24.118: Paradox and Infinity, Spring 2024
Problem Set 5: Newcomb's Problem

Please include your name and list all of your collaborators on the last page of your problem set, preferably a separate page so that no one sees it till the end. *Failing to list collaborators constitutes a violation of academic integrity.*

How your answers will be graded:

- In Part I there is no need to justify your answers. Submit answers in a quiz that you will access on Canvas.
- In Part II you must justify your answers unless stated otherwise in the problem. 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 your justification. 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 comprise more than 180 English words. Longer answers will be penalized. However, showing your work in a calculation or proof does *not* count toward the word limit.
- You may consult published literature and the web, but you must credit all sources where failure to do so constitutes plagiarism and can have serious consequences. For advice about how and when to credit sources see `https://integrity.mit.edu/`. Note that merely citing a source does *not* count as a good justification.

All submissions must be in PDF format. Type-written submissions may be strongly preferred by your TA; handwritten submissions are acceptable only if:

1. Your handwriting is easily legible (as judged by your TA);
2. You produce a clean version of the document (as opposed to the sheet of paper you used to work out the problems); and
3. Your manuscript has been scanned to high enough standards (as judged by your TA). Consider using, e.g. *Scannable* or *Adobe Scan*.

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:

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

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 $A \mathbin{\square\!\!\rightarrow} S_i$ 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, $A \mathbin{\square\!\!\rightarrow} S_i$ 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.

C. Your Value Function

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

Part I (Quiz on Canvas: 28 points)

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

   Left Take the contents of the Left Box only.

   Right Take the contents of 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? (2 points)

   (b) Is the expected value of choosing Right "greater than", "less than', or the "same as" the EV of choosing Left? (2 points)

   (c) Assume that you see yourself as equally likely to choose Left and Right, and therefore that $p(\text{LeftPredicted}) = p(\text{RightPredicted}) = 0.5$.
   What is the expected causal utility of choosing Left? (2 points)

   (d) Is the expected causal utility of choosing Right "greater than", "less than', or the "same as" the ECU of choosing Left? (2 points)

2.–5. : See the Canvas Quiz for Part 1 for questions 2 thru 5!

Part II (Submit PDF on Canvas: 72 points)

6. There are two boxes before you: Open and Closed. Open contains $10. You cannot see the contents of Closed, but you are told that it 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.)

(a) There is no predictor. Instead, a fair coin was flipped. If it landed Heads, Closed was filled with $100; if it landed Tails, Closed was left empty. All 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? (8 points; don't forget to justify your answer.)

(b) Same setup as (6a). According to the Principle of Expected Causal Utility Maximization, should you one-box or two-box? (8 points; justify 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? (8 points; don't forget to justify your answer!)

(d) Same setup as (6c). According to the Principle of Expected Causal Utility Maximization, should you one-box or should you two-box? (8 points; don't forget to justify your answer.)

(e) Same setup as (6c), 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? (10 points; don't forget to justify your answer.)

(f) Same setup as (6c), 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? (10 points; don't forget to justify your answer.)

(g) Same setup as (6c), 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; justify your answer.)

(h) Same setup as (6c), 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 and discuss its significance.)