MS&E 233
Game Theory, Data Science and AI
Lecture 7

Vasilis Syrgkanis

Assistant Professor

Management Science and Engineering

(by courtesy) Computer Science and Electrical Engineering

Institute for Computational and Mathematical Engineering

Computational Game Theory for Complex Games

1
- Basics of game theory and zero-sum games (T)
- Basics of online learning theory (T)
- Solving zero-sum games via online learning (T)
- *HW1: implement simple algorithms to solve zero-sum games*
- Applications to ML and AI (T+A)
- *HW2: implement boosting as solving a zero-sum game*

2
- Basics of extensive-form games
- Solving extensive-form games via online learning (T)
- *HW3: implement agents to solve very simple variants of poker*

3
- **General games, equilibria and online learning (T)**
- Online learning in general games, multi-agent RL (T+A)
- *HW4: implement no-regret algorithms that converge to correlated equilibria in general games*

Data Science for Auctions and Mechanisms

4
- Basics and applications of auction theory (T+A)
- Learning to bid in auctions via online learning (T)
- *HW5: implement bandit algorithms to bid in ad auctions*

5
- Optimal auctions and mechanisms (T)
- Simple vs optimal mechanisms (T)
- *HW6: calculate equilibria in simple auctions, implement simple and optimal auctions, analyze revenue empirically*

6
- Optimizing mechanisms from samples (T)
- Online optimization of auctions and mechanisms (T)
- *HW7: implement procedures to learn approximately optimal auctions from historical samples and in an online manner*

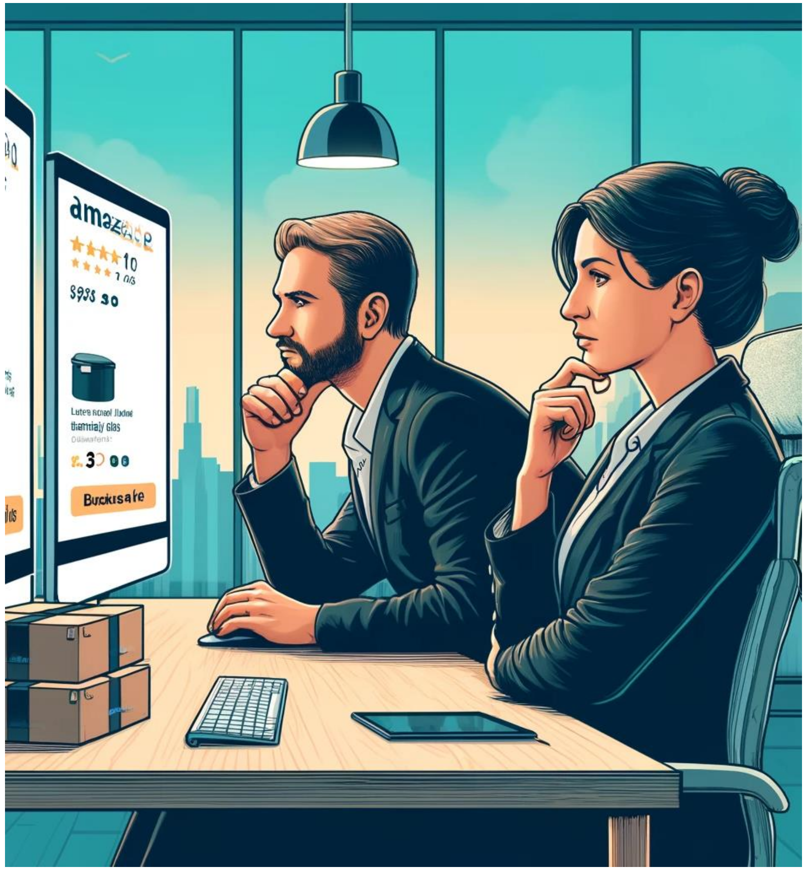
Further Topics

7
- Econometrics in games and auctions (T+A)
- A/B testing in markets (T+A)
- *HW8: implement procedure to estimate values from bids in an auction, empirically analyze inaccuracy of A/B tests in markets*

Guest Lectures

- Mechanism Design for LLMs, Renato Paes Leme, Google Research
- Auto-bidding in Sponsored Search Auctions, Kshipra Bhawalkar, Google Research

General Multiplayer Games



Many real-world games are not zero-sum

Are there simple scalable algorithms that compute Nash equilibria or other reasonable solution concepts in general games?

Learning to Communicate with Deep Multi-Agent Reinforcement Learning

Jakob N. Foerster[1,†]
jakob.foerster@cs.ox.ac.uk

Yannis M. Assael[1,†]
yannis.assael@cs.ox.ac.uk

Nando de Freitas[1,2,3]
nandodefreitas@google.com

Shimon Whiteson[1]
shimon.whiteson@cs.ox.ac.uk

[1]University of Oxford, United Kingdom
[2]Canadian Institute for Advanced Research, CIFAR NCAP Program
[3]Google DeepMind

nature

Explore content ⌄ **About the journal** ⌄ **Publish with us** ⌄

nature > articles > article

Article | Published: 30 October 2019

Grandmaster level in StarCraft II using multi-agent reinforcement learning

OpenAI Five

Our team of five neural networks, OpenAI Five, has started to defeat amateur human teams at Dota 2.

Recent Successes

Much harder to compute equilibria;
theory typically considers relaxed solution concepts that are computationally easy
practice typically uses similar algorithms as in zero-sum games as good heuristics

Equilibria in General Games

Battle of Partners

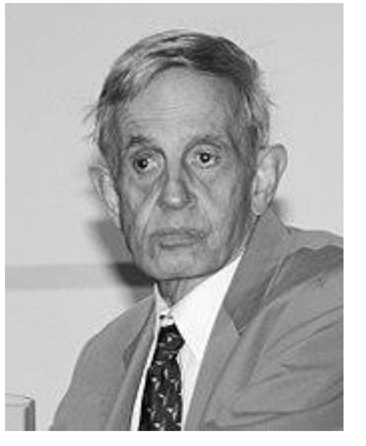
Partner 1 \ Partner 2	Opera (O)	Football (F)
Opera (O)	3, 1	0, 0
Football (F)	0, 0	1, 3

How should partners behave?

Recap: Mixed Nash Equilibrium

- A mixed strategy profile $\sigma = (\sigma_1, \ldots, \sigma_n)$ is a Nash equilibrium if no player is better off in expectation, by choosing another strategy s_i'

$$\forall s_i' \in S_i: E_{s_1 \sim \sigma_1, \ldots, s_n \sim \sigma_n}[u_i(s_1, \ldots, s_n)] \geq E_{s_{-i} \sim \sigma_{-i}}\left[u_i\left(s_i', s_{-i}\right)\right]$$



Recap: Existence of Nash Equilibrium [Nash1950]

Every n player finite action game has at least one mixed Nash equilibrium

EQUILIBRIUM POINTS IN N-PERSON GAMES

BY JOHN F. NASH, JR.*

PRINCETON UNIVERSITY

Communicated by S. Lefschetz, November 16, 1949

One may define a concept of an n-person game in which each player has a finite set of pure strategies and in which a definite set of payments to the n players corresponds to each n-tuple of pure strategies, one strategy being taken for each player. For mixed strategies, which are probability distributions over the pure strategies, the pay-off functions are the expectations of the players, thus becoming polylinear forms in the probabilities with which the various players play their various pure strategies.

Any n-tuple of strategies, one for each player, may be regarded as a point in the product space obtained by multiplying the n strategy spaces of the players. One such n-tuple counters another if the strategy of each player in the countering n-tuple yields the highest obtainable expectation for its player against the $n - 1$ strategies of the other players in the countered n-tuple. A self-countering n-tuple is called an equilibrium point.

The correspondence of each n-tuple with its set of countering n-tuples gives a one-to-many mapping of the product space into itself. From the definition of countering we see that the set of countering points of a point is convex. By using the continuity of the pay-off functions we see that the graph of the mapping is closed. The closedness is equivalent to saying: if $P_1, P_2, \ldots$ and $Q_1, Q_2, \ldots, Q_n, \ldots$ are sequences of points in the product space where $Q_n \rightarrow Q$, $P_n \rightarrow P$ and Q_n counters P_n then Q counters P.

Since the graph is closed and since the image of each point under the mapping is convex, we infer from Kakutani's theorem[1] that the mapping has a fixed point (i.e., point contained in its image). Hence there is an equilibrium point.

In the two-person zero-sum case the "main theorem"[2] and the existence of an equilibrium point are equivalent. In this case any two equilibrium points lead to the same expectations for the players, but this need not occur in general.

Battle of Partners

Partner 2

Partner 1	Opera (O)	Football (F)
Opera (O)	3, 1	0, 0
Football (F)	0, 0	1, 3

How should partners behave?

Choose whether you will go to football or opera favorite activity

Football

Opera

Choose whether you will go to football or opera favorite activity

Football

Opera

Choose whether you will go to football or opera favorite activity

Football	0%
Opera	0%

Battle of Partners

Partner 2

Partner 1		Opera (O)	Football (F)
	Opera (O)	3, 1	0, 0
	Football (F)	0, 0	1, 3

How should partners behave?

Battle of Partners

		Partner 2	
		Opera (O)	Football (F)
Partner 1	Opera (O)	3, 1	0, 0
	Football (F)	0, 0	1, 3

How should partners behave?

For a full support NE both rows need to yield the same utility to row player

$$3y_1 + 0y_2 = 0y_1 + 1y_2 \Rightarrow y_2 = 3y_1$$

and columns need to yield the same utility to column player

$$1x_1 + 0x_2 = 0x_1 + 3x_2 \Rightarrow x_1 = 3x_2$$

Battle of Partners

Partner 2

Partner 1

		1/4	3/4
		Opera (O)	Football (F)
3/4	Opera (O)	3, 1	0, 0
1/4	Football (F)	0, 0	1, 3

How should partners behave?

For a full support NE both rows need to yield the same utility to row player

$$3y_1 + 0y_2 = 0y_1 + 1y_2 \Rightarrow y_2 = 3y_1$$

and columns need to yield the same utility to column player

$$1x_1 + 0x_2 = 0x_1 + 3x_2 \Rightarrow x_1 = 3x_2$$

Battle of Partners

Partner 2

Partner 1		1/4	3/4
		Opera (O)	Football (F)
3/4	Opera (O)	$\frac{3}{4} \cdot \frac{1}{4}$ 3, 1	$\frac{3}{4} \cdot \frac{3}{4}$ 0, 0
1/4	Football (F)	$\frac{1}{4} \cdot \frac{1}{4}$ 0, 0	$\frac{1}{4} \cdot \frac{3}{4}$ 1, 3

What is the expected payoff to each player at the mixed Nash?

Column:

$$\frac{3}{4}\frac{1}{4}1 + \frac{3}{4}\frac{3}{4}0 + \frac{1}{4}\frac{1}{4}0 + \frac{1}{4}\frac{3}{4}\ 3 = \frac{12}{16}$$

Row:

$$\frac{3}{4}\frac{1}{4}3 + \frac{3}{4}\frac{3}{4}0 + \frac{1}{4}\frac{1}{4}0 + \frac{1}{4}\frac{3}{4}\ 1 = \frac{12}{16}$$

Recap: Intractability of Mixed Nash Equilibrium

- If we know the supports of the player strategies then we can easily calculate a mixed Nash equilibrium
- For games with many actions, we cannot enumerate all possible supports (combinatorial explosion)
- Turns out there is no easy way to side-step this

- Computing a mixed NE in two player games is “intractable”

- It is provable as hard as computing a “fixed point” ($f(x) = x$) of an arbitrary function f, which is considered an intractable problem

No learning dynamics will converge to a *Nash Equilibrium*, generically *for every game*, in a reasonable amount of time *in the worst-case*!

Look for other equilibrium concepts

Analyze special classes of games

No learning dynamics will converge to a *Nash Equilibrium* in *every game* in a reasonable time *in the worst-case*!

Develop heuristics that typically converge fast in practice

Correlated equilibrium, coarse correlated equilibrium

Look for other equilibrium concepts

Zero-sum games, potential games, auction games, strictly monotone games...

Analyze special classes of games

No learning dynamics will converge to a *Nash Equilibrium* in *every game* in a reasonable time *in the worst-case*!

Develop heuristics that typically converge fast in practice

Fictitious play, EXP, perturbed fictitious play, best-response dynamics, self-play...

In Search for Other Equilibrium Concepts

What if we can flip public coins?

Partner 2

Partner 1

	Opera (O)	Football (F)
Opera (O)	$\frac{1}{2}$ 3, 1	0 0, 0
Football (F)	0 0, 0	$\frac{1}{2}$ 1, 3

We flip a coin!

Heads we choose (O, O)

Tails we choose (F, F)

Choose whether you will go to football or opera favorite activity

Football

Opera

Choose whether you will go to football or opera favorite activity

Football

Opera

Choose whether you will go to football or opera favorite activity

Football

0%

Opera

0%

What if we can flip public coins?

Partner 2

Partner 1		Opera (O)	Football (F)
	Opera (O)	$\frac{1}{2}$ 3, 1	0 0, 0
	Football (F)	0 0, 0	$\frac{1}{2}$ 1, 3

We flip a coin!
Heads we choose (O, O)
Tails we choose (F, F)

Does P1 gain by not adhering to the protocol if P2 adheres?

What if we can flip public coins?

Partner 2

Partner 1

	Opera (O)	Football (F)
Opera (O)	$\frac{1}{2}$ **3**, 1	0 0, 0
Football (F)	0 **0**, 0	$\frac{1}{2}$ 1, 3

We flip a coin!
Heads we choose (O, O)
Tails we choose (F, F)

Does P1 gain by not adhering to the protocol if P2 adheres?

Heads. P2 chooses (O). If I don't choose (O), I get **0**. Now I get **3**.

What if we can flip public coins?

Partner 2

Partner 1

	Opera (O)	Football (F)
Opera (O)	$\frac{1}{2}$ 3, 1	0 **0**, 0
Football (F)	0 0, 0	$\frac{1}{2}$ **1**, 3

We flip a coin!
Heads we choose (O, O)
Tails we choose (F, F)

Does P1 gain by not adhering to the protocol if P2 adheres?

Heads. P2 chooses (O). If I don't choose (O), I get **0**. Now I get **3**.
Tails. P2 chooses (F). If I don't choose (F), I get **0**. Now I get **1**.

Structure of equilibrium distributions

		P2 (A)	P2 (B)
P1	(A)	$\frac{1}{2}$	0
P1	(B)	0	$\frac{1}{2}$

Consider a new game

You don’t know the utilities

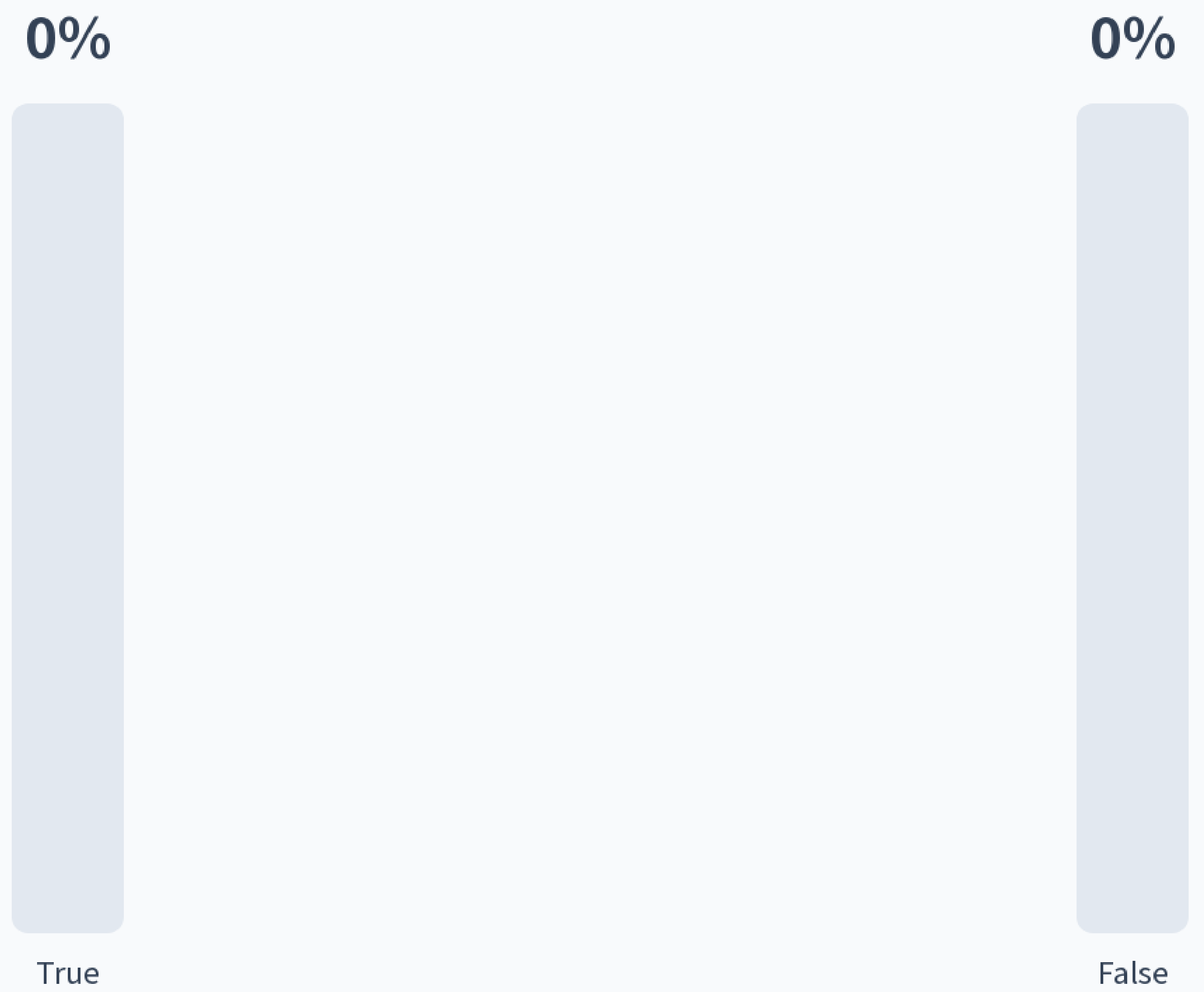
The yellow numbers depict the probability distribution over outcomes (strategy profiles)

This distribution over pairs of strategies can be the result of a mixed Nash equilibrium?

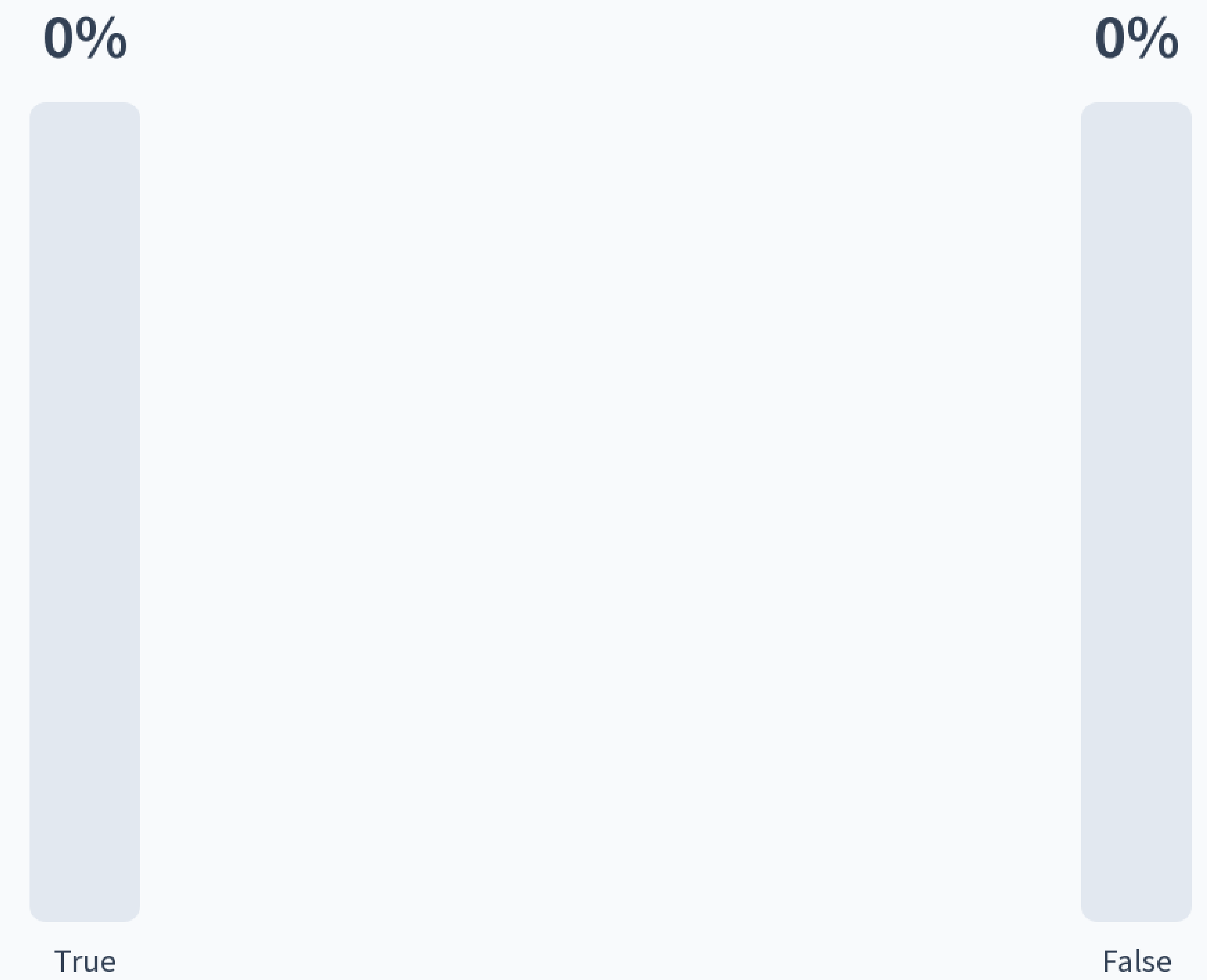
True

False

This distribution over pairs of strategies can be the result of a mixed Nash equilibrium?



This distribution over pairs of strategies can be the result of a mixed Nash equilibrium?



The Junction Game

Driver 2

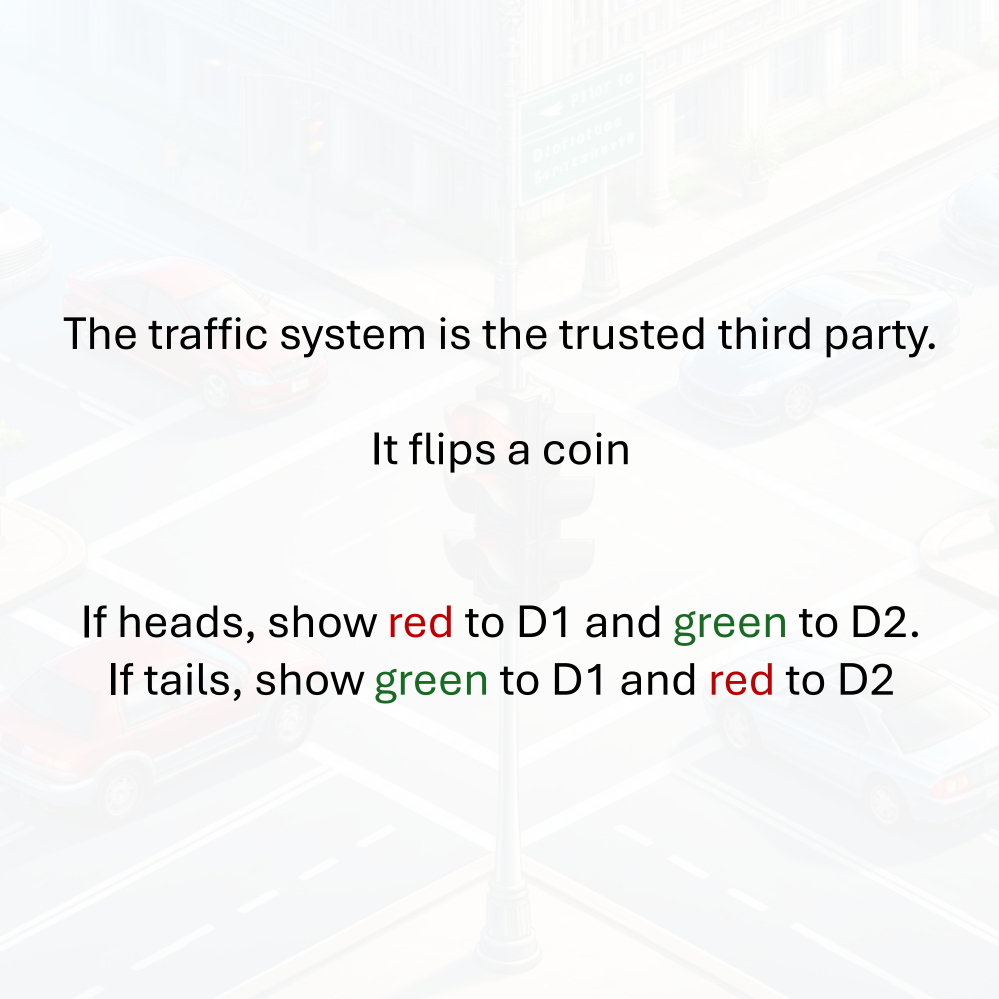
Driver 1		$\frac{1}{101}$ Pass (P)	$\frac{100}{101}$ Not Pass (N)
$\frac{1}{101}$	Pass (P)	-100, -100	1, 0
$\frac{100}{101}$	Not Pass (N)	0, 1	0, 0

Image credits: chat.openai.com

What if we have a trusted third party that can flip coins?

Driver 2

Driver 1	Pass (P)	Not Pass (N)
Pass (P)	-100, -100	1, 0
Not Pass (N)	0, 1	0, 0



The traffic system is the trusted third party.

It flips a coin

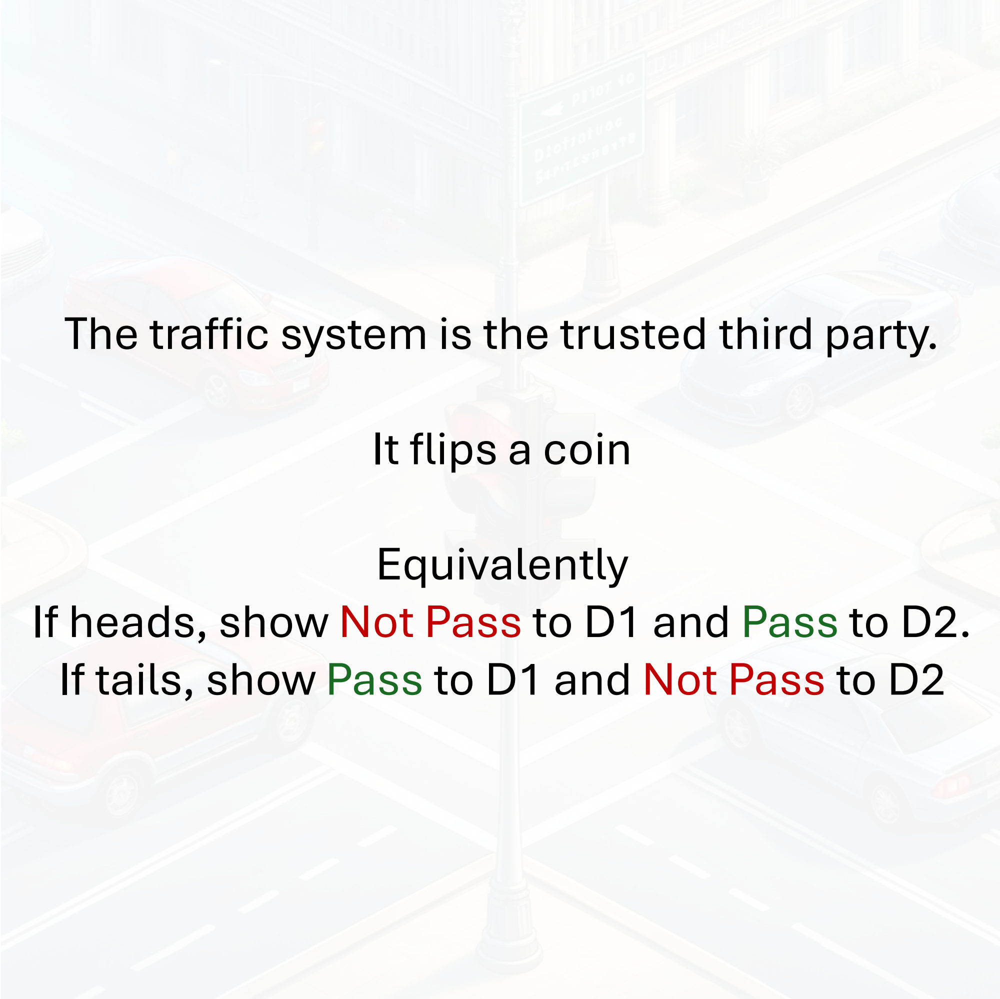
If heads, show red to D1 and green to D2.
If tails, show green to D1 and red to D2

What if we have a trusted third party that can flip coins?

Driver 2

Driver 1

	Pass (P)	Not Pass (N)
Pass (P)	0 -100, -100	$\frac{1}{2}$ 1, 0
Not Pass (N)	$\frac{1}{2}$ 0, 1	0 0, 0



The traffic system is the trusted third party.

It flips a coin

Equivalently
If heads, show Not Pass to D1 and Pass to D2.
If tails, show Pass to D1 and Not Pass to D2

Image credits: chat.openai.com

Correlated Equilibrium

- A trusted third party draws strategy profiles $s = (s_1, \dots, s_n)$ of the game from some distribution D
- Communicates to each participant their part of the profile, i.e., the recommended strategy s_i

- The distribution D is a correlated equilibrium if participants don't have incentive to deviate from their recommended strategy

$$\underbrace{\forall s_i, s_i' \in S_i:}_{\text{For any recommendation } s_i \text{ and possible deviation } s_i'} \quad \underbrace{E_{s\sim D}[\,u(s) \mid s_i\,]}_{\text{Expected utility of choosing } s_i \text{ when recommended } s_i} \quad \geq \quad \underbrace{E_{s\sim D}\left[\,u(s_i', s_{-i}) \mid s_i\,\right]}_{\text{Expected utility of deviating to } s_i' \text{ when recommended } s_i}$$

For any recommendation s_i and possible deviation s_i' | Expected utility of choosing s_i when recommended s_i $\geq$ Expected utility of deviating to s_i' when recommended s_i

Correlated Equilibria are Tractable

- Define a variable $\pi(s)$ for every strategy profile $s \in S_1 \times \cdots \times S_n$
- The variables encode a distribution

$$\sum_s \pi(s) = 1$$

- The distribution π is a correlated equilibrium if participants don't have incentive to deviate from their recommended strategy

$$\forall s_i, s_i' \in S_i: \quad \sum_{s_{-i}} \frac{\pi(s)}{\Pr(s_i)} u(s_i, s_{-i}) \geq \sum_{s_{-i}} \frac{\pi(s)}{\Pr(s_i)} u(s_i', s_{-i})$$

For any recommendation s_i and possible deviation s_i'

Expected utility of choosing s_i when recommended s_i

$\geq$

Expected utility of deviating to s_i' when recommended s_i

Correlated Equilibria are Tractable

- Define a variable $\pi(s)$ for every strategy profile $s \in S_1 \times \cdots \times S_n$
- The variables encode a distribution

$$\sum_s \pi(s) = 1$$

- The distribution π is a correlated equilibrium if participants don't have incentive to deviate from their recommended strategy

$$\forall s_i, s_i' \in S_i: \quad \sum_{\tilde{s}_{-i}} \frac{\pi(s_i, \tilde{s}_{-i})}{\Pr(s_i)} u(s_i, \tilde{s}_{-i}) \geq \sum_{\tilde{s}_{-i}} \frac{\pi(s_i, \tilde{s}_{-i})}{\Pr(s_i)} u(s_i', \tilde{s}_{-i})$$

By Bayes rule this is the conditional distribution of $\tilde{s} \sim \pi | s_i$, i.e. $\Pr_\pi(\, \tilde{s} \mid s_i \,) = \frac{\pi(\tilde{s}) 1\{\tilde{s}_i = s_i\}}{\sum_{s_{-i}} \pi(s_i, s_{-i})}$

Correlated Equilibria are Tractable

- Define a variable $\pi(s)$ for every strategy profile $s \in S_1 \times \cdots \times S_n$
- The variables encode a distribution

$$\sum_{s} \pi(s) = 1$$

- The distribution π is a correlated equilibrium if participants don't have incentive to deviate from their recommended strategy

$$\forall s_i, s_i' \in S_i: \quad \sum_{\tilde{s}_{-i}} \frac{\pi(s_i, \tilde{s}_{-i})}{\cancel{\Pr(s_i)}} u(s_i, \tilde{s}_{-i}) \geq \sum_{\tilde{s}_{-i}} \frac{\pi(s_i, \tilde{s}_{-i})}{\cancel{\Pr(s_i)}} u(s_i', \tilde{s}_{-i})$$

Correlated Equilibria are Tractable

- Define a variable $\pi(s)$ for every strategy profile $s \in S_1 \times \cdots \times S_n$
- The variables encode a distribution

$$\sum_{s} \pi(s) = 1$$

- The distribution π is a correlated equilibrium if participants don't have incentive to deviate from their recommended strategy

$$\forall s_i, s_i' \in S_i: \quad \sum_{\tilde{s}_{-i}} \pi(s_i, \tilde{s}_{-i})\, u(s_i, \tilde{s}_{-i}) \geq \sum_{\tilde{s}_{-i}} \pi(s_i, \tilde{s}_{-i})\, u\left(s_i', \tilde{s}_{-i}\right)$$

Correlated Equilibria are Tractable

- Define a variable $\pi(s)$ for every strategy profile $s \in S_1 \times \cdots \times S_n$
- The variables encode a distribution

$$\sum_{s} \pi(s) = 1$$

- The distribution π is a correlated equilibrium if participants don't have incentive to deviate from their recommended strategy

$$\forall s_i, s_i' \in S_i: \quad \sum_{\tilde{s}_{-i}} \pi(s_i, \tilde{s}_{-i}) \left(u(s_i, \tilde{s}_{-i}) - u\left(s_i', \tilde{s}_{-i}\right) \right) \geq 0$$

Correlated Equilibria are Tractable

- Define a variable $\pi(s)$ for every strategy profile $s \in S_1 \times \cdots \times S_n$
- The variables encode a distribution

$$\sum_s \pi(s) = 1$$

- The distribution π is a correlated equilibrium if participants don't have incentive to deviate from their recommended strategy

$$\forall s_i, s_i' \in S_i: \quad \sum_{s_{-i}} \pi(s) \boxed{\left(u(s) - u\left(s_i', s_{-i}\right)\right)} \geq 0$$

A known quantity $\Delta_i(s, s_i')$: utility gain for player i when switching from $s_i \to s_i'$ when others play s_{-i}

Correlated Equilibria are Tractable

- Define a variable $\pi(s)$ for every strategy profile $s \in S_1 \times \cdots \times S_n$
- The variables encode a distribution

$$\sum_{s} \pi(s) = 1$$

- The distribution π is a correlated equilibrium if participants don't have incentive to deviate from their recommended strategy

$$\forall s_i, s_i' \in S_i: \quad \sum_{s_{-i}} \pi(s)\, \Delta_i(s, s_i') \geq 0$$

- A Linear Program with variables $\pi(s)$

Why do correlated equilibria always exist?

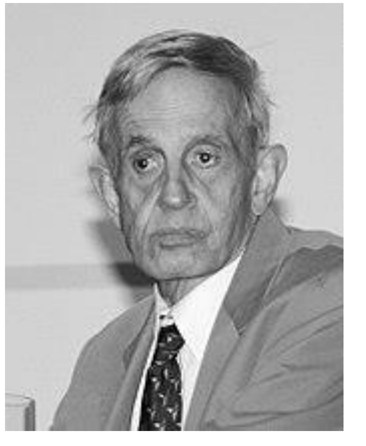
Nobody has responded yet.

Hang tight! Responses are coming in.

Recap: Mixed Nash Equilibrium

- A mixed strategy profile $\sigma = (\sigma_1, \dots, \sigma_n)$ is a Nash equilibrium if no player is better off in expectation, by choosing another strategy s_i'

$$\forall s_i' \in S_i \colon E_{s_i \sim \sigma_i,\, s_{-i} \sim \sigma_{-i}}[u_i(s_i, s_{-i})] \geq E_{s_{-i} \sim \sigma_{-i}}[u_i(s_i', s_{-i})]$$

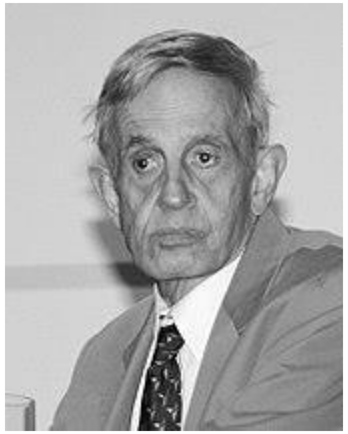


Recap: Mixed Nash Equilibrium

- A mixed strategy profile $\sigma = (\sigma_1, \dots, \sigma_n)$ is a Nash equilibrium if no player is better off in expectation, by choosing another strategy s_i'

$$\forall s_i \in \text{support}(\sigma_i), s_i' \in S_i : E_{s_{-i} \sim \boxed{\sigma_{-i}}}[u_i(s_i, s_{-i})] \geq E_{s_{-i} \sim \boxed{\sigma_{-i}}}[u_i(s_i', s_{-i})]$$

Due to independence of strategies, σ_{-i} is also the conditional distribution $s_{-i} \mid s_i$

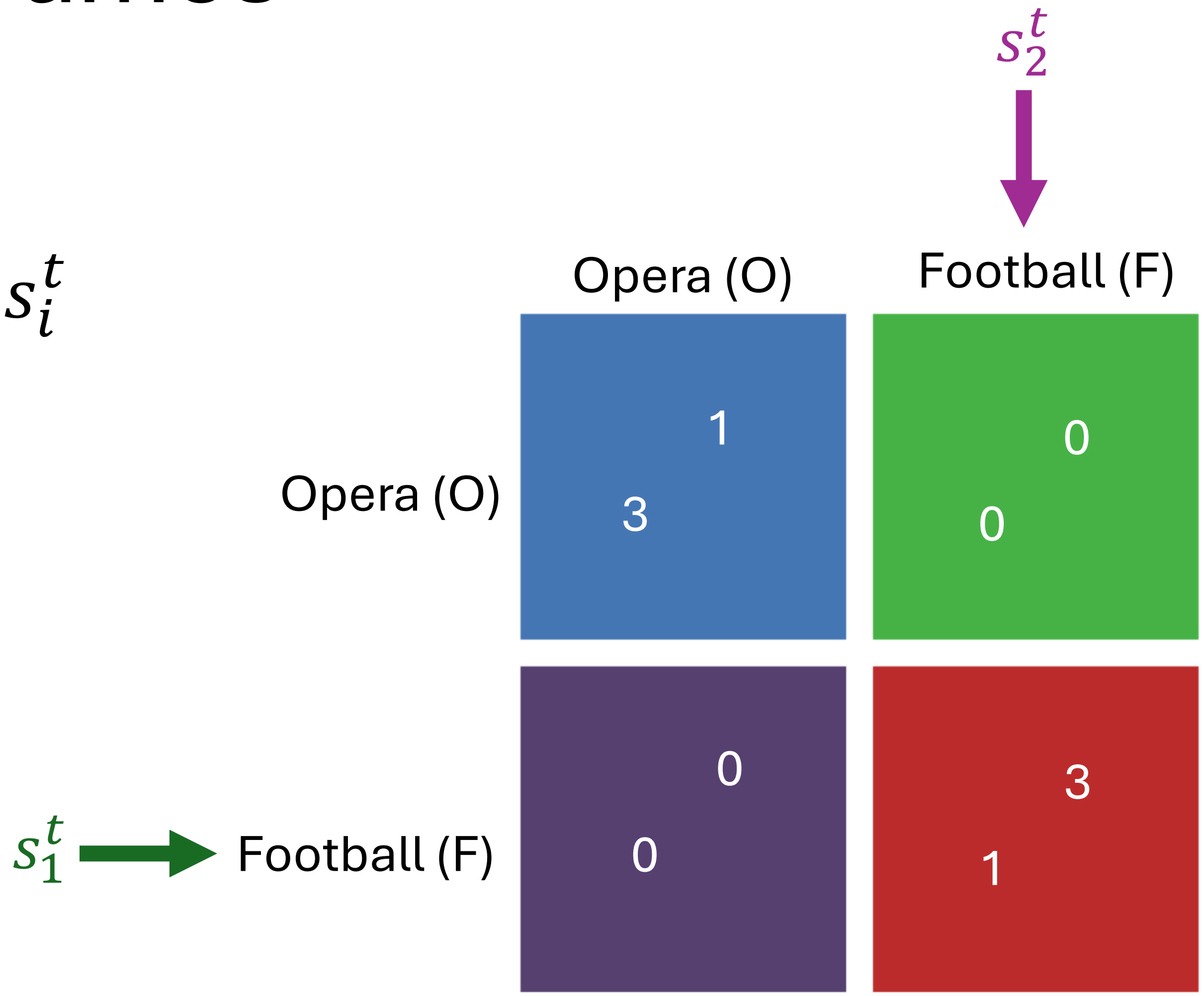


Learning Dynamics and Correlated Equilibria

Learning in General Games

At each period t:

- Each player i picks a strategy s_i^t

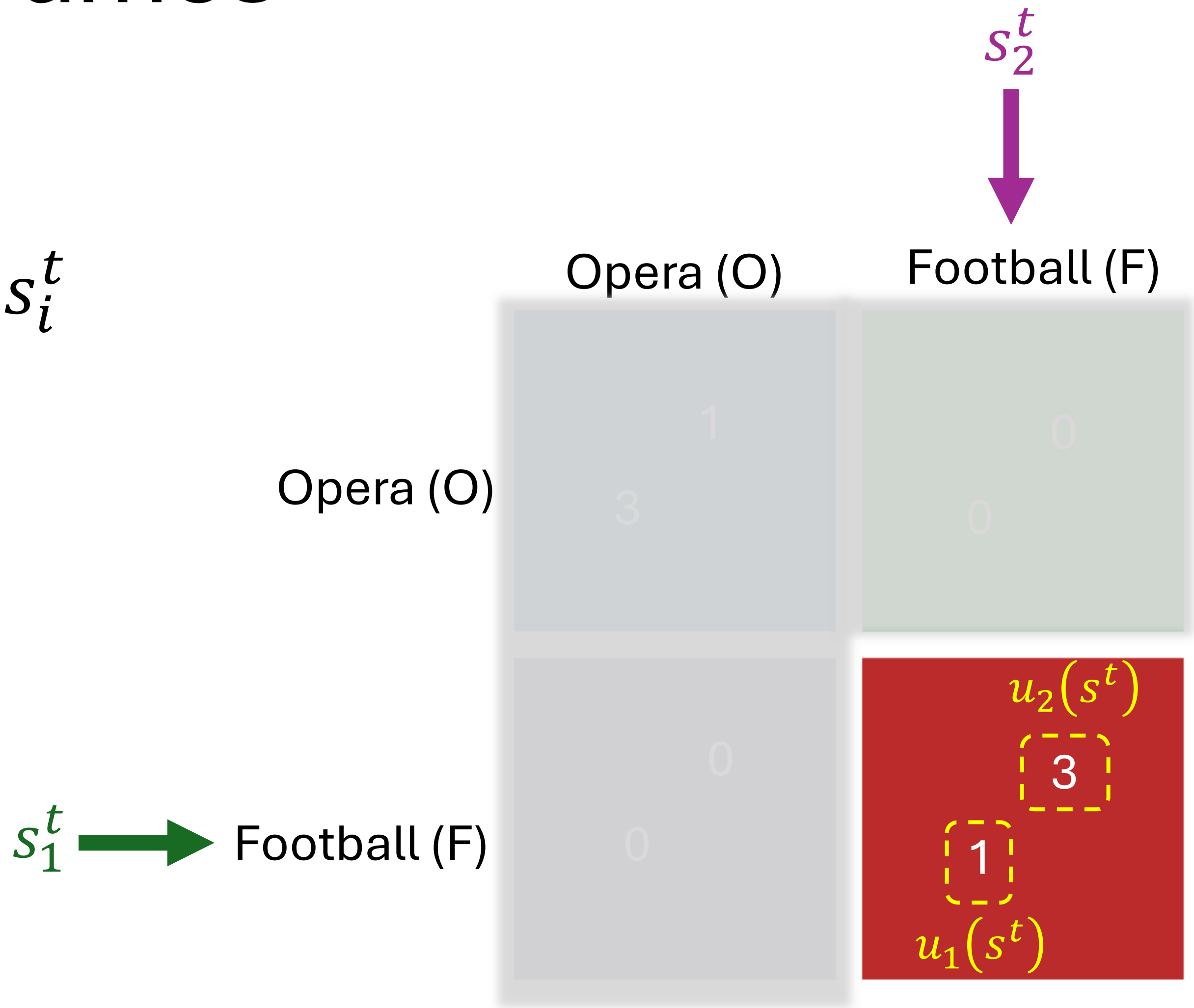


Learning in General Games

At each period t:

- Each player i picks a strategy s_i^t
- Receives a payoff

$$u_i(s^t) = u_i(s_1^t, \ldots, s_n^t)$$



Learning in General Games

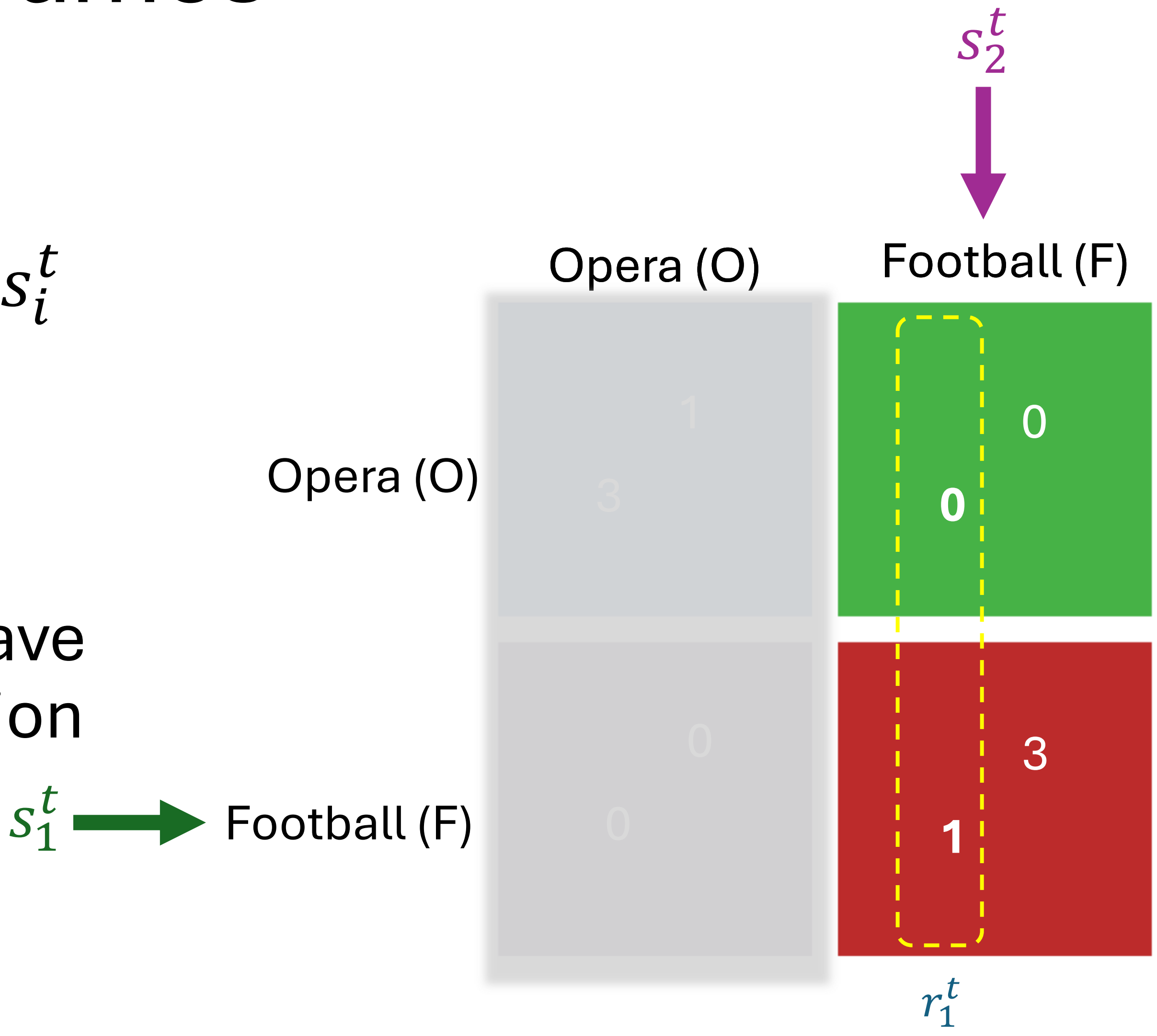
At each period t:

- Each player i picks a strategy s_i^t
- Receives a payoff

$$u_i(s^t) = u_i(s_1^t, \ldots, s_n^t)$$

- Observes utility they would have received from every other action

$$r_i^t = \left(u_i(s_i, s_{-i}^t)\right)_{s_i \in S_i}$$



Learning in General Games

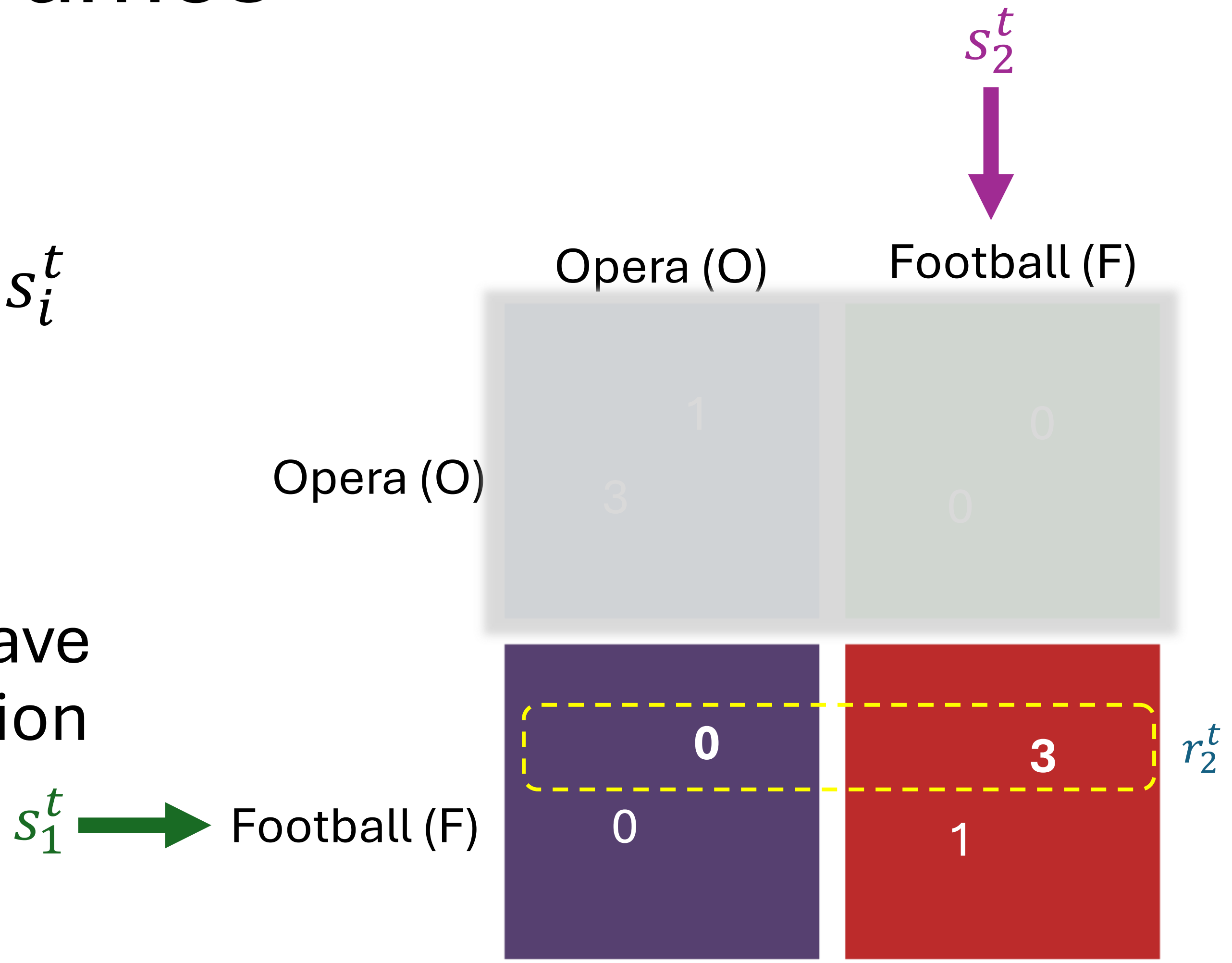
At each period t:

- Each player i picks a strategy s_i^t
- Receives a payoff

$$u_i(s^t) = u_i(s_1^t, \dots, s_n^t)$$

- Observes utility they would have received from every other action

$$r_i^t = \left(u_i(s_i, s_{-i}^t)\right)_{s_i \in S_i}$$



No-Regret Learning in General Games

What if all players use a no-regret algorithm to choose $s_i^t \sim \sigma_i^t$, which guarantees for some $\epsilon(T) \to 0$

$$\frac{1}{T}\sum_{t=1}^{T} E\left[u_i\left(s^t\right)\right] \geq \max_{s_i' \in S_i} \frac{1}{T}\sum_{t=1}^{T} E\left[u_i\left(s_i', s_{-i}^t\right)\right] - \epsilon(T)$$

No-Regret Learning in General Games

What if all players use a no-regret algorithm to choose $s_i^t \sim \sigma_i^t$, which guarantees for some $\epsilon(T) \to 0$

$$\frac{1}{T}\sum_{t=1}^{T} E\left[u_i(s^t)\right] \geq \max_{s_i' \in S_i} \frac{1}{T}\sum_{t=1}^{T} E\left[u_i(s_i', s_{-i}^t)\right] - \epsilon(T)$$

Using standard Martingale concentration inequalities, this also implies that with high probability $1-\delta$, for some $\tilde{\epsilon}(T,\delta) \to 0$:

$$\frac{1}{T}\sum_{t=1}^{T} u_i(s^t) \geq \max_{s_i' \in S_i} \frac{1}{T}\sum_{t=1}^{T} u_i(s_i', s_{-i}^t) - \tilde{\epsilon}(T,\delta)$$

What can we say about the empirical distribution of outcomes of such learning dynamics?

Empirical Distribution of Outcomes

	s^1	s^2	s^3	s^4	s^5	s^6	s^7	s^8	s^9	s^{10}	time...
s_1^t:	O	O	F	O	O	F	F	F	O	O	
s_2^t:	O	F	F	O	O	O	F	O	O	O	

Want to understand the empirical distribution of strategy profiles (aka “outcomes”)

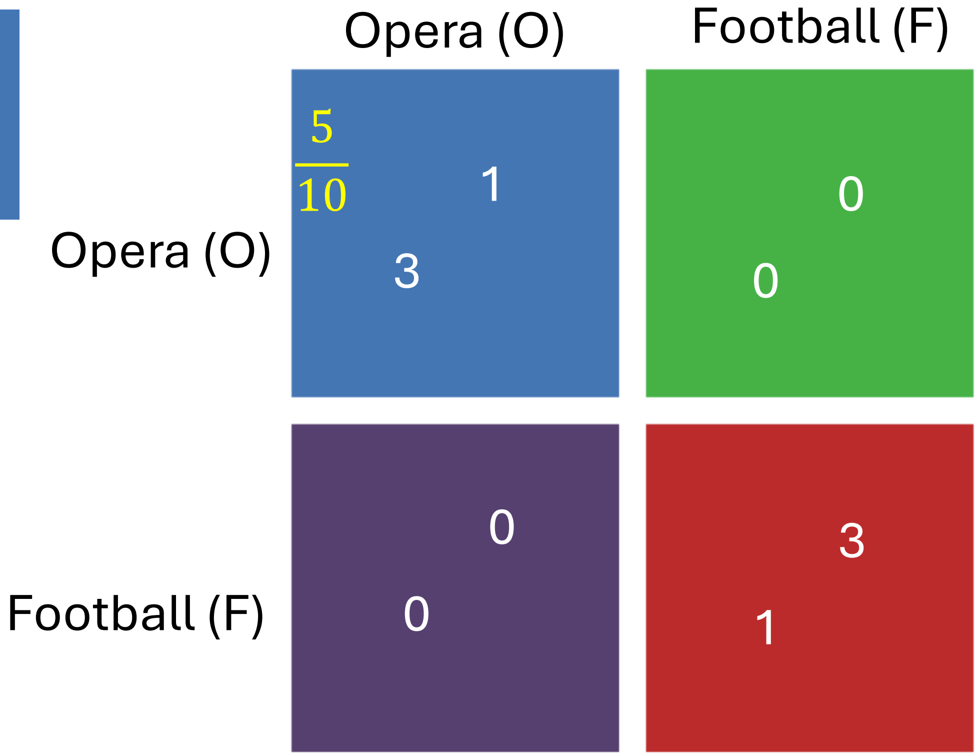
	Opera (O)	Football (F)
Opera (O)	1 3	0 0
Football (F)	0 0	3 1

Empirical Distribution of Outcomes

	s^1	s^2	s^3	s^4	s^5	s^6	s^7	s^8	s^9	s^{10}
s_1^t:	O	O	F	O	O	F	F	F	O	O
s_2^t:	O	F	F	O	O	O	F	O	O	O

time…

Want to understand the empirical distribution of strategy profiles (aka “outcomes”)



Empirical Distribution of Outcomes

	s^1	s^2	s^3	s^4	s^5	s^6	s^7	s^8	s^9	s^{10}
s_1^t:	O	O	F	O	O	F	F	F	O	O
s_2^t:	O	F	F	O	O	O	F	O	O	O

time...

Want to understand the empirical distribution of strategy profiles (aka “outcomes”)

	Opera (O)	Football (F)
Opera (O)	$\frac{5}{10}$ 3, 1	$\frac{1}{10}$ 0, 0
Football (F)	0, 0	1, 3

Empirical Distribution of Outcomes

	s^1	s^2	s^3	s^4	s^5	s^6	s^7	s^8	s^9	s^{10}
s_1^t:	O	O	F	O	O	F	F	F	O	O
s_2^t:	O	F	F	O	O	O	F	O	O	O

time...

Want to understand the empirical distribution of strategy profiles (aka “outcomes”)

	Opera (O)	Football (F)
Opera (O)	$\frac{5}{10}$ 1 3	$\frac{1}{10}$ 0 0
Football (F)	0 0	$\frac{2}{10}$ 3 1

Empirical Distribution of Outcomes

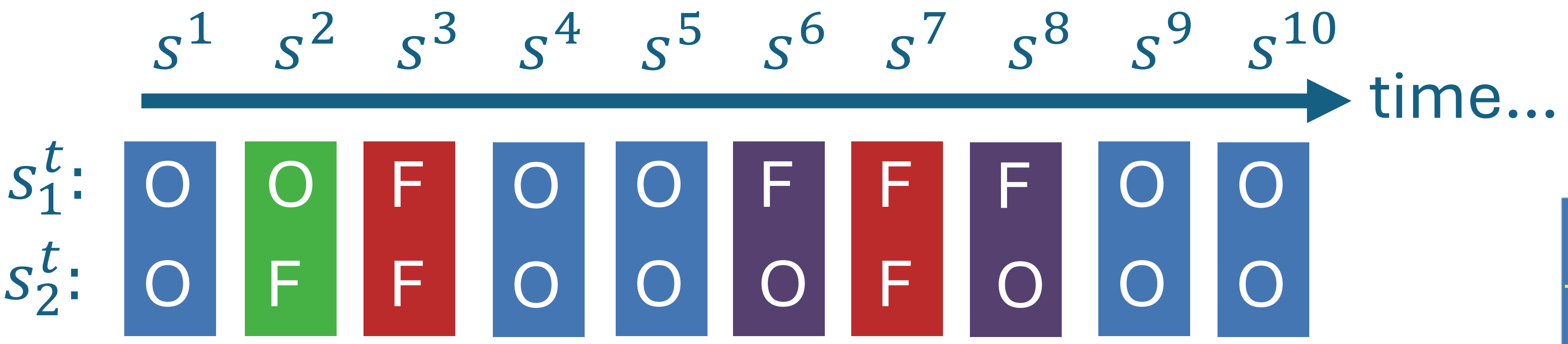
	s^1	s^2	s^3	s^4	s^5	s^6	s^7	s^8	s^9	s^{10}
s_1^t:	O	O	F	O	O	F	F	F	O	O
s_2^t:	O	F	F	O	O	O	F	O	O	O

time...

Want to understand the empirical distribution of strategy profiles (aka “outcomes”)

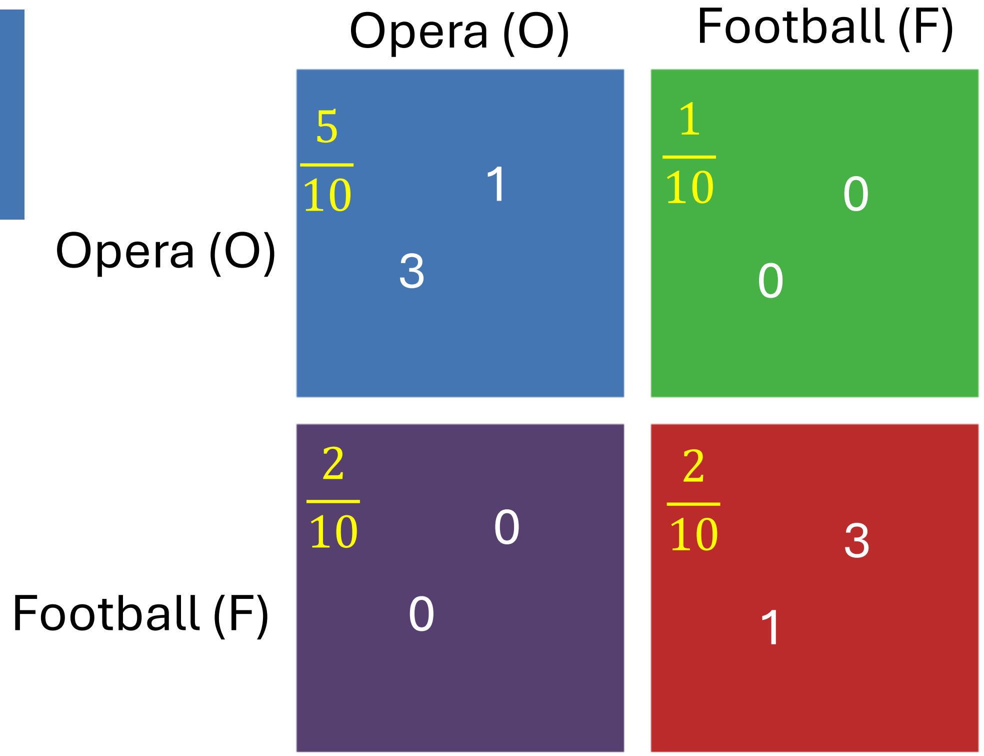
	Opera (O)	Football (F)
Opera (O)	$\frac{5}{10}$ 1 3	$\frac{1}{10}$ 0 0
Football (F)	$\frac{2}{10}$ 0 0	$\frac{2}{10}$ 3 1

Empirical Distribution of Outcomes



Want to understand the empirical distribution of strategy profiles (aka “outcomes”)

$$\pi^T(s) = \frac{|t: s^t = s|}{T}$$



Empirical Distribution of Outcomes

- For zero-sum games, looked at empirical distribution of marginals

$$\rho_i^T(s_i) = \frac{|t : s_i^t = s_i|}{T}$$

- The product of empirical marginals converges to Nash

$$\tilde{\pi}^T(s) = \rho_1^T(s_1) \cdot \rho_2^T(s_2) \rightarrow \text{Nash equilibrium}$$

- Now we look at the empirical joint distribution

$$\pi^T(s) = \frac{|t : s^t = s|}{T}$$

Correlation of Outcomes

- Players observe a shared history, their actions are correlated
- Shared history plays the role of the "correlating public coin flip"
- Maybe in some games, eventually the play de-correlates
- If mixed strategies of the players converge (typically not the case)

$$\sigma_i^T \to \sigma_i^*$$

- *Other players must choose approximate best-response strategies to have vanishing regret*
- Each player's mixed strategy → best response to opponents'
- Empirical distribution $\pi^T \to \sigma_1^* \times \cdots \times \sigma_N^*$ which is a Nash

Correlation of Outcomes

- Players observe a shared history, their actions are correlated
- Shared history plays the role of the "correlating public coin flip"
- Maybe in some games, eventually the play de-correlates

Even if play doesn't decorrelate and the mixed strategies of the players don't converge, can we argue that empirical distribution converges to some nice set?

When all players use no-regret algorithms, the empirical distribution converges to a:

Nash Equilibrium	0%
Correlated Equilibrium	0%
Some other equilibrium concept	0%
Nothing interesting	0%

What does the empirical distribution satisfy?

- No-regret property, for each player i:

$$\frac{1}{T}\sum_{t=1}^{T} u_i\left(s^t\right) \geq \max_{s_i' \in S_i} \frac{1}{T}\sum_{t=1}^{T} u_i\left(s_i', s_{-i}^t\right) - \tilde{\epsilon}(T, \delta)$$

- Re-write no-regret property in terms of the empirical distribution

$$\frac{1}{T}\sum_{s} \sum_{t:s^t=s}^{T} u_i(s) \geq \max_{s_i' \in S_i} \frac{1}{T}\sum_{s} \sum_{t:s^t=s}^{T} u_i\left(s_i', s_{-i}\right) - \tilde{\epsilon}(T, \delta)$$

What does the empirical distribution satisfy?

- No-regret property, for each player i:

$$\frac{1}{T}\sum_{t=1}^{T} u_i(s^t) \geq \max_{s_i' \in S_i} \frac{1}{T}\sum_{t=1}^{T} u_i(s_i', s_{-i}^t) - \tilde{\epsilon}(T, \delta)$$

- Re-write no-regret property in terms of the empirical distribution

$$\sum_{s} u_i(s)\frac{|t: s^t = s|}{T} \geq \max_{s_i' \in S_i} \sum_{s} u_i(s_i', s_{-i})\frac{|t: s^t = s|}{T} - \tilde{\epsilon}(T, \delta)$$

Empirical joint distribution of strategies

What does the empirical distribution satisfy?

- No-regret property, for each player i:

$$\frac{1}{T}\sum_{t=1}^{T} u_i(s^t) \geq \max_{s_i' \in S_i} \frac{1}{T}\sum_{t=1}^{T} u_i(s_i', s_{-i}^t) - \tilde{\epsilon}(T,\delta)$$

- Re-write no-regret property in terms of the empirical distribution

$$\underbrace{\boxed{\sum_{s} \pi^T(s) u_i(s)}}_{\text{Average utility}} \geq \max_{s_i' \in S_i} \underbrace{\boxed{\sum_{s} \pi^T(s)\, u_i(s_i', s_{-i})}}_{\text{Average utility had I always played } s_i'} - \tilde{\epsilon}(T,\delta)$$

Regret Example

	s^1	s^2	s^3	s^4	s^5	s^6	s^7	s^8	s^9	s^{10}
s_1^t:	O	O	F	O	O	F	F	F	O	O
s_2^t:	O	F	F	O	O	O	F	O	O	O

time...

	Opera (O)	Football (F)
Opera (O)	$\frac{5}{10}$ 1 **3**	$\frac{1}{10}$ 0 **0**
Football (F)	$\frac{2}{10}$ 0 **0**	$\frac{2}{10}$ 3 **1**

P1: I got average utility

$$\frac{5}{10}3 + \frac{1}{10}0 + \frac{2}{10}0 + \frac{2}{10}1 = \frac{17}{10}$$

Regret Example

	s^1	s^2	s^3	s^4	s^5	s^6	s^7	s^8	s^9	s^{10}
s_1^t:	O	O	F	O	O	F	F	F	O	O
s_2^t:	O	F	F	O	O	O	F	O	O	O
u_1^t:	3	0	1	3	3	0	1	0	3	3
$\tilde{u}_1^t$:	3	0	0	3	3	3	0	3	3	3

time...

Opera (O)

	Opera (O)	Football (F)
Opera (O)	$\frac{5}{10}$ — 3, 1	$\frac{1}{10}$ — 0, 0
Football (F)	$\frac{2}{10}$ — 0, 0	$\frac{2}{10}$ — 1, 3

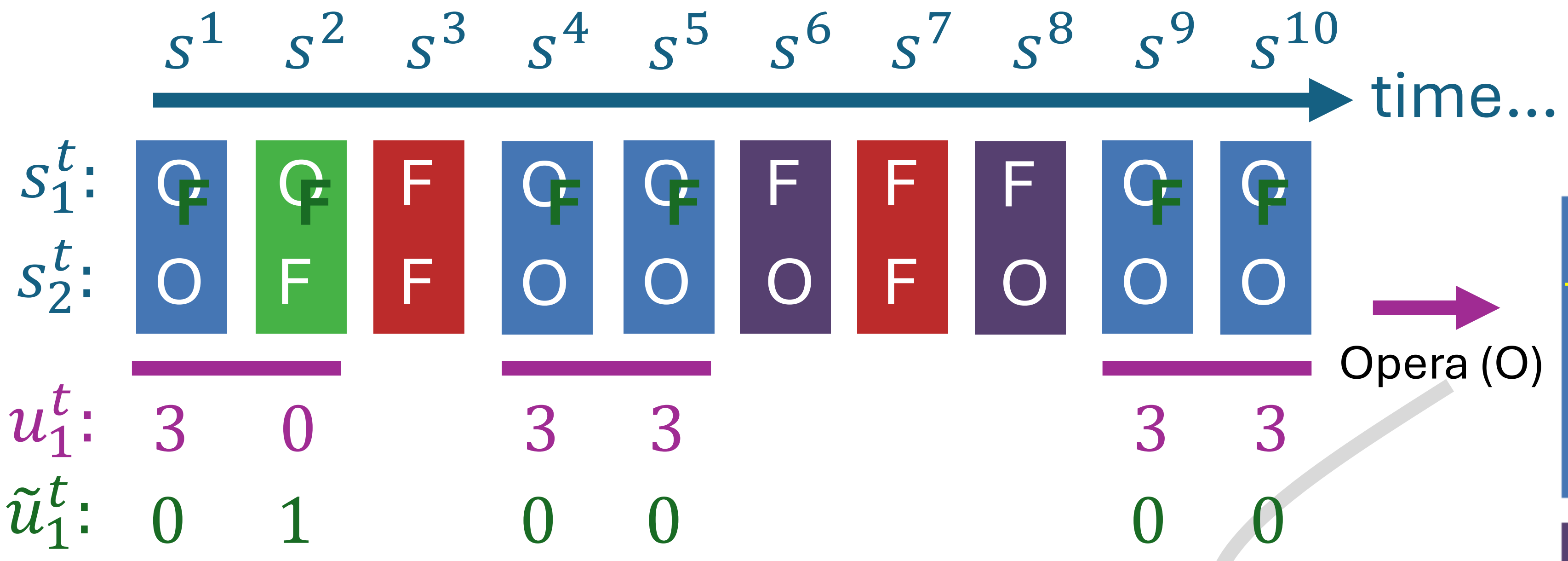
P1: I got average utility

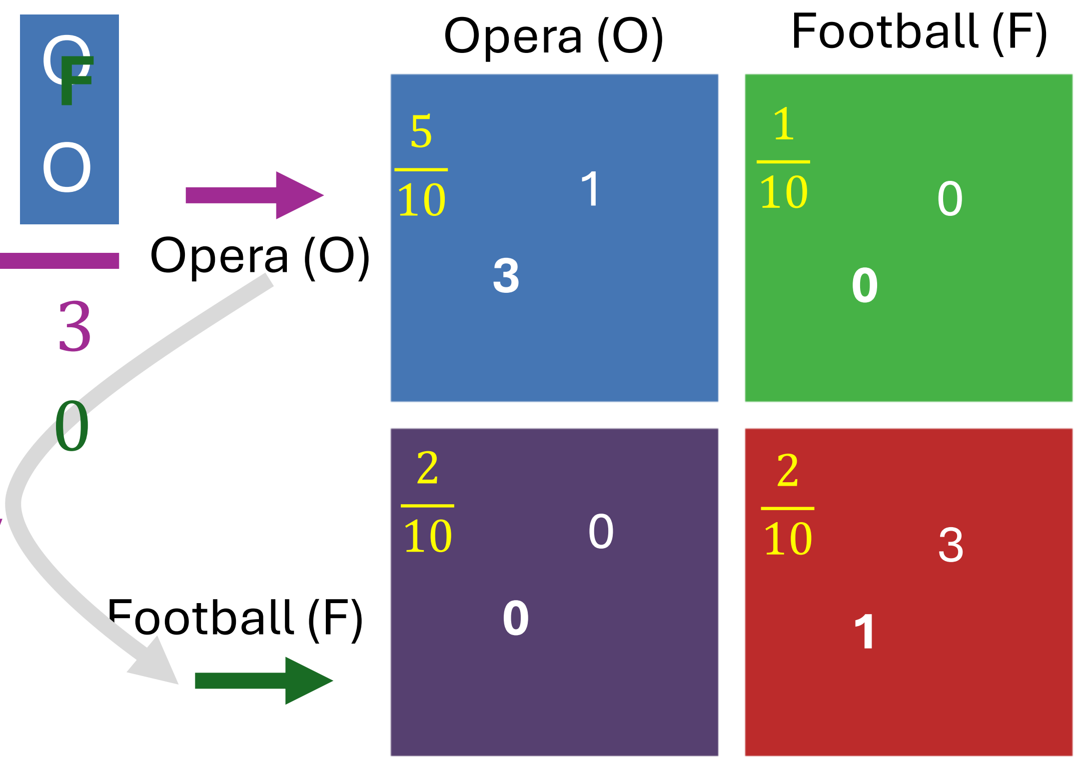
$$\frac{5}{10}3 + \frac{1}{10}0 + \frac{2}{10}0 + \frac{2}{10}1 = \frac{17}{10}$$

If I were to always play O

$$\frac{5}{10}3 + \frac{1}{10}0 + \frac{2}{10}3 + \frac{2}{10}0 = \frac{21}{10}$$

The correlated equilibrium calculation?





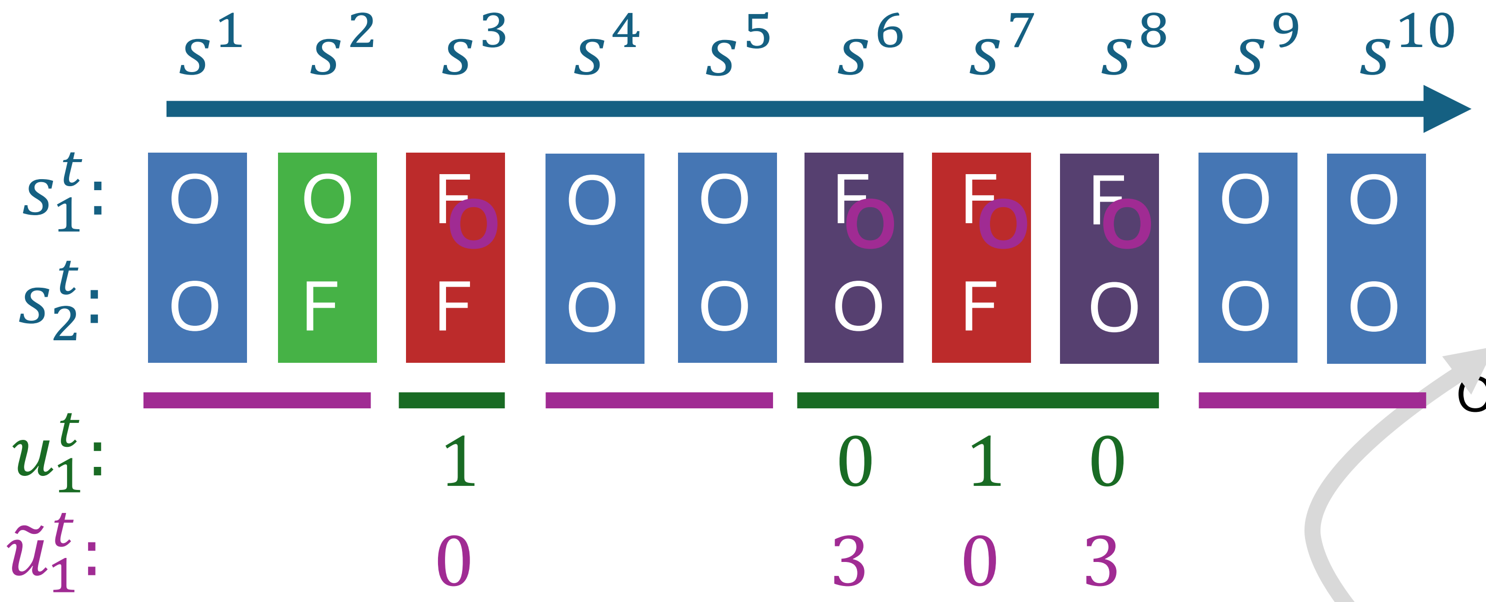
When recommended O, my opponent chooses with probability 5/6 O and with probability 1/10 F, utility from following

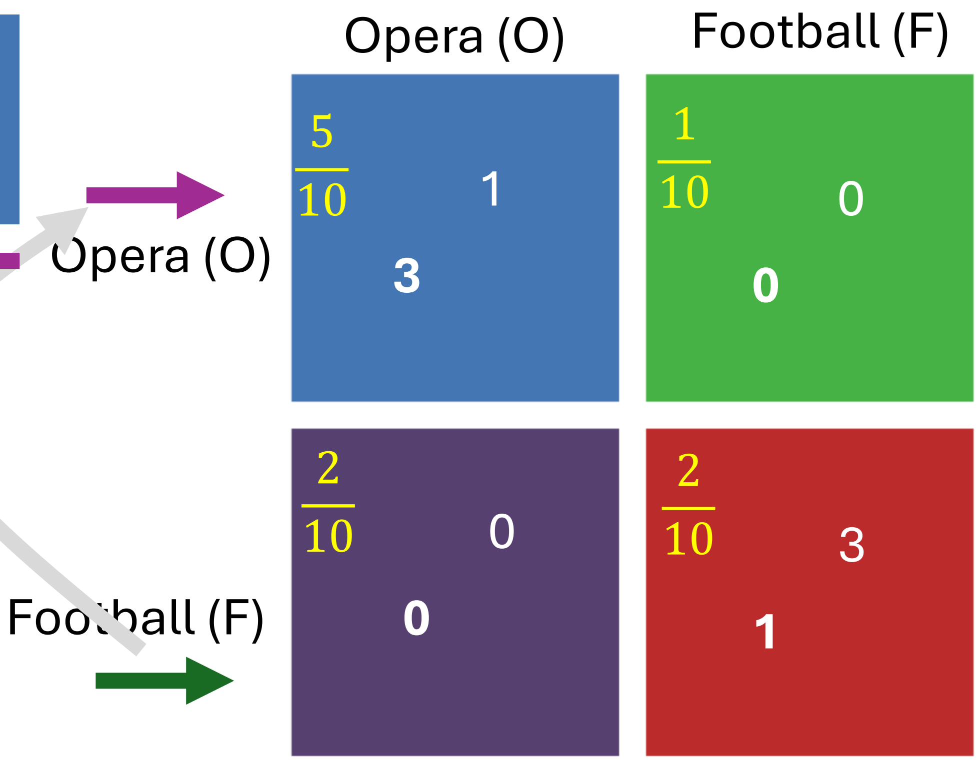
$$\frac{5}{6}3 + \frac{1}{6}0 = \frac{15}{6}$$

My utility from deviating to F

$$\frac{5}{6}0 + \frac{1}{6}1 = \frac{1}{6}$$

The correlated equilibrium calculation?





When recommended F, my opponent chooses with probability 1/2 O and with probability 1/2 F, utility from following

$$\frac{1}{2}0 + \frac{1}{2}1 = \frac{1}{2}$$

My utility from deviating to O

$$\frac{1}{2}3 + \frac{1}{2}0 = \frac{3}{2}$$

Regret vs Correlated Equilibrium

- No-regret property, implies

$$\forall s_i': \sum_{s} \pi^T(s)\left(u_i(s) - u_i\left(s_i', s_{-i}\right)\right) \geq -\tilde{\epsilon}(T, \delta) \to 0$$

- Correlated equilibrium requires conditioning on recommendation

$$\forall s_i^*, s_i': \sum_{s: s_i = s_i^*} \pi^T(s)\left(u_i(s) - u_i\left(s_i', s_{-i}\right)\right) \geq 0$$

At subset of periods when **played s_i^***

You don't regret **switching to s_i'**

Regret vs Correlated Equilibrium

- No-regret property, implies

Distributions that satisfy this are called **Coarse Correlated Equilibria**

$$\forall s_i': \sum_{s} \pi^T(s)\left(u_i(s) - u_i(s_i', s_{-i})\right) \geq -\tilde{\epsilon}(T, \delta) \to 0$$

- Correlated equilibrium requires conditioning on recommendation

$$\forall s_i^*, s_i': \sum_{s:s_i = s_i^*} \pi^T(s)\left(u_i(s) - u_i(s_i', s_{-i})\right) \geq 0$$

At subset of periods when **played s_i^***

You don't regret **switching to s_i'**

Need a New Notion of Regret

Swaps and Correlated Equilibrium

- Correlated equilibrium requires conditioning on recommendation

$$\forall s_i^*, s_i': \sum_{s:s_i=s_i^*} \pi^T(s)\left(u_i(s) - u_i\left(s_i', s_{-i}\right)\right) \geq 0$$

- Equivalently: for any **swap** function ϕ that maps original actions s_i to deviating actions s_i' (potentially different for each original s_i)

$$\sum_{s} \pi^T(s)\left(u_i(s) - u_i(\phi(s_i), s_{-i})\right) \geq 0$$

At all periods

You don't regret **swapping your original action based on the mapping ϕ**

No-Swap Regret!

- No-regret property requires

$$\frac{1}{T}\sum_{t=1}^{T} u_i(s^t) \geq \max_{s_i' \in S_i} \frac{1}{T}\sum_{t=1}^{T} u_i(s_i', s_{-i}^t) - \tilde{\epsilon}(T, \delta)$$

- No-swap regret property requires

$$\forall \phi \colon \frac{1}{T}\sum_{t=1}^{T} u_i(s^t) \geq \frac{1}{T}\sum_{t=1}^{T} u_i(\phi(s_i^t), s_{-i}^t) - \tilde{\epsilon}(T, \delta)$$

Theorem. If all players use no-swap regret algorithms, then the empirical joint distribution converges to a Correlated Equilibrium

Can we construct algorithms with vanishing no-swap regret?