Newcomb's Problem

Summary Sheet – 24.118, Spring 2021

1 The Problem

There are two boxes. The transparent box contains \$1K; you're not sure what the opaque box contains but it's either \$0 or \$1M. You have two choices:

Two-Box Keep both boxes.

One-Box Keep the large box; leave the small box behind.

The boxes were sealed before you entered the room, and your choice will not cause their contents to change. How should you choose?

2 A Predictor

The contents of the box were selected by a predictor, who is known to be 99% accurate:

Prediction	**Opaque Box**	**Transparent Box**
One-Box	\$1M	\$1K
Two-Box	\$0	\$1K

As before, the boxes were sealed before you entered the room, and your choice will not cause their contents to change. How should you choose?

- *Wait!* Could there even be such a predictor?

3 A Case for One-Boxing

- if you one-box, it is almost certain (99%) that the large box will contain a million dollars;
- if you two-box, it is almost certain (99%) that the large box will be empty.

4 A Case for Two-Boxing

- If the large box is empty, you'll be better off if you two-box than if you one-box.
- If the large box is full, you'll be better off if you two-box than if you one-box.

5 Decision Theory

your options + your probabilities + your values $\longrightarrow$ a recommendation

Expected Value Maximization Choose an option whose *expected value* is at least as high as that of any rival option.

The **expected value** of an option A is the weighted average of the value of the outcomes that A might lead to, with weights determined by the probability of the relevant state of affairs, given that you choose A.

5.1 Formally Speaking...

$$EV(A) = v(AS_1) \cdot p(S_1|A) + \ldots + v(AS_n) \cdot p(S_n|A)$$

- $S_1, S_2, \ldots S_n$ is any list of (exhaustive and mutually exclusive) states of the world;
- $v(AS_i)$ is the value of being in a situation in which you've chosen A and S_i is the case;
- $p(S|A)$ is the probability of S, given that you choose A.

6 Independence

- Two events are **causally independent** of one another if neither of them is a cause of the other.

  (Otherwise, the effect is **causally dependent** on the cause.)
- Two events are **probabilistically independent** of one another if the assumption that one of them occurs does not affect the probability that the other one will occur.

  (Otherwise, each of them is **probabilistically dependent** on the other.)

7 Newcomb's Puzzle

- The contents of the opaque box are probabilistically dependent on your decision.

  $p(F|1) > p(F)$
- The contents of the opaque box are causally independent of your decision.

  One-boxing does not cause the opaque box to be full. (And vice-versa.)

Should you one-box, if you want the opaque box to be full?

8 An Analogy

- Wet sidewalks are probabilistically dependent on umbrella use.

  $p(W|U) > p(W)$

- Wet sidewalks are causally independent of umbrella use.

  Bringing your favorite umbrella will not cause the sidewalks to be wet. (And vice-versa.)

Should you refrain from carrying your favorite umbrella, if you want the sidewalks to be dry?

9 Conditionals, Indicative and Subjective

Past Events:

Indicative If Oswald didn't shoot Kennedy, somebody else did.

Subjunctive Had Oswald not shot Kennedy, somebody else would have.

Present Events:

Indicative If people are using their umbrellas, it's raining.

Subjunctive Were people to use their umbrellas, there would be rain.

Indicative If it's raining, people are using their umbrellas.

Subjunctive Were it to rain, people would use their umbrellas.

10 Indicative Conditionals

- "$A \to B$" is shorthand for the indicative conditional"if A, then B".

- The probability of an indicative conditional is the conditional probability of its consequent given its antecedent:*

$$p(A \to B) = p(B|A)$$

11 Subjunctive Conditionals

- "$A \mathrel{\Box\!\!\rightarrow} B$" is shorthand for the subjunctive conditional "were it that A, it would be that B".

- If A causes B, then, typically, $A \mathrel{\Box\!\!\rightarrow} B$.

  Example: Rain causes umbrella use. So: were it to rain, people would use their umbrellas.

---

*Provided that A contains no indicative conditionals.

- If A doesn't cause B, then, typically:
  - If B: $A \mathrel{\Box\!\!\rightarrow} B$ and $\neg A \mathrel{\Box\!\!\rightarrow} B$; or
  - If $\neg B$: $A \mathrel{\Box\!\!\rightarrow} \neg B$ and $\neg A \mathrel{\Box\!\!\rightarrow} \neg B$

  Example: Umbrella use doesn't cause rain. Assuming it's sunny:

  Were people to use their umbrellas, it would (still) be sunny outside; and

  Were people to refrain from using their umbrellas, it would (still) be sunny outside.

12 A Connection with Free Will

As you have breakfast in NYC, Susan tells you that she was considering going on a train trip to Alaska the previous night. She obviously did not make the trip. Was she in a position to do otherwise?

- As judged using the indicative conditional, no:

  If Susan decided to make a train trip to Alaska last night, she failed.

- As judged using the subjunctive conditional, maybe:

  Had Susan decided to make the trip, she would have succeed.

Susan faces a Newcomb scenario. Should she One-Box?

- As judged using the indicative conditional, yes:
  - If Susan one-boxes, she'll almost certainly be rich.
  - If Susan two-boxes, she'll almost certainly be poor.
- As judged using the subjunctive conditional, no:
  - Were Susan to one-box, she would fail to bring about a situation in which she ends up with as much money as is available.
  - Were Susan to two-box, she would succeed in bringing about a situation in which she ends up with as much money as is available.

13 Evidential Decision Theory

$$\begin{aligned} EV(1) &= v(1\,F) \cdot p(F|1) + v(1\,E) \cdot p(E|1) \\ EV(2) &= v(2\,F) \cdot p(F|2) + v(1\,E) \cdot p(E|2) \end{aligned}$$

But $p(A \to B) = p(B|A)$. So:

$$\begin{aligned} EV(1) &= v(1\,F) \cdot p(1 \to F) + v(1\,E) \cdot p(1 \to E) \\ EV(2) &= v(2\,F) \cdot p(2 \to F) + v(1\,E) \cdot p(2 \to E) \end{aligned}$$

14 Causal Decision Theory

$$\begin{aligned} EV(1) &= v(1\,F)\cdot p(1 \;\Box\!\!\rightarrow F) + v(1\,E)\cdot p(1 \;\Box\!\!\rightarrow E) \\ EV(2) &= v(2\,F)\cdot p(2 \;\Box\!\!\rightarrow F) + v(2\,E)\cdot p(2 \;\Box\!\!\rightarrow E) \end{aligned}$$

But $p(1 \;\Box\!\!\rightarrow F) = F$ (and similarly for other cases). So:

$$\begin{aligned} EV(1) &= v(1\,F)\cdot p(F) + v(1\,E)\cdot p(E) \\ EV(2) &= v(2\,F)\cdot p(F) + v(2\,E)\cdot p(E) \end{aligned}$$

Since, $v(2\,F) > v(1\,F)$ and $v(2\,E) > v(1\,E)$:

$$EV(1) < EV(2)$$

15 Prisoner's Dilemma

	You defect	You keep quiet
Jones defects	Jones $\rightarrow -9{,}000$ You $\rightarrow -9{,}000$	Jones $\rightarrow 0$ You $\rightarrow -10{,}000$
Jones keeps quiet	Jones $\rightarrow -10{,}000$ You $\rightarrow 0$	Jones $\rightarrow -1{,}000$ You $\rightarrow -1{,}000$