Games of imperfect recall

Instructor: Vincent Conitzer
Monty Hall problem

• Game show participants can choose one of three doors

• One door has a car, two have a goat
  – Assumption: car is preferred to goat

• Participant chooses door, but not opened yet

• At least one of the other doors contains a goat; the (knowing) host will open one such door (flips coin to decide if both have goats)

• Participant is asked whether she wants to switch doors (to the other closed door) – should she?
Sleeping Beauty problem

• There is a participant in a study (call her Sleeping Beauty)
• On Sunday, she is given drugs to fall asleep
• A coin is tossed (H or T)
• If H, she is awoken on Monday, then made to sleep again
• If T, she is awoken Monday, made to sleep again, then again awoken on Tuesday

Sunday  Monday  Tuesday
H
T

Due to drugs she cannot remember what day it is or whether she has already been awoken once, but she remembers all the rules

You’re SB and you’ve just been awoken. What is your (subjective) probability that the coin came up H?
Information structure

- Nature
  - Heads
  - Tails

Monday: player 1

Tuesday
Dutch book against Halfer [Hitchcock’04]

- A Dutch book is a set of bets that someone with a particular belief system would each accept, but that in combination lead to a sure loss.

- Offer Beauty the following bet whenever she awakens:
  - If the coin landed Heads, Beauty receives 11
  - If it landed Tails, Beauty pays 10

- Argument: Halfer will accept, Thirder won’t.

- Also offer Beauty on Sunday:
  - If the coin lands Heads, Beauty will pay 12
  - If the coin lands Tails, Beauty will receive 13

- Argument: everyone will accept this one.

- If it’s Heads, Halfer Beauty will get $-12 + 11 = -1$

- If it’s Tails, Halfer Beauty will get $13 - 10 - 10 = -7$

- Guaranteed loss! 
  
  Same bet twice!
The betting game
(ignoring the Sunday bet)

Left=accept, Right= decline
Evidential decision theory

• Idea: when considering how to make a decision, should consider what it would tell you about the world if you made that decision.

• EDT Halfer: “With prob. ½, it’s Heads; if I accept, I will end up with 11. With prob. ½, it’s Tails; if I accept, then I expect to accept the other day as well and end up with -20. I shouldn’t accept.”

• As opposed to more traditional causal decision theory (CDT).

• CDT Halfer: “With prob. ½, it’s Heads; if I accept, it will pay off 11. With prob. ½, it’s Tails; if I accept, it will pay off -10. Whatever I do on the other day I can’t affect right now. I should accept.”

• EDT Thirder can also be Dutch booked.

• CDT Thirder and EDT Halfer cannot
  – [Draper & Pust’08, Briggs’10]

• EDTers arguably can in more general setting
  – [Conitzer’15]
Philosophy of “being present” somewhere, sometime

1: world with creatures simulated on a computer

2: displayed perspective of one of the creatures

simulated light (no direct correspondence to light in our world)

To get from 1 to 2, need additional code to:

A. determine in which real-world colors to display perception

B. which agent’s perspective to display

Is 2 more like our own experience than 1? If so, are there further facts about presence, perhaps beyond physics as we currently understand it?

See also: [Hare 2007-2010, Valberg 2007, Hellie 2013, Merlo 2016, …]