Lecture 7: More on equilibria and regret

Admin

HW1: email me your teammates if your submission did not include it

- Students who submitted individually will get a penalty of 20%
- Students who submit individually for HW2 will receive a score of **zero**

Quiz 1 is due **today**

Project topics are due in a few days (assignment opened on canvas)

PSA: Possible internship opportunity at MIT-IBM Watson AI lab (not affiliated with me!): https://forms.gle/H6dNSywXCjDDyBsq7

The story so far…

One-shot

or

Sequential

Perfect

or

Imperfect information

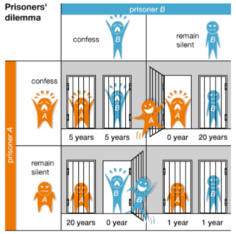
Deterministic/random environment?
Perfect/imperfect recall?
…

Game Structure

Zero-sum / Competitive

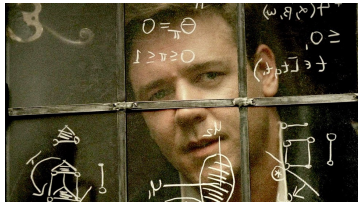
Cooperative

General Sum (i.e., everything else)

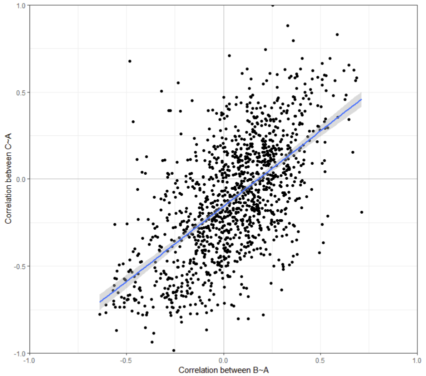


Payoff Structure

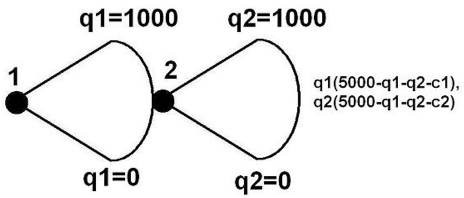
Nash



Correlated



Stackelberg (leader-follower)



Equilibrium/Solution Concept

Recall the definition of general-sum Nash

If there are n players, profile $(x_1, x_2, \ldots x_n) \in \Delta_{m_1} \times \Delta_{m_2} \times \cdots \Delta_{m_n}$

- $u_i(x_i, x_{-i}) \geq u_i(x_i', x_{-i})$ for all $x_i' \in \Delta_{m_i}$ (equivalently, for any pure action)
- No player can do strictly better by unilaterally deviating

Weaknesses

- Hard to compute → do people really play according to Nash?

Correlated Equilibrium

If there is intelligent life on other planets, in a majority of them they would have discovered correlated equilibrium before Nash equilibrium.

-Roger Myerson (2017)

Chicken Game

	Dare	Chicken out
Dare	0, 0	7, 2
Chicken out	2, 7	6, 6

A game of Chicken

Exactly three Nash

- (C, D), (D, C) and (C=2/3, D=1/3)

What is the social welfare for each of them?

- Want to avoid (D, D) but get (C, C) as much as possible,
- But incentive constraints make this hard
  - If you knew your opponent was chickening out, you will want to dare

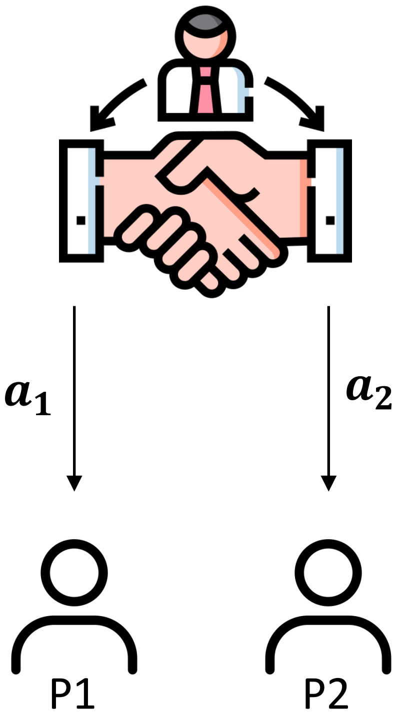
What if there was some kind of device/mechanism that assures us that if we were to chicken out, our opponent is also likely to chicken?

Enter the mediator

Assume access to a **trusted mediator**

	Dare	Chicken out
Dare	0, 0	7, 2
Chicken out	2, 7	6, 6

A game of Chicken



"I will sample joint actions (a_1, a_2) from this publicly known **joint distribution** and **suggest** to you **privately** what to do"

0	**1/3**
1/3	**1/3**

Observe: this cannot be written as the outer product of 2 distributions

Both P1 and P2: Given I was told to play a_i, should I stick to the recommendation? Or play something else?

Q: Are there joint distributions over actions that players are incentivized to play if recommended? (incentive compatible). Yes! Nash would do it. But are there more?

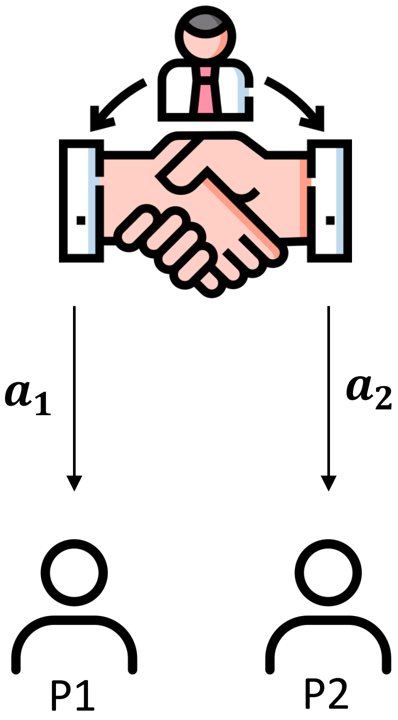
Mediator can **correlate** actions!

A possible equilibrium

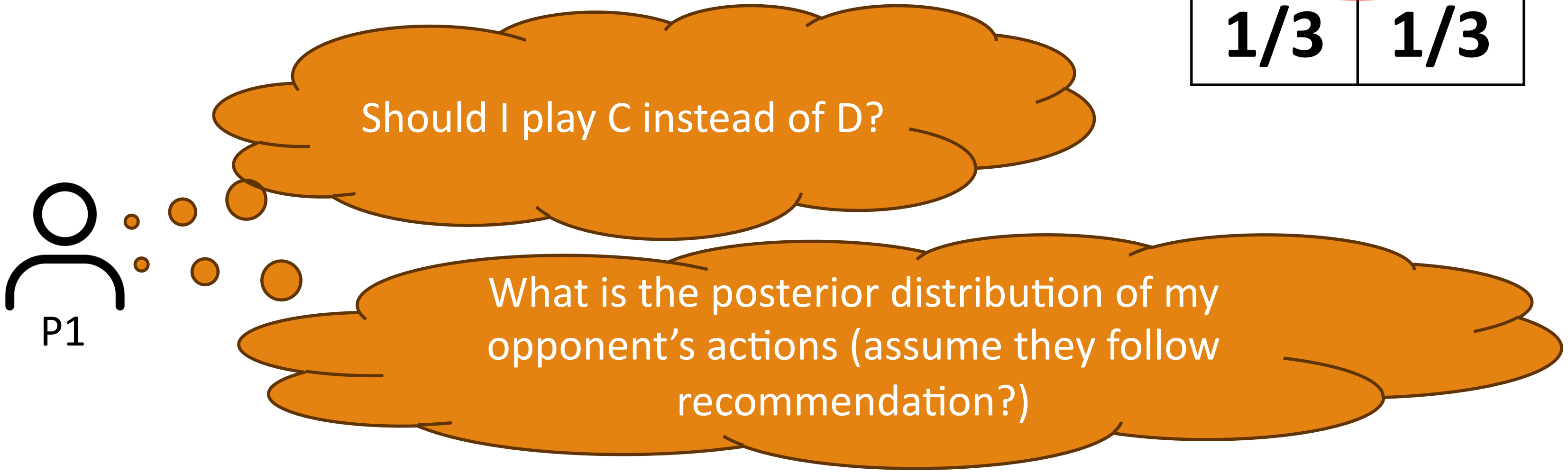
	Dare	Chicken out
Dare	0, 0	7, 2
Chicken out	2, 7	6, 6

A game of Chicken

0	**1/3**
1/3	**1/3**



Suppose $a_1 = D$ was drawn



$P(a_2 = C) = 1$

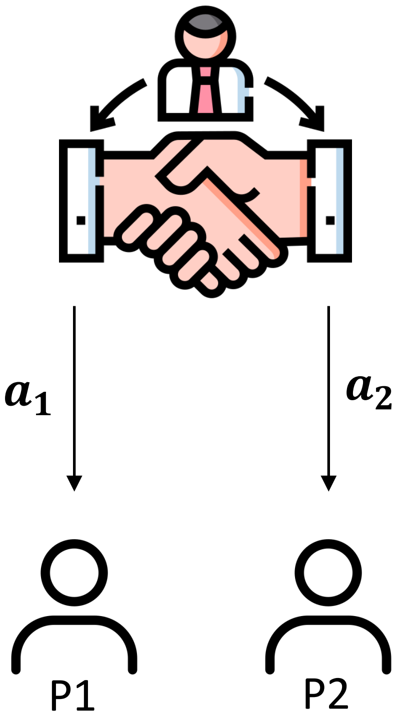
$\Rightarrow$ If don't swap, guaranteed to get 7 👍

$\Rightarrow$ If swap, then get 6

A possible equilibrium

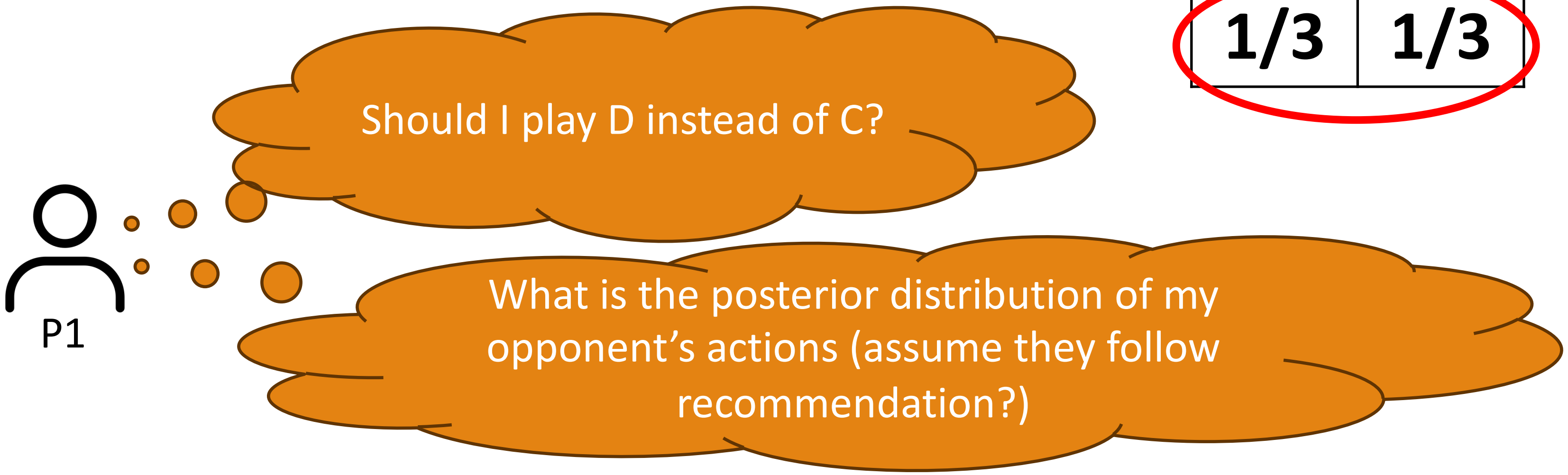
	Dare	Chicken out
Dare	0, 0	7, 2
Chicken out	2, 7	6, 6

A game of Chicken



Suppose $a_1 = C$ was drawn

0	**1/3**
1/3	**1/3**



$P(a_2 = C) = P(a_2 = D) = 0.5$

$\Rightarrow$ If don’t swap, get 4

$\Rightarrow$ If swap, get 3.5

By symmetry, P2 is also incentivized to stick to recommendation

A social welfare optimum solution

	Dare	Chicken out
Dare	0, 0	7, 2
Chicken out	2, 7	6, 6

A game of Chicken

0	**1/4**
1/4	**1/2**

Exercise: show that both players are incentivized to stick to their recommended actions

Correlated Equilibrium (2 players)

$$\sum_{j\in[m]} u_1(i,j)\cdot z(i,j) \geq \sum_{j\in[m]} u_1(k,j)\cdot z(i,j) \qquad \forall i\in[n], k\in[n]$$

$$\sum_{i\in[n]} u_2(i,j)\cdot z(i,j) \geq \sum_{i\in[n]} u_2(i,k)\cdot z(i,j) \qquad \forall j\in[m], k\in[m]$$

$$\sum_{i\in[n]}\sum_{j\in[m]} z(i,j) = 1$$

Not $\Delta_n \times \Delta_m$!!!

$$z_{i,j} \geq 0 \qquad \forall i\in[n], j\in[m]$$

Dividing by $z(i) > 0$

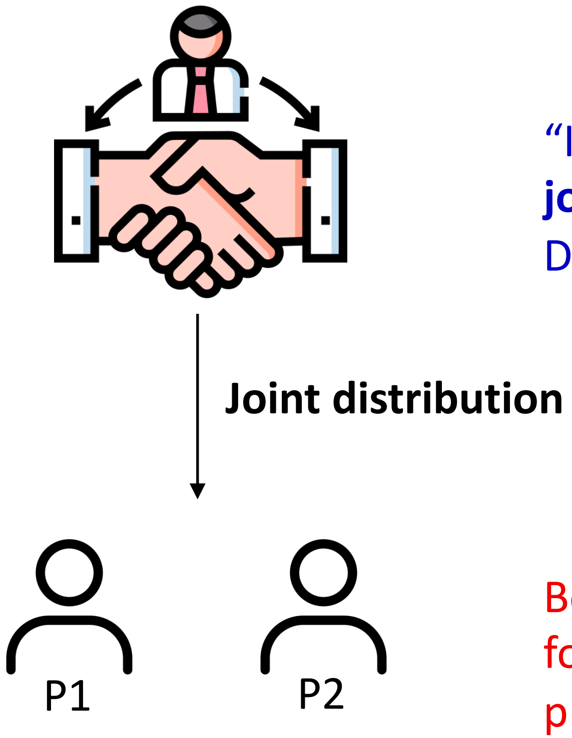
$$\sum_{j\in[m]} u_1(i,j)\cdot \underbrace{z(j|i)} \geq \sum_{j\in[m]} u_1(k,j)\cdot \underbrace{z(j|i)} \qquad \forall i\in[n], k\in[n]$$

Posterior distribution over opponent playing j if i was recommended

Coarse Correlated Equilibrium

Same as CE, but only deviate **before** observing recommendation

- Imagine players having to sign a binding contract to accept the recommended action drawn from some joint distribution



"I will sample joint actions (a_1, a_2) from this publicly known **joint distribution** and **suggest** to you what to do"
Do you pledge to stick to this joint distribution I am proposing?

Both P1 and P2: Assuming the other player pledges to follow the distribution, should I play something else over pledging to follow this distribution?

Formula for CCE

$$\sum_{i\in[n]}\sum_{j\in[m]} u_1(i,j)\cdot z(i,j) \geq \sum_{i\in[n]}\sum_{j\in[m]} u_1(k,j)\cdot z(i,j) \qquad \forall k\in[n]$$

$$\sum_{i\in[n]}\sum_{j\in[m]} u_1(i,j)\cdot z(i,j) \geq \sum_{i\in[n]}\sum_{j\in[m]} u_2(i,k)\cdot z(i,j) \qquad \forall k\in[m]$$

$$\sum_{i\in[n]}\sum_{j\in[m]} z(i,j) = 1$$

$$z_{i,j} \geq 0 \qquad \forall i\in[n], j\in[m]$$

Nash as a rank-constrained CE

Another way of writing a Nash (x^*, y^*)

- Solution can be interpreted as finding the **joint** distribution via the outer product z $= x^* y^T$
- Consider the following **structural** constraints

$$z \in \mathbb{R}^{n \times m}$$

Structural constraints for independent play

$$\left\{ \begin{aligned} z &\geq 0 \\ \sum_{i \in [n], j \in [m]} z_{ij} &= 1 \\ z \text{ is rank } 1 \end{aligned} \right.$$

*This is non-convex

- We need to add **incentive constraints** (next slide)

What if we just drop the rank 1 constraint?

Gives rise to **correlated equilibrium (CE)**

If we were to deviate to action k when recommended i

$$\sum_{j \in [m]} u_1(i,j) \cdot z(i,j) \geq \sum_{j \in [m]} u_1(k,j) \cdot z(i,j) \quad \forall i \in [n], k \in [n]$$

Don't deviate

Deviate to action k

Suppose P1 was recommended action i

Or **Coarse Correlated Equilibrium (CCE)**

$$\sum_{i \in [n]} \sum_{j \in [m]} u_1(i,j) \cdot z(i,j) \geq \sum_{i \in [n]} \sum_{j \in [m]} u_1(k,j) \cdot z(i,j) \quad \forall k \in [n]$$

Just deviate to action k

Don't deviate

Deviate to action k

Of course, need to consider other player's incentive (symmetric)

Properties of correlated eqm

$CCE \supseteq CE \supseteq Nash$

- Every Nash is a CE which is also a CCE
- Nash is CE with rank 1 constraint on z
- CCE constraints can be derived from CE

$$\sum_{j\in[m]} u_1(i,j)\cdot z(i,j) \geq \sum_{j\in[m]} u_1(k,j)\cdot z(i,j) \qquad \forall i\in[n], k\in[n]$$

$$\Big\downarrow \text{ sum over all possible i}$$

$$\sum_{i\in[n]}\sum_{j\in[m]} u_1(i,j)\cdot z(i,j) \geq \sum_{i\in[n]}\sum_{j\in[m]} u_1(k,j)\cdot z(i,j) \qquad \forall k\in[n]$$

CE/CCE can be found in poly time

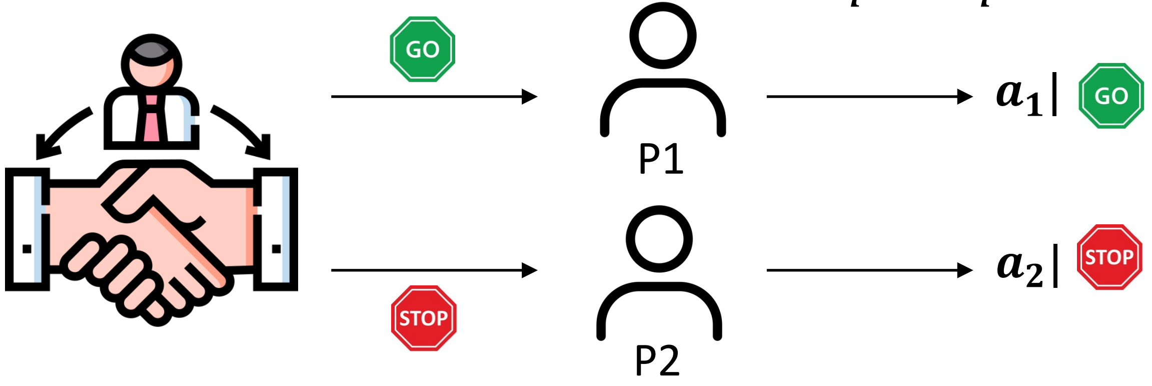
- Just linear feasibility constraint
- Objective is still “free”, can optimize for e.g., social welfare

Optional:

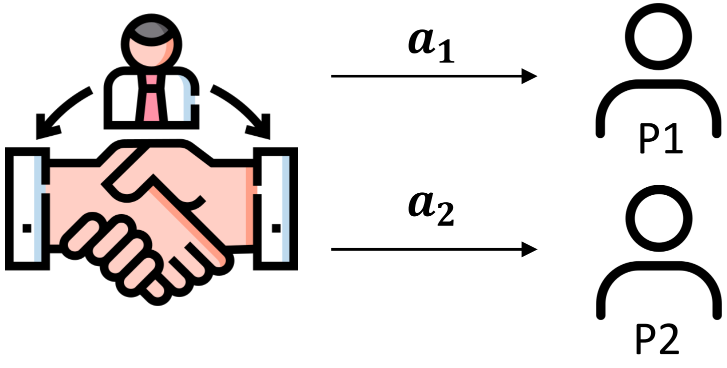
(C)CE can be seen as a zero-sum game between mediator and deviators

Classical definition of a CE based on signals from mediator

- Players receive signals from mediator (signals between players are correlated), conditions their action (possible randomized) on signal
- Strategically equivalent to mediator recommending the action directly
  - Bypass mapping from signal to conditional strategy.
  - Instance of the *revelation principle*



versus



Mediator need not be an “individual”, can be a signaling device

- Examples: Traffic lights(?)

Equilibrium and Regret

"What did it cost?" — "Everything."

-Thanos, Avengers: Infinity War

Recall: no-regret learning in 2p0s games

Average iterates converge to NE in 2p0s games

- Basic argument: saddle-point residual approaches 0 as $T \to \infty$

We **don't** get a NE in general-sum games. Why?

$$\max_y \sum_{\tau=1}^{t} \langle x^{(\tau)}, Ay \rangle - \langle x^{(\tau)}, Ay^{(\tau)} \rangle \leq R_1$$

$$\min_x \sum_{\tau=1}^{t} \langle x^{(\tau)}, Ay^{(\tau)} \rangle - \langle x, Ay^{(\tau)} \rangle \leq R_2$$

These look approximate like incentive compatible constraints
Why can't we use the same argument to obtain Nash?

Taking averages and summing

$$\max_y \langle \frac{\sum_{\tau=1}^{t} x^{(\tau)}}{t}, Ay \rangle - \min_x \langle x, A\frac{\sum_{\tau=1}^{t} y^{(\tau)}}{t} \rangle \leq \frac{R_1 + R_2}{t}$$

Assume P2 and P1 play via no-regret algorithms. From P2's perspective since it is no-regret (P1 is the same), we have

$$\max_y \sum_{\tau=1}^{t} \langle x^{(\tau)}, By\rangle - \langle x^{(\tau)}, By^{(\tau)}\rangle \le R$$

We want to try to get $\frac{x^{(\tau)}}{t} \approx x^*, \frac{y^{(\tau)}}{t} \approx y^*$

$$\max_y \sum_{\tau=1}^{t} \underbrace{\langle \frac{x^{(\tau)}}{t}, By\rangle}_{1/t} - \underbrace{\langle \frac{x^{(\tau)}}{t}, B\frac{y^{(\tau)}}{t}\rangle}_{1/t^2} \le R$$

Doesn't quite work..., isn't a straightforward division by t.

For 2p0s games we overcame this by summing this with the *other* player's incentive compatibility to give saddle point residual, requires payoff matrix being the same (or negated)

What if we talk about average of joint strategies?

$$\max_y \sum_{\tau=1}^{t} \underbrace{\langle \frac{x^{(\tau)}}{t} y^{\mathrm{T}}, B\rangle}_{\text{Payoff if we switched to play } y\text{, P1 stays in } \frac{\sum x^{(\tau)}}{t}} - \underbrace{\langle \frac{x^{(\tau)} {y^{(\tau)}}^{\mathrm{T}}}{t}, B\rangle}_{\text{Average payoff we got, also equal to average joint strategy}} \le \frac{R}{t}$$

Payoff if we switched to play y, P1 stays in $\frac{\sum x^{(\tau)}}{t}$

Average payoff we got, also equal to average **joint** strategy

Seems to work!

$\frac{R}{t} \to 0 \Rightarrow$ almost no incentive to deviate

Self-play →average strategy gives CCE!

Average of product distributions give CCE (this is in general NOT rank 1)

$$\max_y \langle \underbrace{\sum_{\tau=1}^{t} \frac{x^{(\tau)} y^{\mathrm{T}}}{t}}, B\rangle - \langle \sum_{\tau=1}^{t} \frac{x^{(\tau)} {y^{(\tau)}}^{\mathrm{T}}}{t}, B\rangle \le \frac{R}{t}$$

$$\sum_{\tau=1}^{t} x^{(\tau)} = \sum_{\tau=1}^{t} \sum_{j\in[m]} x^{(\tau)} {y^{(\tau)}}^{\mathrm{T}}$$

Because $y^{\{(\tau)\}}$ is a distribution (sums to 1)

When R equal 0, exactly the CCE constraints!

Can we get CE instead of just CCE?

Long story short, **yes**, just strengthen the regret minimizer

- So far, we have been dealing with **external** regret
- For CE we will require zero **internal** regret

External regret means we choose some a' for all our past actions

- Rock, Rock, Scissors, Paper, Rock →Rock, Rock, Rock, Rock, Rock

Internal regret means we choose a single $a \rightarrow a'$

- Only change to a' if a was chosen initially
- Rock, Rock, **Scissors**, Paper, Rock →Rock, Rock, **Paper**, Paper, Rock

Can also switch to randomized actions, but often unnecessary to include. Why?

Swap regret means we have a mapping $\phi(a) = a'$

- Every time we played a in the past, we play $\phi(a)$ instead
- Rock, Rock, Scissors, Paper, Rock →Paper, Paper, Rock, Scissors, Paper
  - For $\phi =$ Rock→Paper, Scissors→Rock, Rock→Paper

From External to swap regret minimizers

Exists a *generic* method that uses external regret minimizers to give swap regret minimizers

- Incurs swap regret an extra factor of $|A|$

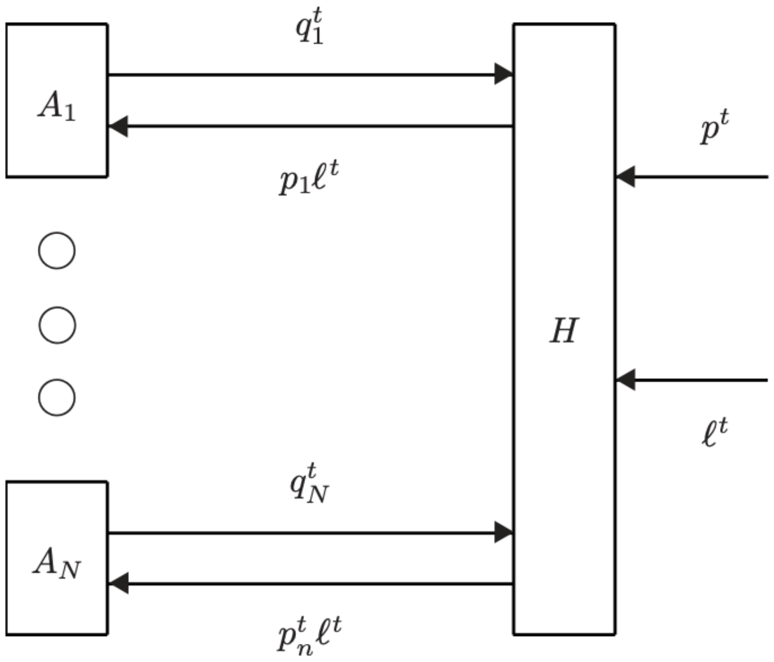


Figure 4.1. The structure of the swap regret reduction.

Source: *Algorithmic Game Theory*

What about (C)CE in EFGs?

Many different variants defined differently

- NF(C)CE: Convert to normal form and solve for CE the usual way
  - Note: naively doing this is exponential sized!
- EF(C)CE: Extensive form (C)CE (Von Stengel and Forges, 2008)
  - Mediator reveals recommended actions incrementally at infosets (sealed recommendations)
  - Don't know what recommendations you can receive in future infosets
- In general, an entire family of equilibria can be constructed based on the concept of *hindsight rationality* (Morrill et. al., 2020)

Computationally, finding a CCE is easy

- Just use CFR

Finding a CE is a different story…

Leadership and Stackelberg Equilibrium

Let's go back to general-sum games…

Stackelberg equilibrium

Heinrich Freiherr von Stackelberg

Studied by von Stackelberg in the early-mid 1900s

Models leader-follower dynamics, handles general-sum games
- Players no longer make “moves” simultaneously
- In a general sense bilevel optimization
- *Very* frequently confused with EFGs, despite being very different

Unfortunately, different communities have different definitions
- Originally intended for market analysis
- “Appropriated” by the modern game theory community
- This portion will focus on the latter

Lots of applications
- Security games
- Also, an alternative to Nash or Correlated equilibrium for general-sum games

Stackelberg equilibrium

Two players, leader and follower

- Leader commits to a distribution, follower best responds

$$\max_x u_1(x, y)$$

*commitment by leader

$$\text{such that } y \in BR_2(x)$$

$$y \in \text{Argmax}\{u_2(x, y)\}$$

Both players maximizing can have differing utilities

Could have multiple best-responses (in fact, usually, at equilibrium)

- Choice of tiebreaking can make a big difference, we usually tiebreak in favour of the leader, also known as the *Strong* Stackelberg Equilibrium

Pure Strategy Commitments

5	**-4**
-5	**2**

-3	**1**
4	**-1**

If Leader player action 1, Follower plays action 2 (-3 < 1)

- Leader gets payoff of -4

If Leader plays action 2, Follower plays action 1(4 > -1)

- Leader gets payoff of -5

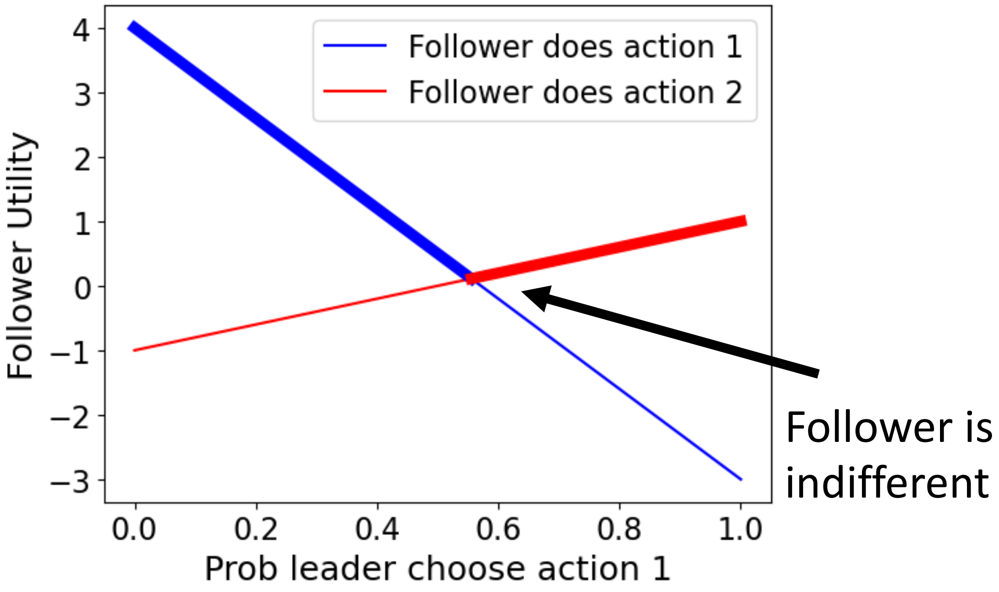
Best pure strategy commitment is to play <u>action 1</u>

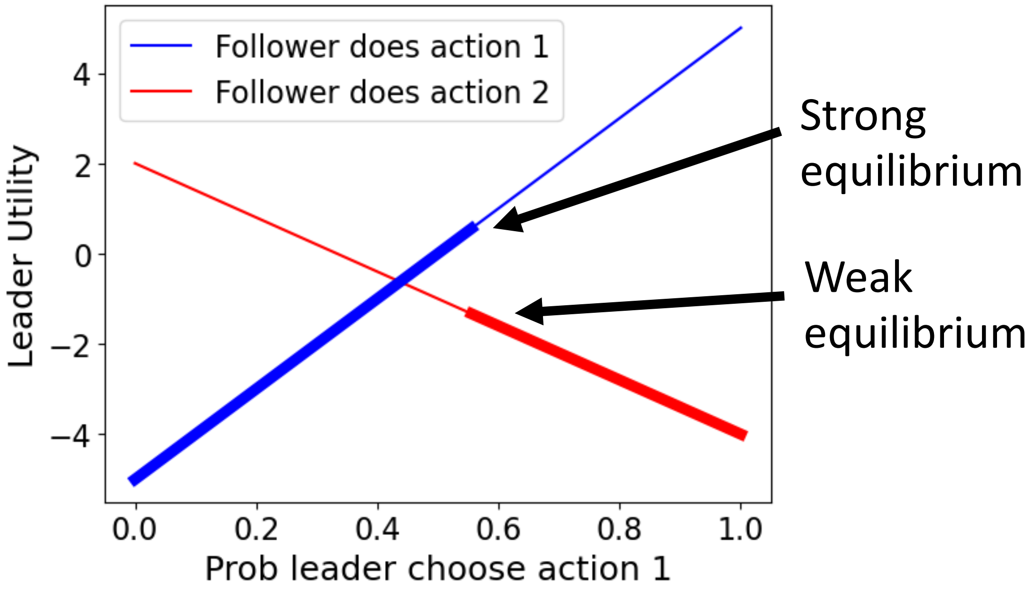
Leader can do better by randomizing!

Mixed Strategy Commitments

5	-4
-5	2

-3	1
4	-1





We typically assume strong equilibrium, i.e., tiebreaks in favour of leader

Special cases of SSE

Fully cooperative

- Highest possible payoff, no coordination issues (why?)

Fully competitive

- Leader's strategy is same as zero-sum game (why?)

Relationship to Nash

- Leader's payoff is no less than **any** mixed NE, but could be strictly higher (why?)

A small nuance: tiebreaking rules

Strong Stackelberg eqm: break ties based on best leader payoff

- How is this justified?

What about the weak Stackelberg equilibrium?

- Why does or doesn’t this make sense?
- [Illustrate by drawings]
- What about others? E.g., averaging?

Multiple LP method

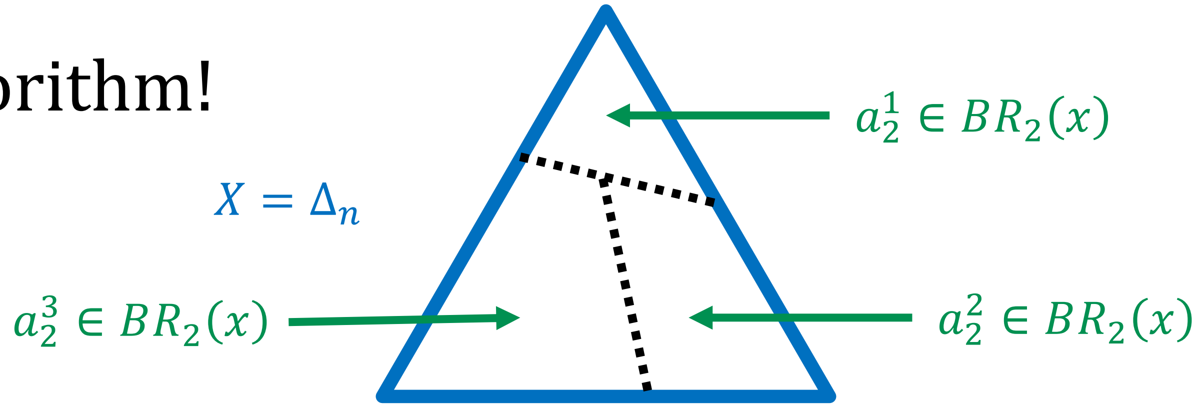
Assuming Strong Stackelberg equilibrium

- Follower best response is always deterministic (up to payoff equivalence)
- In matrix games, there are a finite number of deterministic best responses

Idea: split leader (mixed) strategy space into finitely many regions, one for each best response a_2

- Each region is either empty or a polytope in leader's strategy space
- Polytope is defined by constraints on x such that some $a_2 \in BR_2(x)$
- For each region, set objective to maximize leader payoff, subject to x lying inside polytope → solve an LP

Polynomial time algorithm!



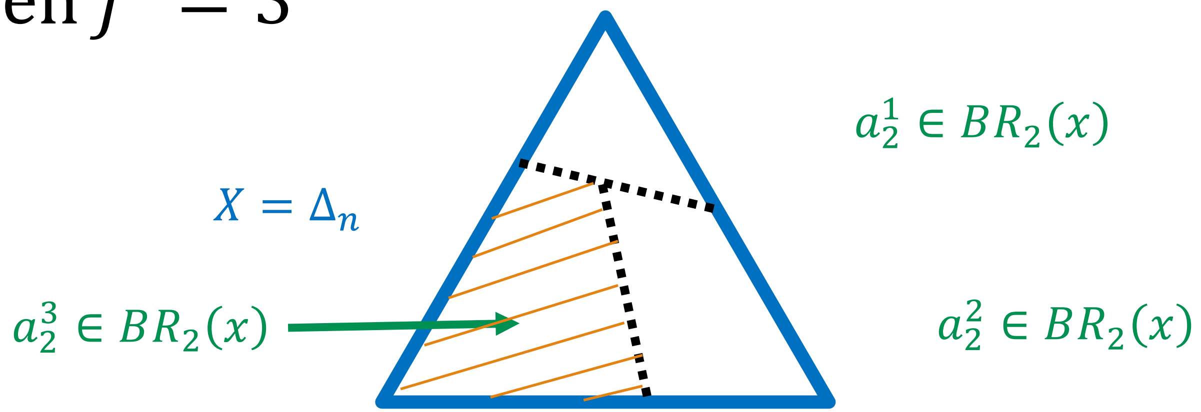
Multiple LP method

Repeat for all $j^* \in [m]$

$$\begin{aligned} \max \quad & x^T A e_j^* \\ x^\mathrm{T} B e_{j^*} \geq x^\mathrm{T} B e_j & \qquad \forall j \in [m] \\ x \in & \Delta_n \end{aligned}$$

Best achievable commitment from leader that indices j^* as best response, **could be infeasible**!

Example when $j^* = 3$



Take maximums right at the end

- Guaranteed that at least one j^* has a feasible LP (why?)

Stackelberg Security Games

r security resources, t targets, assign resources to targets optimally assuming attacker best responds to assignment

- Combinatorial explosion of number of possible assignments
- Idea: work in the space of *coverage probabilities* rather than individual actions

Table 1: Example payoffs for an attack on a target.

	Covered	Uncovered
Defender	5	–20
Attacker	–10	30

ORIGAMI/ERASER algorithm

- Essentially water filling

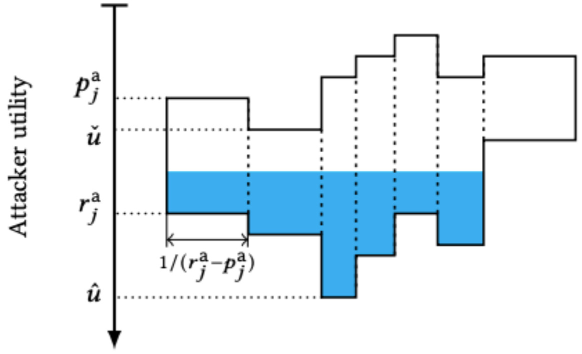
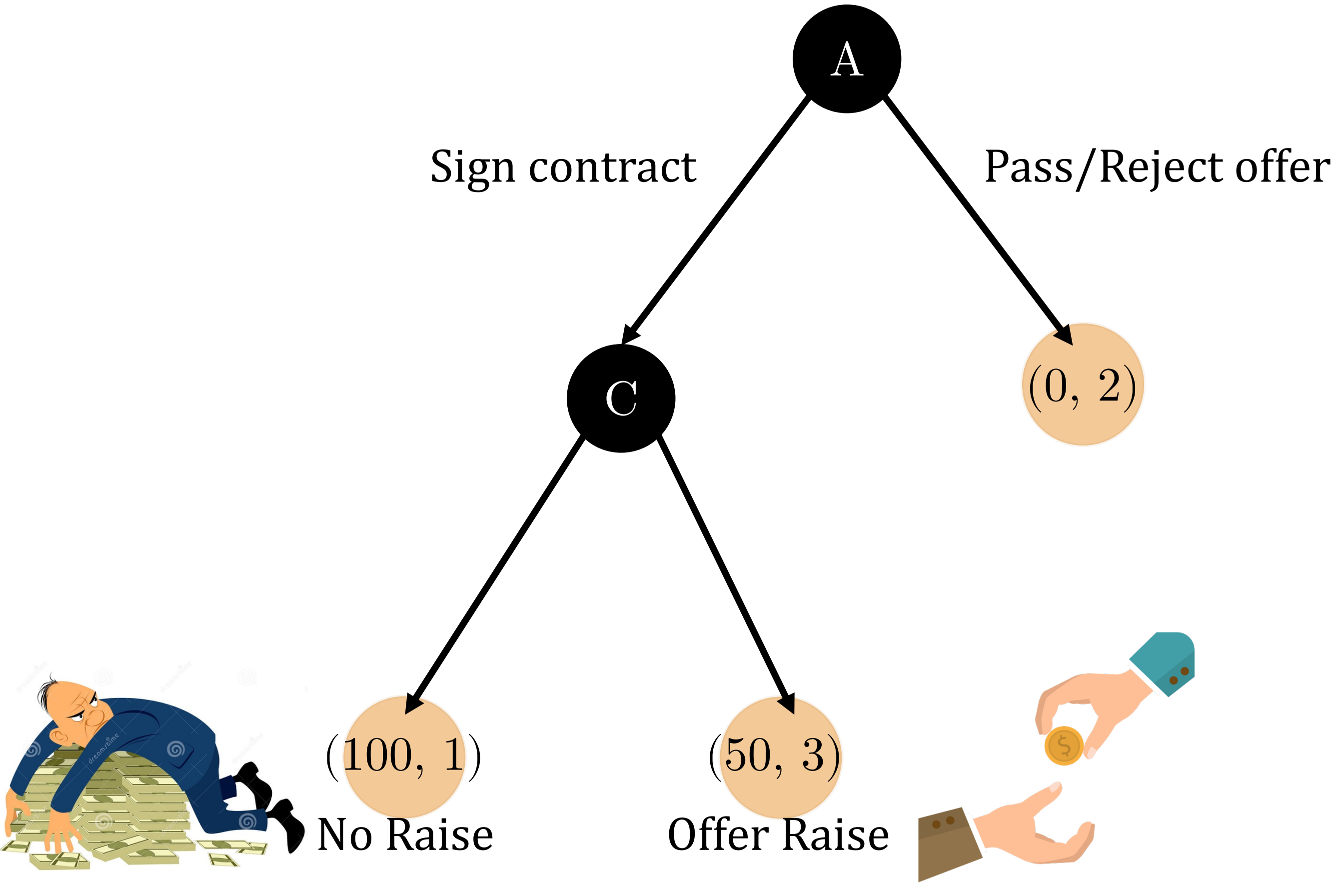


Figure 1: Visualizing a level coverage.

https://projects.iq.harvard.edu/files/teamcore/files/2009_7_teamcore_computing_aamas_09.pdf

Stackelberg Eqm in EFGs (I)

Payoffs are shown as (Company, Applicant), or (Leader, Follower)



*AKA the Leader

Played between **Company** (C) and **Applicant** (A) *AKA the follower

Applicant has an option of signing a 6-year contract with the company

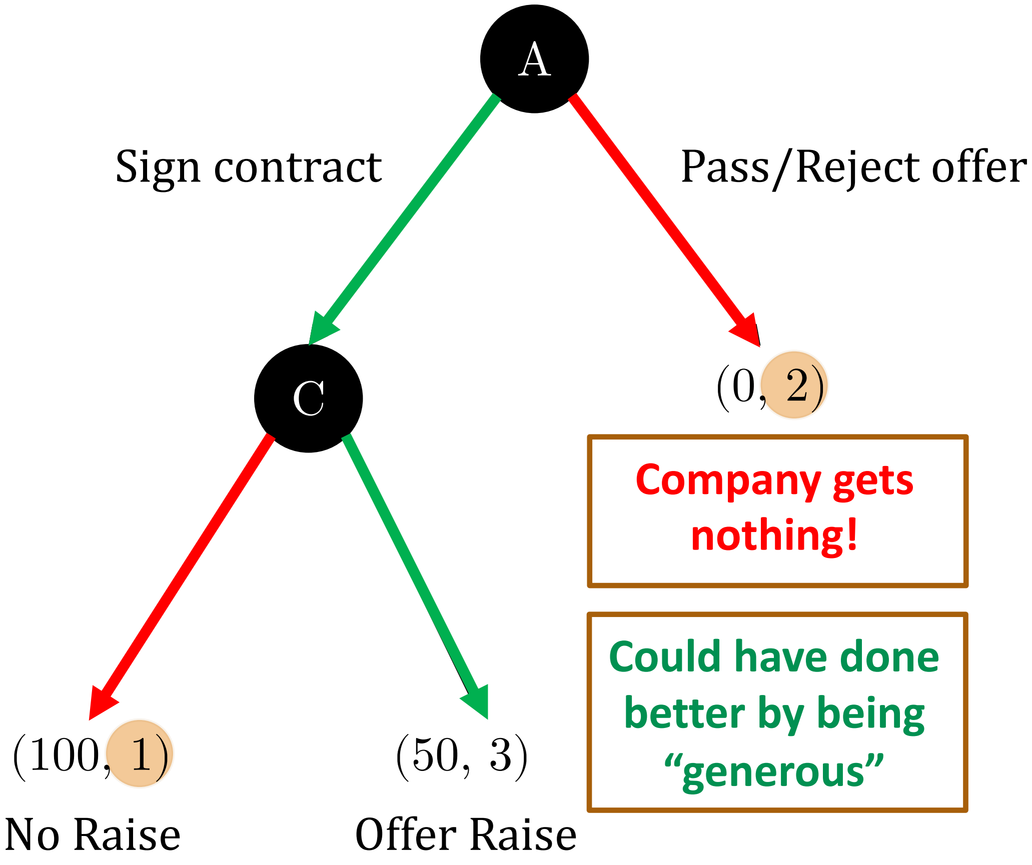
After 3 years, company can decide to give a raise or otherwise

Image sources: https://www.pngitem.com/ https://www.dreamstime.com/

Solving the Hiring Game

Payoffs are shown as (Company, Applicant), or (Leader, Follower)

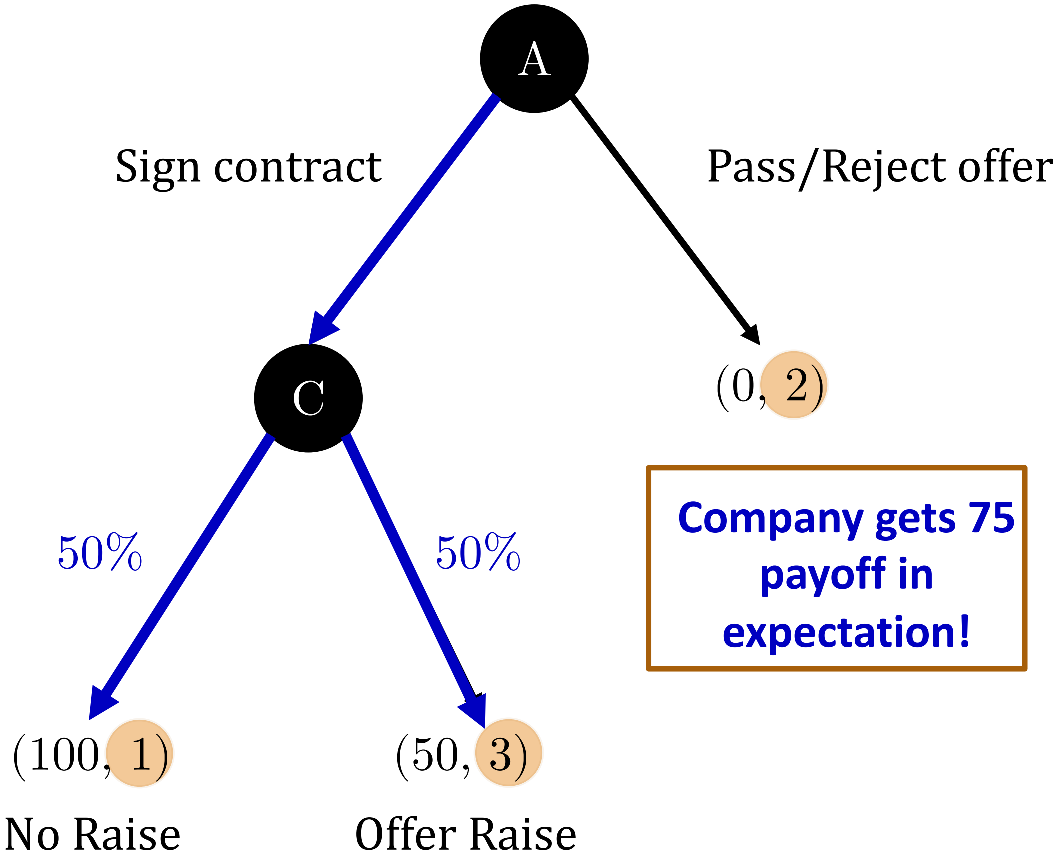
A naïve solution



Company gets nothing!

Could have done better by being "generous"

An optimal solution



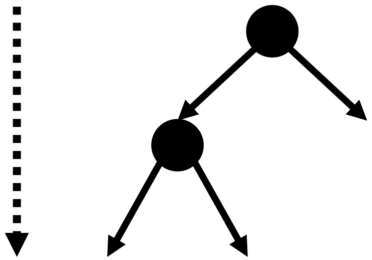
Company gets 75 payoff in expectation!

Stackelberg Eqm in EFGs (II)

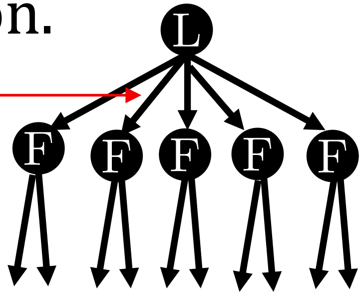
Potential confusion: why follower moves “first” in the game tree?

- Yes, but leader **commits to a random strategy**.
- Follower best responds by choosing the optimal (deterministic) path through the tree given leader’s randomization.

*game is played as physical time passes



*infinitely many possible commitments



*Order that commitment to entire strategies are made

- Some people, e.g., economists feel SE is subsumed by EFGs for this reason

Computational Complexity

- Perfect information without chance: poly-time algorithm using DP
- With chance: NP-hard, poly-time **if** allow for correlation between players
- Imperfect information games: NP-hard

*Why won’t the multiple LP method run in poly time?

- For general SSE in EFGs, can use integer programs