edX

The Problem
The Problem

You are led into a room, and shown two boxes: a large one and a small one. You're told that the small box contains a thousand dollars. You're not sure what the large one contains, but you're told that it either contains a million dollars or is completely empty. You are then offered two choices:

Two-Box
Keep the contents of both boxes.

One-Box
Keep the contents of the large box, but leave the contents of 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 proceed? Should you one-box or should you two-box? The answer seems totally obvious. You should take both boxes! How could you possibly benefit from leaving a thousand dollars behind? Whether or not the large box contains a million dollars, you'll end up with more money if you take the small box as well.

A Twist

Wait! There's a twist. Let me tell you about what happened *before* you entered the room. A couple of weeks ago, a personality expert was summoned, and was handed as much information about you as could be gathered. The expert was then asked to predict, on the basis of that information, whether you would one-box or two-box. If the expert concluded that you would one-box, the large box was filled with a million dollars. If she concluded that you would two-box, the large box was left empty. In other words:

Expert's Prediction	Contents of Large Box	Contents of Small Box

Expert's Prediction	Contents of Large Box	Contents of Small Box
You One-Box	$1,000,000	$1,000
You Two-Box	$0	$1,000

The expert issued her prediction last night, and the boxes have been sealed since then. If the large box was filled with a million dollars last night, it will continue to hold a million dollars, regardless of what you decide. And if it was left empty last night, it will remain empty regardless of what you decide.

One final point: the predictor is known to be highly reliable. She has participated in many thousands of experiments of this kind, and has made accurate predictions 99% of the time. There is nothing special about your case, so you should think that the expert is 99% likely to correctly predict whether you will one-box or two-box.

How should you proceed now that you know about the procedure that was used to fill the boxes? Is it still obvious that you should two-box? (Keep in mind that the predictor knows that you'll be told about how the experiment works, and about the method that was used to decide how to fill the boxes. So she knows that you'll be engaging in just the kind of reasoning that you are engaging in right now!)

The Predicament

The fact that the predictor is 99% likely to predict whether you will one-box or two-box gives you the following two pieces of information:

1. if you decide to take both boxes, it is almost certain (99%) that the large box will be empty;

2. if you decide to leave the small box behind, it is almost certain (99%) that the large box will contain a million dollars.

Before learning about the expert, it seemed clear that two-boxing was the right thing to do. But now you know that if you one-box you are almost certain to end up with a million dollars, and that if you two-box you are almost certain to end up with just a thousand. Should you be a one-boxer after all? What to do?

Video Review: The Newcomb Problem

So today I want to talk about Newcomb's problem.

So imagine that we have a blue box and a red box.

And you can see inside the blue box

and you can tell that it has $100.

The red box is sealed and you can't look inside,

but you have been told by

▶   0:00 / 0:00   ▶ 1.50x   🔊   ⛶   CC   ❝

Video
Download video file

Transcripts
Download SubRip (.srt) file
Download Text (.txt) file

Discussion

[Hide Discussion]

Topic: Week 5 / The Problem

Add a Post

| Show all posts ⌄ | by recent activity ⌄ |

💬 I find the wording of the problem absurd
The concept of probability is related to *random mass phenomena*. It is based on the fact that r...   3

💬 Is there something circular going on?
I haven't read up on the problem, but I suspect that their may be a circular reasoning path since ...   2

💬 **Incomplete description of the problem**   8
Never stated in the problem is 1. Was the predictor told how the boxes would be filled based on ...

? **Choose randomly**   3
Do you learn about this expert before or after you make your choice? Perhaps the expert has **...

💬 **Initial guess**   1
At first it seems obvious. Thinking like a Poker player using expected value. Forget the entire stor...