# Learning and Games, day 3

## Price of Anarchy and Game Dynamics

Éva Tardos, Cornell

# Learning and Games

## Price of Anarchy and Game Dynamics

Day 3:

- Learning in changing environments

Next: Can learning do better than Nash?

# Summary from last two days

simple games and variants:

- matching pennies,
- coordination,
- prisoner's dilemma,
- Rock-paper-scissor

Learning algorithms

- Fictitious play, and smoothed versions

No-regret as outcome of learning or as a behavioral model

Price of Anarchy and learning outcomes in

- Congestion games, such as traffic routing
- Auction games

Learning in multi-item auctions is hard,

Alternate learning we can do instead

# Quality of Learning Outcome

Price of Anarchy [Koutsoupias-Papadimitriou'99]

$$PoA = \max_{a\ Nash} \frac{cost(a)}{Opt}$$

Assuming **no-regret learners** in fixed game: [Blum, Hajiaghayi, Ligett, Roth'08, Roughgarden'09]

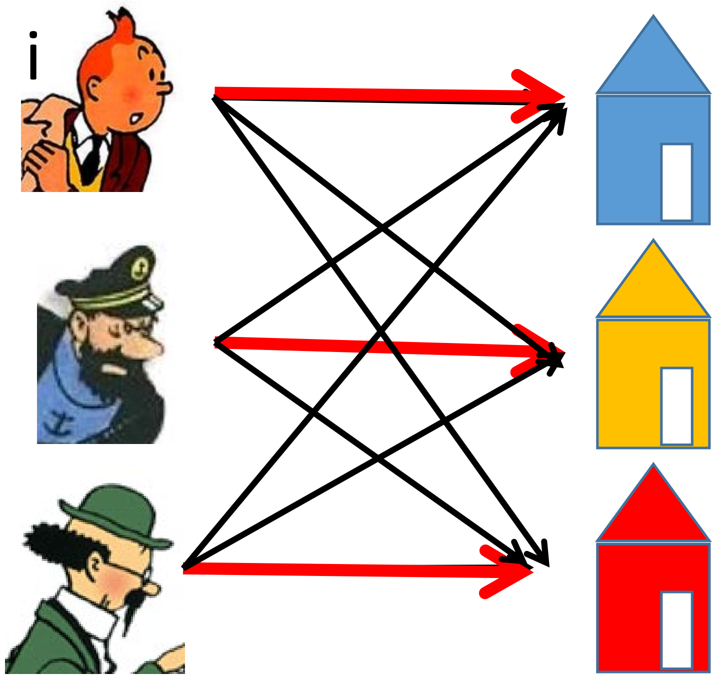
$$PoA = \lim_{T \to \infty} \frac{\sum_{t=1}^{T} cost(a^t)}{T\ Opt}$$

[Lykouris, Syrgkanis, T. 2016] dynamic population

$$PoA = \lim_{T \to \infty} \frac{\sum_{t=1}^{T} cost(a^t, v^t)}{\sum_{t=1}^{T} Opt(v^t)}$$

where $v^t$ is the vector of player types at time t

# Today's context: unit demand bidders in second price auction




Value if $i$ gets subset $S$ is $v_i(S)$

for example: $v_i(S) = \max_{j \in S} v_{ij}$

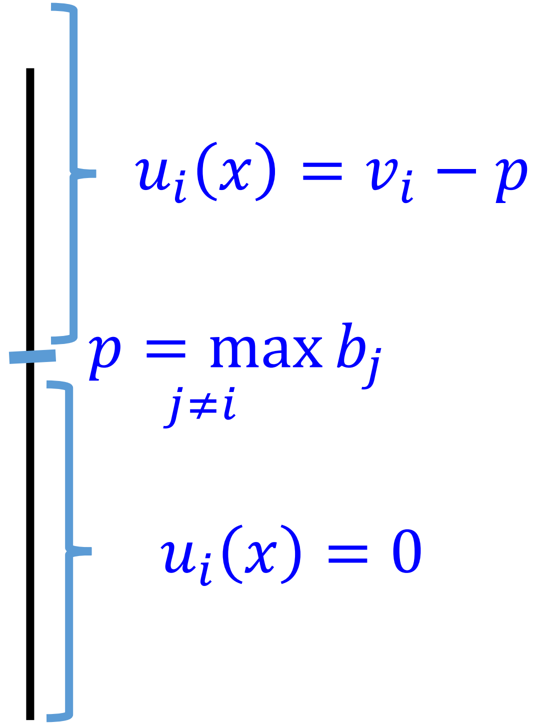
Optimum is max value matching!

$$\max_{M^*} \sum_{ij \in M^*} v_{ij}$$

Second price:

- Bid vector $(b_{i1}, b_{i2}, \ldots, b_{in})$,
- Each item sold on 2nd price (max wins pays next price)

# Second Price Auction (Vickrey)




- Bidding the true value $b_i = v_i$ is dominant strategy
  - $u_i(v_i, b_{-i}) \geq u_i(b)$ for any bid vector $b$
- Yet: there are many other equilibria.
  - Example: values $100, 5, 4, 3, 2, 1$
  - Bids : $99+, 99, 4, 3, 2, 1$ are full information Nash with bidder 1 winning
  - Bids 0, 101, 4, 3, 2, 1 are full information Nash with bidder 2 winning
    - Is either likely?
    - Bidding $b_i > v_i$ is dominated strategy!!! $b_i = v_i$ is better

# $2^{nd}$ price multi-item, unit demand

- Learning to get no-regret is NP-hard (low regret)

Can learn if

- Bid always only on one item: $(0,..,0,b_{ij},0,..,0)$

    Why? Bidding $v_{ij}$ on selected item $j$ is dominant strategy!

    # strategies is n=#items, and we get

$$\sum_\tau u_i(s^\tau) \geq (1-\epsilon)\max_x \sum_\tau u_i(x, s_i^\tau) - O(\frac{\log n}{\epsilon})$$

- Bid is either 0 or $v_{ij}$ on all items (last time)

$$\sum_\tau u_i(s^\tau) \geq (1-\epsilon)\max_x \sum_\tau u_i(x, s_i^\tau) - O(\frac{n}{\epsilon})$$

# $2^{nd}$ price and Price of Anarchy

- Is no-regret enough?
  - **No!** recall example with bid 101
  - This is not a problem with $1^{st}$ price. Why?
- No overbidding assumption: $\sum_{j \in S} b_{ij} \leq v_i(S)$ for all $S$
- Dominant strategy if bidding for one item only
- Not true always!

# Price of Anarchy with second price with no overbidding

Recall Roughgarden smoothness version:

- $Rev(s) + \sum_i u_i(s_i^*, s_{-i}) \geq \lambda \sum_i v_i(s^*) - \mu \sum_i v_i(s)$

implies PoA of $\frac{\mu+1}{\lambda}$

Claim: unit demand buyers with no-overbidding, 2nd price item auction is (1,1)-smooth

- Unit demand: optimum $s^*$ bid only on item j assign in opt, and bid $v_{ij}$ on this item.

$$u_i(s_i^*, s_{-i}) \geq v_{ij} - \max_k b_{kj}$$

Summing over players this gives us

$$\sum_i u_i(s_i^*, s_{-i}) \geq Opt - \sum_j \max_i b_{ij} \geq OPT - \sum_i v_i(s)$$

Value of optimum matching

No over-bidding

(1,1) smooth implying price of anarchy of 2

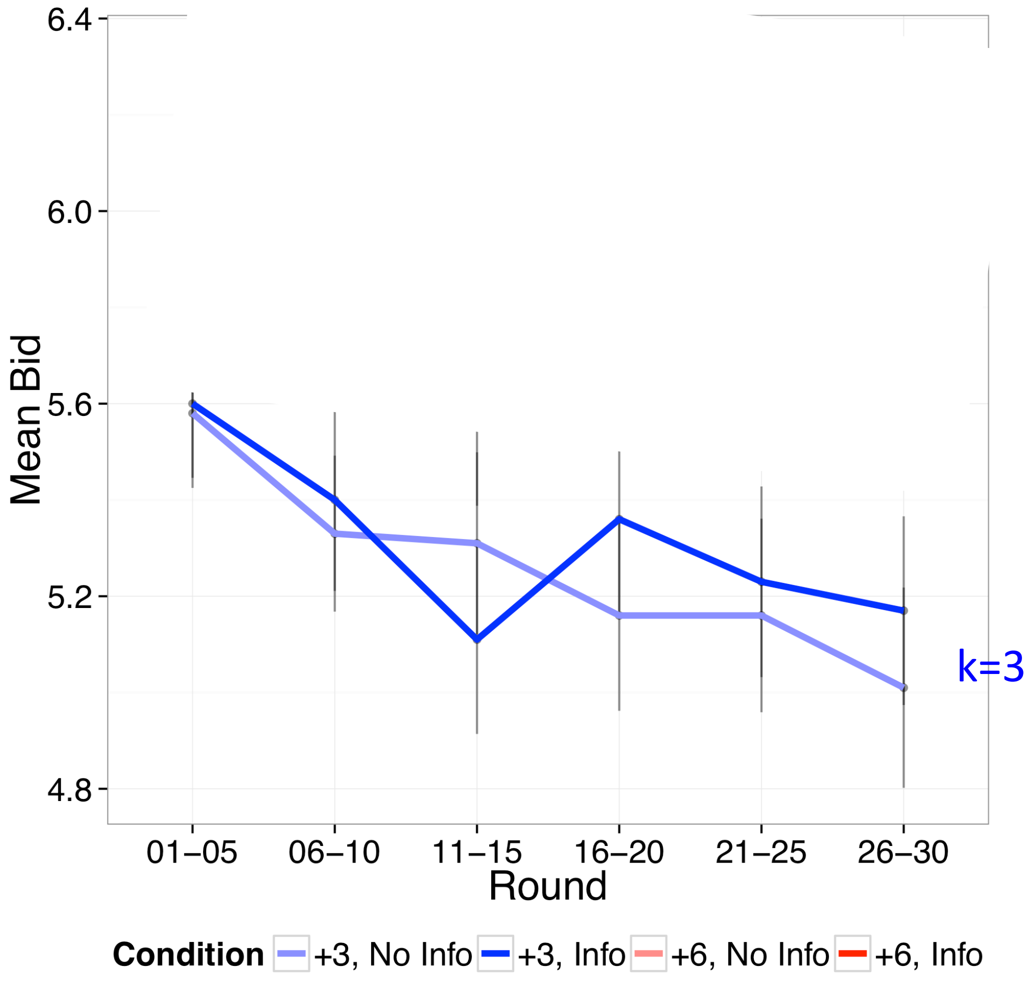
# Do people actually learn?

Buyer-seller game [Fudenberg-Peysakhovich'14]:

- Seller has a used car of value $v \in [0,10]$ integer, unif. random, she knows the value
- Buyer has value $v + k$ for the car. He knows $k$, but doesn't know $v$.
- offers a bid $b$, and gets the item for price $b$ if $v \leq b$, his value is then $v + k - b$ (quasi-linear value)

Experiment: $k = 3$, after bid, inform buyer of value $v$ (in any case)

# Equilibrium outcome and optimum bid




- Equilibrium with v ∈ [0,1] real.
- Bid $b$ maximized expected value

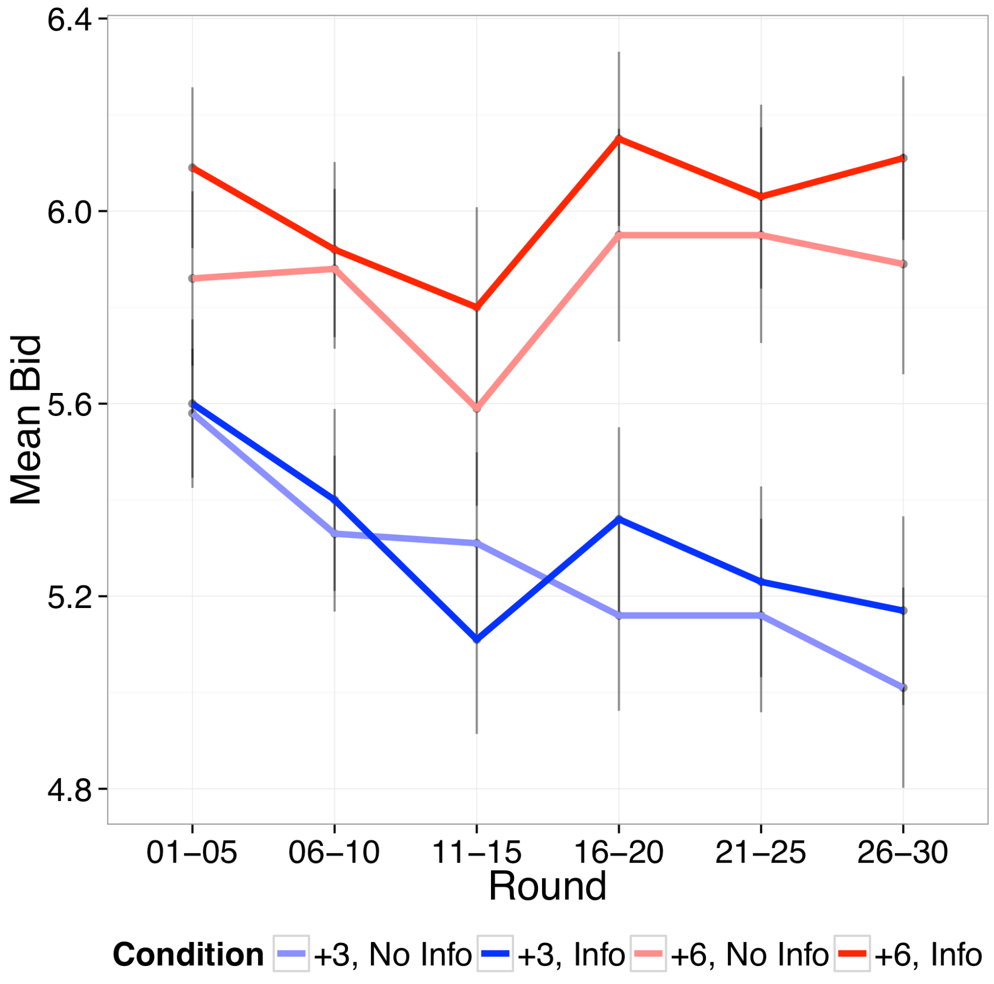
$$\Pr(v \leq b)\,[E(v|b \geq v) + k - b]$$
$$= b\left(\frac{b}{2} + k - b\right)$$
$$= bk - \frac{1}{2}b^2$$

Minimum when derivative =0

Derivative $= k - b$

Optimum bid: $b = k$

# Equilibrium outcome and recency bias




k=6

k=3

- Learning 0: best respond to the most recent information
- Best response to hearing value $v$ is
- Bid $b = v$
- Behavior closer to best response to last value than proper learning!

# Repeated game that is (slowly) changing [Lykouris, Syrgkanis, T.'16]

Dynamic population model:

At each step t each player i

is replaced with an arbitrary new player with probability p

In a population of n players (on m node graph), each step, Np players replaced in expectation

- Population changes all the time: need to adjust! ($p \approx \frac{1}{\log m}$)
- players stay long enough to be able to learn ($\frac{1}{\mathrm{p}} \approx \log m$ steps)

# Learning in Dynamic Game: [Lykouris, Syrgkanis, T. '16]

Dynamic population model:

At each step t each player i

is replaced with an arbitrary new player with probability p

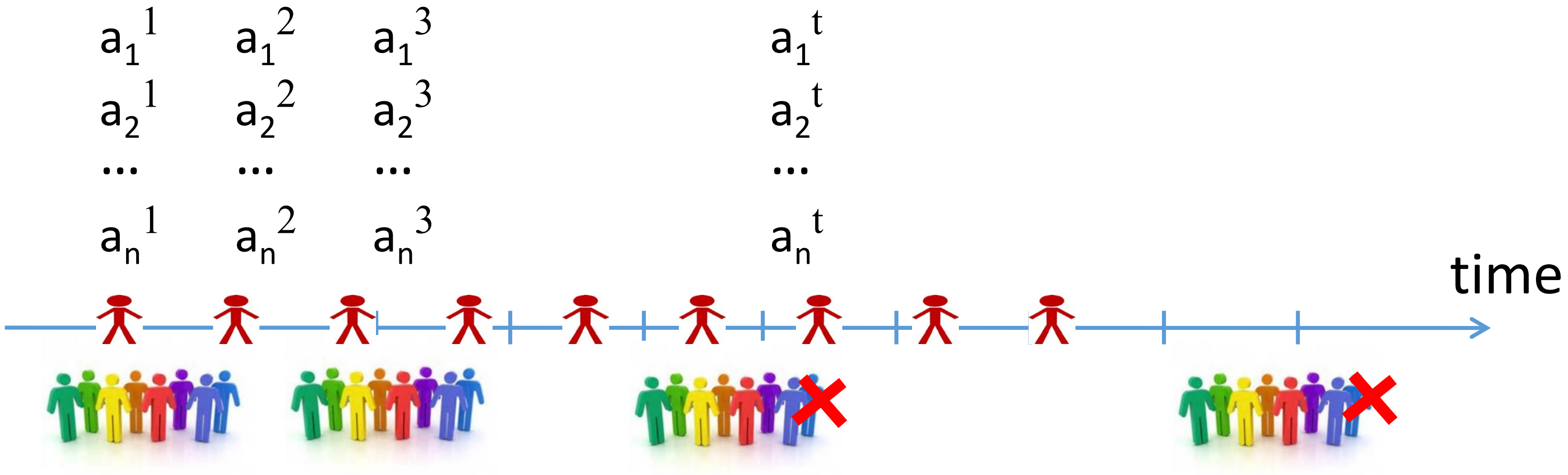
In a population of n players on m items, each step, np players replaced in expectation

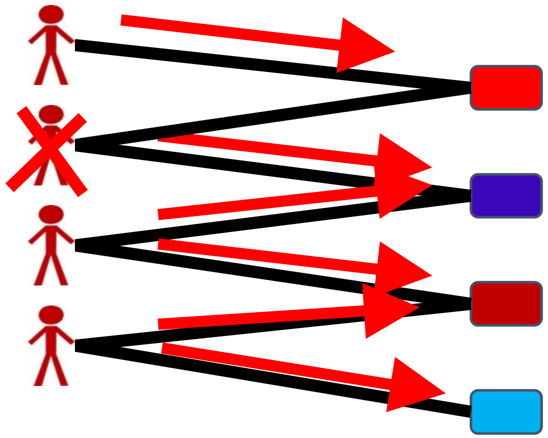
What should they learn from data?

*No regret good enough?*

$$\sum_t u_i(s^t) \geq (1+\epsilon) \sum_t u_i(s_i^*, s_{-i}^t) + R$$

# Need for adaptive learning

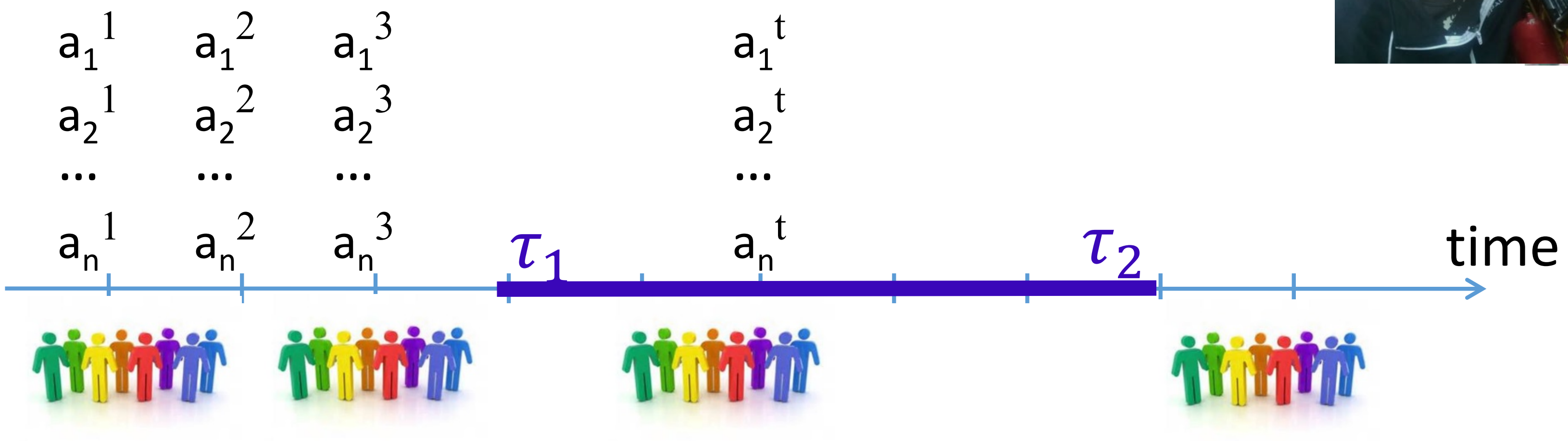


## Example unit demand

- Strategy = item to bid on
- Best “fixed” strategy in hindsight too weak in changing environment
- Learners need to adapt to the changing environment

# Adaptive Learning



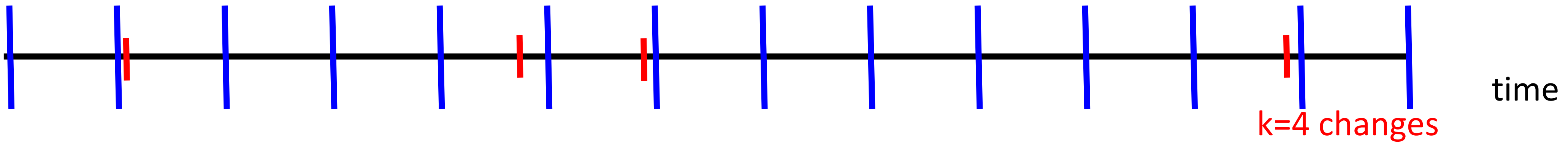

Theorem Approximate Regret [Foster,Li,Lykouris,Sridharan,T. NIPS'16]

for all player i, strategy $x^\tau$ sequence that changes k times

$$\sum_\tau u_i(s^\tau, v^\tau) \geq \sum_\tau (1+\epsilon)\, u_i(x^\tau, s_{-i}^\tau; v^\tau) + O\left(\frac{k}{\epsilon}\log m\right)$$

Using any classical learning mixed with a bit of recency bias

# Adaptive Learning (sketch of weaker bound)




- Restart at roughly event $\frac{\epsilon T}{k}$ steps, so have $\frac{k}{\epsilon}$ intervals.
- Only $k$ intervals can have change. No guarantee on these intervals, but that is a total of $\mathrm{k}\,\epsilon T/k = \epsilon T$ steps
- Remaining intervals we do get learning! Each having a regret error of at most $(\log m)/\epsilon$ for a total of $\mathrm{k}\,(\log \mathrm{m})/\epsilon^2$.
- Total guarantee this gives:

$$\sum_\tau u_i(s^\tau, v^\tau) \geq \sum_\tau (1+\epsilon) \sum_\tau u_i(x^\tau, s_{-i}^\tau; v^\tau) + O(\frac{\mathrm{k}}{\epsilon} \log m)$$

# Adapting result to dynamic populations

Inequality we "wish to have"

$$\sum_t cost_i(s^t; v^t) \leq \sum_t cost_i(s_i^{*t}, s_{-i}^t; v^t)$$

where $s_i^{*t}$ is the optimum strategy for the players at time t.

with stable population = no regret for $s_i^*$ : optimal solution

Too much to hope for in dynamic case?

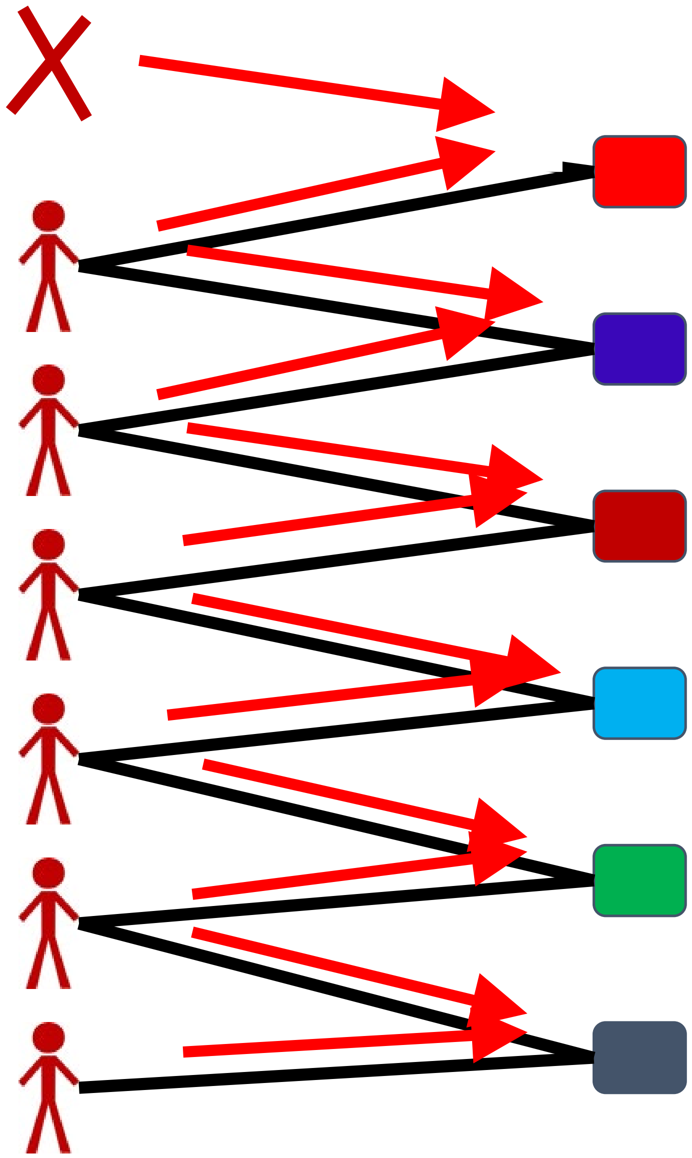
- sequence $s^{*t}$ of optimal solutions changes too much.
- No hope of learners not to learn this well!

# Change in Optimum Solution

True optimum is too sensitive

- Example using matching
- The optimum solution
- One person leaving
- Can change the solution for everyone

- Np changes each step → No time to learn!! (we have p>>1/n)



# Theorem (high level)

If a game satisfies a “smoothness property”

The welfare optimization problem admits an approximation algorithm whose outcome $\tilde{s}^*$ is stable to changes in one player’s type

Then any adaptive learning outcome is approximately efficient

$$\text{PoA} = \lim_{T \to \infty} \frac{\sum_{t=1}^{T} Opt(v^t)}{\sum_{t=1}^{T} SW\,(a^t, v^t)} \text{ close to PoA}$$

Proof idea: use this approximate solution as $\tilde{\boldsymbol{s}}^*$ in Price of Anarchy proof

With $\tilde{\boldsymbol{s}}^*$ not changing much, learners have time to learn not to regret following $\tilde{\boldsymbol{s}}^*$

# Result (Lykouris, Syrgkanis, T'16) :

In many smooth games welfare close to Price of Anarchy even when the rate of change is high, $\boldsymbol{p} \approx \frac{\mathbf{1}}{\log \boldsymbol{m}}$ with n players, assuming **adaptive** no-regret learners

- Worst case change of player type $\Rightarrow$ need for learning players
- Bound $\boldsymbol{\alpha} \cdot \boldsymbol{\beta} \cdot \boldsymbol{\gamma}$ depends on
  - $\boldsymbol{\alpha}$ price of anarchy bound — as game gets large, goes to 1 in auctions, goes to 4/3 in linear congestion games
  - $\boldsymbol{\gamma}$ loss due to regret error — goes to 1 as $p \to 0$
  - $\boldsymbol{\beta}$ loss in opt for stable solutions — goes to 1 as $p \to 0$ & game is large

# Proof (of a bit weaker version)

Assume we have matching sequence $M^\tau$ such that

1. # times player or assigned item changes $\leq k$

for each of the n sequences of players

2. total value of $\mathrm{v}(\mathrm{M}^\tau, \mathrm{v}^\tau) = \sum_\tau \sum_{ij \in M^\tau} v_{ij}^\tau \geq \beta OPT^\tau = \beta \sum_\tau \max_M \sum_{ij} v_{ij}^\tau$

Then total social welfare $\geq \frac{\beta}{2}(1-\epsilon) \sum_\tau Opt^\tau - \mathrm{nk}\frac{\log \mathrm{m}}{\epsilon^2}$

Proof: let $\tilde{s}_i^*$ be that $i$ bids on her assigned item in $M^\tau$

$\sum_\tau u_i(s^\tau, v^\tau) \geq (1-\epsilon) \sum_\tau u_i(\tilde{s}_i^*, s_{-i}^\tau, v^\tau) - k\frac{\log m}{\epsilon^2}$ learning

$u_i(\tilde{s}_i^*, s_{-i}^\tau, v^\tau) \geq v_{ij}^\tau - \max_k b_{kj}^\tau$ where $(i,j) \in M^\tau$ smoothness

# Proof outline(cont)

So far we have

$\sum_\tau u_i(s^\tau, v^\tau) \geq (1-\epsilon) \sum_\tau u_i(\tilde{s}_i^*, s_{-i}^\tau, v^\tau) - k \frac{\log m}{\epsilon^2}$ learning

$u_i(\tilde{s}_i^*, s_{-i}^\tau, v^\tau) \geq v_{ij}^\tau - \max_k b_{kj}^\tau$ where $(i,j) \in M^\tau$ smoothness

Summing over all players and using the above we get

$\sum_\tau \sum_i u_i(s^\tau, v^\tau) \geq (1-\epsilon) \sum_\tau v(M^\tau, v^\tau) - \sum_\tau \sum_j \max_k b_{kj}^\tau$ $\leq \sum_i v_i(s)$

No over-bidding

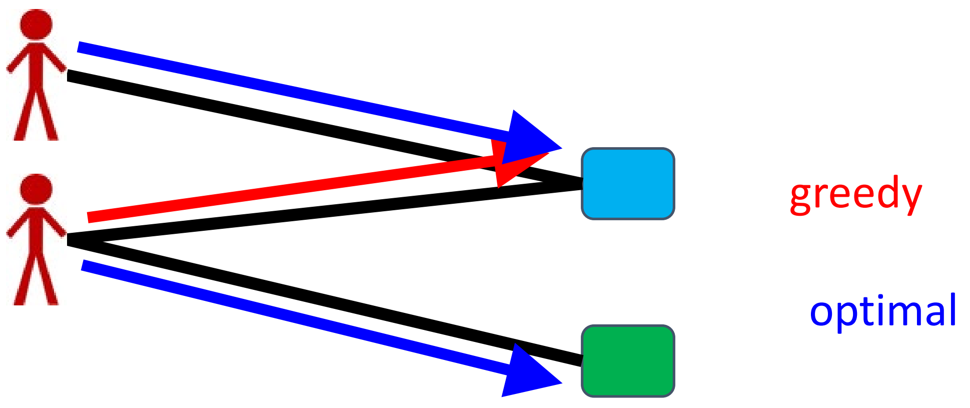
So we get

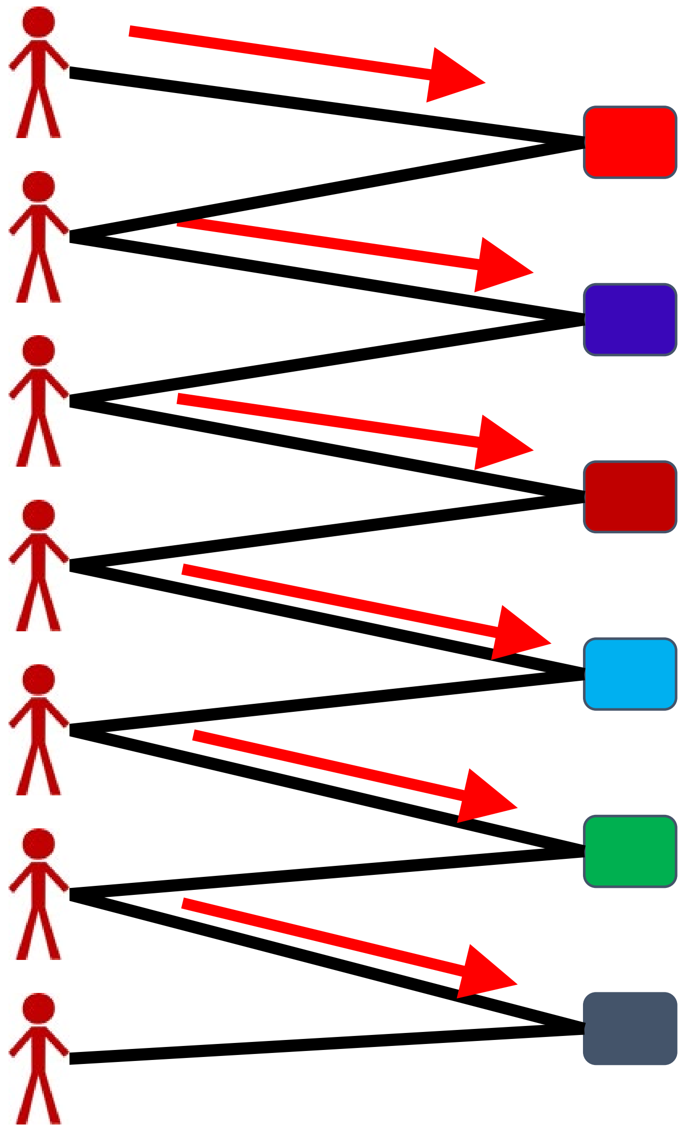
$2 \sum_\tau \sum_i v_i(s) \geq (1-\epsilon)\beta \sum_\tau Opt^\tau - nk \frac{\log m}{\epsilon^2}$

# Stable ≈ Optimum in Matching

True optimum is too sensitive

- Round all values to powers of 2. Values in range [1,v] then only $\log v$ values (loss of factor of 2)
- Use greedy allocation: assign large values first (loss of factor of 2)



# Stable ≈ Optimum in Matching

Not too many changes of assignments:

Potential function argument:

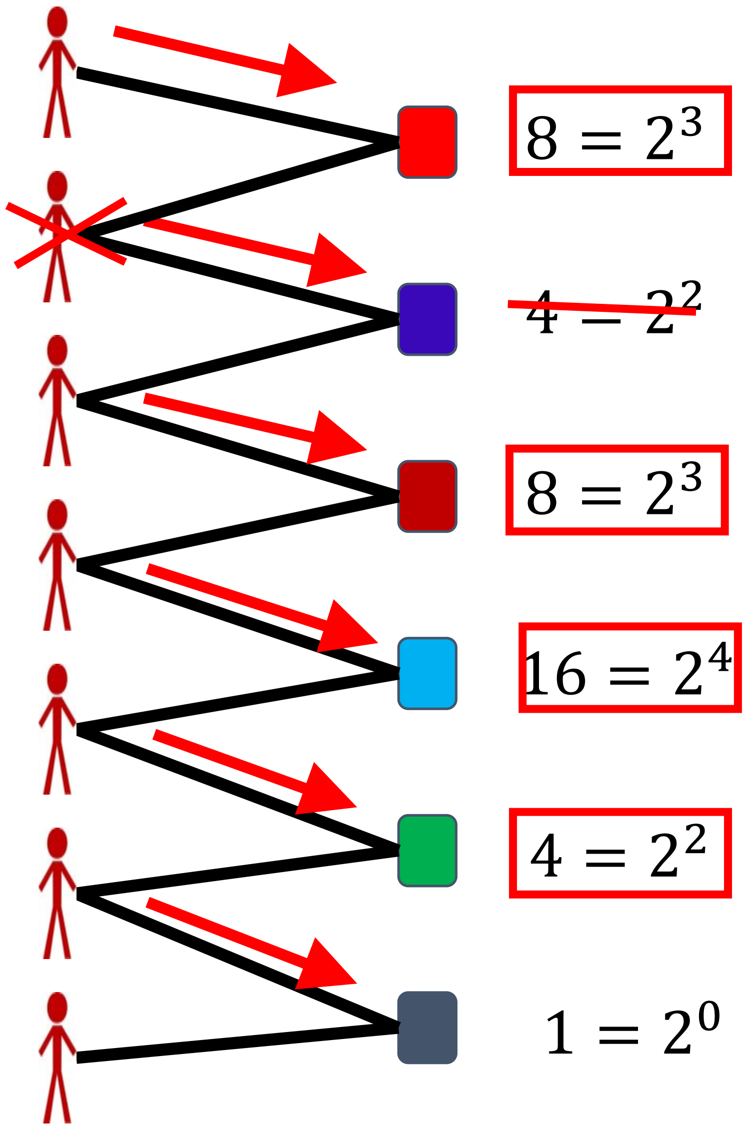
$\Phi$ =sum of the powers of assignment values

In example $\Phi = 3 + 2 + 3 + 4 + 2 + 0 = 14$

Range of $0 \leq \Phi \leq m \log v$

- decrease only due to departures, $mpT \log v$ in expectation
- Increase due to improved allocation or new arrival

So total change per player $\mathrm{k} = \frac{mpT \log v}{n}$ (on average)



# Use Differential Privacy $\rightarrow$ Stable Solutions

Joint privacy [Kearns et al. ’14, Dwork et al. ‘06]

A randomized algorithm is jointly differentially private if

- when input from player **i** changes
- the probability of change in solution of players other than **i** is smaller than $\epsilon$

- Turn a sequence of randomized solutions to a randomized sequence with small number of changes using Coupling Lemma
- and handling “failure probabilities” of private algorithms

# Open problem: Auctions with budgets?

Values $v_{i1}, v_{i2}, \ldots, v_{im}$ and a budget $B$.

Version 1 (no learning). There are m items, and need to submit a single bid: $\alpha_i$ meaning

- Bid vector $\mathrm{b}_{\mathrm{i1}} = \alpha_i v_{i1}, b_{i2} = \alpha_i v_{i2}, \ldots, b_{im} = \alpha_i v_{im}$
- Give each bidder a subset of items where he is the max bidder (with fractional allocation OK) on first/second price.
- Equilibrium if items with positive bid are fully allocated, no player exceeds their budget, and all players either have $\alpha_i = 1$ or fully spend their budget

Theorem: there is Nash equilibrium of this game with all budgets exhausted.
First price: defines a market equilibrium!

Open: can the players learn to bid such an $\alpha_i$? When small items arrive online

# Exercises 1

1. can learning algorithms, such as MW or FPL put > 0 probability on a strictly dominated strategy x ?

   Strictly dominated = for some y we have $u(y, s_{-i}) > u(x, s_{-i})$ for all strategies $s_{-i}$ of other players.

2. In a coarse correlated equilibrium can a player play a strictly dominated strategy x with probability >0?

# Main question:
# Quality of Selfish outcome

Selfish outcome = result of Learning behavior

Our Question: quality of learning outcomes?

which correlated equilibrium do users coordinate on?

Answer: depends on which learning…

Theorem: $\forall$ correlated equilibrium is the limit point of no-regret play