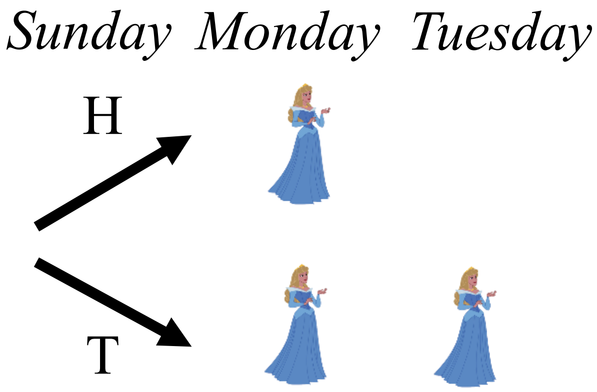
CS 15-784: Cooperative AI
Games of imperfect recall

The Sleeping Beauty problem [Elga'00]

- 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
- Due to drugs she **cannot remember what day it is or whether she has already been awoken once**, but she remembers all the rules
- Imagine **you** are SB and you've just been awoken. What is your (subjective) probability that the coin came up H?



note @24

25 views

Actions

(lack of) equilibria of imperfect recall games

We had a quick discussion about this in class today but it's maybe an interesting one to expand on a bit. Recapping a bit from class, consider the following game.

Tuomas is taking a penalty kick, Vince is the goalkeeper; except, unlike the regular (matching pennies) version, in the first round, Tuomas needs to decide whether to wear his left-kicking shoe or his right-kicking shoe; in the second round, Tuomas takes the penalty kick, having forgotten what kind of shoe he is wearing. Vince jumps left or right without knowing anything about either of Tuomas' choices. If Tuomas is wearing the wrong kind of shoe for the direction he chooses to kick in, he misses regardless of what Vince does (let's say he shoots over the goal). Otherwise, it's regular matching pennies: Tuomas scores if and only if he chooses the direction that Vince does not jump to.

What is an equilibrium of this game? In some sense, an equilibrium strategy for Tuomas is to half the time to be a left-kicking person, wearing his left-kicking shoe and kicking left; and half the time to be a right-kicking person. This is a mixed strategy (randomization over pure strategies), but not a behavioral strategy (there's no way to put probabilities *at the information sets* that make this work – Kuhn's theorem does not apply because it's an imperfect recall game). Indeed in some way it feels like the mixed strategy is "cheating" and not in the spirit of the imperfect recall of the problem.

Meanwhile, though, we can't get an equilibrium in behavioral strategies. The reason is that whatever Vince's strategy is, there is at least one pure strategy for Tuomas that scores with probability at least 1/2 – and pure strategies are still behavioral strategies. So in equilibrium, Tuomas would need to score at least 1/2 of the time, or he would deviate to that pure strategy. Meanwhile, whatever Tuomas' strategy is, it results in some distribution (l,r,o) reflecting how often he shoots left, right, over; WLOG $l \geq r$ and Vince chooses left, so that Tuomas scores only r of the time. But note that r can be $1/2$ only if $(l,r,o) = (1/2,1/2,0)$, and that is simply not possible with a behavioral strategy. So there is no equilibrium in behavioral strategies.

So how should you play?

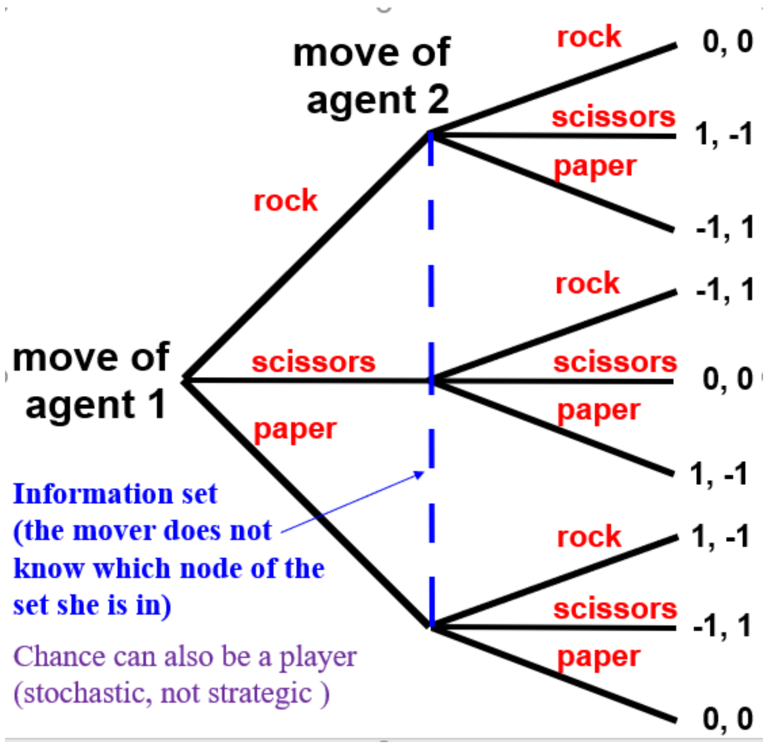
(One way to think about it is that Tuomas also shouldn't be able to deviate in a coordinated way across both of his two information sets, to the pure strategy, because that's also intuitively violating imperfect recall.)

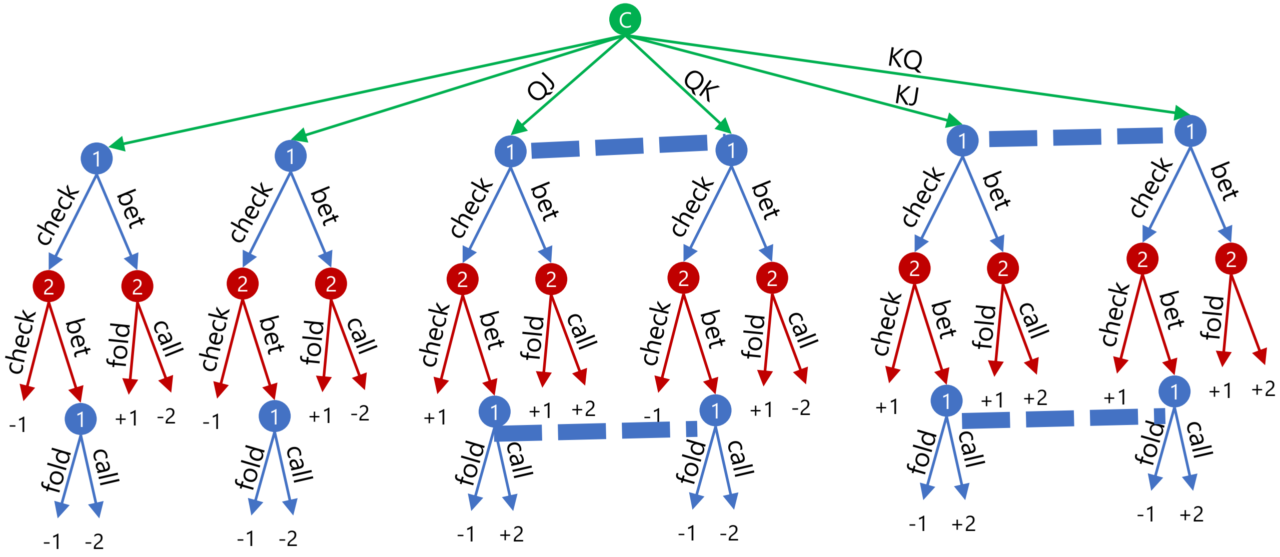
other

Edit good note | 1

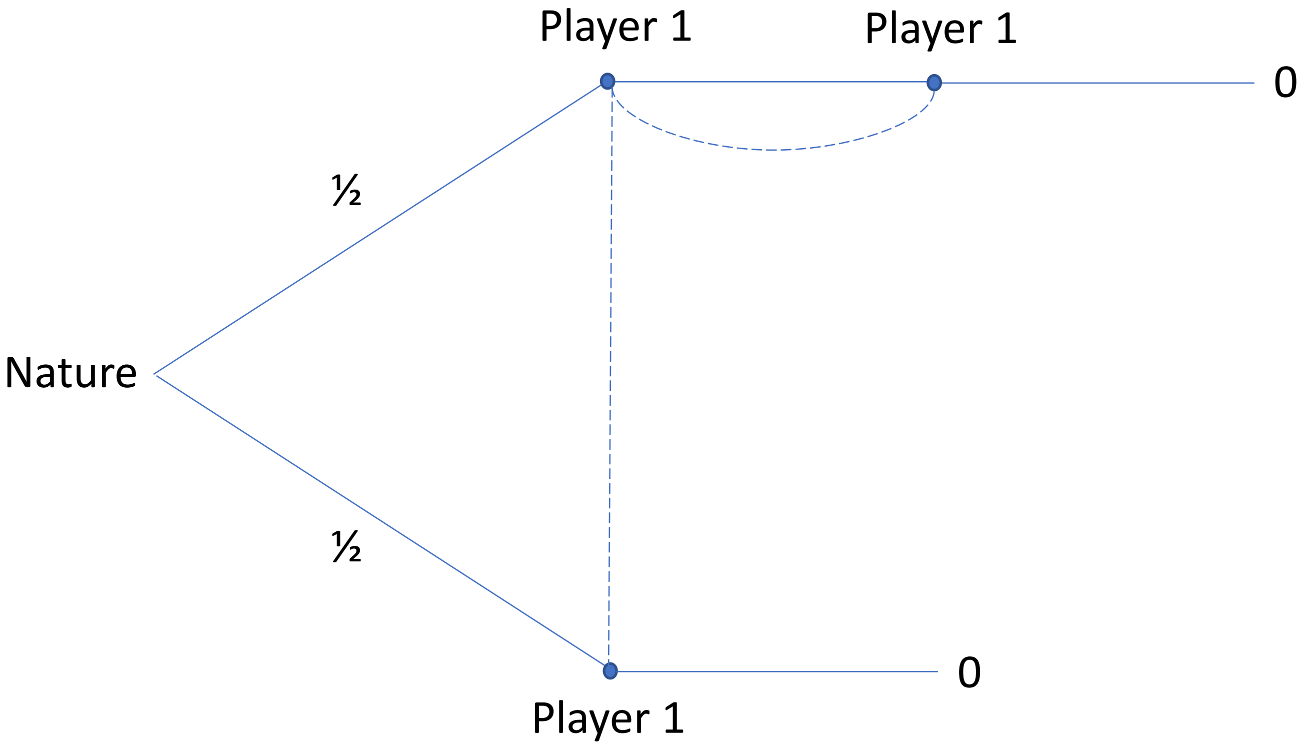
Updated 1 month ago by Vincent Conitzer

Recall: Extensive-form games



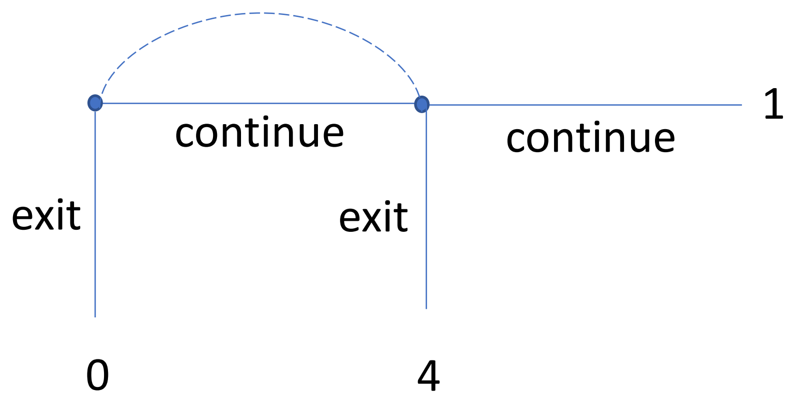


Sleeping Beauty as an extensive-form game



The Absent-Minded Driver (Piccione and Rubinstein 1997)

An individual is sitting late at night in a bar planning his midnight trip home. In order to get home he has to take the highway and get off at the second exit. Turning at the first exit leads into a disastrous area (payoff 0). Turning at the second exit yields the highest reward (payoff 4). If he continues beyond the second exit, he cannot go back and at the end of the highway he will find a motel where he can spend the night (payoff 1). The driver is absentminded and is aware of this fact. At an intersection, he cannot tell whether it is the first or the second intersection and he cannot remember how many he has passed (one can make the situation more realistic by referring to the 17th intersection). While sitting at the bar, all he can do is to decide whether or not to exit at an intersection. We exclude at this stage the possibility that the decision maker can include random elements in his strategy. Example 1 describes this situation.



(How does this relate to Sleeping Beauty?)

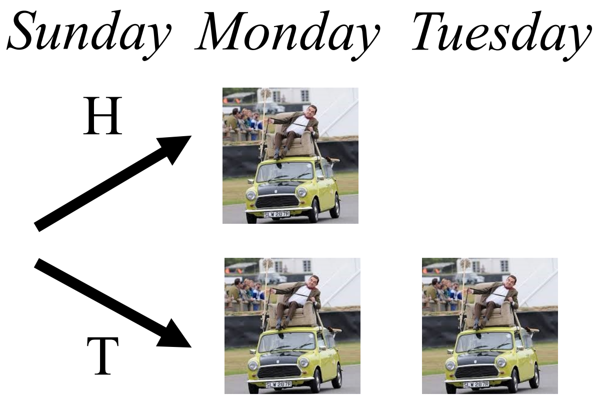
Agenda

- Why care
- Principles of decision making in games of imperfect recall
- Compatibility results

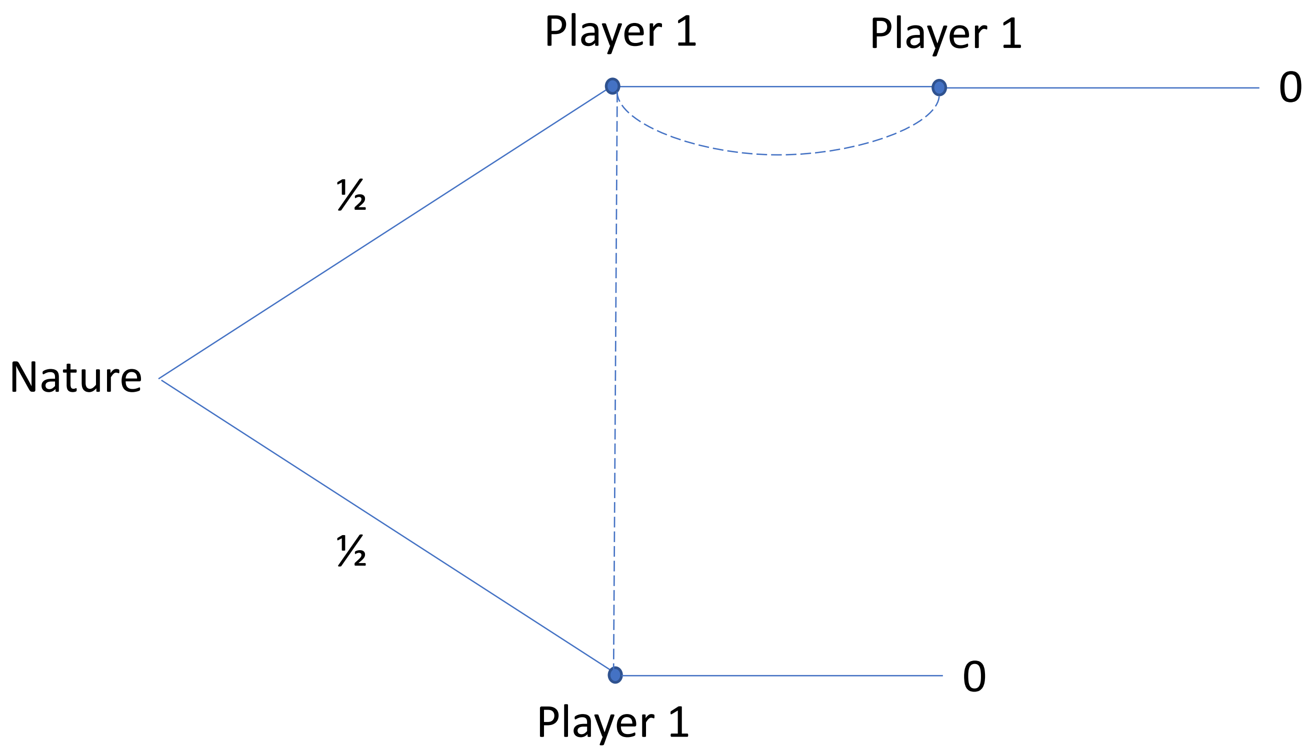
Why care?

Modern version

- Low-level autonomy cars with AI that intervenes when driver makes major error
- Does not keep record of such event
- Two types of drivers: Good (1 major error), Bad (2 major errors)
- Upon intervening, what probability should the AI system assign to the driver being good?
- (Similarly: half of households install a given AI system on two devices – with what probability does the AI system think it is alone? And what about simulation case from before?)



Games of imperfect recall without imperfect recall!

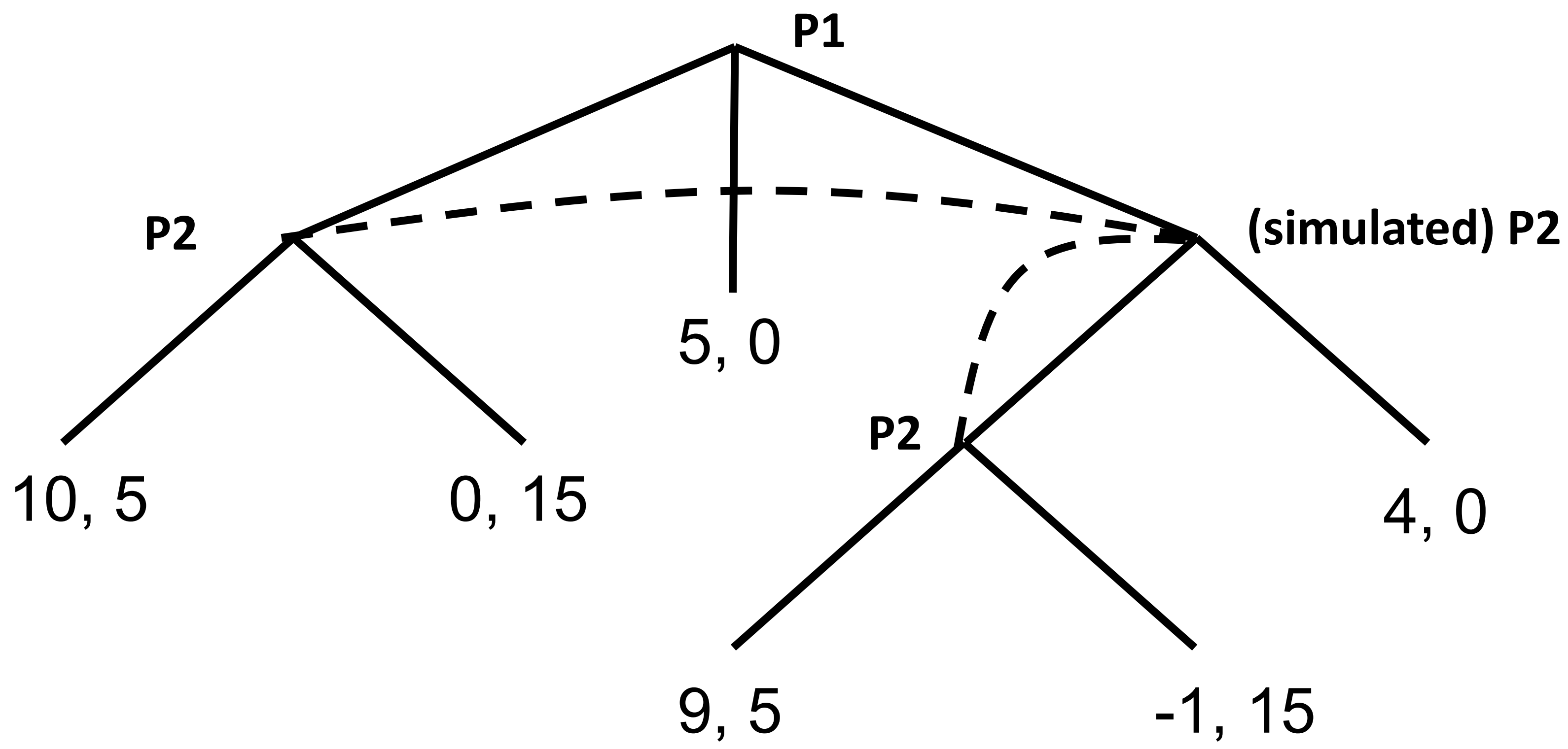


Anthropic arguments

- Bostrom's simulation argument
- The Doomsday argument
- Fine-tuning arguments
- Boltzmann brains
- Is the universe large or small?
(Many copies of Earth or just us?)

Simulating our way to cooperation?

- Restricted trust game: P1 can give 5 which would be tripled, or 0; after receiving 15, P2 can give back 10, or 0
- Twist: P1 can *simulate* P2 first, at a cost of 1



As (AI system) P2, how likely is it you're now running *as a simulation*? → *self-locating belief*

What happens in equilibrium?

Cooperation via ε-grounded simulation

(Oesterheld 2019)

ϵ-grounded Fair Bot (ϵGFB):
Input: opponent program p_{-i}, this program ϵGFB
Output: Cooperate or Defect

1: With probability ϵ:
2: **return** Cooperate
3: **return** $p_{-i}(\epsilon\text{GFB})$

For ε>0, εGFB cooperates against εGFB with probability 1.

(εGFB,εGFB) is a Nash equilibrium for sufficiently small ε.

Solution concepts for single-player games of imperfect recall

Ex ante-optimal policies for single-player games of imperfect recall

For any policy π and node s with predecessors $s_0 \ldots s_{n-1}$,

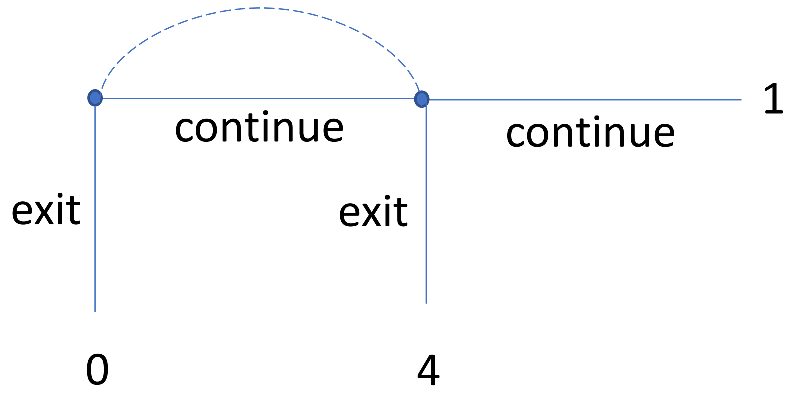
$$P(s \mid \pi) = \prod_{i=0}^{n-1} P(s_{i+1} \mid s_i, \pi) = \prod_{i=0}^{n-1} \sum_a P(s_{i+1} \mid s_i, a)\pi(a \mid s_i).$$

Thus,

$$E[u \mid \pi] = \sum_{s_t} P(s_t \mid \pi) u(s_t).$$

The *ex ante*-optimal strategies are

$$argmax_\pi \, E[u \mid \pi].$$



$$E[u \mid p_C] = p_C^2 + 4p_C(1 - p_C)$$

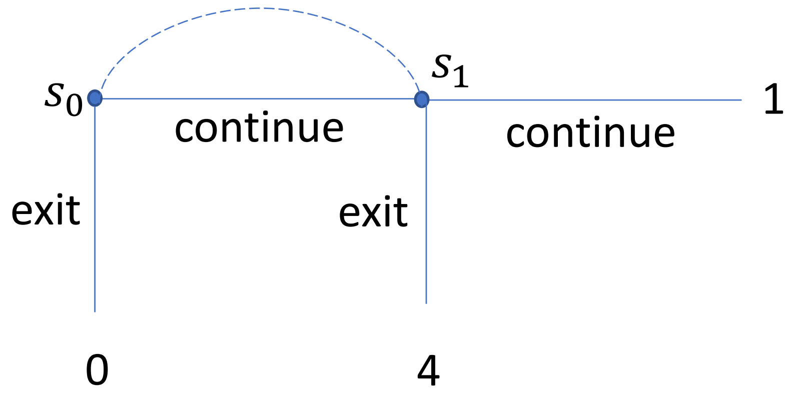
This is maximized at $p_C = \frac{2}{3}$.

Assigning probabilities *a la* Sleeping Beauty: *Generalized Thirding* (a.k.a. self-indication assumption)

If $s \notin I$, then $P_{GT}(\, s \mid I, \pi \,) = 0$. Else

$$P_{GT}(\, s \mid I, \pi \,) = \frac{P(\, s \mid \pi \,)}{\sum_{s' \in I} P(s' \mid \pi)}.$$

Intuition:
Imagine you play the game repeatedly (with π).
Then of all the times I is observed, what fraction is s.

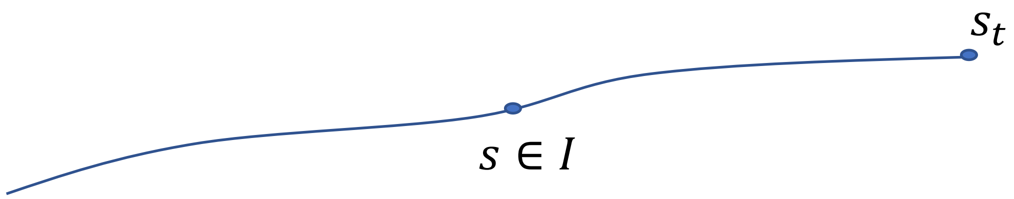


$P_{GT}(s_1 \mid p_C, I) = \frac{p_C}{1+p_C}$.

For $p_C = \frac{1}{2}$ (as in Sleeping Beauty),

$$P_{GT}(\, s_1 \mid p_C, I \,) = \frac{1}{3}.$$

Assigning probabilities – (double-)halfing
(a.k.a. the minimum-reference class self-sampling assumption)



Let s_t be a leaf and $s \in I$ be a player node on the way to s_t. Then

$$P_{GDH}(s_t \mid I, \pi) := P(s_t \mid \pi)/P(I \mid \pi).$$

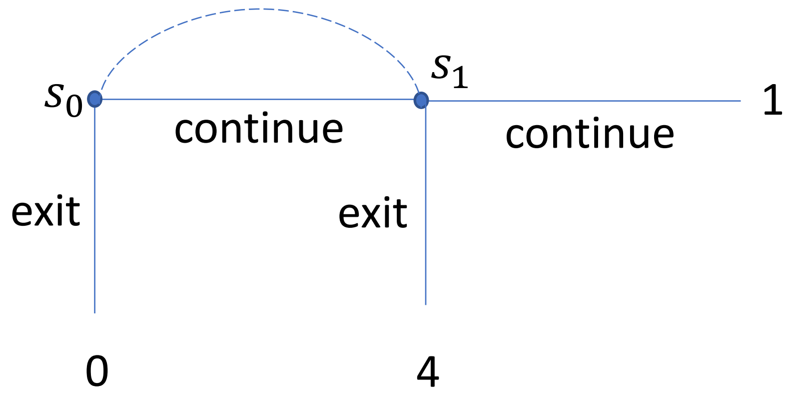
The probability of being at s in particular can be defined by

$$P_{GDH}(s, s_t \mid I, \pi) \coloneqq \frac{P_{GDH}(s_t \mid I, \pi)}{\#(I, s_t)}$$

and

$$P_{GDH}(s \mid I, \pi) \coloneqq \sum_{s_t} P_{GDH}(s, s_t \mid I, \pi)$$

Intuition:
Update about full histories (s_t) by updating only on the fact that I is observed at all.



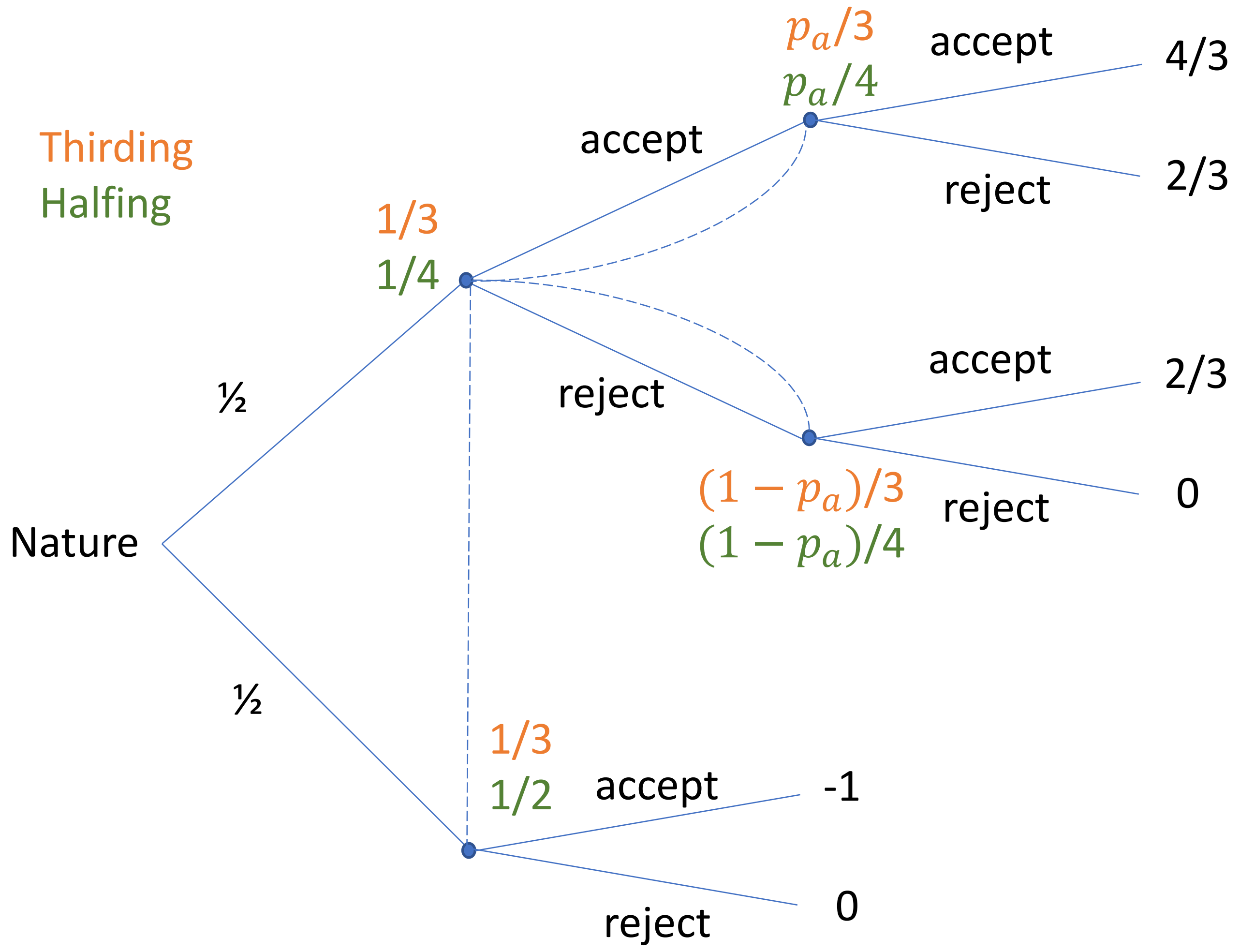
$P_{GDH}(s_1 \mid p_C, I) \sim \frac{p_C^2}{2} + \frac{p_C(1-p_C)}{2} = \frac{p_C}{2}$.
(Because $P(I) = 1$, we don't need to renormalize.)

For $p_C = \frac{1}{2}$ (as in Sleeping Beauty),

$$P_{GDH}(s_1 \mid p_C, I) = \frac{1}{4}.$$

Betting in Sleeping Beauty Problem

Imagine you are in the Sleeping Beauty scenario, but:
Every time you wake up, you are offered a bet that loses \$1 if the coin came up Heads and pays \$2/3 if the coin came up Tails.



Ex ante-optimal: accept

$$EU_{CDT+GT}(accept \mid \pi) - EU_{CDT+GT}(reject \mid \pi) = \frac{1}{3}\cdot(-1) + \frac{2}{3}\cdot\frac{2}{3} = \frac{1}{9}$$

$$EU_{CDT+GDH}(accept \mid \pi) - EU_{CDT+GDH}(reject \mid \pi) = \frac{1}{2}\cdot(-1) + \frac{1}{2}\cdot\frac{2}{3} = -\frac{1}{6}$$

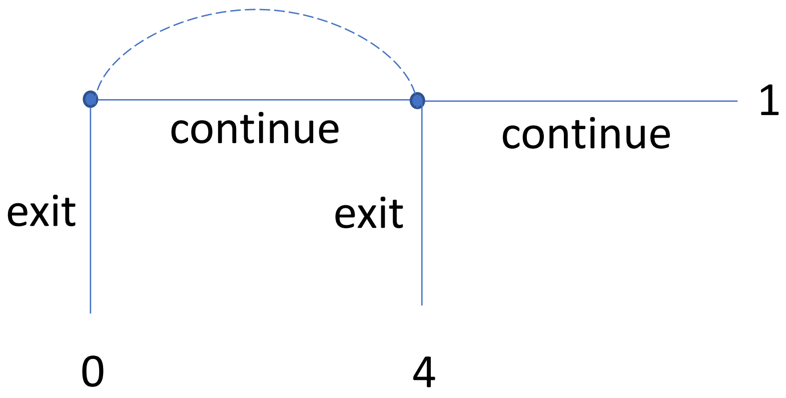
$$EU_{EDT+GT}(accept \mid \pi) = \frac{1}{3}(-1) + \frac{2}{3}\cdot\frac{4}{3} = \frac{5}{9}$$

$$EU_{EDT+GDH}(accept \mid \pi) = \frac{1}{2}(-1) + \frac{1}{2}\cdot\frac{4}{3} = \frac{1}{6}$$

A version of evidential decision theory for games of imperfect recall

Upon observing I maximize over α:

$$EU_{EDT+x}(\alpha \mid \pi, I) = \sum_{s_t} P_x(s_t \mid I, \pi_{I\to\alpha}) u(s_t)$$

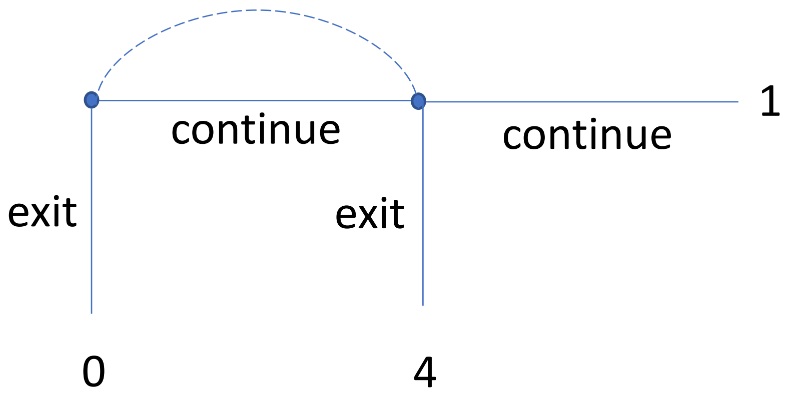


$$\begin{aligned} &EU_{EDT+GDH}(\alpha \mid \pi, I) \\ &= \sum_{s_t} P_{GDH}(s_t \mid I, p_c)u(s_t) \\ &= \sum_{s_t} P(s_t \mid I, p_c)u(s_t) \\ &= E[u \mid p_C]. \end{aligned}$$

A version of causal decision theory for single-player extensive-form games of imperfect recall

Upon observing I maximize over α:

$$EU_{CDT+x}(a \mid \pi, I) = \sum_s P_x(s \mid I, \pi)\, P(s_t \mid s, a, \pi) u(s_t)$$



$$EU_{CDT+GT}(c \mid \pi, I) = \frac{1}{1+p_C} \cdot (p_C + 4(1-p_C)) + \frac{p_C}{1+p_C}$$

$$EU_{CDT+GT}(e \mid \pi, I) = \frac{p_C}{1+p_C} \cdot 4$$

$EU_{CDT+GT}(c \mid \pi, I) = EU_{CDT+GT}(e \mid \pi, I)$ solves to $p_C = \frac{2}{3}$.

Compatibility

Theorem (Piccione and Rubinstein 1997): In any single-player extensive-form game, the *ex ante*-optimal policy is compatible with CDT + GT.

Theorem (Oesterheld and Conitzer 2022): In any single-player extensive-form game, the *ex ante*-optimal policy is compatible with EDT + GDH.

The other combinations don’t work (e.g., Briggs 2010)!

Even for CDT+GT and EDT+GDH (as defined here) they may be many compatible policies.

References

- Piccione and Rubinstein. 1997. On the Interpretation of Decision Problems with Imperfect Recall. In: Games and Economic Behavior 20, pp. 3–24.
- Briggs. 2010. Putting a value on Beauty. In: Oxford Studies in Epistemology. Vol. 3. Oxford University Press, pp. 3–24.
- Oesterheld. 2019. Robust Program Equilibrium. Theory and Decision, 86(1): 143–159.
- Oesterheld and Conitzer. 2022. Can de se choice be ex ante reasonable in games of imperfect recall? https://www.andrew.cmu.edu/user/coesterh/DeSeVsExAnte.pdf