Admin

Homework 2 posted, deadline Friday of reading week

- **Register** and Submit as a **team** of 2-3 students.
- No individual submissions or teams larger than 3 people allowed without my explicit permission
- Let me know if you cannot find a team member

Quiz 2 is in progress

- Do people prefer an in-class quiz instead…?

Course feedback

- You should have gotten email about this by now
- Please try to complete, this affects my teaching reviews ☹

Lecture 11: Blackwell Approachability, Correlated Equilibria in EFGs

Agenda

Blackwell Approachability

Correlated equilibria in EFGs

~~Stackelberg Equilibria in EFGs~~

Part 1: Blackwell Approachability

We are going to explain why regret matching works

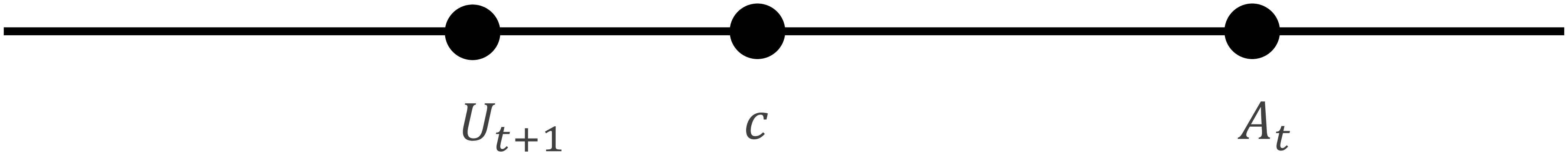
Approachability in Scalars

Sequence of **bounded** scalars $\{U_t\}, U_t \in \mathbb{R}$

Let average be $A_T = \frac{1}{T}\sum_{t=1}^{T} U_t$

Let $c \in \mathbb{R}$ be a target.

Assume $\{U_t\}$ is constrained such that $(U_{T+1} - c)(A_T - c) \leq 0$



Then $\lim_{T\to\infty} A_T = c$

Intuition: being on the “opposite” side gives enough “power” to reach c, boundedness of U ensures no oscillations.

Approachability in Vectors

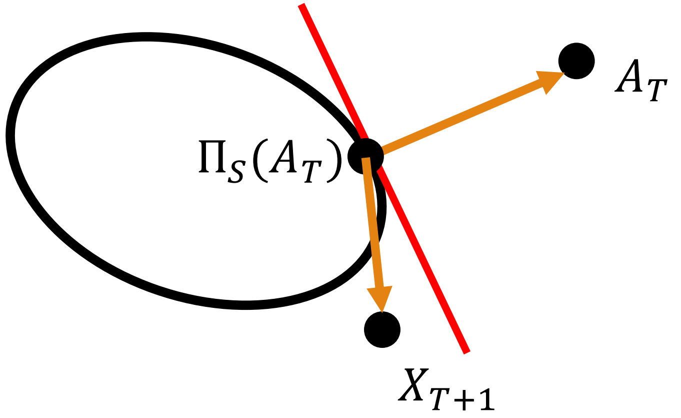
Sequence of **bounded** vectors $\{U_t\}, U_t \in \mathbb{R}^K$

Let average be $A_T = \frac{1}{T}\sum_{t=1}^{T} U_t$

Let $S \in \mathbb{R}$ be a **convex target set**.

- Let $\Pi_S(A_t)$ be the closest point (projection) of A_t onto S

Assume $\{U_t\}$ is such that $(U_{T+1} - \Pi_S(A_T)) \cdot (A_T - \Pi_S(A_T)) \leq 0$



Then $d(A_T, S) \to 0$

Intuition: Always walking "towards" the tangent hyperplane with enough "power"

Approachability in Vectors in Expectation

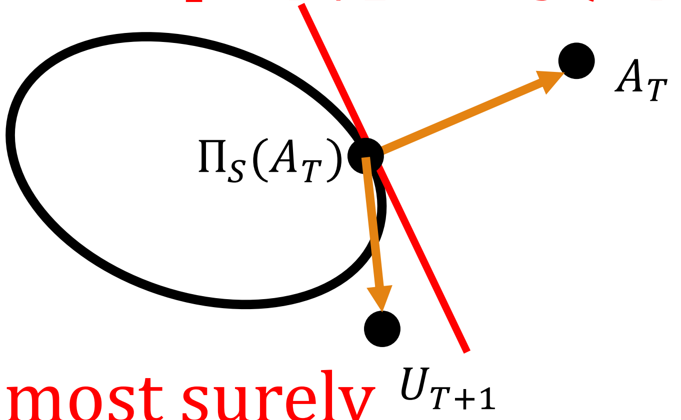
Sequence of **bounded** random vectors $\{U_t\}$, $U_t \in \mathbb{R}^K$

Let average be $A_T = \frac{1}{T}\sum_{t=1}^{T} U_t$

Let $S \in \mathbb{R}$ be a **convex target set.**

- Let $\Pi_S(A_t)$ be the closest point (projection) of A_t onto S

Assume $\{U_t\}$ is such that $\mathrm{E}[(U_{T+1} - \Pi_S(A_T)) \cdot \big(A_T - \Pi_S(A_T)\big)] \leq 0$



Then $d(A_T, S) \rightarrow 0$ almost surely

U_t's do not have to be iid. In fact, the expectation doesn't even have to be conditioned on the past!

Blackwell Approachability Game

First, P1 selects action $x_t \in \mathcal{X}$

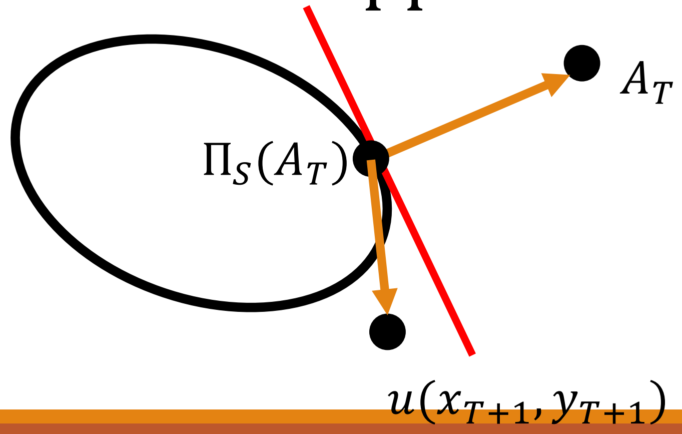
Then, P2 selects action $y_t \in \mathcal{Y}$, adversarial w.r.t. all x_t thus far

P1 incurs a **vector-valued** payoff $u(x_t, y_t)$. Typically, u is biaffine.

P1's goal is to force the average u's to converge to target set S

$$\min_{\hat{s} \in S} \left\| \hat{s} - \frac{1}{T} \sum_{t=1}^{T} u(x_t, y_t) \right\| \to 0 \text{ as } T \to \infty$$

Idea: Let's use Blackwell approachability



*Want to be able to choose x_T such that no matter how y_{T+1} is chosen, $u(x_{T+1}, y_{T+1})$ will always be on left side of hyperplane!

Forcing Halfspaces and Actions

Convex sets can be difficult to deal with: lets work with halfspaces

Let's consider halfspaces tangent to S: call it $\mathcal{H}$

$$\mathcal{H} = \{x \in \mathbb{R}^K | a^T x \leq b\}$$

$\mathcal{H}$ is forceable if there exists a strategy in x^* such that $u(x^*, y) \in \mathcal{H}$ for all possible choices of y

- x^* is called a **forcing action**

Blackwell: P1's goal will if every halfspace $H \supseteq S$ is forceable

Constructive Proof:

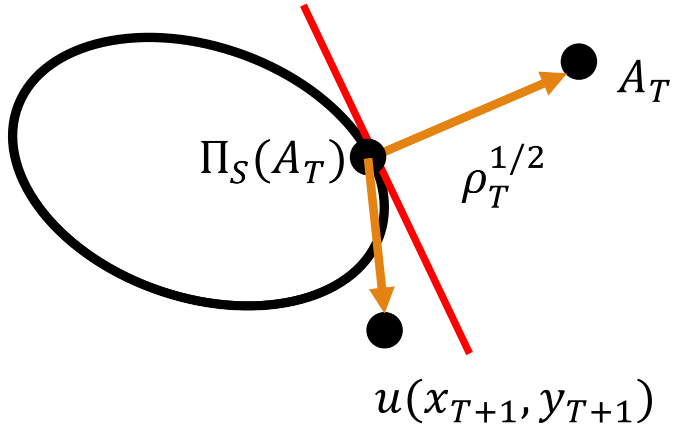
- At T , if $A_T \in S$, choose any $x^* \in \mathcal{X}$
- If not, let $\mathcal{H}$ be halfspace tangent to S containing $\Pi_S(A_T)$, choose x^* to be forcing action of $\mathcal{H}$.

Some derivations (optional)

We could just use Blackwell's theorem, but since this is deterministic it is easy to explicitly show that $d(A_T, S)$ decreases at rate of $1/\sqrt{T}$

$$A_{T+1} = \frac{1}{T+1}\sum_{t=1}^{T+1} u(x_t, y_t) = \frac{T}{T+1}A_T + \frac{1}{T+1}u(x_{T+1}, y_{T+1})$$

$$\rho_T = ||\Pi_S(A_T) - A_T||^2 = \min_{\hat{s}\in\mathcal{S}} ||\hat{s} - A_T||^2$$



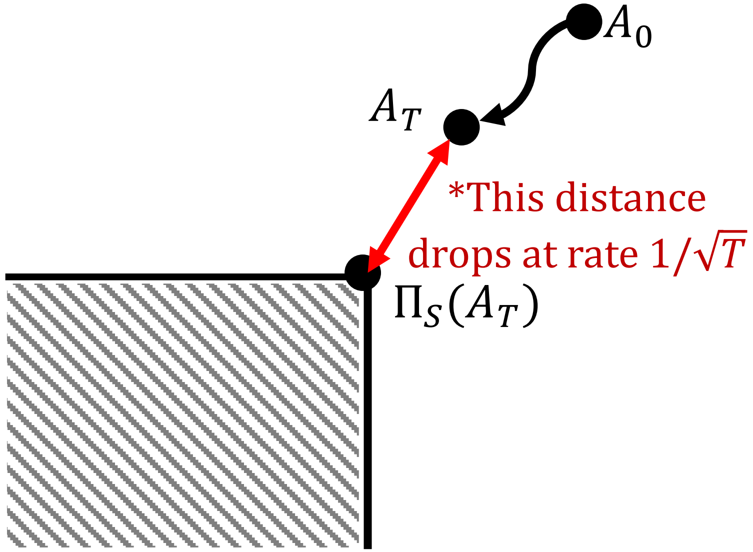
$$\begin{aligned}
\rho_{T+1} &= ||\Pi_S(A_{T+1}) - A_{T+1}||^2 \\
&\leq ||\Pi_S(A_T) - A_{T+1}||^2 && \text{Projection must be shortest distance} \\
&= ||\Pi_S(A_T) - \frac{T}{T+1}A_T - \frac{1}{T+1}u(x_{T+1}, y_{T+1})||^2 && \text{Rewrite} \\
&= ||\frac{T}{T+1}(\Pi_S(A_T) - A_T) + \frac{1}{T+1}(\Pi_S(A_T) - u(x_{T+1}, y_{T+1}))||^2 && \text{Expand} \\
&= \left(\frac{T}{T+1}\right)^2 \rho_T + \left(\frac{1}{T+1}\right)^2 \underbrace{||\Pi_S(A_T) - u(x_{T+1}, y_{T+1})||^2}_{\text{Bounded by Diameter } \Omega^2} + \frac{2T}{(T+1)^2}\underbrace{\langle \Pi_S(A_T) - A_T, \Pi_S(A_T) - u(x_{T+1}, y_{T+1})\rangle}_{\leq 0 \text{ because forcing action}}
\end{aligned}$$

$$(T+1)^2\rho_{T+1} - T^2\rho_T \leq \Omega^2 \implies \rho_{T+1} \leq \frac{\Omega^2}{T+1} \implies \min_{\hat{s}\in\mathcal{S}} ||\hat{s} - A_T||_2 \leq \frac{\Omega}{\sqrt{T}}$$

No-regret as a Blackwell Game

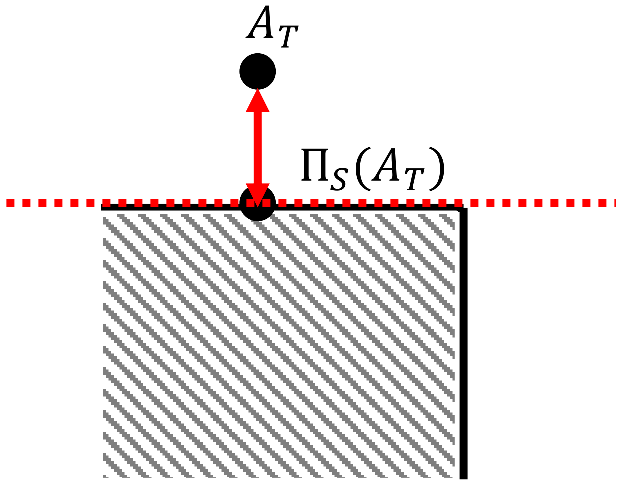
Instantiate

- $u(x_t, y_t) = \ell_t - \langle \ell_t, x_t \rangle \mathbf{1}$, i.e., regret incurred at t
- Hence, $A_T = \frac{1}{T}\sum_{t=1}^{T} u(x_t, y_t) = R_T/T$ gives average regret up till T
- $S = \{s \in \mathbb{R}^k | s \leq 0\}$, i.e., nonpositive quadrant
- Hence, if A_T tends to S then we are no-regret (roughly speaking)!

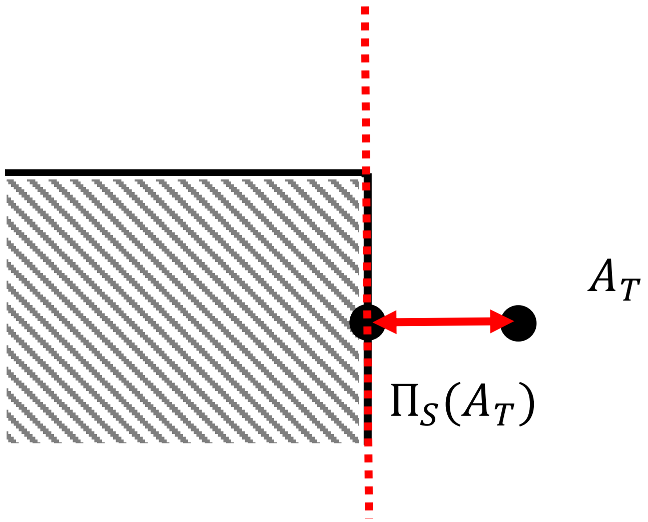


Theorem: The average regret is no greater than $d(A_T, S)$

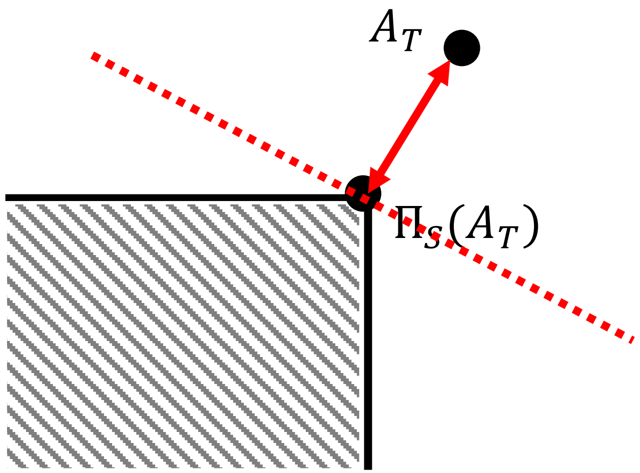
Regret Matching (RM)



Always play action corresponding to vertical-axis



Always play action corresponding to horizontal-axis



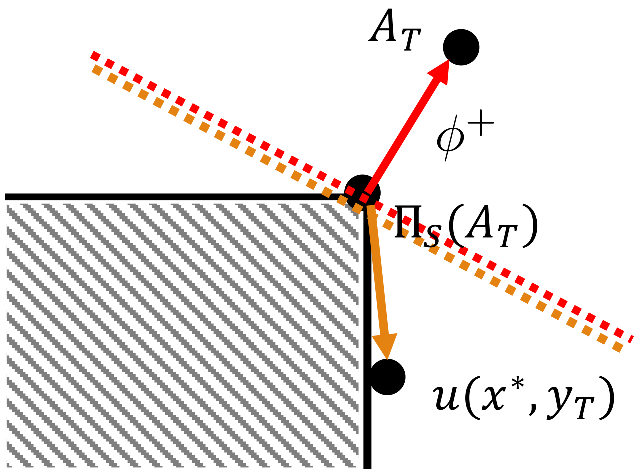
Play according to ratio of nonnegative average regrets (?)

Regret Matching Proof

Projection onto nonpositive orthant

$$A_T = [-2, 5, 2, -4] \implies \Pi_S(A_T) = [-2, 0, 0, -4], \underbrace{A_T - \Pi_S(A_T)}_{\triangleq\, \phi^+} = [0, 5, 2, 0]$$

*Assume $\neq 0$



$$\mathcal{H} = \{x \in \mathbb{R}^K | \langle \phi^+, z \rangle \leq 0\}$$

$$u(x^*, y_T) \in \mathcal{H} \qquad \forall y_T$$

$$\iff \langle \phi^+, u(x^*, y_T) \rangle \leq 0 \qquad \forall y_T \quad \text{Definition}$$

$$\iff \langle \phi^+, \ell_T - \langle \ell_T, x^* \rangle 1 \rangle \leq 0 \qquad \forall \ell_T \in \mathbb{R}^K \quad \text{Definition}$$

$$\iff \langle \phi^+, \ell_T \rangle - \langle \ell_T, x^* \rangle ||\phi^+||_1 \rangle \leq 0 \qquad \forall \ell_T \in \mathbb{R}^K \quad \text{Rearrange}$$

$$\iff \langle \ell_T, \frac{\phi^+}{||\phi^+||_1} - x^* \rangle \leq 0 \qquad \forall \ell_T \in \mathbb{R}^K$$

Forcing action: Just choose $x^* = \frac{\phi^+}{||\phi^+||_1}$

RM and RM+

= reward vector Py_t

$\textsc{ObserveUtility}(\boldsymbol{\ell}_t)$

$$\underbrace{A_{T+1}}_{\text{New average regret}} = \left[\frac{T}{T+1} \underbrace{A_T}_{\text{Old average regret}} + \frac{1}{T+1} \underbrace{(\ell_T - \langle \ell_T, x_T \rangle 1)}_{\text{Regret to accumulate for this round}} \right]^{+}$$

RM+: change average/cumulative regrets to 0 if negative

$\textsc{NextStrategy}()$

If $\phi^+ = 0$ just choose x^* uniformly at random

$$\underbrace{A_{T+1} = [-2, 5, 2, -4]}_{\text{Average regret}} \implies \underbrace{\phi^+ = [0, 5, 2, 0]}_{\text{Truncate negative regrets}} \implies \underbrace{x^* = [0, 5/7, 2/7, 0]}_{\text{Renormalize}}$$

Note: To make things simpler we could just work with cumulative regret all the way

Recall: convergence at rate $1/\sqrt{T}$

Summary

Blackwell approachability

- Guarantees that average iterate gets “closer and closer” to target set S
- Regret matching corresponds to Blackwell approachability with the target set of the nonnegative quadrant
- Extensions exist for almost every imaginable case
  - Continuous time, infinite dimension, different distance metrics, etc
- Lots of applications beyond game-playing

Interesting connection generally between Blackwell approachability and *any* regret minimization algorithm:

- Blackwell Approachability and Low-Regret Learning are Equivalent (Abernathy, Barlett and Hazan, 2010)

Part 2: Correlated Equilibria and EFGs

Example: Battleship

Each player possesses some ships of size $k \times 1$. Board of size $N * M$

Two phases

- Phase 1: players take turns placing ships one at a time unknown to the other player
- Phase 2: players take turns shooting at each other for T timesteps.
  - Players decide which coordinate to fire at. Only notified if a ship is hit or miss. Cannot fire at the same spot over and over.
  - A ship is sunk if all of the cells it contains is destroyed. Game ends if either player loses all ships (or time limit reached)

Zero-sum variant

- Each ship is worth some value (say 1)

General sum variant

- Players lose γ per ship lost, obtain 1 per ship destroyed
- Typically risk adverse, $\gamma \geq 1$

3x1 Battleship…

Board size 3x1, one ship each of size 1x1, T=2, $\gamma = 1$ (zero-sum)

How would you place your ship?

How would you fire?

Who has the advantage?

What if $\gamma > 1$

Turns out, NE is also to randomize uniformly (why?)

- In the end, players still end up trying to kill each other optimally
- One would think that if γ sufficiently high, then players would try to avoid getting at each other's throats (think mutually assured destruction)

Is any peaceful resolution possible with the help of a mediator?

- **Yes!** Same as how (C)CE can help to resolve the chicken game in a more socially acceptable way.

But what does correlation mean in an EFG?

- In normal form, the CE is just a distribution over joint actions
- What is the analogue for EFGs? Turns out there are **many ways**

Some ways to define CE

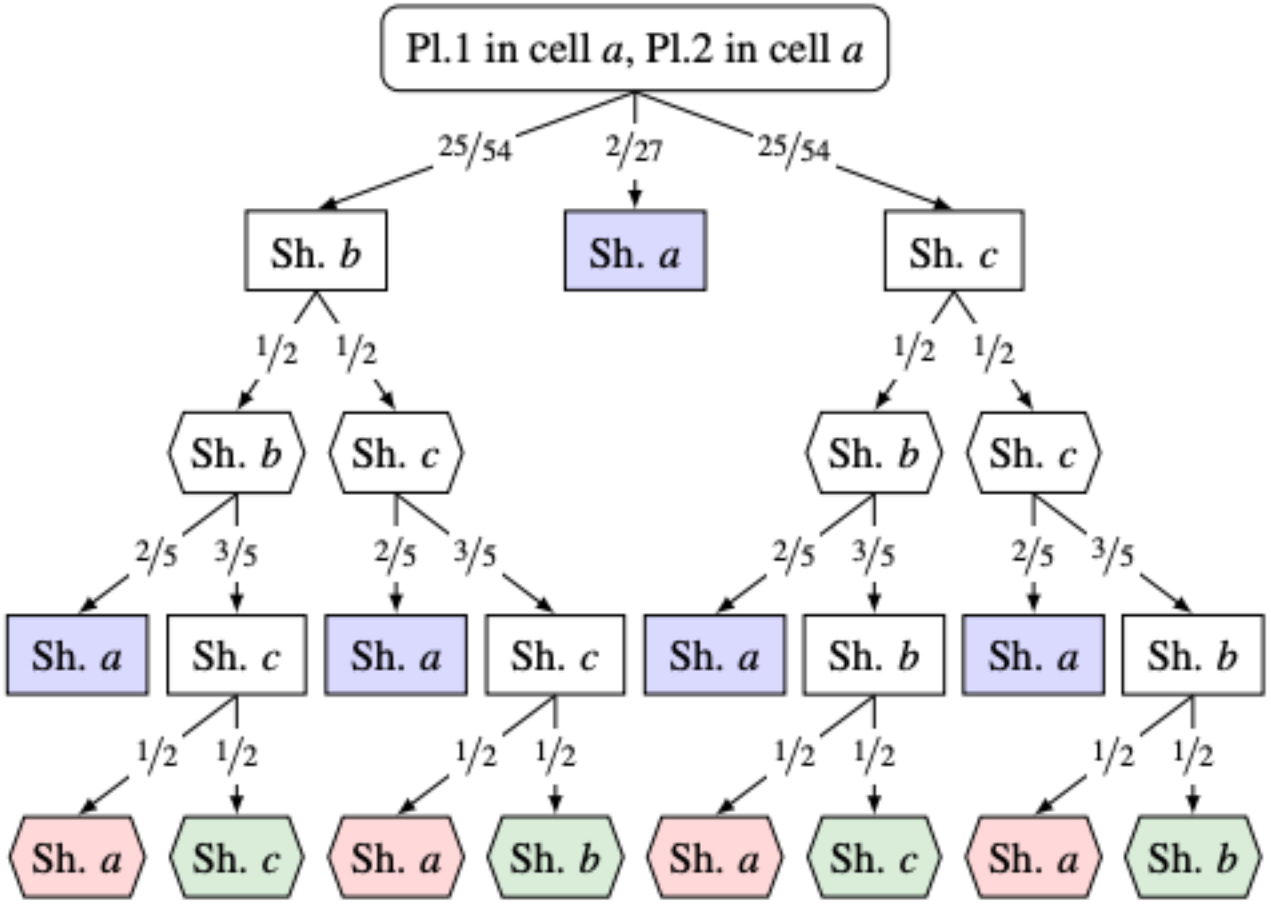
NFCE/NFCCE

- Convert all strategies to normal form → recover matrix game
- Use our standard definitions to define correlations (see previous lecture)

EFCE (this lecture)

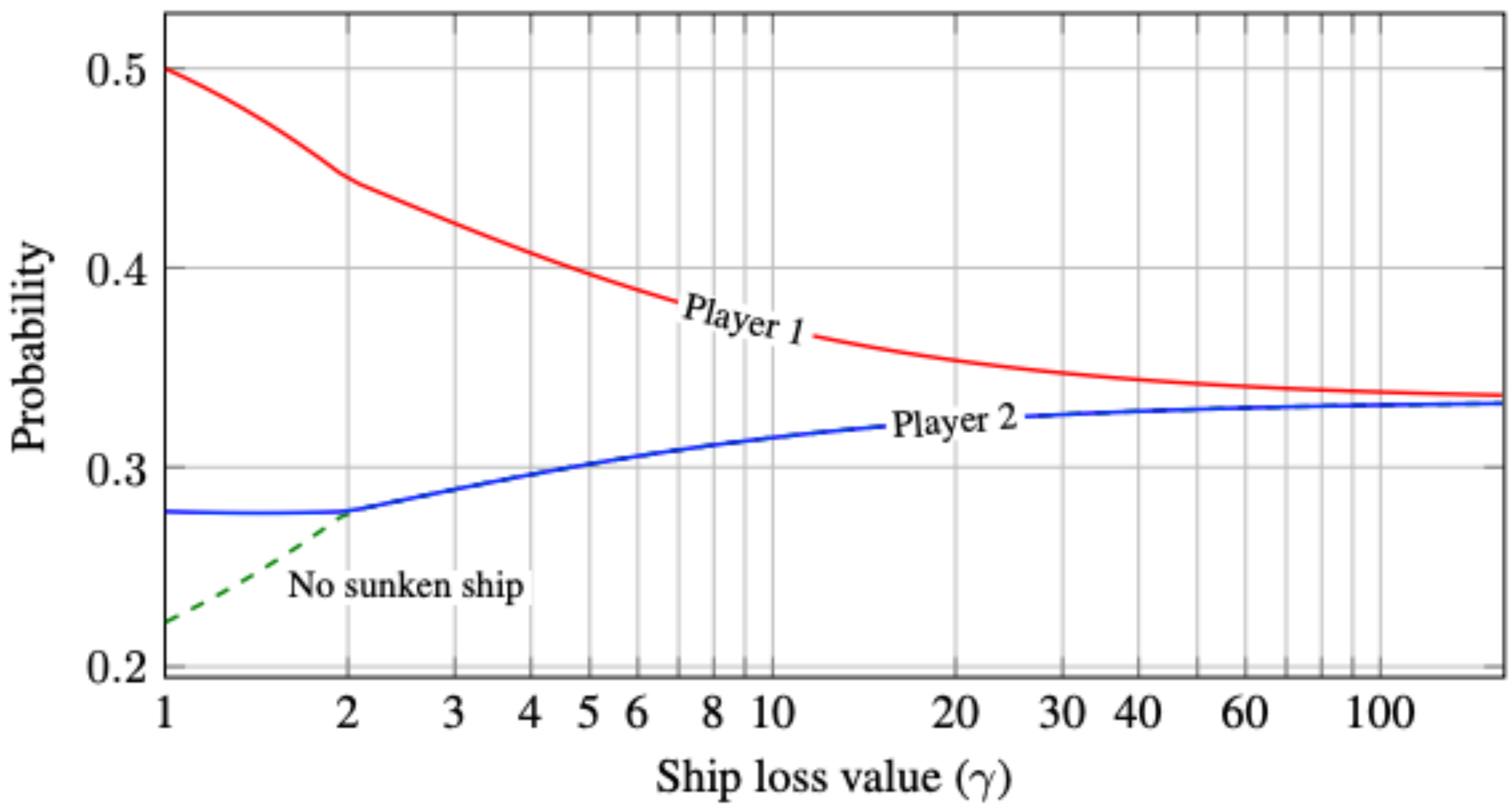
- Same as NFCE but…
- Mediator gives you recommendations “on-demand”, **only for the infoset that you are currently in**
- That means that if you are at infoset I (preceding I’), you will only know what the mediator wants you to do at I, but not I’ → Less information than CE!
- What about deviations?
  - If you do not follow a recommendation at infoset I, then **you will not get recommendations for the rest of the game**. Your opponent is still assumed to get recommendations

Battleship playthrough



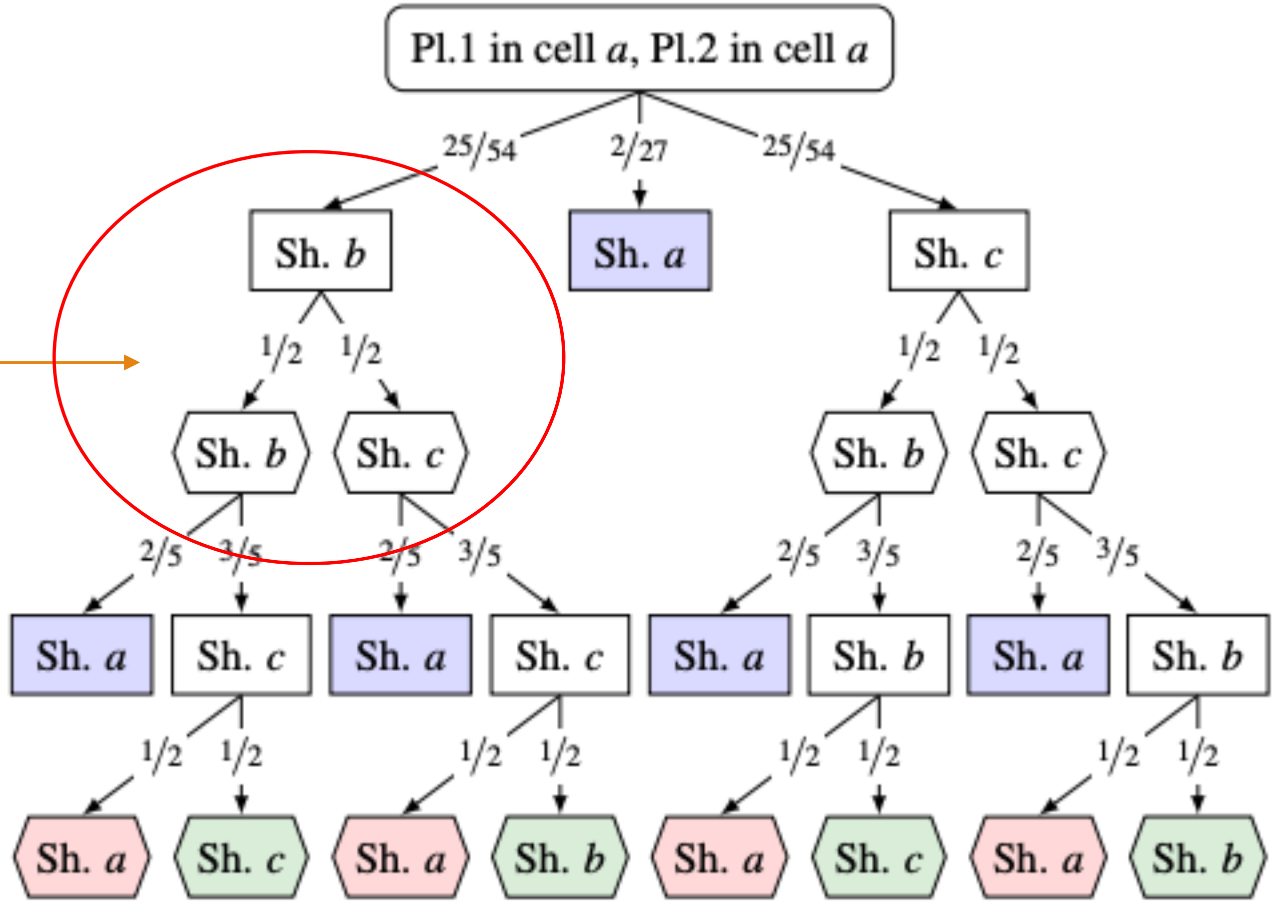
NOTE: this is not all there is to an EFCE!

Distribution over outcomes as γ changes



Wait, this doesn't make sense...

Player 2 shoots to miss?!



Why should P2 follow the recommendation if it knows it is being recommended to miss on purpose?

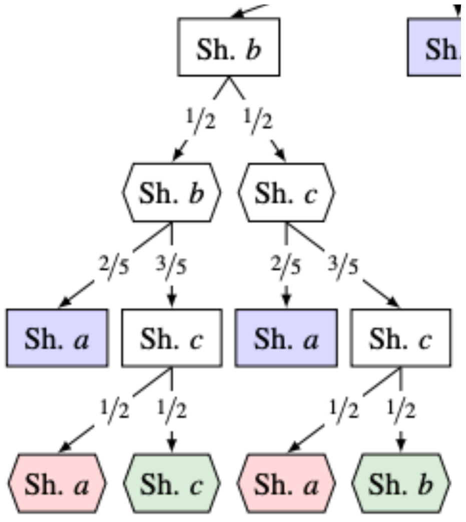
Outside the equilibrium path

Suppose player 2 was recommended to shoot at (b) but chooses to deviate by shooting at (a) or (c)

- If it chooses (a) then it is lucky and won (remember this is not guaranteed because the mediator places ships uniformly at random)
  - Payoff = 1, happens with probability 0.5
- But if it chose (c) which is incorrect, the mediator can punish him by telling P1 the location of P2's ship! **Sure to lose next turn**
  - Payoff = -2 happens with probability 0.5
- Total expected utility from deviating: -0.5

If stick to recommendation

- P1 wins with probability 2/5 = 4/10
- P2 wins with probability 3/5 * 1/2 =3/10
- Expected util = -2 * 4/10 + 3/10 = -0.5 →SAME as deviating!



What is going on?

Mediator "threatens" player with future, gets P2 to toe the line

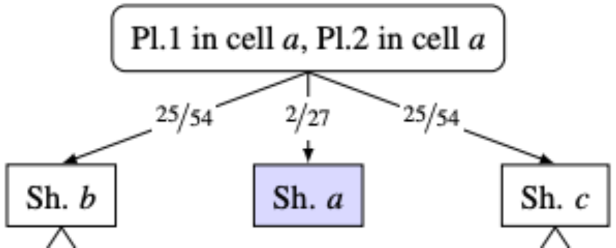
- My opinion: somewhat reminiscent of real life?
- Think about all the "peace deals" which have been "brokered" by mediators

Wait… how does the mediator know our (P2's) ship location?

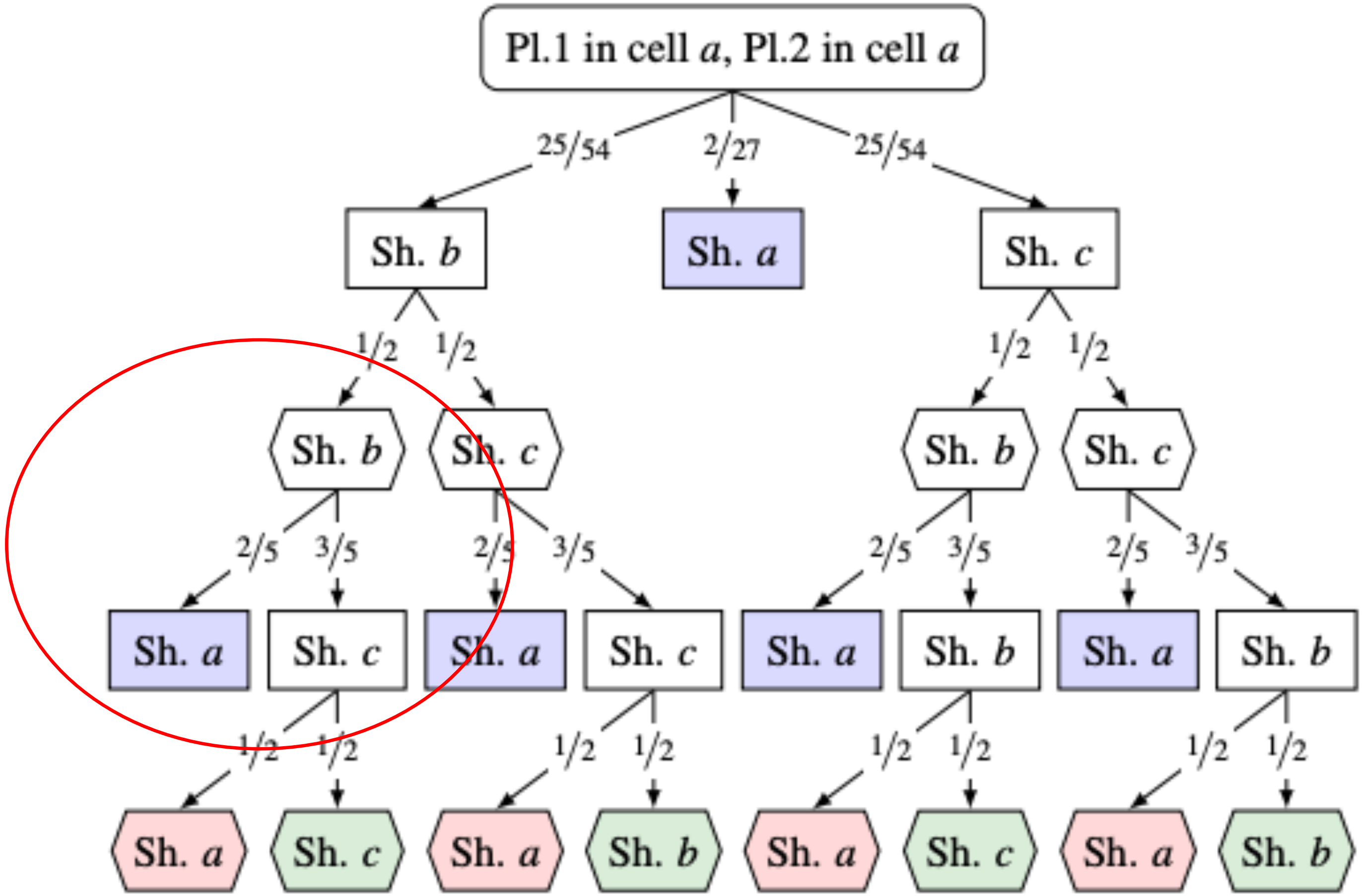
- By definition of EFCE, mediator knows it because **it recommended P2 to place the ship there anyway**, and P2 hasn't deviated until the shooting phase
- But… if P2 knows that the mediator is going to potentially threaten him by revealing his ship's location in the future, then maybe he **should have deviated during the placement phase** to avoid being threatened!

Turns out it's okay!

- If P2 deviates and places in (b) or (c), Player 1 has high chance (~50%) of shooting him in first shot!



Exercise:

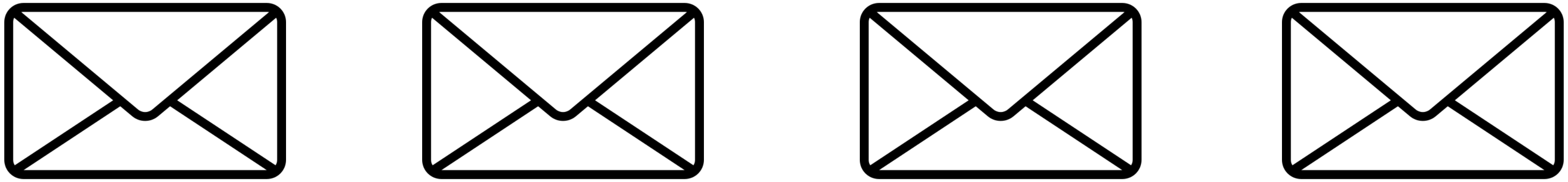


What happens here if P1 deviates?

How would you even implement this?

Classic method: sealed envelopes

- Choose CE over **normal form strategies**
- Put action into each envelope based on normal form strategy
- Contains what action to play at each infoset
- Can only open envelope for the infoset you are in



a_1 a_2 a_3 a_4

But what happens when you deviate?

- Couldn't you still open envelopes?
- Trick: use reduced normal form instead→envelopes could be "empty"

Another method, cryptographic protocols

Example 2: Sheriff of Nottingham

Smuggler must choose n $\in \{0, \ldots, n_{max}\}$ illegal items to smuggle

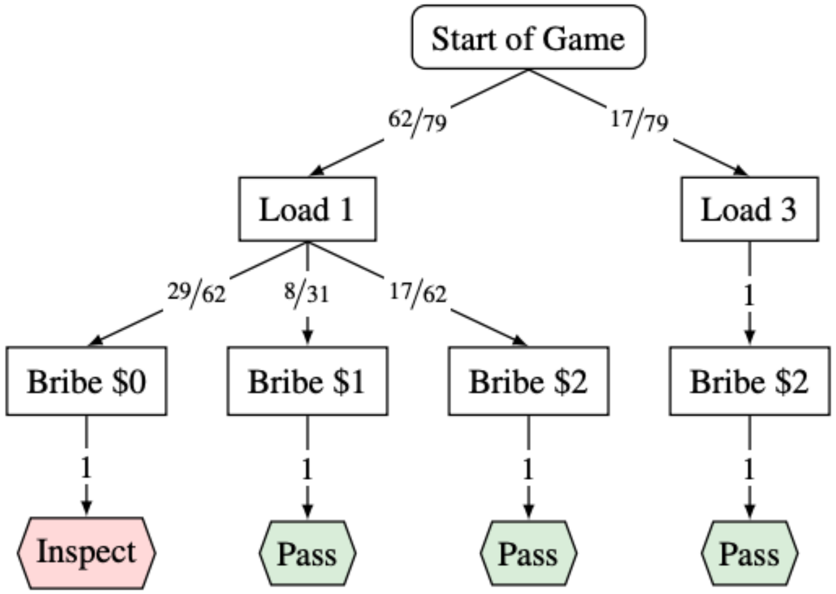
Sheriff must decide whether to inspect goods or not

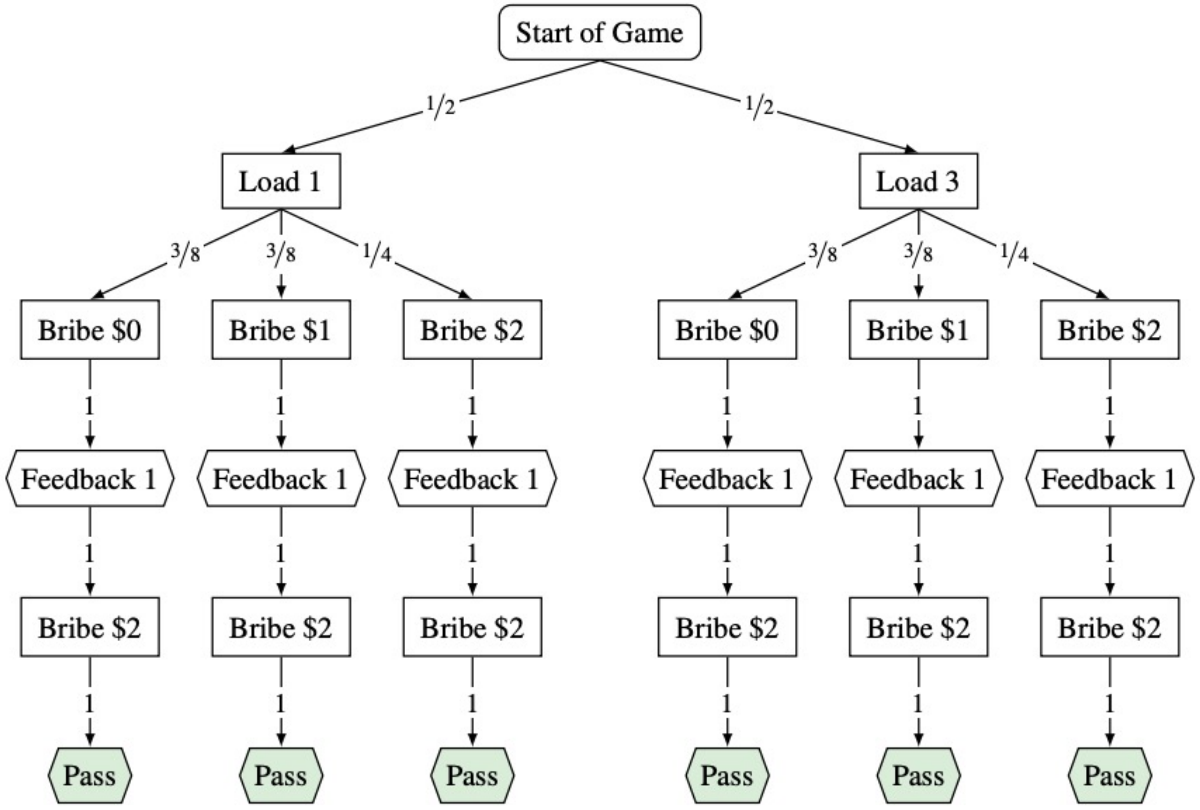
- If inspect and finds illegal goods, fine of $n \cdot p$
- If inspect no illegal goods, sheriff must compensate by c
- If no inspection, smuggler gets $n \cdot v$, sheriff gets 0

However, smuggler can **bribe** sheriff to not inspect

- Multiple rounds of offering bribes $b_i \in \{0, \ldots, b_{max}\}$
- Sheriff can accept or reject bribe, only the last offer matters

Example playthroughs





Qualitative trends

$v = 5, p = 1, s = 1, n_{max} = 10, b_{max} = 2$, number of rounds r = 2

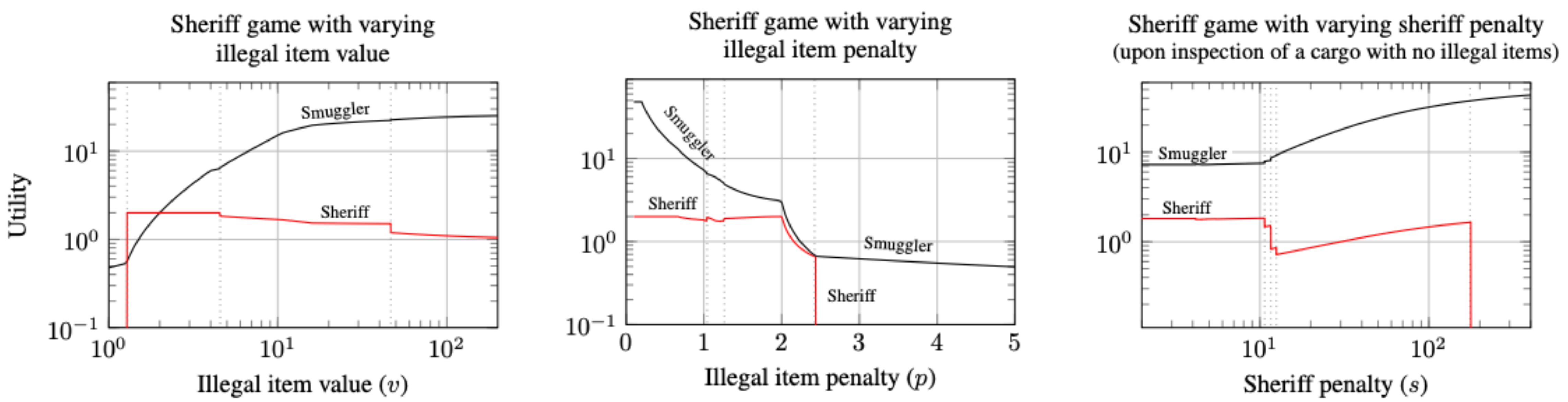


Figure 2: Utility of players with varying v, p and s for the SW-maximizing EFCE. We verified that these plots are not the result of equilibrium selection issues.

Mechanism: codebook to threaten player

Properties of EFCE

We know NFCE $\subseteq NFCCE$ (why?)

Is there any subset/superset relationship between EFCE, CE, CCE?

NFCE $\subseteq EFCE \subseteq NFCCE$

- Compared to NFCCE, players who are thinking about deviating have *more* information from mediator, need to "satisfy" them, hence set is *smaller*
- Compared to NFCE, players who are thinking about deviating have *less* information from mediator (why?), hence set is *larger*

Computing EFCE

Solving for EFCE: the classical setting

How do we even represent an EFCE?

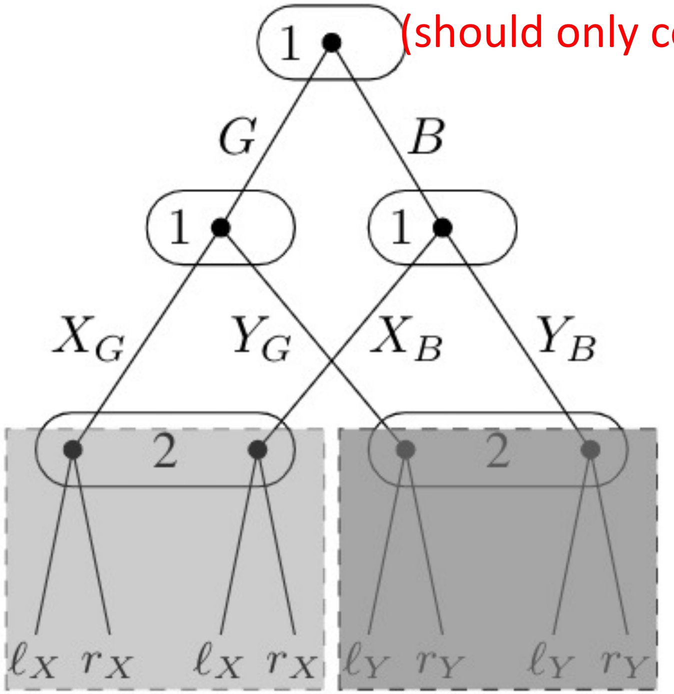
- Normal form: just a 2d matrix (for 2 player games)
- Can we avoid converting to normal form?
  - That would incur exponentially large matrices (why?)

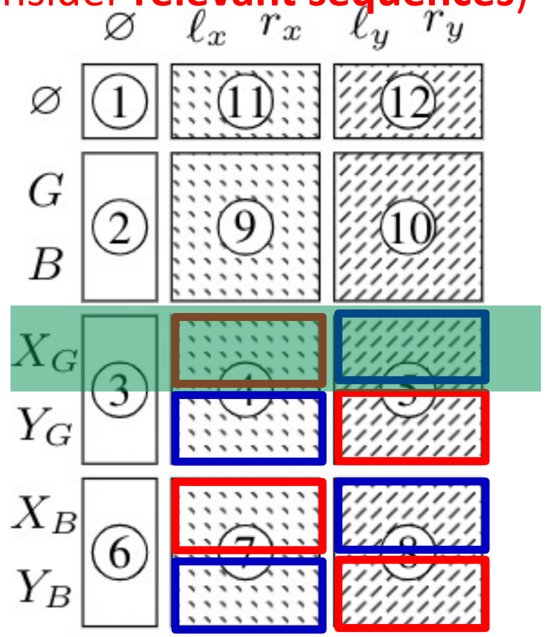
For certain special games we get compact formulations for the **resultant correlation plan**

- Example: games without chance, games that are "triangle free"
- 2-dimensional "sequence form"

Correlation plans

Technical detail: not every entry is necessary
(should only consider **relevant sequences**)





$$\xi(\emptyset,\emptyset) = 1$$

$$\xi(\emptyset,\emptyset) = \xi(G,\emptyset) + \xi(B,\emptyset)$$

$$\xi(G,\emptyset) = \xi(X_G,\emptyset) + \xi(Y_G,\emptyset)$$

$$\xi(B,\emptyset) = \xi(X_B,\emptyset) + \xi(Y_B,\emptyset)$$

$$\xi(\emptyset,\emptyset) = \xi(\emptyset,\ell_x) + \xi(\emptyset,r_x)$$

$$\xi(\emptyset,\emptyset) = \xi(\emptyset,\ell_y) + \xi(\emptyset,r_y)$$

[+constraints for other rows, cols]

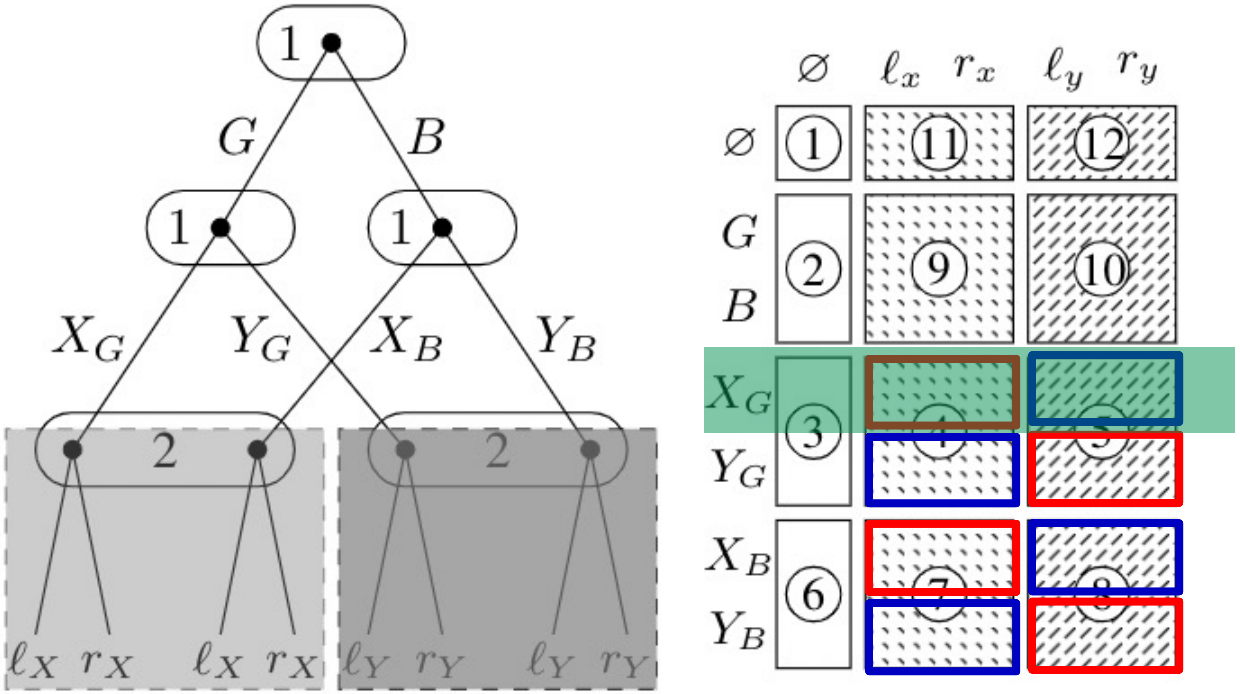
Probs of leaves of the game

Counterfactuals (don't correspond to nodes in tree)

Incentive constraints: If P1 was recommended X_G and is considering deviating to Y_G, it will consider probability that P2 plays ℓ_y, r_y (given in blue)

Counterfactuals specify behavior of other player if a player deviates

Incentive compatibility



Probs of leaves of the game

Counterfactuals (don't correspond to nodes in tree)

Incentive constraints:

(A) If P1 was recommended X_G and is considering deviating to Y_G, will consider probability that P2 plays ℓ_y, r_y (given in blue)

VERSUS

(B)expected payoff against not deviating for rest of game

Computing (A)

- Get best response towards row/column indexed by σ (corresponds to expected opponent strategy)

Computing (B)

- Iterate over **leaves** underneath σ
- Leaves correspond to **cells** in correlation plan

LP solver via compact representation

Initialize LP with "rectangular correlation plan"

- Note: usually **not explicitly** a 2d matrix since set of relevant sequences is **sparse**

Enforce Treeplex constraints on every **row** and **column**

Enforce incentive compatibility constraints for every sequence for every player

- Based on previous slide
- How to write best response to expected opponent strategy as a set of linear inequalities?
- Linear inequalities by recursively traversing your treeplex strategy bottom-up

Same as NFCE, objective is open

- This was how we computed social welfare optimal strategies in previous slides

What about self-play?

We know self-play is a lot more efficient in practice

- LPs are slow and inefficient

What if it was not one of the "nice" games with compact representation?

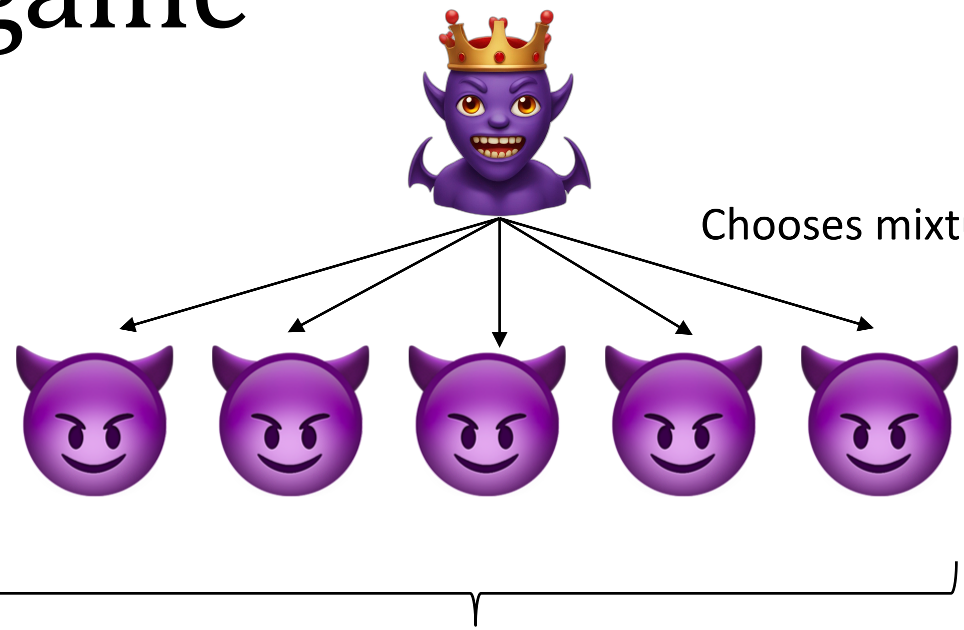
What if there are more than 2 players?

This class: two **older** methods based on online learning(~5 years old)

- Convert to zero-sum game between *mediator and deviator*
- Sample access solver based on *trigger regret*
- Other methods exist, e.g., ellipsoid against hope, more general phi regret minimizers

Method 1: zero-sum game

Does regret minimization over **correlation plans**



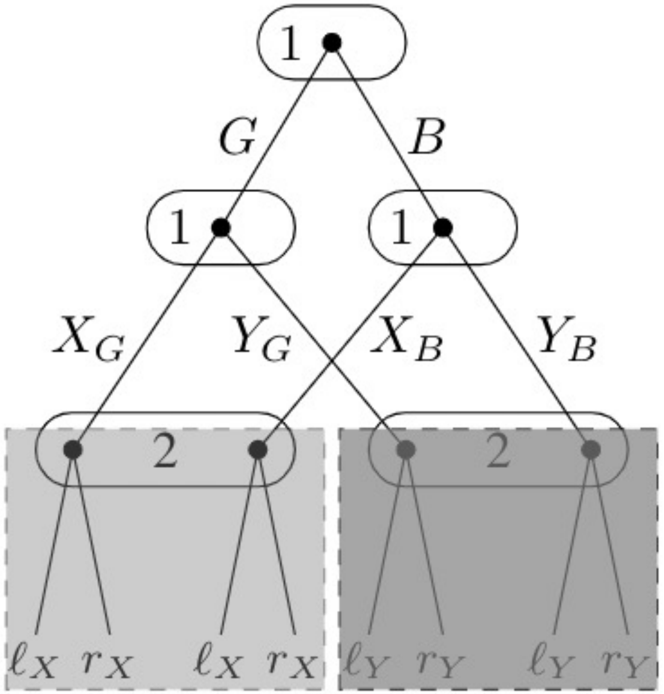
Chooses mixture of…

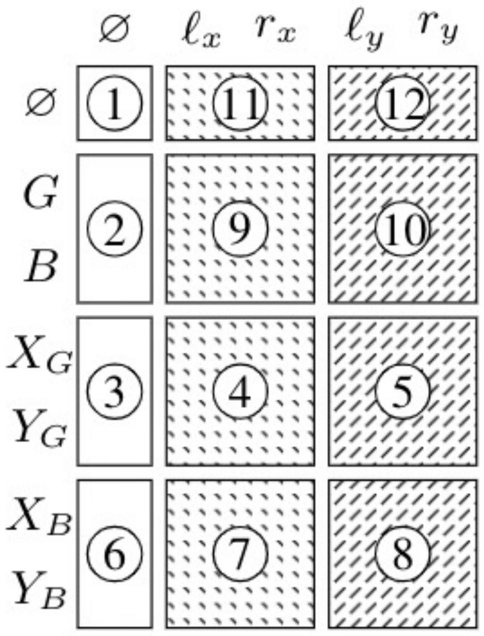
Collection of deviators, one for each **sequence** of **each player**, performs regret minimization starting over the sub-treeplex starting from infoset containing that sequence (best responder)

Intuition: this is a **zero-sum** game where payoffs are equal to weighted average of benefits (potentially negative) from deviating

At equilibrium, should be non-positive!

Regret minimizer over correlation plan





Seems to have interlaced constraints, cannot be done by treeplex only. No clear DAG structure either (why?)

Good news! For games without chance (and a few others), this structure can be reduced to DAG structure→structural constraints can be encoded by **scaled extensions** (previous lecture)

Not obvious

Experimental Results

Board size	Num turns	Ship length	$\lvert\Sigma_1\rvert$	$\lvert\Sigma_2\rvert$	Num. rel. seq. pairs
(3, 2)	3	1	15k	47k	3.89M
(3, 2)	4	1	145k	306k	26.4M
(3, 2)	4	2	970k	2.27M	111M

Table 1: Game metrics for the different instances of the Battleship game we test on.

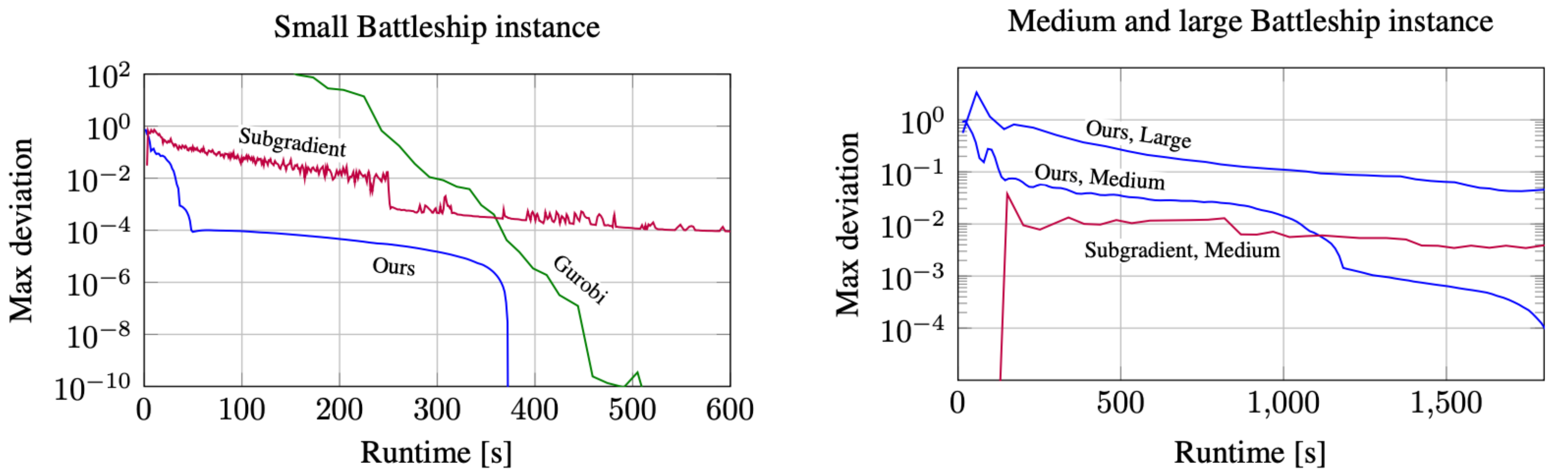


Figure 3: Experimental results. The y-axis shows the maximum utility increase upon deviation.

Subgame solving for EFCE

Applying resolving to EFCEs?

Correlation plan gives recommendations to both players

- Players do not play independently
- Who is the mediator `siding' with, if at all?
- Solution: mediator optimizes for social welfare (or some linear objective)

What is the metric for quality of solution?

- No longer a single player's payoff under the best-response of the other
- Solution: combination of exploitability (no worse than blueprint) and social welfare (also not worse than blueprint)

Correlation plan does not have a clear (one-dimensional) hierarchy over sequences/infosets

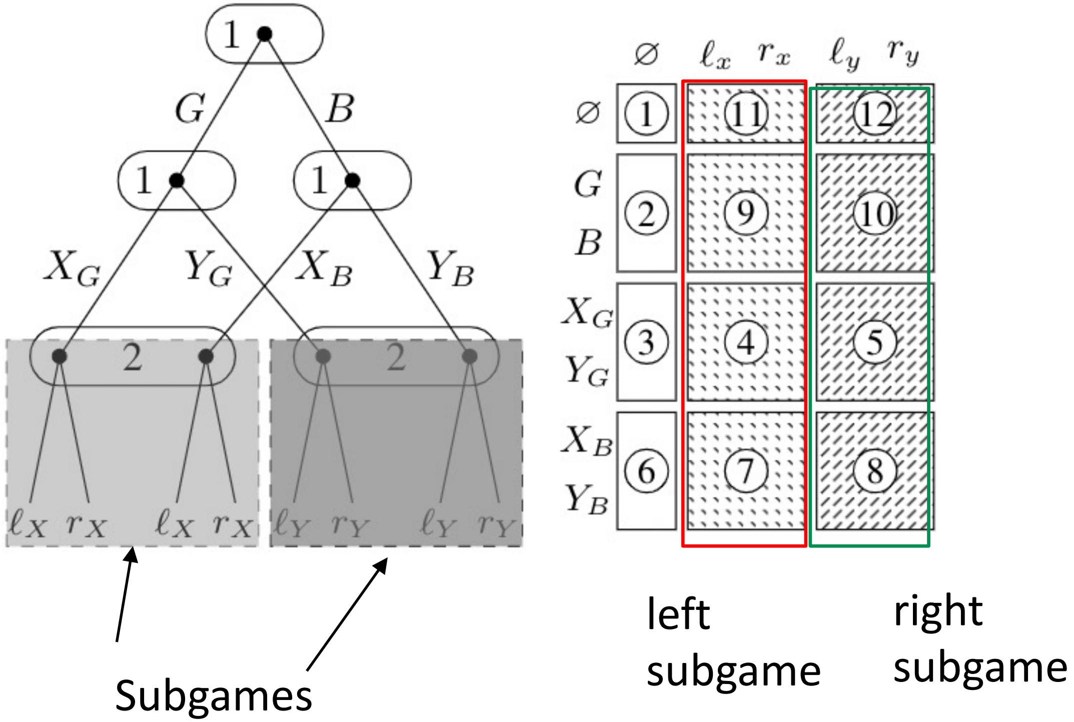
- Correlation plan indexed by *sequence pairs*
- How to decide which subgame does a sequence pair belong to?

Decomposition of correlation plan

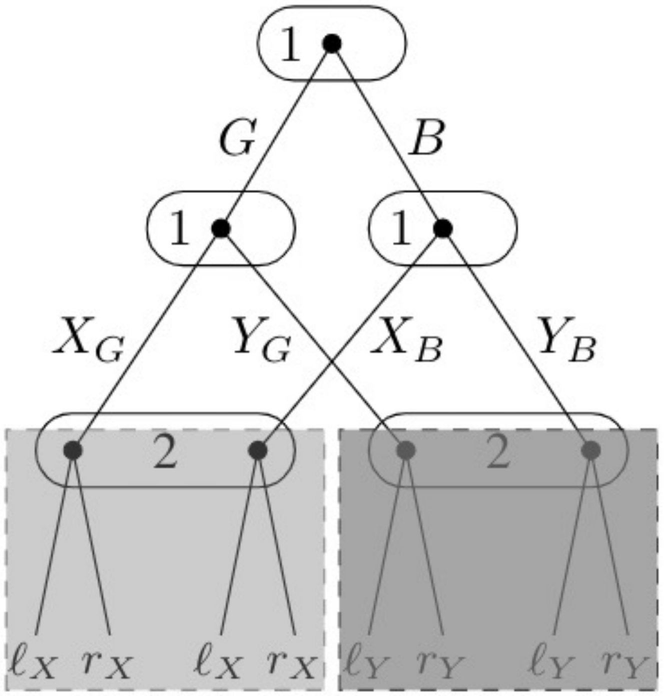
Correlation plan can be decomposed into non-overlapping parts corresponding to subgames

- Relevant sequence pairs belong to a subgame iff at least one sequence belongs an infoset in a subgame

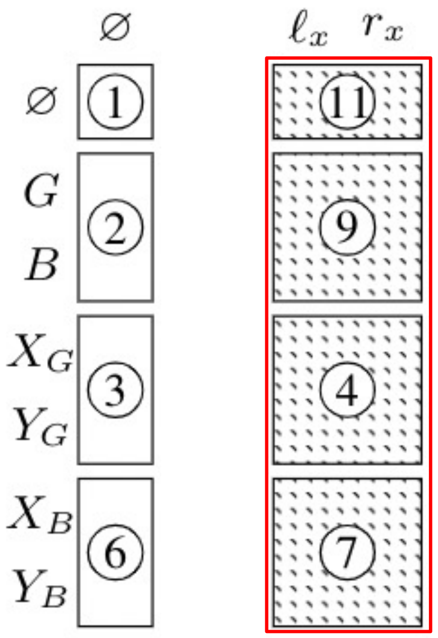
Refinement can be taken as solving the green or red columns *only*



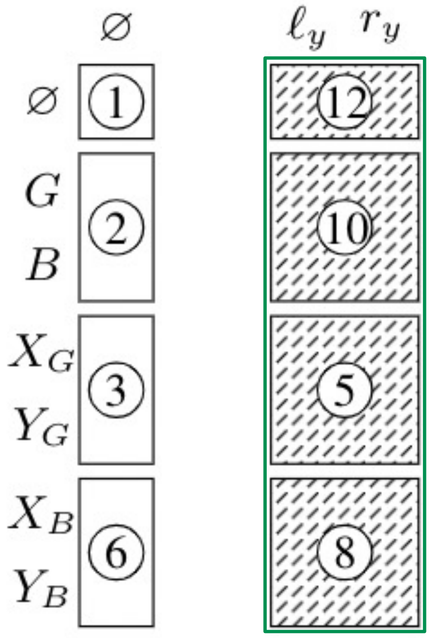
Example



If left subgame is entered, we solve for:



If right subgame is entered, we solve for:



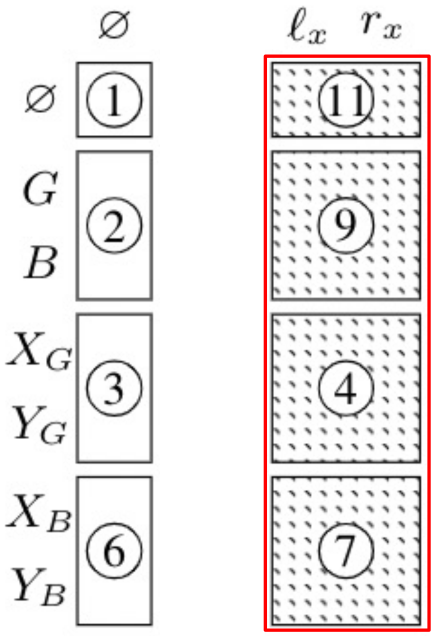
Never have to solve for both subgames in a single playthrough

Safety and Complete refinements

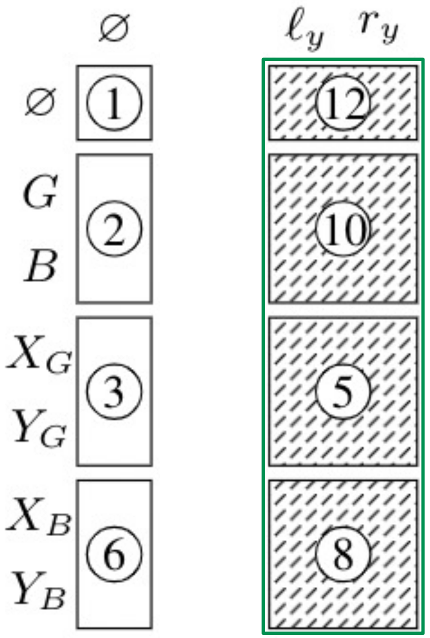
From its perspective, strategy refinement occurs for players whichever subgame was reached

- When players are considering deviations, they would consider that refinement is done for *all* subgames

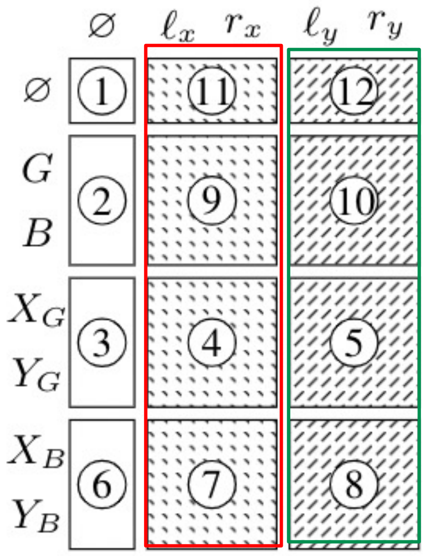
Partial refinement (actual computation)



OR



Complete refinement (what players "see")



Safety guarantees

Ideally, complete refinement should satisfy EFCE constraints

In reality, the blueprint can be `bad' enough that every refined strategy cannot satisfy EFCE incentive constraints

Our guarantee: refinement is no more exploitable than blueprint

- If player cannot improve by more than ϵ by deviating under blueprint, then player will not improve by more than ϵ under refinement
- Social welfare is *at least* that of blueprint
- But can be (i) strictly less exploitable or (ii) extract more social welfare

Implication

- We will do no worse (in exploitability and in social welfare) than blueprint
- If blueprint was the best we can do with an approximate offline solver, there is no harm in resolving apart from extra computational cost

Our algorithms

Safety achieved by preprocessing and enforcing upper and lower bounds on player payoffs in leading infosets of subgames

Linear program similar to that of Von Stengel and Forges

- Safety bounds enforced directly via constraints
- Size of LP quadratic in size of subgame (not full game)

Regret Minimization (Farina et. al., 2019)

- Two player zero-sum game between the mediator and deviator; the latter of chooses the best deviation strategy for each recommended sequence
- Solve by self-play with regret minimizer for mediator and deviator
- Safety achieved by adding ‘escape’ values for deviator and mediators, reflecting upper and lower bounds.

Summary

Extension of CE/CCE in EFGs

- Examples, benchmarks
  - How mediation correspond to “intuitive” mechanisms “in real life” to incentivize players to toe the line
- LP solver
- Efficient self-play solution via Mediators vs Deviators & scaled extensions

Turns out there are **many more** such equilibria once we define equilibria **in terms of regret** (next week)

The end!