

Newcomb's Problem

PARADOX AND INFINITY

Benjamin Brast-McKie

April 24, 2024

Green Grass

Drought: The town of Bayes has seen a terrible drought and the grass is dying.

- $P(\text{WetGrass} \mid \text{Umbrellas}) = .99$ and $P(\text{DryGrass} \mid \text{Umbrellas}) = .01$.
- $P(\text{WetGrass} \mid \neg \text{Umbrellas}) = .1$ and $P(\text{DryGrass} \mid \neg \text{Umbrellas}) = .9$.
- Assume that $v(\text{WetGrass}) = 10$ and $v(\text{DryGrass}) = -10$.
- $EV(\text{Umbrellas}) = .99 \times v(\text{WetGrass}) + .01 \times v(\text{DryGrass}) = 9.8$.
- $EV(\neg \text{Umbrellas}) = .1 \times v(\text{WetGrass}) + .9 \times v(\text{DryGrass}) = -8$.

Solution: Upon learning of these numbers, the mayor calls for the Bayesians to go outside with their umbrellas so that the grass can get some water.

- Something has gone wrong, but what is it?

Epilepsy

Lassie: Angela has a dog Lassie which can reliably predict her seizures.

- $P(\text{Seizure} \mid \text{Bark}) = .99$ and $P(\neg \text{Seizure} \mid \text{Bark}) = .01$.
- Assume that $v(\text{Seizure} + \text{Meds}) = 0$ and $v(\text{Seizure} + \neg \text{Meds}) = -10$.
- Lassie is barking and so Angela is sure to take her medication.
- But given what was said above, has she made a mistake?

Two Conditionals

Indication: Umbrellas on the streets and Lassie's barks are reliable indicators.

- If you only knew Umbrellas/Barks you could make a safe bet.
- Umbrellas and Barks indicate a cause, but are not causes themselves.
- There is a common cause of Umbrellas/WetGrass and Barks/Seizure.

Indicatives: Indicative conditionals can be used to assert conditional knowledge.

- If Barks, then Seizure.
- If Umbrellas, then WetGrass.

Subjunctives: Subjunctive conditionals can be used to track causal connections.

- If the Bayesians *were* to go out with umbrellas, the grass *would* be wet.
- If Lassie *were* to bark, then Angela would have a seizure.

Two Box

Action: Faced with two boxes, the question is what are you in a position to do.

- LIKELY: If you choose OneBox, the big box will be Full.
- UNLIKELY: If you were to choose OneBox, the big box would be Full.
- Choosing OneBox isn't going to fill it with money (compare Umbrellas).
- So you might as well take what is there, and hence TwoBox.

Independence: When can we use an expected utility calculation as before?

- Is the outcome causally or probabilistically dependent on the action?
- An outcome S is *counterfactually independent* of an action A iff either:
 1. $A \Box \rightarrow S$ and $\neg A \Box \rightarrow S$.
 2. $A \Box \rightarrow \neg S$ and $\neg A \Box \rightarrow \neg S$.
- $P(\text{OneBox} \Box \rightarrow \text{Full}) = P(\text{Full})$ and $P(\text{TwoBox} \Box \rightarrow \text{Full}) = P(\text{Full})$.
- $P(\text{OneBox} \Box \rightarrow \text{Empty}) = P(\text{Empty})$ and $P(\text{TwoBox} \Box \rightarrow \text{Empty}) = P(\text{Empty})$.
- By exclusivity and exhaustivity, $P(\text{Empty}) = 1 - P(\text{Full})$.

Causal Decision Theory: To be rational, maximize expected causal utility if possible.

- Weighting utilities in proportion to their likelihood is not the problem.
- The problem is mistaking probabilistic for counterfactual dependence.
- $ECU(A) = \sum_{i \in I_A} v(S_i^A)P(A \Box \rightarrow S_i^A)$ instead of $EV(A) = \sum_{i \in I_A} v(S_i^A)P(S_i^A|A)$.
- $ECU(\text{OneBox}) = \$1,000,000 \times P(\text{OneBox} \Box \rightarrow \text{Full}) + \$0 \times P(\text{OneBox} \Box \rightarrow \text{Empty})$
 $= \$1,000,000 \times P(\text{Full})$.
- $ECU(\text{TwoBox}) = \$1,001,000 \times P(\text{TwoBox} \Box \rightarrow \text{Full}) + \$1,000 \times P(\text{OneBox} \Box \rightarrow \text{Empty})$
 $= \$1,001,000 \times P(\text{Full}) + \$1,000 \times P(\text{Empty})$
 $= \$1,001,000 \times P(\text{Full}) + \$1,000 \times (1 - P(\text{Full}))$
 $= \$1,000,000 \times P(\text{Full}) + \$1,000$
 $= ECU(\text{OneBox}) + \$1,000$
- Two boxing is better independent of the value of $P(\text{Full})$.

Short

Bonus: Choose, but before opening, bet about the total value for a bonus.

- Your choice may indicate what you were inclined to choose in the past.
- And your past inclinations indicate the prediction made about you.
- Bet Empty if and only if you chose TwoBox.
- You might feel that you have been punished for your rationality.
- So it goes in mischievous though experiments!