Decision theory

Example (Newcomb's problem)

A genie, which is able to predict your behavior with high accuracy, places two boxes in front of you. You can choose to take either

(1) Box B only, or

(2) Both boxes.

The boxes contain:

Box A	Box B
\$ 1000	\$1M, if the genie predicted you'd choose (1); \$0, if —— " —— (2)

The genie made its predictions yesterday and is now inside its lamp, unable to interfere. What do you pick?

Three ways of thinking about this

1. Naive classical game theory

You \ Genie	Predict (1)	Predict (2)
(1)	\$1M	\$0
(2)	\$1M + \$1000	\$1000

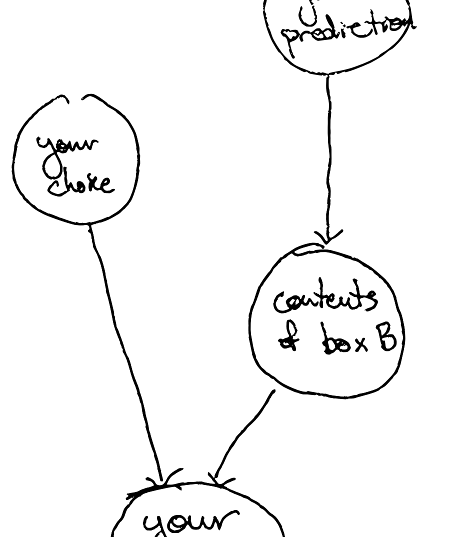
(2) is a dominant strategy, so pick (2).

Issue: why are we modelling this like a 2-player game? This fails to capture an essential feature of the situation: your choice is correlated with the genie's choice. But more fundamentally: why are we treating the genie like a player instead of like a feature of the environment?

2. Causal decision theory

Model causal effects of actions.

We imagine this with a causal network:

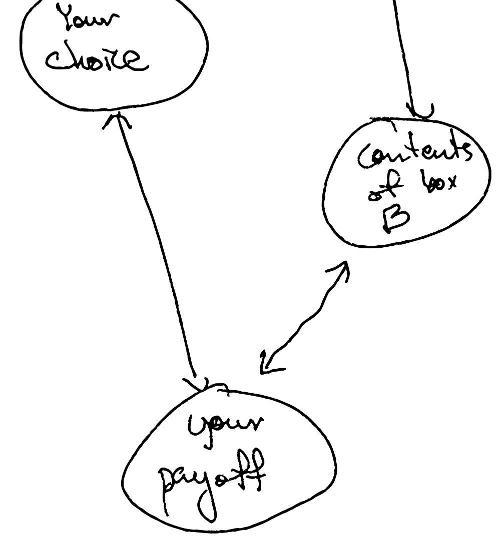


Again we see that no matter what the "genie's prediction" node is setting "Your choice" = (2) leads to a higher value of "Your payoff." So this analysis suggests option (2).

We'll see how to do this analysis more formally later in the course.

3. Evidential decision theory

Same as above, but now model evidence that nodes give about each other.



Setting "your choice" to (2) gives evidence that "genie's prediction" = (2), which gives evidence that "your payoff" is low. Thus, which properly done, this analysis will recommend choosing (1).

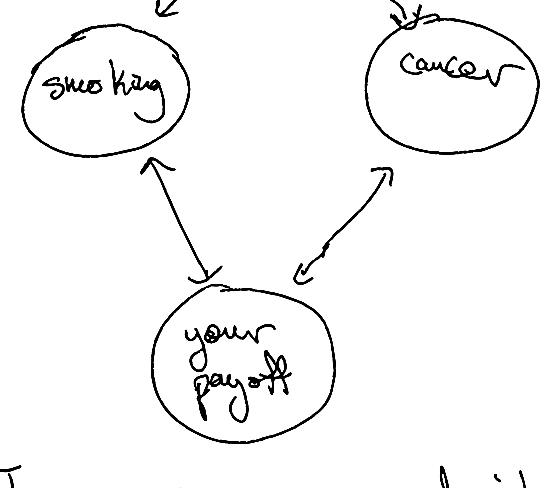
Example (Smoking lesion)

Suppose:

- Smoking is strongly correlated with lung cancer
- But this correlation is the result of a genetic brain lesion, which both causes cancer and makes people want to smoke.
- After controlling for the presence of the lesion, there is no further correlation b/t smoking and cancer.
- You enjoy smoking, but not as much as you dislike cancer.

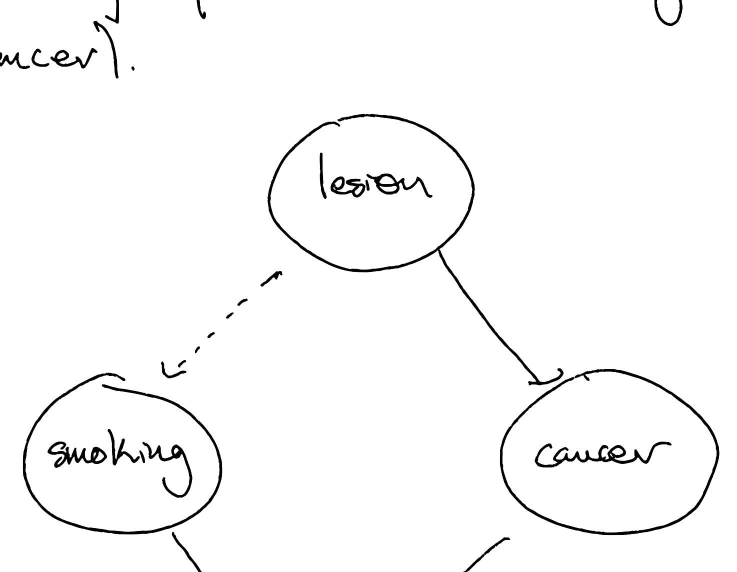
Should you smoke?

EDT



EDT says no, don't smoke (since smoking provides evidence you have cancer).

CDT



where we shouldn't consider the dotted line when making a decision, since intervening on the "smoking" node breaks the causal connection.

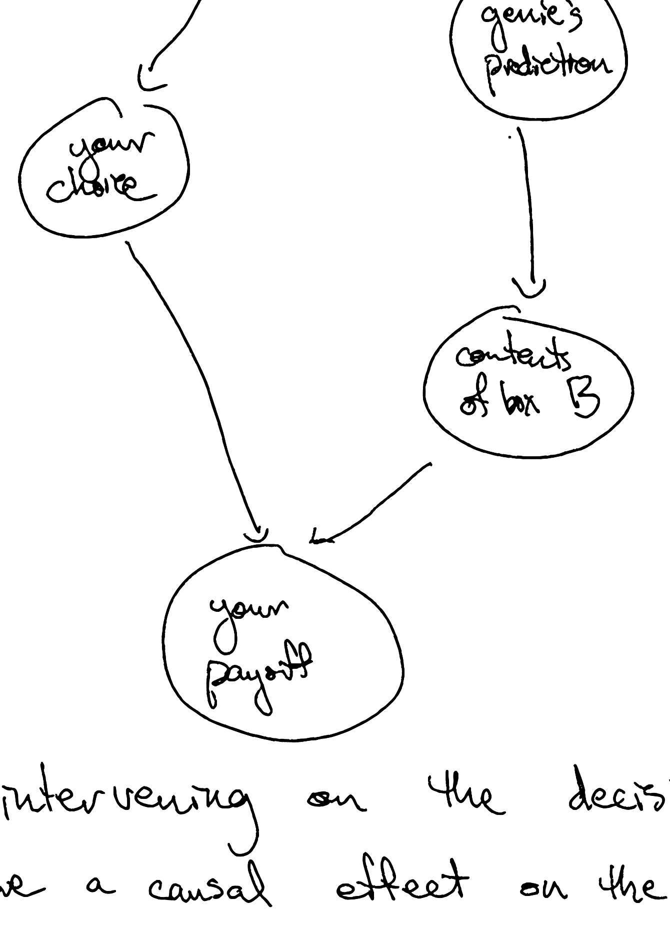
Thus CDT says yes to smoking, since it does not cause you to get cancer.

Rmk Intuitively, it seems that CDT gets Newcomb's problem wrong and smoking lesion right, and EDT does the reverse. We would like a decision theory that works in both cases

4. Secret option: causal effects of your policy

We'll get to this later in the course.

Here, we treat your decision node as being your choice of policy.



Here, intervening on the decision node does have a causal effect on the contents of box B.