6.S890: Topics in Multiagent Learning

Lecture 16 – Prof. Farina

Equilibrium Perfection and Team Strategies

Fall 2023

Are all Nash equilibria equally good? Or should we aim for some but not other?

Not all Nash equilibria are equally sensible, especially in sequential games!

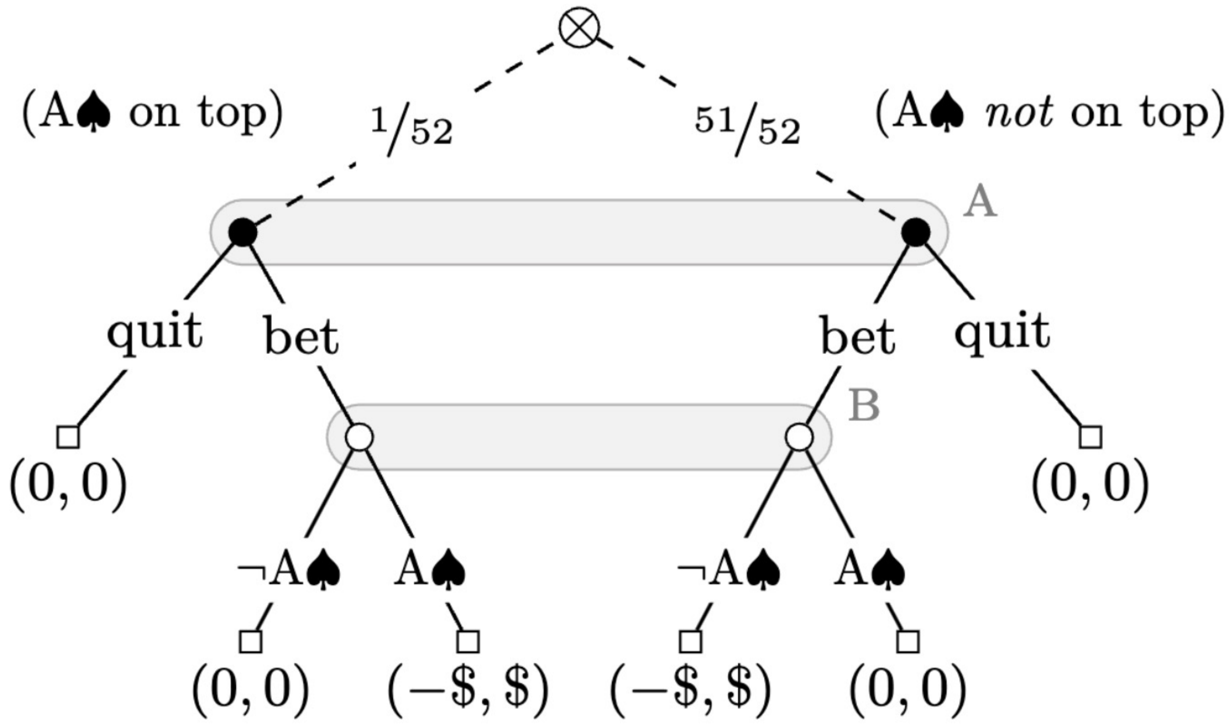
Intuition: Nash equilibria stem from the idea that the opponent is as strong as possible, and might therefore be completely unprepared to handle the case of an imperfect opponent

Very relevant when playing with humans!

Guess-the-Ace game

To make the discussion more concrete, consider the following game (due to Miltersen and Sorensen)

- At the start a standard 52-card deck is perfectly shuffled, face down, by a dealer
- Then, Player 1 decides whether to immediately end the game (no money transfer), or offer $1000 to Player 2 if they can correctly guess whether the top card of the shuffled deck is the ace of spaces or not.
- If Player 2 guesses correctly, the $1000 get transferred from Player 1 to Player 2; if not, no money is transferred

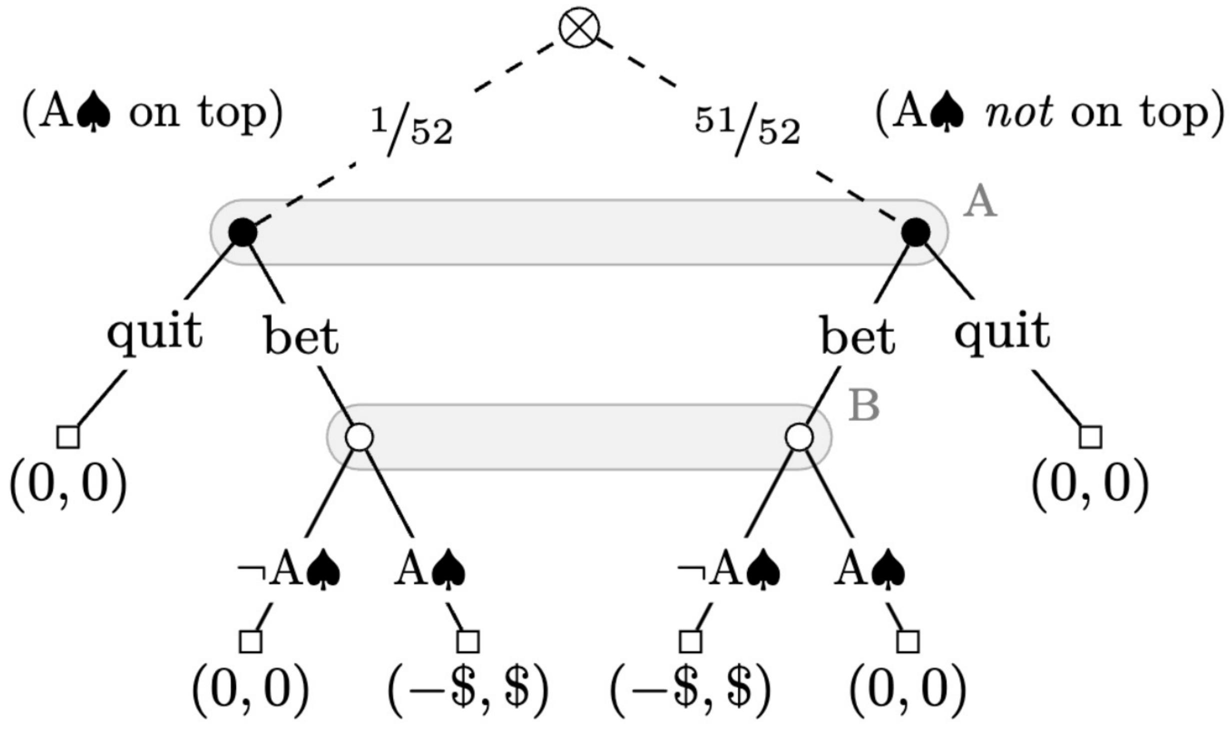


Guess-the-Ace game

To make the discussion more concrete, consider the following game (due to Miltersen and Sorensen)

- At the start a standard 52-card deck is perfectly shuffled, face down, by a dealer
- Then, Player 1 decides whether to immediately end the game (no money transfer), or offer $1000 to Player 2 if they can correctly guess whether the top card of the shuffled deck is the ace of spaces or not.
- If Player 2 guesses correctly
trans
mo

Q: As Player 1, what is the only sensible way to play the game?



Answer: the only sensible thing for Player 1 to do is to quit immediately
(anything else loses money to Player 1 in expectation)

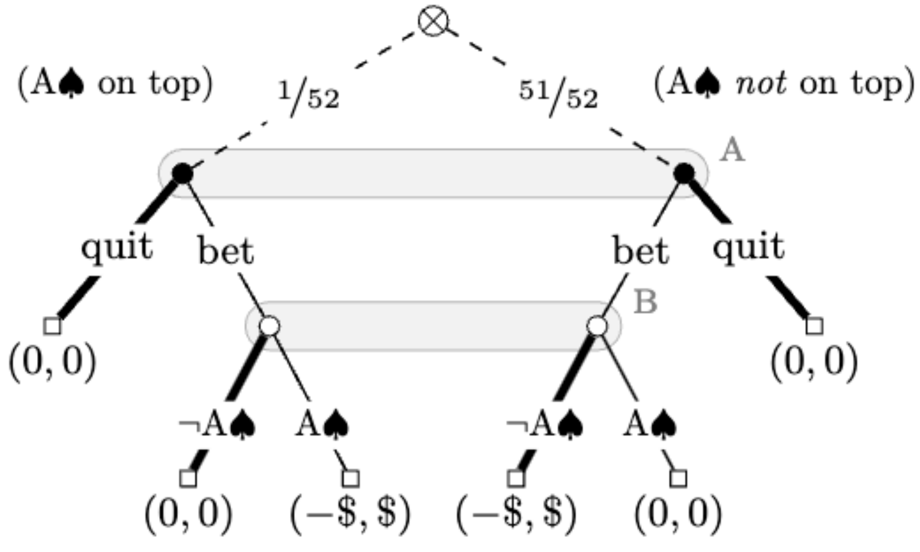
Indeed, that is the only Nash equilibrium strategy for Player 1.

Guess-the-Ace game

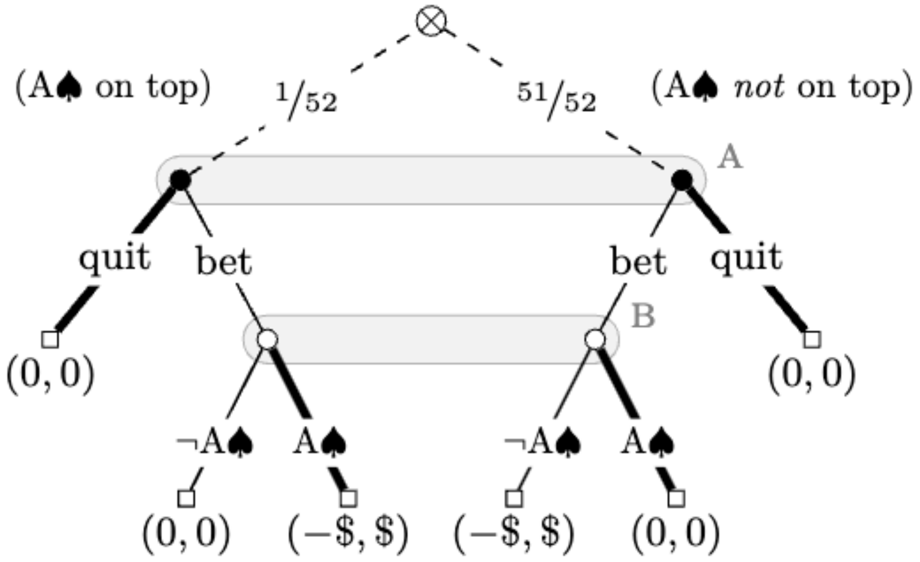
But then, Player 2 does not get to play. From the point of view of the definition of Nash equilibrium, anything that Player 2 does is a Nash equilibrium strategy

Yet, huge difference between the strategies. Only one of the two approaches can be called "rational"

Both of these are Nash equilibria. Nash eq. does not distinguish between the two



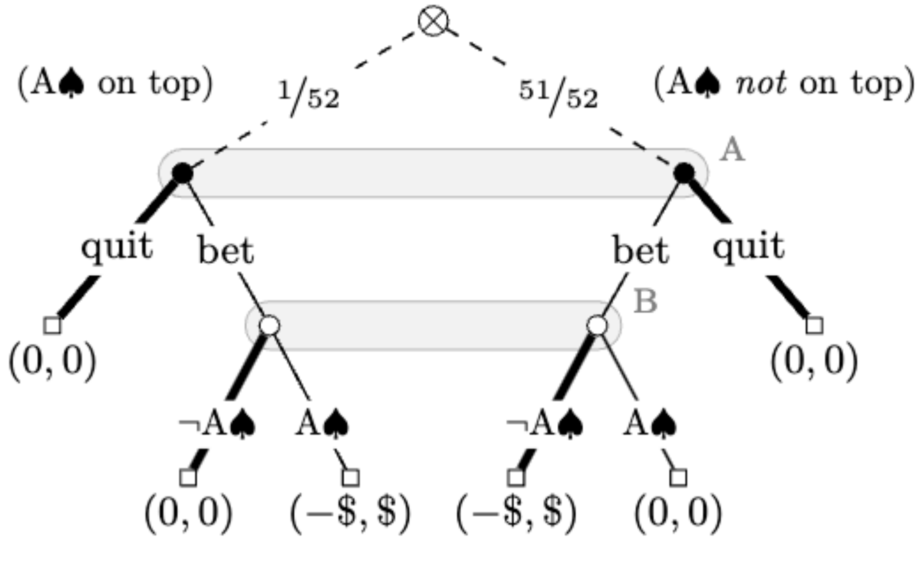
Sensible Nash equilibrium



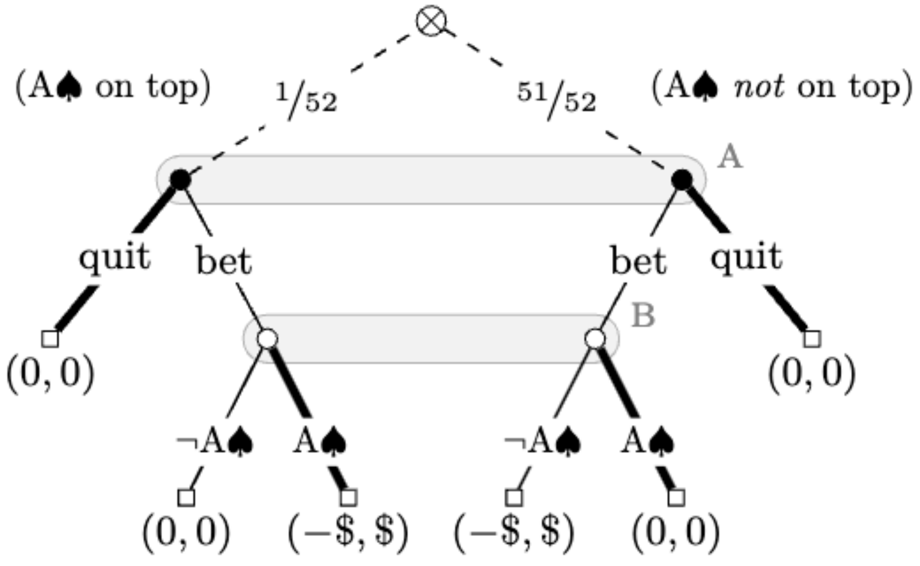
Questionable Nash equilibrium

Imagine that Player 2 is a **bot** playing against opponents in the real world, blindly following the Nash equilibrium strategy it has precomputed

If Player 1 makes a mistake and decides to offer the $1000 instead of immediately quitting, the Nash equilibrium that bets that the top card is not the ace of space has an expected utility of > $980 whereas the Nash equilibrium that bets that the top card is the ace of spades only has an expected utility of < $20.

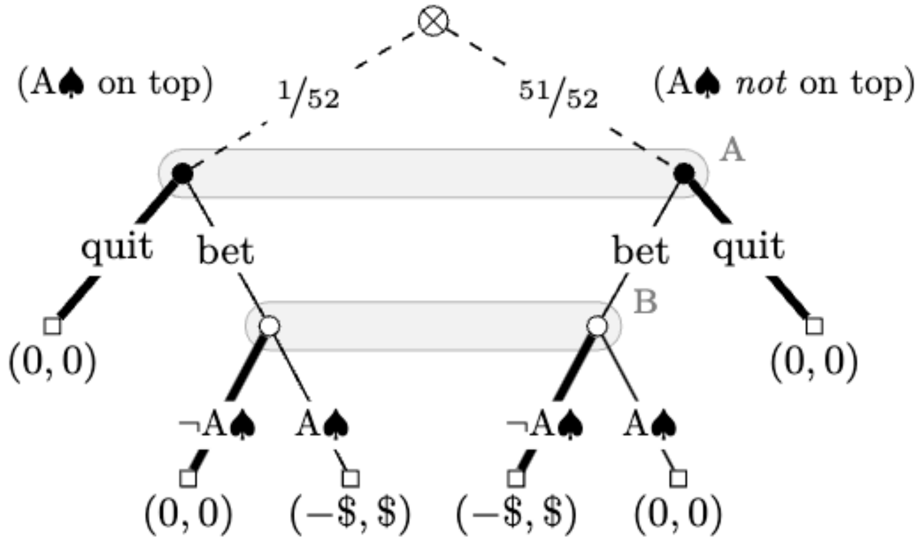


Sensible Nash equilibrium

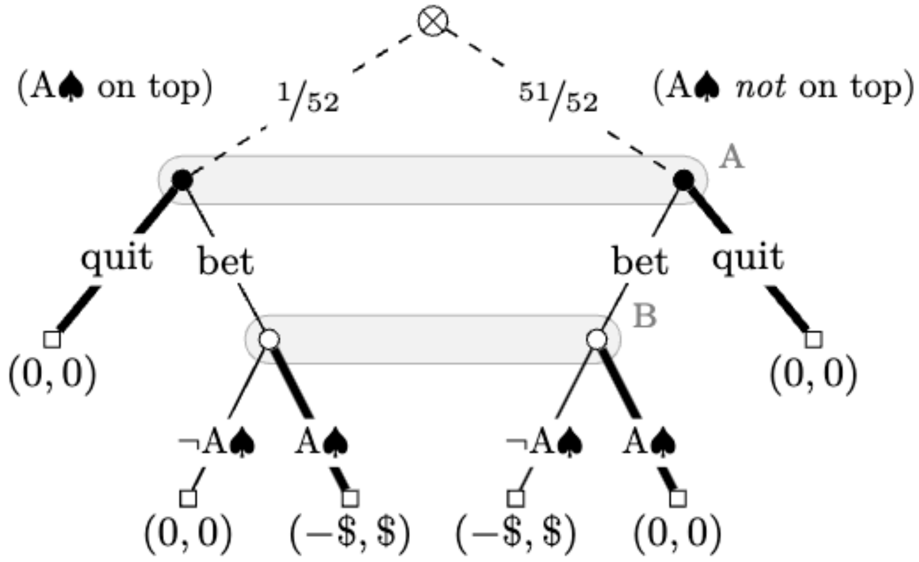


Questionable Nash equilibrium

Formalizing this subtle notion of rationality within the set of Nash equilibria has been a major endeavor for the game-theoretic literature in the 70s and 80s. Today, we say that the equilibrium in Figure 1 (Left) is **sequentially irrational**, while the one on the right is sequentially rational.



Sensible Nash equilibrium



Questionable Nash equilibrium

Formalizing this subtle notion of rationality within the set of Nash equilibria has been a major endeavor for the game-theoretic literature in the 70s and 80s. Today, we say that the equilibrium in Figure 1 (Left) is **sequentially irrational**, while the one on the right is sequentially rational.

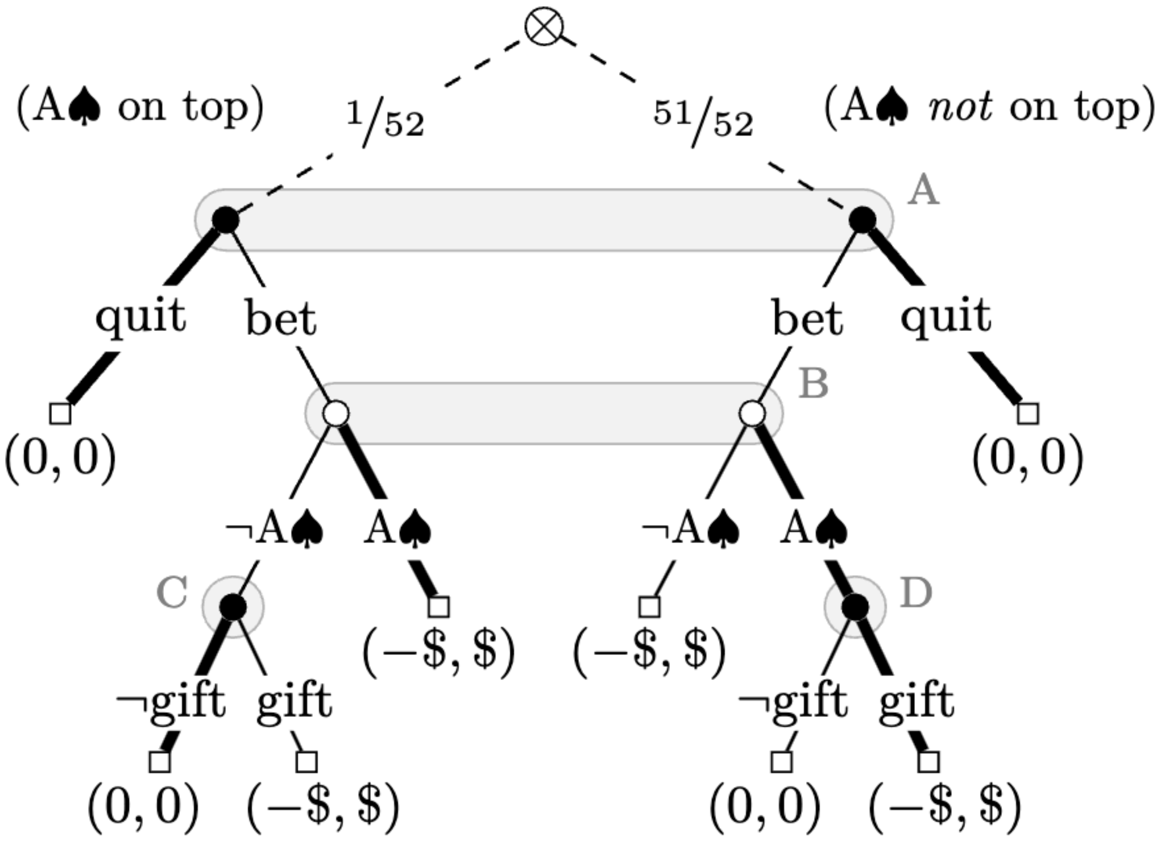
Not all Nash equilibria are equally "good" when the agents can make mistakes.
Sequentially-irrational Nash equilibria might leave value on the table, by being incapable of capitalizing on opponents' mistakes

Trivia: this kind of surprising behavior kicked in during the poker tournament with the pros, and people were worried there was possibly a bug in the bot. Instead, it was likely the pro that had made a mistake and entered an off-equilibrium part of the tree

Undomination

One might believe that the problem of sequential irrationality is that of picking **dominated** strategies

While it is true that restricting to undominated strategies fixes the previous example, this is not a general fix!

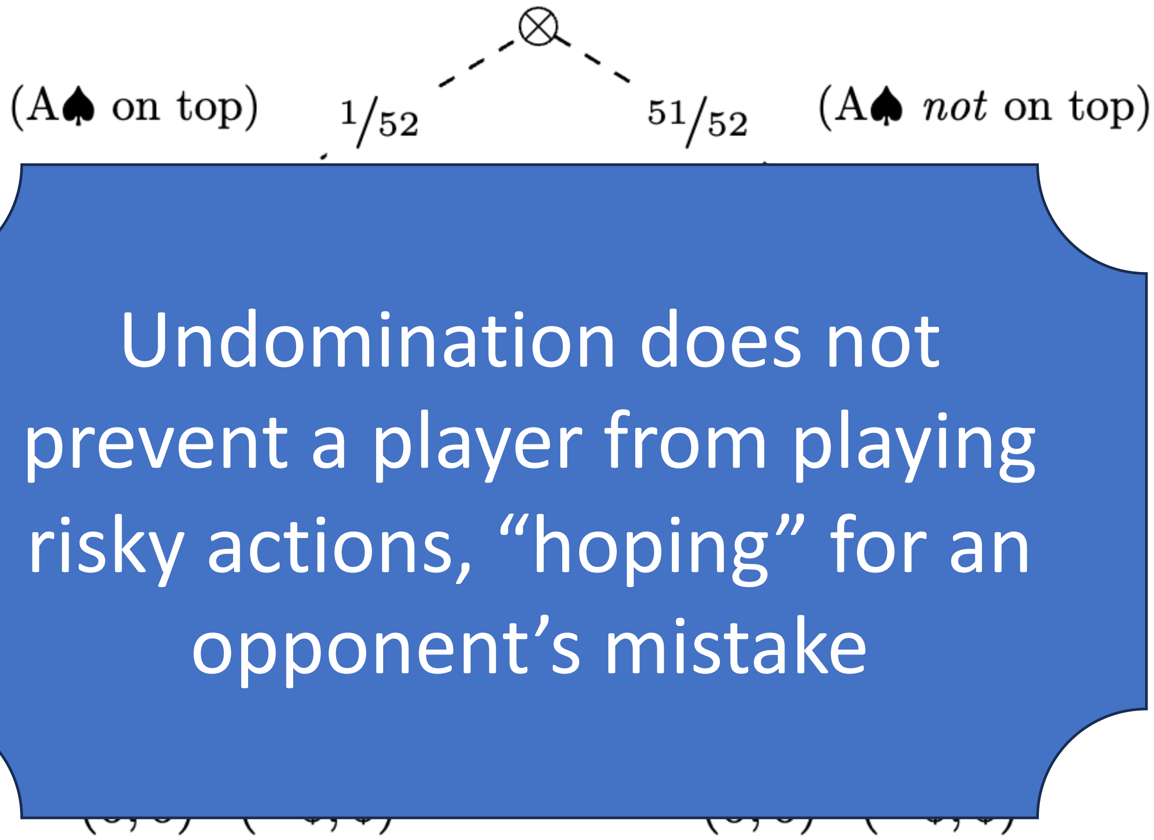


Questionable *undominated* Nash equilibrium

Undomination

One might believe that the problem of sequential irrationality is that of picking **dominated** strategies

While it is true that restricting to undominated strategies fixes the previous example, this is not a general fix!



Undomination does not prevent a player from playing risky actions, “hoping” for an opponent’s mistake

Questionable *undominated* Nash equilibrium

Nonetheless, **undomination tends to perform well,** better than unrefined Nash equilibrium in practice

Furthermore, undominated equilibrium is not very expensive to compute!

Computation of Undominated Nash Equilibrium

- Idea: two-step solution
- First step: compute expected utility at equilibrium (aka game value)
- Second step: out of all strategies that guarantee utility >= game value, select the one that performs the best against the uniform strategy of the opponent

Second linear program to search on the set of equilibrium strategies to find a “robust” one

Computation of Undominated Nash Equilibrium

- Idea: two-step solution
- First step: compute expected utility at equilibrium (aka game value)
- Second step: out of all strategies that guarantee utility >= game value, select the one that performs the best against the uniform-random strategy of the opponent
  - Why the uniform strategy?
  - Idea: uniform strategies explore all of the game tree. Furthermore, they are probably pretty bad strategies.
  - We are counterbalancing the fact that we are looking for equilibrium strategies, with the fact that we want the strategy that does best against a uniform-random opponent.

Recall: Nash as LP

$$Q_1 = \begin{cases} F_1 x = f_1 \\ x \geq 0 \end{cases} \qquad Q_2 = \begin{cases} F_2 y = f_2 \\ y \geq 0 \end{cases}$$

$$\max_{x \in Q_1} \min_{y \in Q_2} x^T A y \xrightarrow{1} \begin{cases} \max & \boxed{\begin{cases} \min & x^T A y \\ \text{s.t.} & y \in Q_2 \end{cases}} \\ \text{s.t.} & x \in Q_1 \end{cases}$$

$$\xrightarrow{2} \begin{cases} \max & \boxed{\begin{cases} \min & x^T A y \\ \text{s.t.} & F_2 y = f_2 \\ & y \geq 0 \end{cases}} \\ \text{s.t.} & F_1 x = f_1 \\ & x \geq 0 \end{cases}$$

3 Dualize!

$$\begin{cases} \max & \boxed{\begin{cases} \max & f_2 v \\ \text{s.t.} & F_2^T v \leq A^T x \\ & v \in \mathbb{R} \end{cases}} \\ \text{s.t.} & F_1 x = f_1 \\ & x \geq 0 \end{cases}$$

$$\xrightarrow{4} \begin{cases} \max & f_2 v \\ \text{s.t.} & F_1 x = f_1 \\ & F_2^T v \leq A^T x \\ & x \geq 0 \\ & v \in \mathbb{R} \end{cases}$$

Single linear program!

We compute the game value by solving this LP

Undominated Nash equilibrium (for Pl. 1)

Step 1

$$\gamma := \begin{cases} \max & f_2 v \\ & F_1 x = f_1 \\ \text{s.t.} & F_2^T v \le A^T x \\ & x \ge 0 \\ & v \in \mathbb{R} \end{cases}$$

Game value

Step 2

$$\gamma := \begin{cases} \max & x^\top (Au) \\ & f_2 v \ge \gamma \\ & F_1 x = f_1 \\ \text{s.t.} & F_2^T v \le A^T x \\ & x \ge 0 \\ & v \in \mathbb{R} \end{cases}$$

Uniform strategy of Player 2

Second linear program to search on the set of equilibrium strategies to find a "robust" one

If Undominated Nash Equilibrium is not a solution to sequential irrationality, what is?

Fundamentally, the issue of sequential irrationality stems from the fact that **some parts of the game tree are unreachable at equilibrium**.

IDEA: to avoid sequential irrationality, **force all players to explore the whole game** tree by imposing that they **must pick all actions with some positive, vanishing** probability

"Trembling-hand" equilibria

These equilibria guarantee sequential rationality

Two approaches

Trivia: "quasi-perfect" turns out to have stronger properties than "perfect" equilibrium. Naming is hard!

Extensive-Form Perfect Equilibrium	Quasi-Perfect equilibrium
Force **every action in the game** to be played with probability at least $\epsilon > 0$	For any d, force **every sequence of d actions from the root** to be played with probability at least ϵ^d for some $\epsilon > 0$
Take **any limit point** of Nash equilibria of the constrained games as $\epsilon \downarrow 0$	Take **any limit point** of Nash equilibria of the constrained games as $\epsilon \downarrow 0$

Relationship between the equilibria

We have already observed that **undomination does not imply sequential rationality**

Interestingly, the converse is also not true in general, so **sequential rationality and undomination are incomparable** (neither implies the other)

Natural question: is there an equilibrium that achieves both sequential rationality and undomination?

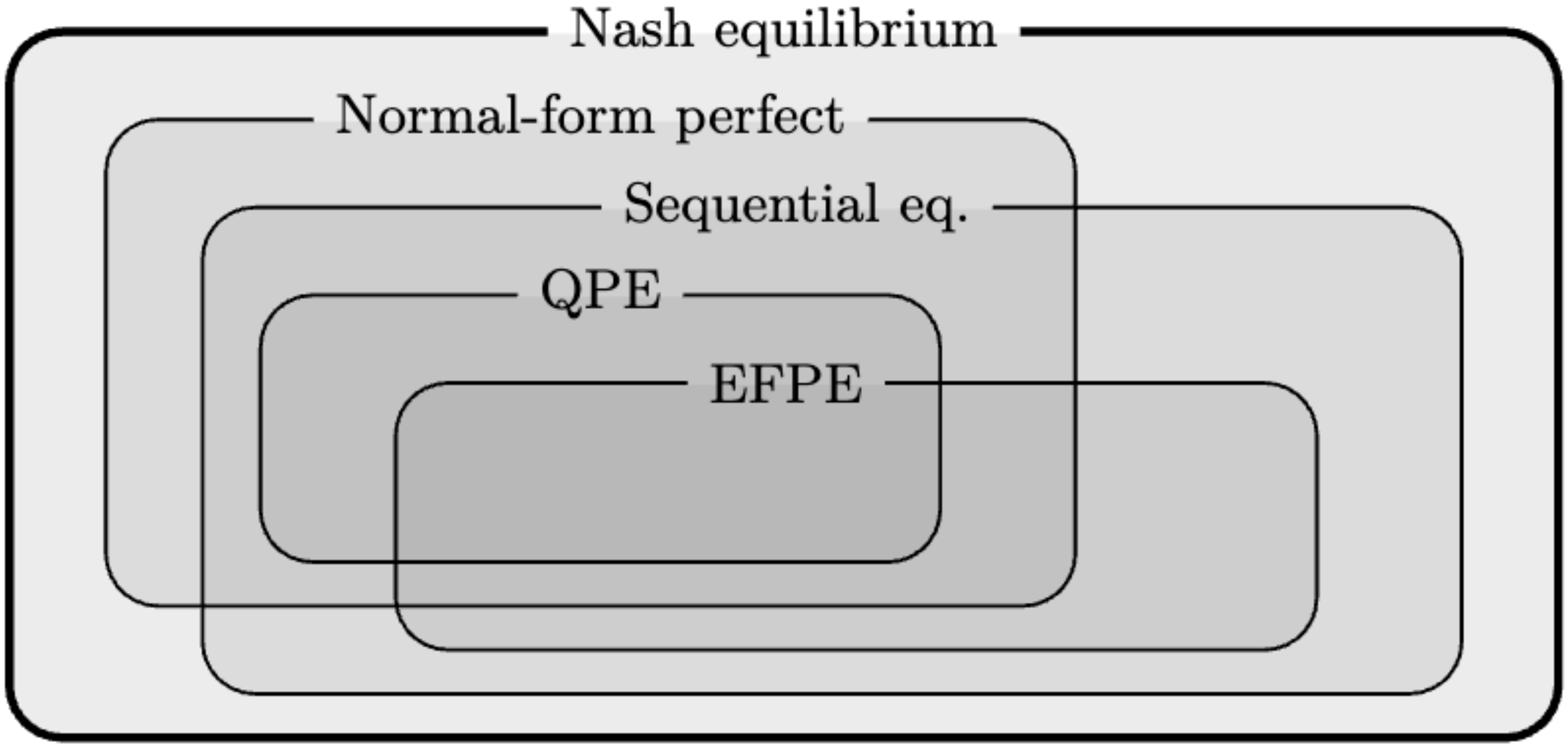
Is there an equilibrium that achieves both sequential rationality and undomination?

YES! A nice property of quasi-perfect equilibrium is that not only it is sequentially rational, but it is also undominated

For this reason, as Mertens noted in 1995, **quasi-perfect equilibrium is nowadays considered superior to extensive-form perfect equilibrium**

> *Observe that the "quasi-perfect" equilibria [..] are still sequential—and sequential equilibria have all backward-induction properties (e.g., Kohlberg and Mertens, 1986)—but are at the same time normal form perfect—which can be viewed as the strong version of undominated. (And every proper equilibrium is quasi-perfect.) Thus, by some irony of terminology, the "quasi"-concept seems in fact far superior to the original unqualified perfection itself.*

Venn diagram of equilibria



How hard is it compute a sequentially—rational equilibrium?

Cool fact: in theory, it's not harder than Nash!

Solution concept	General-sum	Zero-sum
Nash (NE)	PPAD-complete [Daskalakis et al., 2009]	FP [Romanovskii, 1962] [von Stengel, 1996]
Quasi Perfect (QPE)	PPAD-complete [Miltersen and Sørensen, 2010]	FP [Miltersen and Sørensen, 2010]
Extensive-Form Perfect (EFPE)	PPAD-complete [Farina and Gatti, 2017]	FP [Farina and Gatti, 2017]

In practice, unfortunately we are still far behind, due to the added practical intricacy of forcing exploration and taking the limit as exploration vanishes

Intuition for the computation

Extensive-Form Perfect Equilibrium

Force **every action in the game** to be played with probability at least $\epsilon > 0$

Quasi-Perfect equilibrium

For any d, force **every sequence of d actions from the root** to be played with probability at least ϵ^d for some $\epsilon > 0$

Computation of Equilibrium Refinements

Computing extensive-form perfect equilibria and quasi-perfect equilibria requires solving games in which the strategy sets are perturbed

Luckily, we can still use the sequence-form representation to capture these constraints

Extensive-form perfect equilibrium

$$\boldsymbol{x}[ja] \geq \begin{cases} \epsilon & \text{if } p_j = \varnothing \\ \epsilon \cdot \boldsymbol{x}[p_j] & \text{otherwise.} \end{cases}$$

Quasi-perfect equilibrium

$$x[ja] \geq \epsilon^{\mathrm{depth}(ja)}$$

Hence, for both equilibrium types, we need to compute the limit as $\epsilon \to 0$ of solutions to problems of the form

$$\max_{x \in Q_1(\epsilon)} \min_{y \in Q_2(\epsilon)} x^T A y$$

Polytopes that depend on the "trembling" amount (perturbation) ϵ

For any given ϵ, we can still use linear programming to compute those solutions

Problem: how do we compute the limit?

Computation of Equilibrium Refinements

We need to compute a limit as $\epsilon \to 0$ of solutions to a **trembling LP**

$$P(\epsilon):\begin{cases}\max & \boldsymbol{c}(\epsilon)^\top \boldsymbol{x}\\ \text{s.t.} & \mathbf{A}(\epsilon)\,\boldsymbol{x}=\boldsymbol{b}(\epsilon)\\ & \boldsymbol{x}\geq \mathbf{0},\end{cases}$$

Idea: there exists $\epsilon^* > 0$—called a negligible positive perturbation (NPP)—such that for all $0 < \epsilon \leq \epsilon^*$, the optimal vertex for $P(\epsilon)$ remains the same. Furthermore, such a value ϵ^* can be computed in polynomial time in the input size

Conceptual algorithm

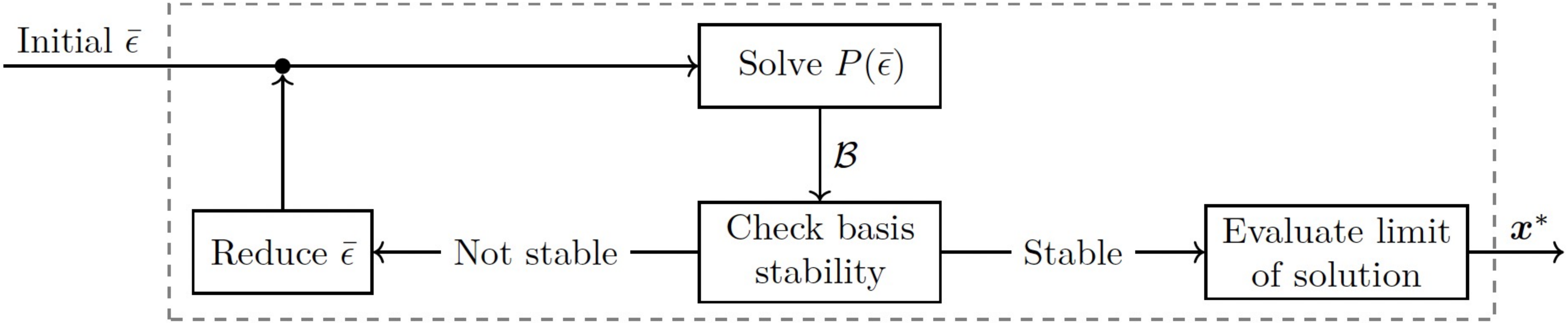
$$P(\epsilon) : \begin{cases} \max & \boldsymbol{c}(\epsilon)^\top \boldsymbol{x} \\ \text{s.t.} & \mathbf{A}(\epsilon)\, \boldsymbol{x} = \boldsymbol{b}(\epsilon) \\ & \boldsymbol{x} \geq \mathbf{0}, \end{cases}$$

- First, compute the value of ϵ^*
- Then, solve the numerical linear program $P(\epsilon^*)$ to optimality. Since the bit complexity of ϵ^* is polynomial in the size of the trembling LP, the numerical LP can be solved to optimality in polynomial time, and a vertex (basis) can be extracted.
- Finally, extract the limit solution to the trembling LP by evaluating the vertex in $\epsilon = 0$.

In practice, the perturbation ϵ^* is impractically small -> needs rational-precision simplex algorithm

A More Scalable Algorithm

$$P(\epsilon) : \begin{cases} \max & \boldsymbol{c}(\epsilon)^\top \boldsymbol{x} \\ \text{s.t.} & \mathbf{A}(\epsilon)\,\boldsymbol{x} = \boldsymbol{b}(\epsilon) \\ & \boldsymbol{x} \geq \mathbf{0}, \end{cases}$$



Team equilibria in two-player zero-sum games

So far, we have always looked at individual players that seek to maximize their own utility

On the other hand, many realistic interactions require studying correlated strategies.

As an example, consider *a team of players that collude at a poker table*: the optimal strategy for the team is to strategically coordinate their moves, and as a result the colluding players will not play according to independent strategies

What does equilibrium computation look like in team settings?

Three Models of Team Coordination

Case 1:

Team members can freely (and privately) communicate during play

…the team effectively becomes a single player. Hence, all the tools we've seen so far (for example, learning an optimal strategy using CFR or linear programming) directly apply. We will refer to this equilibrium as "Team Nash equilibrium"

Three Models of Team Coordination

Case 2:

Team members cannot communicate at all

Then, the strategies of the team members should be picked as the pair of strategies that maximizes the expected utility of the team against a best-responding agent. This solution concept is called a team maxmin equilibrium (TME)

Three Models of Team Coordination

<u>Case 2:</u>

Team members cannot communicate at all

Then, the strategies of the team members sh
of strategies that maximizes th
best-responding agent. This

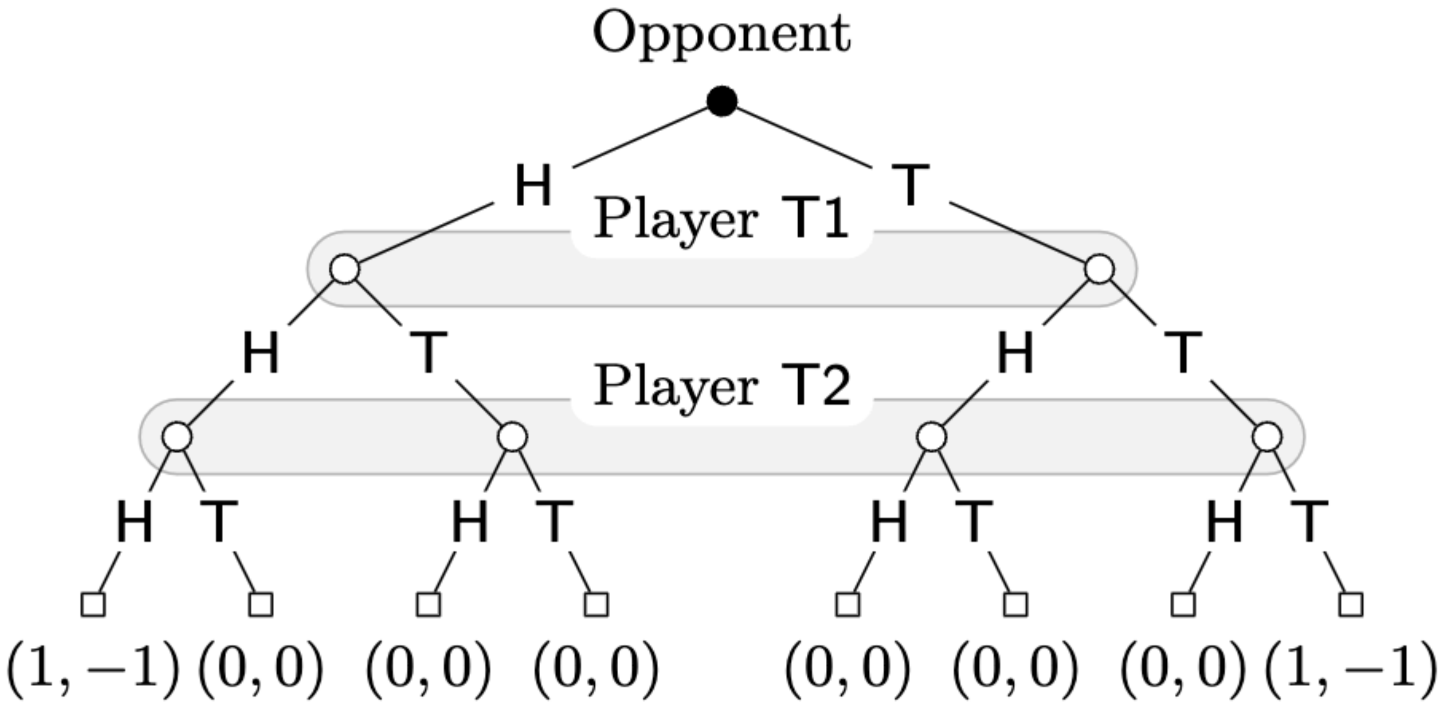
Danger zone:

The minmax theorem does not hold in general for TME. So, perhaps, the term "equilibrium" should be used carefully when referring to TME

The Minimax Theorem Fails for TMECor

- Matching pennies game

If all three pennies match, the team wins a payoff of 1 and the opponent suffers a loss of -1



Maxmin value (team goes first): 1/4

Minmax value (team goes second): 1/2

Three Models of Team Coordination

<u>Case 3:</u>

Team members have an opportunity to discuss and agree on tactics before the game starts, but are otherwise unable to communicate during the game, except through their publicly-observed actions

Equilibrium known as “TMECor”

TMECor

- You can visualize TMECor as the following process:
  - before playing, the team members get together in secret, and discuss about tactics for the game
  - They come up with m possible plans, each of which specifies a deterministic strategy for each team members, and write them down in m separate envelopes
  - Then, just before the game starts, they pick one of the m envelopes according to a shared probability distribution, and play according to the plan in the chosen envelope
- Note: the sampling of the envelope can happen even if the team members cannot communicate before the game starts, as long as they can agree on some shared signal, such as for example a common clock.

Example of TMECor-type coordination

- 100 prisoners on an island (you know where this is going…)
  - Numbered 1, …, 100, each knows their number and everyone else's
- Prisoner numbers {1,…,100} shuffled into 100 closed box in a room
- Each prisoner in order will enter the room one after another
  - Each prisoner can open up to 50 boxes, looking for their number
  - In between prisoners entering the room, the boxes are closed
- If all prisoners find their number, they all survive. Else, they all die.
- Prisoners can discuss tactics before the game starts, but cannot communicate after they have been in the room
- What is the optimal strategy? What is the probability of success?

Comparison of Equilibria

	TME no communication *ever*	**TMECor** no communication *during play*	**Team Nash Eq.** private communication during play
Convex problem	✗	✓	✓
Bilinear saddle-point problem	✗	✓	✓
Low-dimensional strat. polytope	✗	✓	✓
Minmax theorem	✗	✓	✓
Team utility	low	higher	highest
Complexity	very hard	sometimes hard	polynomial