24.118: Paradox and Infinity, Spring 2019
Problem Set 5: Newcomb's Problem

How these problems will be graded:

- In Part I there is no need to justify your answers. Assessment will be based on whether your answers are correct.
- In Part II you must justify your answers. Assessment will be based both on whether you give the correct answer and on how your answers are justified. (In some problem sets I will ask you to answer questions that don't have clear answers. In those cases, assessment will be based entirely on the basis of how your answer is justified. Even if it is unclear whether your answer is correct, it should be clear whether or not the reasons you have given in support of your answer are good ones.)
- *No answer may consist of more than 150 words.* Longer answers will not be given credit. (Showing your work in a calculation, a proof, or a computer program does not count towards the word limit.)
- You may consult published literature and the web. You must, however, credit all sources. Failure to do so constitutes plagiarism and can have serious consequences. For advice about when and how to credit sources see: https://integrity.mit.edu/ (You do not need to credit course materials.)

Notation:

A. Evidential Decision Theory

According to Evidential Decision Theorists, the **expected value** of an option A is the weighted average of the values of A's possible outcomes, with weights determined by the probability of the relevant state of affairs, given that you choose A. Formally:

$$EV(A) = v(AS_1) \cdot p(S_1|A) + v(AS_2) \cdot p(S_2|A) + \ldots + v(AS_n) \cdot p(S_n|A)$$

where $S_1, S_2, \ldots S_n$ 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, and $p(S|A)$ is the probability of S, given that you choose A.

Evidential Decision Theorists endorse the following principle:

Expected Value Maximization In any decision problem, you ought to choose an option whose expected value is at least as high as that of any rival option.

B. Causal Decision Theory

Causal Decision Theorists prefer a different way of calculating expected value. To avoid confusion, I shall refer to the causalist's notion as **expected causal utility**.

The expected causal utility of an option A is the weighted average of the values of A's possible outcomes, with weights determined by the probability of the following subjunctive conditional: *were you to perform A, the outcome would come about.* Formally:

$$ECU(A) = v(AS_1) \cdot p(A \mathrel{\Box\!\!\rightarrow} S_1) + v(AS_2) \cdot p(A \mathrel{\Box\!\!\rightarrow} S_2) + \ldots + v(AS_n) \cdot p(A \mathrel{\Box\!\!\rightarrow} S_n)$$

where $S_1, S_2, \ldots S_n$ 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, and $A \mathrel{\Box\!\!\rightarrow} S_i$ is the claim that S_i would come about were you to perform A.

Note that in the special case in which each S_i ($i \leq n$) is a state of the world whose obtaining or not is causally independent of action A, $A \mathrel{\Box\!\!\rightarrow} S_i$ is equivalent to S_i. So we get:

$$ECU(A) = v(AS_1) \cdot p(S_1) + v(AS_2) \cdot p(S_2) + \ldots + v(AS_n) \cdot p(S_n)$$

Causal Decision Theorists endorse the following principle:

Expected Causal Utility Maximization In any decision problem, you ought to choose an option whose expected causal utility is at least as high as that of any rival option.

(For further discussion of Causal Decision Theory see Section 5.4 of the class textbook.)

C. Your Value Function

Throughout the problem set we will assume that you value only money (and value it linearly). Accordingly, you assign value n to a situation in which you net \$$n$.

Part I

1. There are two boxes before you, Left Box and Right Box. You have two options:

   Left Take the contents the Left Box only.

   Right Take the contents the Right Box only.

   Predictor has placed \$100 in one of the boxes, but you don't know which. What you do know is that last night Predictor made a prediction about whether you would choose Left or Right. The predictor is your friend: if she predicted Left, she put the money in Left Box; if she predicted Right, she put the money in Right Box. Predictor is 90% reliable:

   $$p(\text{LeftPredicted}|\text{Left}) = p(\text{RightPredicted}|\text{Right}) = 0.9$$

   The boxes have been filled ahead of time and your choice will not cause their contents to change. (So, for example, if Predictor placed the \$100 in Left Box, were you to choose Left, the money would still be in Left Box, and were you to choose Right, the money would still be in Left Box.)

(a) What is the expected value of choosing Left? (5 points)

(b) Assume that you see yourself as equally likely to choose Left and Right, and therefore that $p(\text{LeftPredicted}) = p(\text{RightPredicted}) = 0.5$.[1] What is the expected causal utility of choosing Left? (5 points)

Part II

2. There are two ways of getting to the concert:

Taxi Take a taxi.

Car Drive your car.

Taking a taxi is expensive (value = -41). Driving there would cost only gas money (value = -2), but there might be a robber at the parking lot. If he's there, he'll steal your stereo (value = -130). At home, your stereo is safe.

(a) Assume that your two options have the same expected value and that the robber's presence in the parking lot is independent of whether you drive (i.e. $p(\text{RobberInLot}) = p(\text{RobberInLot}|\text{Car})$). What is the probability that the robber will be at the parking lot? (10 points; don't forget to justify your answer.)

(b) Now suppose that the robber has a (perfectly accurate) informant who will tell him whether you'll be driving to the concert. If he's told you're going, he'll be there with probability 80%. If not, he'll be there with probability 50%.

According to the Principle of Expected Value Maximization, should you drive or should you take a cab? (10 points; don't forget to justify your answer.)

3. There are two boxes before you: Open and Closed. Open contains \$10 but you cannot see the contents of Closed. You are told, however, that Closed is either completely empty or contains \$100. You have two options:

One-Box Take the contents of Closed and leave the contents of Open behind

Two-Box Take the contents of both boxes.

The boxes have been filled ahead of time and your choice will not cause their contents to change. (So, for example, if Closed contains \$100, were you to One-Box, it would still contain \$100, and were you to Two-Box, it would still contain \$100.)

[1] I'll ask you to prove this entailment in Problem Set 6.

(a) There is no predictor. Instead, a fair coin was flipped. If it landed Heads, Closed was filled with the \$100; if it landed Tails, Closed was left empty. All of this happened yesterday and your choice will not cause the contents of the boxes to change.

According to the Principle of Expected Value Maximization, should you one-box or two-box? (10 points; don't forget to justify your answer.)

(b) Same setup as (3a). According to the Principle of Expected Causal Utility Maximization, should you one-box or should you two-box? (10 points; don't forget to justify your answer.)

(c) Now assume there is a predictor. Yesterday evening, Predictor was enlisted to make a prediction about whether you would one-box or two-box. If Predictor predicted that you would one-box, the \$100 dollars was placed in Closed. Otherwise, Closed was left empty. The probability that Predictor guesses correctly is 60%. The boxes have now been sealed and their contents will not be changed.

According to the Principle of Expected Value Maximization, should you one-box or should you two-box? (10 points; don't forget to justify your answer.)

(d) Same setup as (3c). According to the Principle of Expected Causal Utility Maximization, should you one-box or should you two-box? (10 points; don't forget to justify your answer.)

(e) Same setup as (3c), except that this time you learn that Closed has \$100.

According to the Principle of Expected Value Maximization, should you one-box or should you two-box? (5 points; don't forget to justify your answer.)

(f) Same setup as (3c), except that this time you learn that Closed has \$0.

According to the Principle of Expected Value Maximization, should you one-box or should you two-box? (5 points; don't forget to justify your answer.)

(g) Same setup as (3c), except for the following. It is now time t_0 and you have no idea whether Closed contains \$100 or \$0. At a later time t_2 you must decide whether to one-box or two-box. At time t_1 between t_0 and t_2 you will learn the contents of Closed.

Assume that you're certain that you always choose in accordance with the Principle of Expected Value Maximization. What should you believe at time t_0 about what your decision at time t_2 will be? (10 points; don't forget to justify your answer.)

(h) Same setup as (3c), except for the following. It is now time t_0 and you have no idea whether Closed contains \$100 or \$0. At a later time t_2 you must decide whether to one-box or two-box. At time t_1 between t_0 and t_2 you will be offered the chance to learn the contents of Closed.

Assume that you're certain that you always choose in accordance with the Principle of Expected Value Maximization. According to the Principle of Expected Value Maximization, should you choose, at t_1, to learn the contents of Closed or should you choose to remain ignorant? Is that answer intuitively correct? (10 points; don't forget to justify your answer.)

4. **Extra Credit:**[2] There are two boxes before you, Left Box and Right Box. You have three options:

   Left Take the contents the Left Box only.

   Right Take the contents the Right Box only.

   Coin Pay \$1 for the privilege of flipping a coin. If it lands Heads you keep the contents of Left Box; if it lands Tails, you keep the contents of Right Box.

   Predictor has placed \$100 in one of the boxes, but you don't know which. What you do know is that last night Predictor made a prediction about whether you would choose Left, Right, or Coin. Unfortunately for you, **the predictor is evil**: If she predicted Left, she put the money in Right Box; if she predicted Right, she put the money in Left Box. If she predicted Coin, she flipped a fair coin. If it landed Heads she put the money in Right Box, if it landed Tails, she put the money in Left Box. Finally, Predictor is 100% reliable:

   $$p(\text{LeftPredicted}|\text{Left}) = p(\text{RightPredicted}|\text{Right}) = p(\text{CoinPredicted}|\text{Coin}) = 1$$

   The boxes have been filled ahead of time, and their contents will not be changed. (So, for example, if Predictor placed the \$100 in Left Box, were you to choose Left, the money would still be in Left Box, were you to choose Right, the money would still be in Left Box, and were you to pick Coin, the money would still be in Left Box.)

   (a) According to the Principle of Expected Value Maximization, which of the three options should you choose? (5 points; don't forget to justify your answer.)

   (b) Assume that you see yourself as equally likely to choose Left, Right, and Coin, and therefore that $p(\text{LeftPredicted}) = p(\text{RightPredicted}) = p(\text{CoinPredicted}) = 1/3$. According to the Principle of Expected Causal Utility Maximization, which of the three options should you choose? (5 points; don't forget to justify your answer.)

   (c) *Optional:* Think about whether the answer to (4b) is intuitively correct.

[2] This example is due to Arif Amhed.