \wedge \neg K_1)$;
- $\{(A^2), (B^2), (C^2)\} \vdash_{\mathbf{KJ_1}} \perp$ and $\{(A^2), (B^2), (C^2)\} \vdash_{\mathbf{KT_1}} \perp$;

Summary

- $\{(A^2), (B^2), (C^2)\} \vdash_{\mathbf{K}} K_1(p_1 \wedge \neg K_1)$;
- $\{(A^2), (B^2), (C^2)\} \vdash_{\mathbf{KJ_1}} \bot$ and $\{(A^2), (B^2), (C^2)\} \vdash_{\mathbf{KT_1}} \bot$;
- $\{(A^3), (B^3), (C^3)\} \nvdash_{\mathbf{S5}} \bot$.

Summary

- $\{(A^2), (B^2), (C^2)\} \vdash_{\mathbf{K}} K_1(p_1 \wedge \neg K_1)$;
- $\{(A^2), (B^2), (C^2)\} \vdash_{\mathbf{KJ_1}} \bot$ and $\{(A^2), (B^2), (C^2)\} \vdash_{\mathbf{KT_1}} \bot$;
- $\{(A^3), (B^3), (C^3)\} \nvdash_{\mathbf{S5}} \bot$.
- $\{(A^3), (B^3)\} \vdash_{\mathbf{K4_1^<}} K_1(p_1 \wedge \neg K_1)$;
- $\{(A^3), (B^3)\} \vdash_{\mathbf{KJ_14_1^<}} \bot$ and $\{(A^3), (B^3)\} \vdash_{\mathbf{KT_14_1^<}} \bot$;

Summary

- $\{(A^2), (B^2), (C^2)\} \vdash_{\mathbf{K}} K_1(p_1 \wedge \neg K_1)$;
- $\{(A^2), (B^2), (C^2)\} \vdash_{\mathbf{KJ_1}} \bot$ and $\{(A^2), (B^2), (C^2)\} \vdash_{\mathbf{KT_1}} \bot$;
- $\{(A^3), (B^3), (C^3)\} \nvdash_{\mathbf{S5}} \bot$.
- $\{(A^3), (B^3)\} \vdash_{\mathbf{K4_1^<}} K_1(p_1 \wedge \neg K_1)$;
- $\{(A^3), (B^3)\} \vdash_{\mathbf{KJ_14_1^<}} \bot$ and $\{(A^3), (B^3)\} \vdash_{\mathbf{KT_14_1^<}} \bot$;

With these facts, one can make a strong case that the culprit behind the paradoxes is the (mistaken) $4_1^<$ axiom, $K_1\varphi \rightarrow K_1K_i\varphi$ $(i > 1)$....

Wes Holliday. *"Simplifying the Surprise Exam."*. UC Berkeley Working paper in Philosophy, 2016.

The paradox of the undiscoverable position

R. Sorensen. *Blindspots*. Oxford: Clarendon Press, 1988.

1	2	3
4	5	6
7	8	9

You can move one position and if you bump the edge it is recorded: e.g., UUL^* means you move up twice then bumped the edge when trying to move left.

1	2	3
4	5	6
7	8	9

You can **discover** you are in position 7 using the moves UUL^*.

1	2	3
4	5	6
7	8	9

Suppose you are limited to only two moves. Say a position is **undiscoverable** if it cannot be uniquely discovered in two moves.

1	2	3
4	5	6
7	8	9

Position 4 is **undiscoverable**.

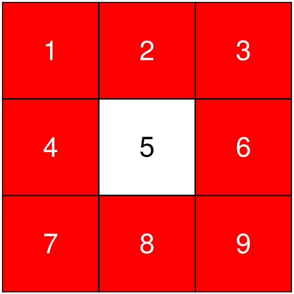
1	2	3
4	5	6
7	8	9

"You are in an undiscoverable position".

1	2	3
4	5	6
7	8	9

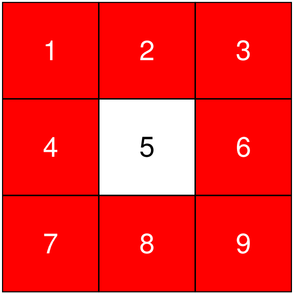
"You are in an undiscoverable position".

1. I can't be in the corners since two bumps will discover each one.



“You are in an undiscoverable position”.

1. I can’t be in the corners since two bumps will discover each one.
2. I can’t be in positions 2, 4, 6 or 8 since one bump will discover those positions (after remove the corners)



"You are in an undiscoverable position".

1. I can't be in the corners since two bumps will discover each one.
2. I can't be in positions 2, 4, 6 or 8 since one bump will discover those positions (after remove the corners)
3. So, I must be in position 5.

1	2	3
4	5	6
7	8	9

But, there 8 similar arguments showing that I'm in each of the other 7 positions. (e.g., remove the corners, then remove 2 and 4, then remove 5, so I must be in position 6).

There are many ways to interpret the announcement: “You are in an undiscoverable position”.

W. Holliday. *On Being in an Undiscoverable Position*. Thought: A Journal of Philosophy, 5(1), 33 - 40, 2016.

Probability and Beliefs

Conceptions of Belief

Binary: “all-out” belief. For any statement *p*, the agent either does or does not believe *p*. It is natural to take an *unqualified* assertion as a statement of belief of the speaker.

Conceptions of Belief

Binary: “all-out” belief. For any statement *p*, the agent either does or does not believe *p*. It is natural to take an *unqualified* assertion as a statement of belief of the speaker.

Graded: beliefs come in degrees. We are *more confident* in some of our beliefs than in others.

Conceptions of Belief

Binary: “all-out” belief. For any statement p, the agent either does or does not believe p. It is natural to take an *unqualified* assertion as a statement of belief of the speaker.

Graded: beliefs come in degrees. We are *more confident* in some of our beliefs than in others.

Eric Schwitzgebel. *Belief*. In The Stanford Encyclopedia of Philosophy.

Franz Huber. *Formal Theories of Belief*. In The Stanford Encyclopedia of Philosophy.

Probability

Kolmogorov Axioms:

1. For each E, $0 \leq p(E) \leq 1$
2. $p(W) = 1$, $p(\emptyset) = 0$
3. If $E_1, \ldots, E_n, \ldots$ are pairwise disjoint ($E_i \cap E_j = \emptyset$ for $i \neq j$), then $p(\bigcup_i E_i) = \sum_i p(E_i)$

Probability

Kolmogorov Axioms:

1. For each E, $0 \leq p(E) \leq 1$
2. $p(W) = 1$, $p(\emptyset) = 0$
3. If $E_1, \ldots, E_n, \ldots$ are pairwise disjoint ($E_i \cap E_j = \emptyset$ for $i \neq j$), then $p(\bigcup_i E_i) = \sum_i p(E_i)$

- $p(\overline{E}) = 1 - p(E)$ ($\overline{E}$ is the complement of E)
- If $E \subseteq F$ then $p(E) \leq p(F)$
- $p(E \cup F) = p(E) + p(F) - p(E \cap F)$

Bridge Principles

Probability 1: $Bel(A)$ iff $P(A) = 1$

Bridge Principles

Probability 1: $Bel(A)$ iff $P(A) = 1$

The Lockean Thesis: $Bel(A)$ iff $P(A) > r$

Bridge Principles

Probability 1: $Bel(A)$ iff $P(A) = 1$

The Lockean Thesis: $Bel(A)$ iff $P(A) > r$

Decision-theoretic accounts: $Bel(A)$ iff
$\sum_{w \in W} P(\{w\}) \cdot u(bel\ A, w)$ has such-and-such property

Bridge Principles

Probability 1: $Bel(A)$ iff $P(A) = 1$

The Lockean Thesis: $Bel(A)$ iff $P(A) > r$

Decision-theoretic accounts: $Bel(A)$ iff
$\sum_{w \in W} P(\{w\}) \cdot u(bel\ A, w)$ has such-and-such property

The Nihilistic proposal: "...no explication of belief is possible within the confines of the probability model."

Preface Paradox

D. Makinson. *The Paradox of the Preface*. Analysis, 25, 205 - 207, 1965.

Preface Paradox

Suppose that in the course of his book an author makes a great many assertions: $s_1, s_2, \dots, s_n$.

Preface Paradox

Suppose that in the course of his book an author makes a great many assertions: $s_1, s_2, \ldots, s_n$.

Given each one of these, he believes that it is true (for each i, $B_A(s_i)$)

Preface Paradox

Suppose that in the course of his book an author makes a great many assertions: $s_1, s_2, \ldots, s_n$.

Given each one of these, he believes that it is true (for each i, $B_A(s_i)$)

If he has already written other books, and received corrections from readers and reviewers, he may also believe that not everything he has written in his latest book is true.

Preface Paradox

Suppose that in the course of his book an author makes a great many assertions: $s_1, s_2, \ldots, s_n$.

Given each one of these, he believes that it is true (for each i, $B_A(s_i)$)

If he has already written other books, and received corrections from readers and reviewers, he may also believe that not everything he has written in his latest book is true.

$B_A(\neg(s_1 \wedge s_2 \wedge \cdots \wedge s_n))$

Preface Paradox

Suppose that in the course of his book an author makes a great many assertions: $s_1, s_2, \ldots, s_n$.

Given each one of these, he believes that it is true (for each i, $B_A(s_i)$)

If he has already written other books, and received corrections from readers and reviewers, he may also believe that not everything he has written in his latest book is true.

$B_A(\neg(s_1 \wedge s_2 \wedge \cdots \wedge s_n))$

But $\{s_1, \ldots, s_n, \neg(s_1 \wedge \cdots \wedge s_n)\}$ is logically inconsistent.

Preface Paradox

A philosopher who asserts “all of my present philosophical positions are correct” would be regarded as rash and over-confident

A philosopher who asserts “at least some of my present philosophical beliefs will turn out to be incorrect” is simply being sensible and honest.

Preface Paradox

1. each belief from the set $\{s_1, \ldots, s_n, s_{n+1}\}$ is rational
2. the set $\{s_1, \ldots, s_n, s_{n+1}\}$ of beliefs is rational.

1\. does not necessarily imply 2.

Preface Paradox: The Problem

"The author of the book is being rational even though inconsistent. More than this: he is being rational even though he believes each of a certain collection of statements, which *he knows* are logically incompatible....this appears to present a living and everyday example of a situation which philosophers have commonly dismissed as absurd; that it is sometimes rational to hold incompatible beliefs."

D. Makinson. *The Paradox of the Preface*. Analysis, 25, 205 - 207, 1965.

Lottery Paradox

H. Kyburg. *Probability and the Logic of Rational Belief*. Wesleyan University Press, 1961.

G. Wheeler. *A Review of the Lottery Paradox*. Probability and Inference: Essays in honor of Henry E. Kyburg, Jr., College Publications, 2007.

Lottery Paradox

Consider a fair lottery with 1,000,000 tickets and one prize.

Lottery Paradox

Consider a fair lottery with 1,000,000 tickets and one prize.

The probability that a given ticket will win is 0.000001 ($1/1,000,000$) and the probability that it will not win is 0.999999.

Lottery Paradox

Consider a fair lottery with 1,000,000 tickets and one prize.

The probability that a given ticket will win is 0.000001 $(1/1,000,000)$ and the probability that it will not win is 0.999999.

"Surely if a sheer probability is ever sufficient to warrant the acceptance of a hypothesis, this is a case"

Lottery Paradox

Consider a fair lottery with 1,000,000 tickets and one prize.

The probability that a given ticket will win is 0.000001 $(1/1,000,000)$ and the probability that it will not win is 0.999999.

"Surely if a sheer probability is ever sufficient to warrant the acceptance of a hypothesis, this is a case"

For each lottery ticket t_i ($i = 1, \ldots, 1000000$), the agent believes that t_i will loose $B_A(\neg$'t_i will win'$)$

Lottery Paradox

A **rule of acceptance**: If S and T are acceptable statements, their conjunction is also acceptable.

Lottery Paradox

A **rule of acceptance**: If S and T are acceptable statements, their conjunction is also acceptable.

So, the conjunction $\bigwedge_{i=1}^{1000000}$ 't_i will not win' should be accepted. That is, the agent should rationally accept that *no lottery ticket will win*.

Lottery Paradox

A **rule of acceptance**: If S and T are acceptable statements, their conjunction is also acceptable.

So, the conjunction $\bigwedge_{i=1}^{1000000}$ 't_i will not win' should be accepted. That is, the agent should rationally accept that *no lottery ticket will win*.

But, this is a fair lottery, so at least one ticket is *guaranteed* to win!

The Lottery Paradox

Kyburg: The following are inconsistent,

1. It is rational to accept a proposition that is very likely true,
2. It is not rational to accept a propositional that you are aware is inconsistent
3. It is rational to accept a proposition P and it is rational to accept another proposition P' then it is rational to accept $P \wedge P'$

$$EU(A) = \sum_{o \in O} P_A(o) \times U(o)$$

$$EU(A) = \sum_{o \in O} P_A(o) \times U(o)$$

Expected utility of action A

$$EU(A) = \sum_{o \in O} P_A(o) \times U(o)$$

Expected utility of action A

Utility of outcome o

$$EU(A) = \sum_{o \in O} P_A(o) \times U(o)$$

Expected utility of action A

Utility of outcome o

Probability of outcome o conditional on A

$P_A(o)$: probability of o conditional on A — how likely it is that outcome o will occur, on the supposition that the agent chooses act A.

$P_A(o)$: probability of o conditional on A — how likely it is that outcome o will occur, on the supposition that the agent chooses act A.

Evidential: $\quad P_A(o) = P(o \mid A) = \frac{P(o \ \& \ A)}{P(A)}$

$P_A(o)$: probability of o conditional on A — how likely it is that outcome o will occur, on the supposition that the agent chooses act A.

Evidential: $P_A(o) = P(o \mid A) = \frac{P(o \,\&\, A)}{P(A)}$

Classical: $P_A(o) = \sum_{s \in S} P(s) f_{A,s}(o)$, where

$$f_{A,s}(o) = \begin{cases} 1 & A(s) = o \\ 0 & A(s) \neq o \end{cases}$$

$P_A(o)$: probability of o conditional on A — how likely it is that outcome o will occur, on the supposition that the agent chooses act A.

Evidential: $P_A(o) = P(o \mid A) = \frac{P(o \,\&\, A)}{P(A)}$

Classical: $P_A(o) = \sum_{s\in S} P(s) f_{A,s}(o)$, where

$$f_{A,s}(o) = \begin{cases} 1 & A(s) = o \\ 0 & A(s) \neq o \end{cases}$$

Causal: $P_A(o) = P(A \mathrel{\Box\!\!\rightarrow} o)$

P("if A were performed, outcome o would ensue")

(Lewis, 1981)

Ellsberg Paradox

	30	60	
Lotteries	Blue	Yellow	Green
L_1	$1M$	0	0
L_2	0	$1M$	0

Ellsberg Paradox

	30	60	
Lotteries	Blue	Yellow	Green
L_3	$1M$	0	$1M$
L_4	0	$1M$	$1M$

Ellsberg Paradox

	30	60	
Lotteries	Blue	Yellow	Green
L_1	$1M$	0	0
L_2	0	$1M$	0
L_3	$1M$	0	$1M$
L_4	0	$1M$	$1M$

$$L_1 \succeq L_2 \text{ iff } L_3 \succeq L_4$$

Ambiguity Aversion

I. Gilboa and M. Marinacci. *Ambiguity and the Bayesian Paradigm*. Advances in Economics and Econometrics: Theory and Applications, Tenth World Congress of the Econometric Society. D. Acemoglu, M. Arellano, and E. Dekel (Eds.). New York: Cambridge University Press, 2013.

Flipping a fair coin vs. flipping a coin of unknown bias: “The probability is 50-50”...

Flipping a fair coin vs. flipping a coin of unknown bias: “The probability is 50-50”...

- Imprecise probabilities
- Non-additive probabilities
- Qualitative probability