**Newcomb’s Problem I**

- What’s essential to Newcomb’s problem?
  1. You must choose between two particular acts: $A_1$ = you take just the opaque box; $A_2$ = you take both boxes, where the two states of nature are: $S_1$ = there’s $1M in the opaque box, $S_2$ = there’s $0 in the opaque box.
  2. Your choice of $A_i$ is causally irrelevant to $S_i$, since the contents of the opaque box ($S_i$) are determined before you choose $A_i$.
  3. $A_2$ dominates $A_1$. That is, $(\forall i)[u(S_i, A_2) > u(S_i, A_1)]$. Here, $u$ is your utility function over outcomes (assume $u$ is linear in $\$, for simplicity).
  4. The evidential expected utility of $A_1$ is greater than the evidential expected utility of $A_2$: $\sum_i \Pr(S_i|A_1) \cdot u(A_1, S_i) > \sum_i \Pr(S_i|A_2) \cdot u(A_2, S_i)$.\(^a\)

- Note: (2) and (3) entail that the Principle of Dominance (POD) applies and prescribes act $A_2$ as the rational act. If (2) fails, then (POD) need not apply.
- So, (PMEU) and (POD) seem to come into conflict in Newcomb’s problem.

\(^a\)I follow Joyce in writing evidential probability as $\Pr(\cdot|\cdot)$ and causal probability as $\Pr(\cdot)$. 

---

**Newcomb’s Problem II**

- Note: (1)–(4) entail that your act confirms the salient state of nature (but is causally irrelevant to it). That is, $A_i$ is merely symptomatic of $S_i$.
- What is inessential to Newcomb’s Problem?
  1. $A_i$ verifies $S_i$ (i.e., perfect evidential correlation between $A_i$ and $S_i$). This is not part of the original statement of NP, and it is inessential to it.
  2. That there is a predictor of your choice whose reliability (and money placing habits) sets-up the evidential correlation between the $A_i$ and the $S_i$. This is part of the original statement of NP, but it is inessential to it.
- What’s crucial here is the causal structure of the problem. Presumably (a la Reichenbach), if (1)–(4) hold, then there is a common cause $CC$ of $A_i$ and $S_i$.

\[ A_i \rightarrow CC \rightarrow S_i \]

- In Presting’s Problem, you must choose a decision algorithm $D_i$, and your “opponent” (the predictor) must choose a prediction algorithm $P_j$.
- The pair $(D_i, P_j)$ then determines which act $A_i$ is performed (if any!), and which state of nature $S_i$ obtains, where the states and acts are as above, in NP.
  - $\langle D_i, P_j \rangle$ does not halt. [no outcome, $0\$?]
  - $\langle D_i, P_j \rangle$ halts, $P_j$ predicts that $D_i$ recommends $A_1$, $D_i$ recommends $A_1$. [$S_1 & S_1$]
  - $\langle D_i, P_j \rangle$ halts, $P_j$ predicts that $D_i$ recommends $A_1$, $D_i$ recommends $A_2$. [$S_2 & S_1$]
  - $\langle D_i, P_j \rangle$ halts, $P_j$ predicts that $D_i$ recommends $A_2$, $D_i$ recommends $A_2$. [$S_2 & S_2$]
  - $\langle D_i, P_j \rangle$ halts, $P_j$ predicts that $D_i$ recommends $A_2$, $D_i$ recommends $A_2$. [$S_2 & S_1$]
- Both “players” have common knowledge of the set-up of the “game”, and also common knowledge of each other’s rationality, etc.
- This is a rule-consequentialist version of the problem. Instead of choosing between two acts, we are choosing between $\mathcal{N}_0$ decision rules (algorithms).
Presting's Problem II

- Presting: there is no effective (general) way of determining the salient utilities \( u(D_i \& P_j) \), since there is no effective way to determine if \( \langle D_i, P_j \rangle \) halts.
- Questions: What are the evidential probabilities \( \Pr(P_j/D_i) \)? Are the \( P_j \) and the \( D_i \) evidentially correlated? Note: assigning equal conditional probabilities to the \( P_j \) would violate countable additivity. We need a Pr-model here!
- And, how can this be a Newcomb Problem? Its causal structure seems to be:

\[
\begin{align*}
&\quad A_i \quad S_i \\
\downarrow< D_i P_j > &\quad \downarrow D_i \\
D_i &\quad P_j
\end{align*}
\]

- In Presting's Problem, your choice of decision algorithm \( D_i \) is prior to the determination of the state \( S_i \).
- Moreover, it appears that your choice of \( D_i \) may be causally positive for \( S_i \).
- Recall that in the NP, your choice of act \( A_i \) is after the salient state \( S_i \) is determined.

Presting's Problem III

- This does seem to be an (effectively) unsolvable problem in the general case.
- But, consider the following pair of constant (hence, trivial) decision algorithms: \( D_1 = \) take only the opaque box, and \( D_2 = \) take both boxes.
- Assuming that all prediction algorithms \( P_j \) can determine the behavior of constant (trivial) decision algorithms like these, we will have the following:

\[
(\forall j)[u(P_j \& D_1) > u(P_j \& D_2)] \quad (\text{since } \$1M > \$1K)
\]

- In other words, \( D_1 \) dominates \( D_2 \). It seems quite clear that \( D_1 \) is to be strictly preferred to \( D_2 \) as a decision algorithm in Presting's Problem.
- While the two-box act is dominant over the one-box act in NP, the one-box (constant) rule is dominant over the two-box rule in Presting's Problem!

\[\text{\( \text{Does } \text{(PDOM) apply here? After all, it seems that the } D_i \text{ are not causally irrelevant to the } S_i. \text{ This is true, but } D_1 \text{ seems causally positive for } S_1, \text{ which makes the preference } D_1 \succ D_2 \text{ even more clear!} \text{ } \)}\]