Learning Collusion in Episodic, Inventory-Constrained Markets

Paul Friedrich, Barna Pásztor, Giorgia Ramponi

AAMAS, 22 May 2025

University of Zurich[UZH]

ETH AI CENTER

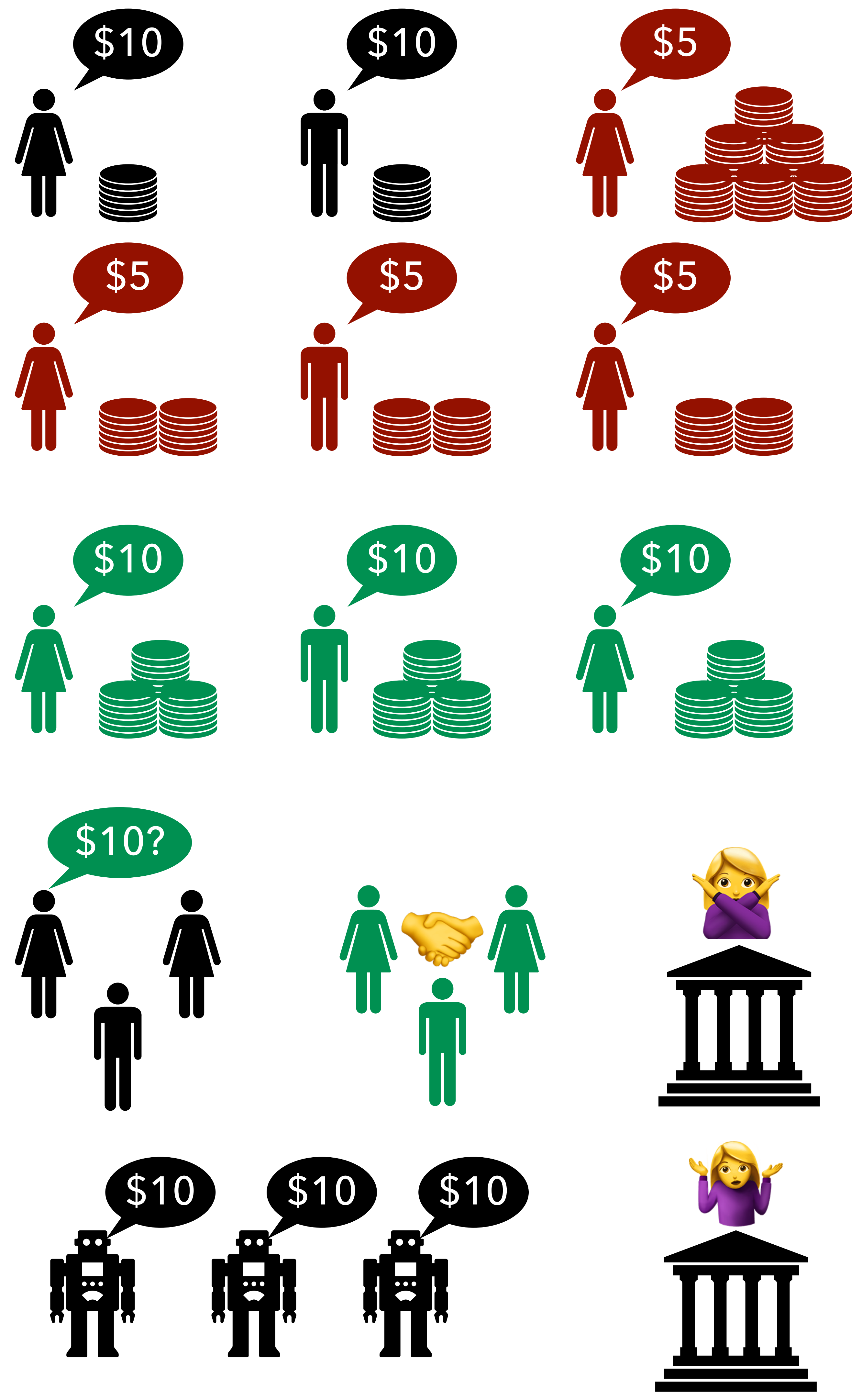
Learning collusion in markets

Selfish agents drive prices down to a low, **competitive (Nash) equilibrium price**

Collusion is when everyone agrees to stay at a high, **supra-competitive price**

Competition law can only target **explicit agreements…**

…but what if AIs can learn to collude without agreements?



The problem: tacit collusion

“Tacit" collusion (without explicit agreements) is **probably* legal!**

Agents **learn individual policies** that collectively

- **create** high prices
  - via random discovery
  - via communication through price signals
- **maintain** high prices
  - by jointly punishing deviations

➜ How does this happen in realistic markets like airline ticketing, hotels, perishables?

➜ How do we mitigate it?

Markets as a Markov game

$\cdots$

$\downarrow$

State
previous prices
current inventories

$$s_t = (\boldsymbol{p}_{t-1}, \boldsymbol{x}_t)$$

Actions
pick a price

$$\underbrace{p_t^{(1)}\ p_t^{(2)}\ p_t^{(3)}}_{\boldsymbol{p}_t}$$

$\downarrow$

Transition
market demand
sold quantities
update inventories

$$d_t^{(i)} = MNL^{(i)}(\boldsymbol{p}_t)$$

$\downarrow$

Rewards
profit: sales - cost

$$R_t^{(i)}$$

Next state
current prices
next inventories

$$s_{t+1} = (\boldsymbol{p}_t, \boldsymbol{x}_{t+1})$$

$\downarrow$

$\cdots$

Markets as a Markov game

State
previous prices
current inventories

$s_t = (\boldsymbol{p}_{t-1}, \boldsymbol{x}_t)$

Actions
pick a price

$p_t^{(1)}\ p_t^{(2)}\ p_t^{(3)}$

$\boldsymbol{p}_t$

Transition
market demand
sold quantities
update inventories

$d_t^{(i)} = MNL^{(i)}(\boldsymbol{p}_t)$

Rewards
profit: sales - cost

$R_t^{(i)}$

Next state
current prices
next inventories

$s_{t+1} = (\boldsymbol{p}_t, \boldsymbol{x}_{t+1})$

Infinite horizon
(prior work)

1-shot $(\pi_1, \pi_2)^{(1)}$

1-shot $(\pi_1, \pi_2)^{(2)}$

...

1-shot $(\pi_1, \pi_2)^*$

1-shot $(\pi_1, \pi_2)^*$

Once policies converge,
consider only 1-shot game

Can analyze policies π^*, prices p^* in the limit

Characterize equilibria by prices:

- low (Nash): p^{Nash}
- high (monopolistic): p^{Coll}
  ➜ calc. with implicit formulae

Markets as a Markov game

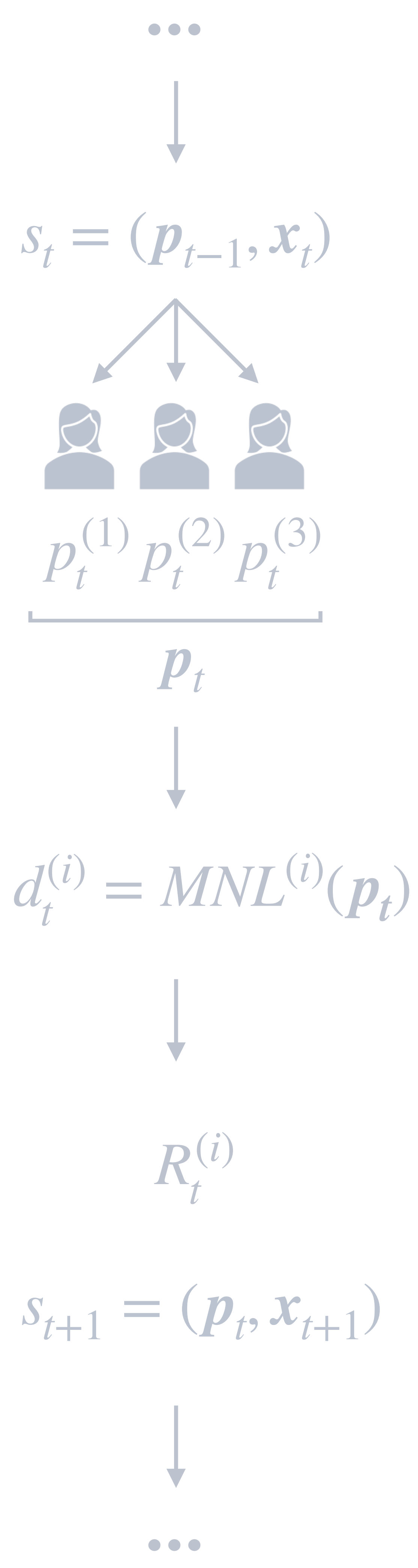
State
previous prices
current inventories

Actions
pick a price

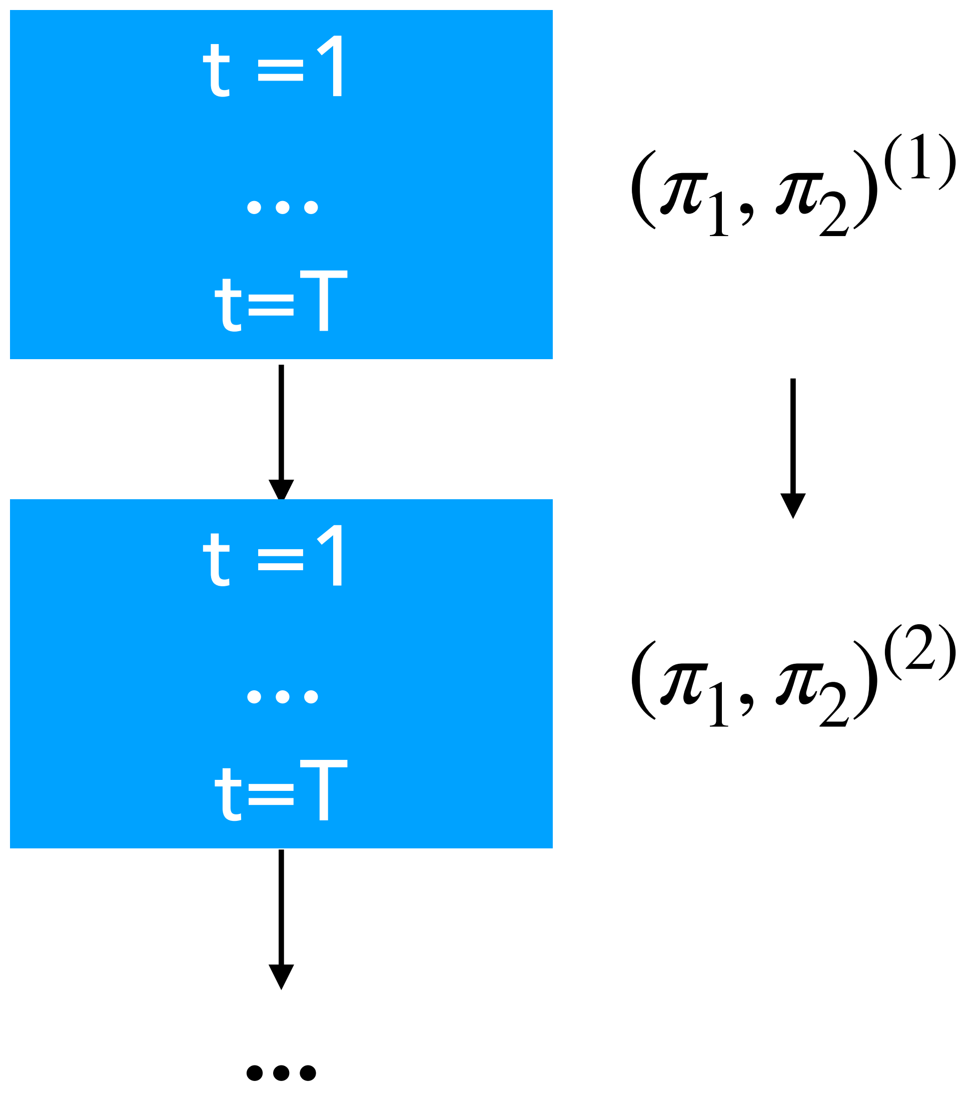
Transition
market demand
sold quantities
update inventories

Rewards
profit: sales - cost

Next state
current prices
next inventories



Finite horizon
(ours)



Episodic nature:

- only Nash-Eq is competition
- no reward-punishment
  ➜ collusion is unstable

Inventory constraint:

- demand function $d_t(\cdot)$ non-convex and non-smooth
  ➜ no formula
  for equilibrium prices
- complex state-action and strategy spaces

Defining collusion measure using equilibria

Obtain equilibrium price levels via simultaneous-move game formulation & mixed-int. non-lin. programming (MINLPs)

Supra-competitive profit per agent:

$$\Delta_i = \frac{1}{T}\sum_{t=1}^{T}\frac{R_t^{obs} - R_t^{Nash}}{R_t^{Coll} - R^{Nash}}$$

Define collusion index:

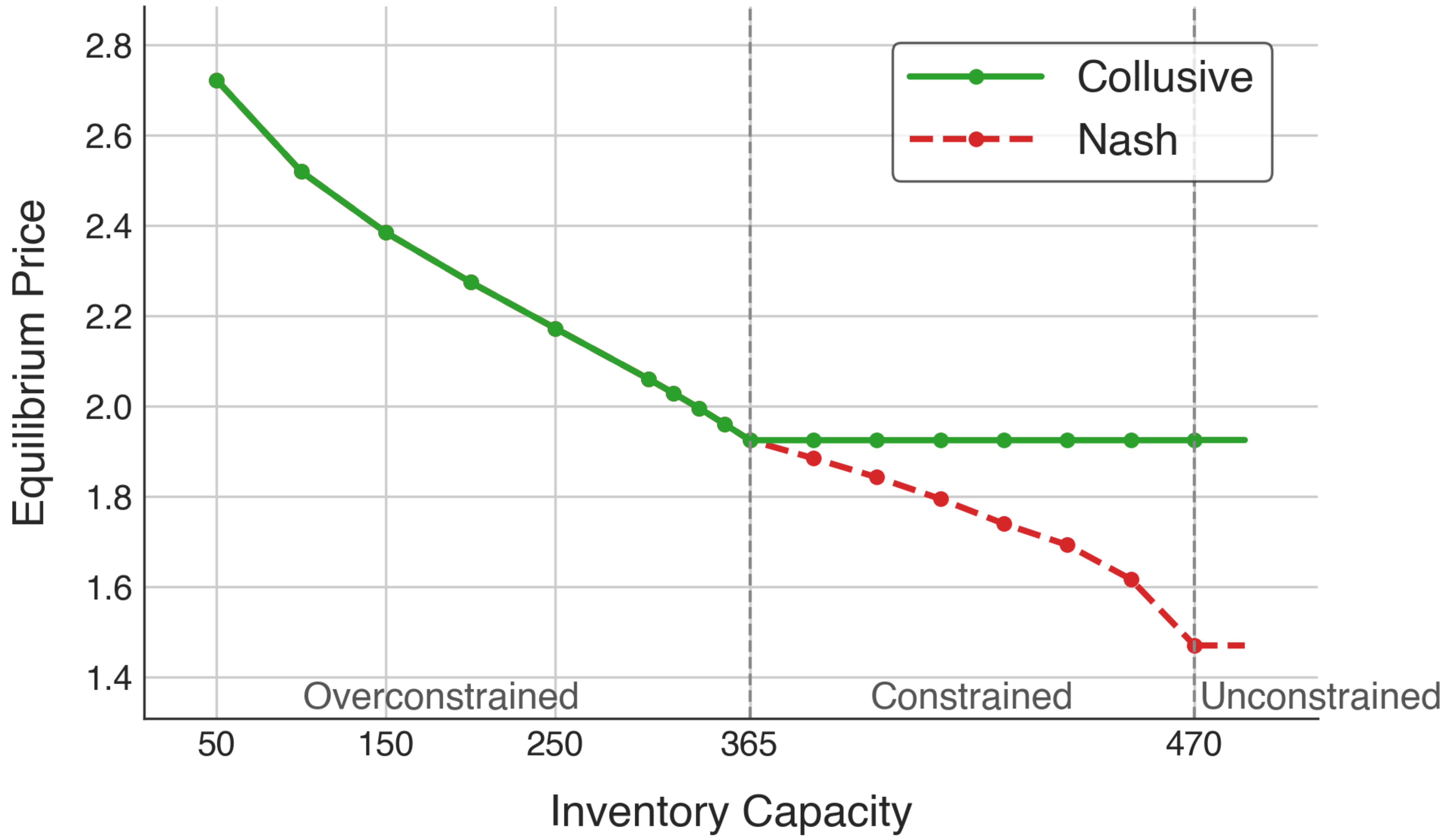
generalized mean of Δ_i over agents

=0: competitive

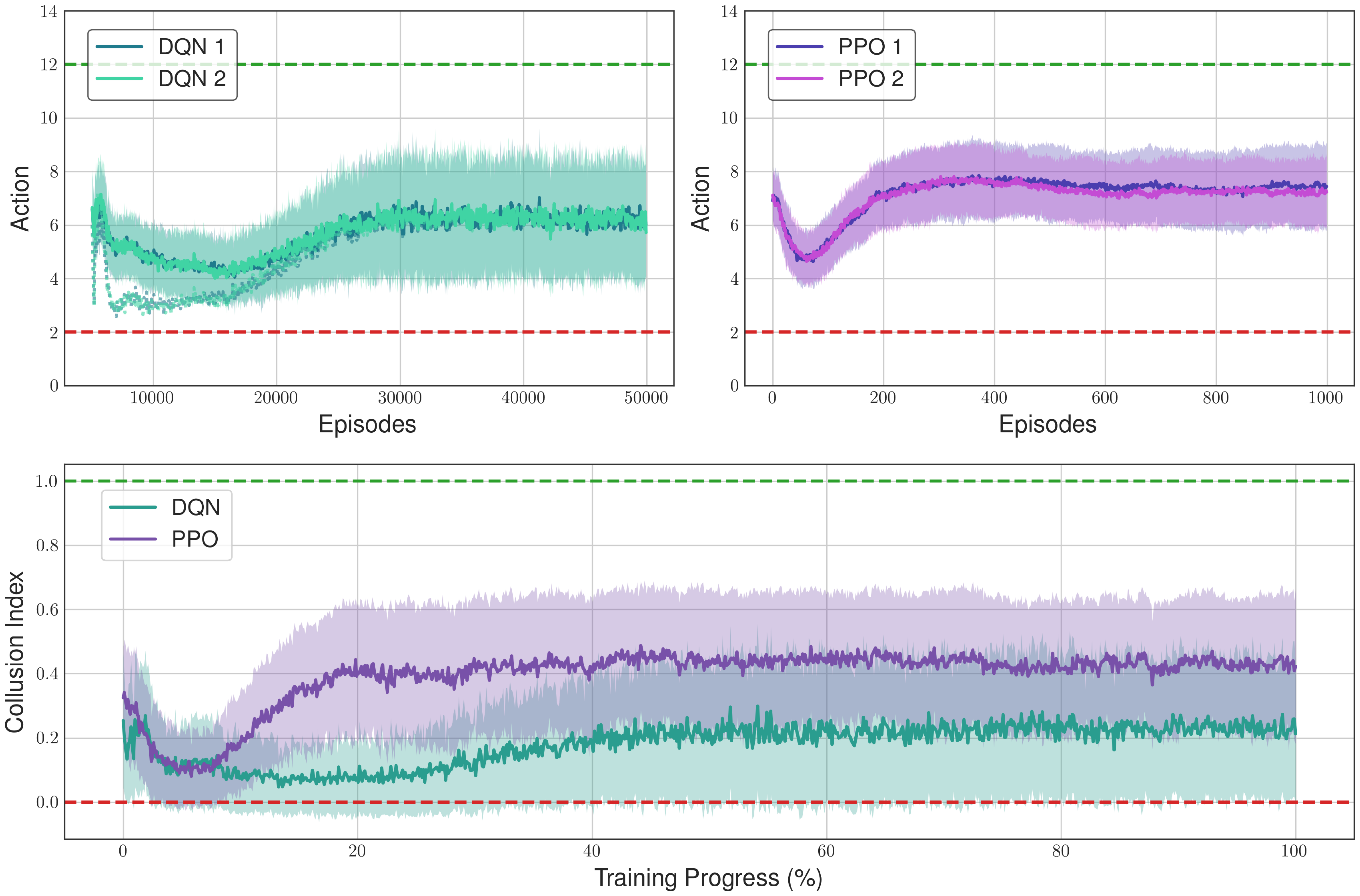
=1: collusive

Other measures?

➜ Social choice theory?



Results I: PPO and DQN collude



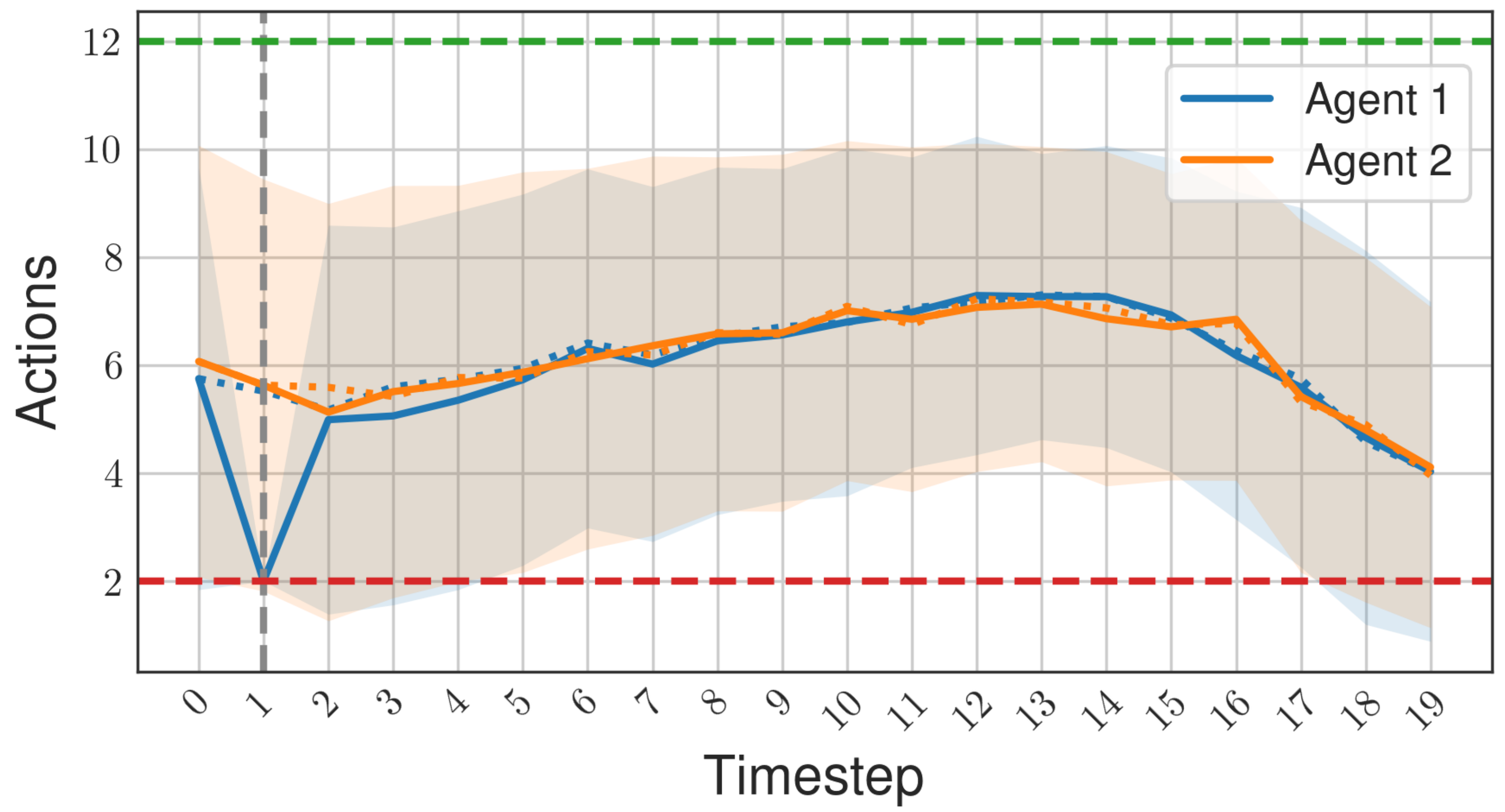
Setting: 2 homogeneous agents

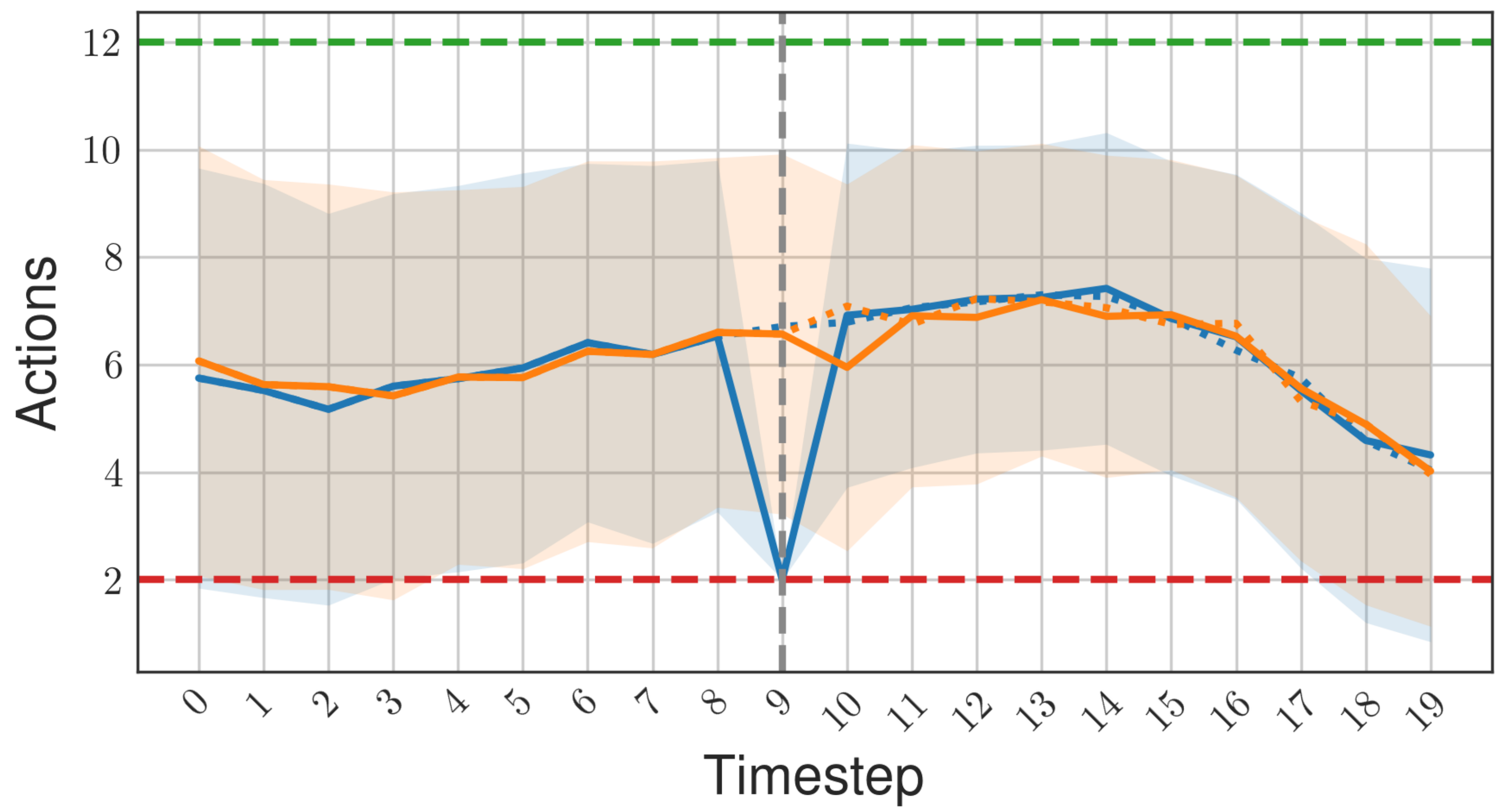
Clear learned collusion
➜ **Not** a Q-learning specific problem!

Explore competition first, then establish collusion
➜ Collusion emergence **not due to insufficient exploration!**

PPO colludes much sooner

Results II: analysing reward-punishments



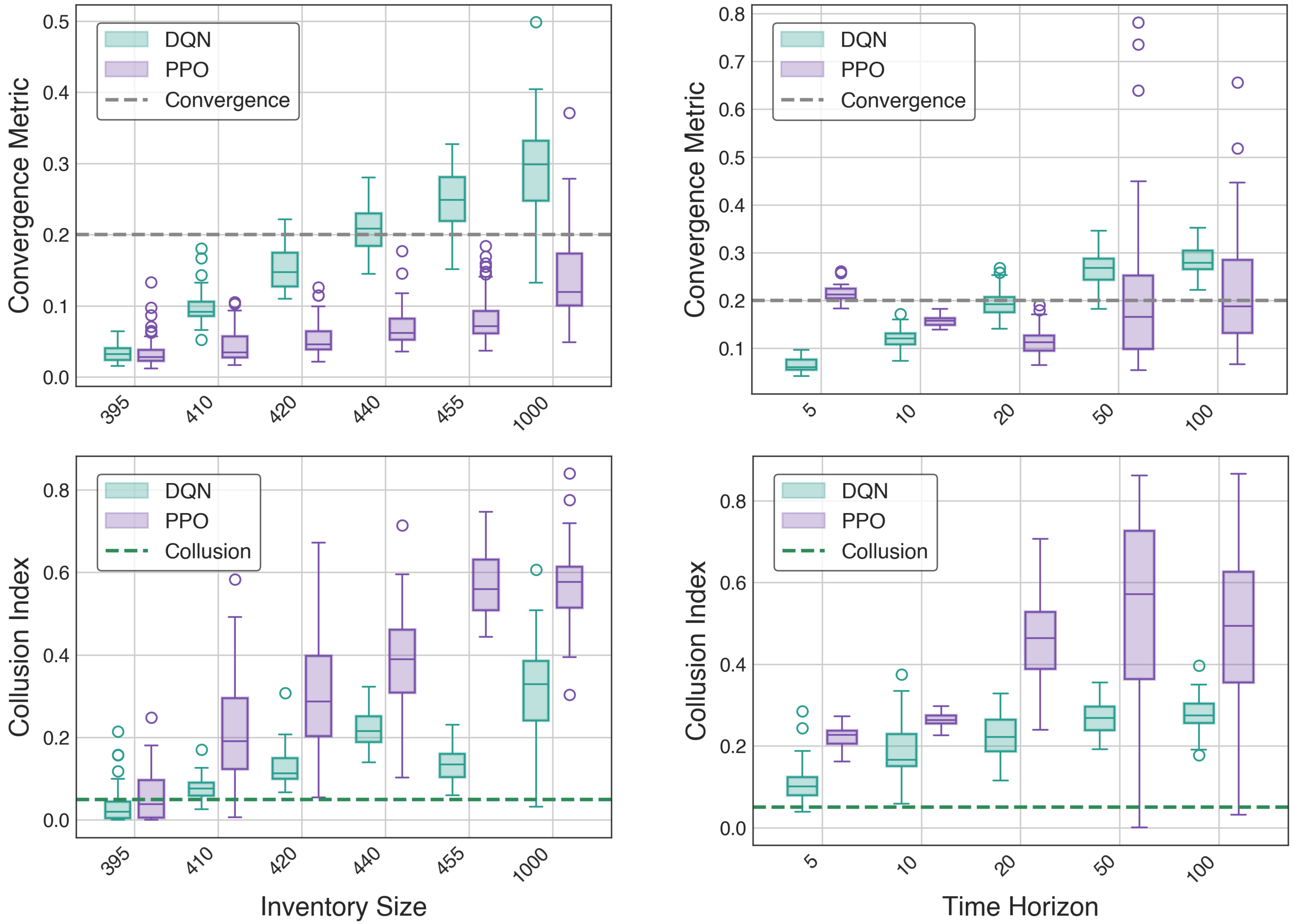


Analysis: force 1 agent to defect

Opponent retaliates (punishment) then both return to collusion (reward)

Both defect near end of episode
➜ Learning backward induction

Results III: collusion is robust



Collusion robust to inventory & episode length

PPO converges better, colludes stronger

Takeaways

Contributions:

- **Model episodic, inventory-constrained markets** via Markov game
- **Find unknown equilibria** via MINLP formulation
- **Learn robust collusion** with Deep RL (PPO & DQN)

In realistic markets, RL agents can achieve collusion without agreements and circumvent competition laws

Future:

- **Prevent and mitigate** learned collusion
- Can **opponent-shaping** methods exploit opponents and circumvent mitigation?
- Can more sophisticated algorithms use **price signaling to communicate**?

Thank you!

"Learning Collusion in Episodic, Inventory-Constrained Markets"

Paul Friedrich, Barna Pásztor, Giorgia Ramponi (Uni & ETH Zurich)