Newcomb's Problem

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.