

**CS 332 Fall 2025**

Project #3

# 1 Outcomes from No-regret Learning in Games

## Methods

Repeated first-price auction (2 bidders, fixed values, full information + full feedback). Utility if win  $u = v - b$ , else 0; ties split 0.5.

*Algorithms*: Myopic (one-step expected utility using empirical  $P(\text{win} \mid b)$ ) vs Flexible (Exponential Weights/Hedge over  $k$  bids with full-feedback updates,  $\epsilon = \sqrt{\log k/n}$ ).

Setup:  $(v_1, v_2) \in \{(10, 10), (12, 8)\}$ ,  $k \in \{10, 100\}$ , rounds  $n \in \{100, 1000\}$ , Monte Carlo  $n_{mc} = 10$ . NE proxies computed at  $(12, 8)$  with  $k=100$ ,  $n=1000$  and a longer run  $n=5000$ .

## Results

Myopic vs Flexible  $(10, 10; k=10; n=1000; n_{mc}=10)$ : Myopic mean utility  $\approx 1022.8$ , win rate  $\approx 0.426$ ; Flexible  $\approx 1027.7$ ,  $\approx 0.574$ . Regret vs best fixed bid is not uniformly lower for Flexible with coarse  $k=10$ ; Flexible bids are more stable (lower variance). Robustness across  $n, k$ , values preserves the ranking (Flexible  $>$  Myopic); scaling  $\epsilon$  in  $\{0.5, 1, 2\} \times$  default changes variance, not ranking.

NE proxies (no pure NE for complete-info FPA with  $v_1 > v_2$ ; mixed equilibrium predicts efficient allocation and price near  $v_L$ ):  $(12, 8)$ ,  $k=100$ : efficiency  $0.946$  (all rounds,  $n=1000$ )  $\rightarrow 0.959$  (last 500 of  $n=5000$ ); average winning price  $7.80 \rightarrow 7.89$  vs  $v_L=8.0$ .

## Takeaways

Flexible algorithm consistently achieves higher average utility and win rate than Myopic; Myopic overreacts to recent bids. Outcomes move toward mixed-equilibrium predictions (high-value wins; price near  $v_L$ ) as  $n$  (and  $k$ ) increase, but do not converge to an exact equilibrium strategy profile.

# 2 Manipulability of No-regret Learners in Games

## Model

Same repeated FPA setting. Opponent: Flexible learner with  $v_1=9$ ,  $k=100$ . Our player (exploiter) has  $v_2=3$  and: observes for a few rounds (bid  $\approx 0.2v$ ); predicts the opponent's next bid via exponential smoothing ( $\alpha \in \{0.2, 0.4, 0.6\}$ ); when predicted bid falls below a threshold ratio ( $\in \{0.5, 0.6, 0.7\}$ ), bids to win with positive margin on the grid. Main run  $n=100$ ,  $n_{mc}=1$ ; robustness with  $n_{mc}=5$ . Full feedback enables computing  $P(\text{win} \mid b)$  for all grid bids and timing the exploit when the flexible learner's sampling puts mass on low bids.

## Results

Baseline (obs=5): Flexible win rate  $\approx 0.91$ , mean utility  $\approx 618$ ; exploiter win rate  $\approx 0.09$ , mean utility  $\approx 11$ . Robustness (obs  $\in \{5, 10, 15\}$ ,  $\alpha \in \{0.2, 0.4, 0.6\}$ , threshold  $\in \{0.5, 0.6, 0.7\}$ ;  $n_{mc}=5$ ): exploiter win rate  $\approx 0.08$ – $0.10$ , mean utility  $\approx 9$ – $12$ ; Flexible remains  $\approx 0.90$ – $0.92$  wins with high utility.

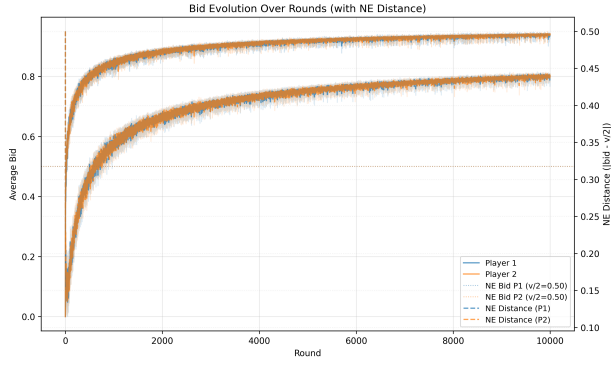
## Takeaways

Under full feedback, prediction-based exploitation yields occasional low-price wins for a low-value bidder, but gains are small; the flexible learner remains dominant. The effect is robust to reasonable observation/smoothing/threshold choices.

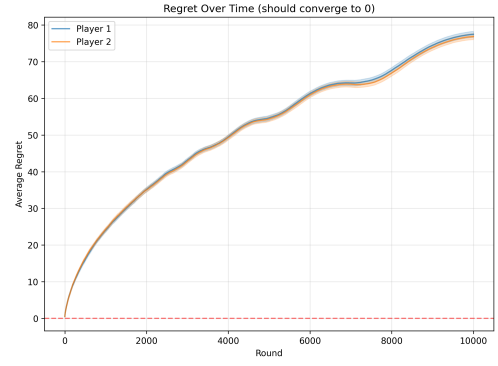
# AI and Collaboration

Both authors jointly designed experiments and interpreted results;

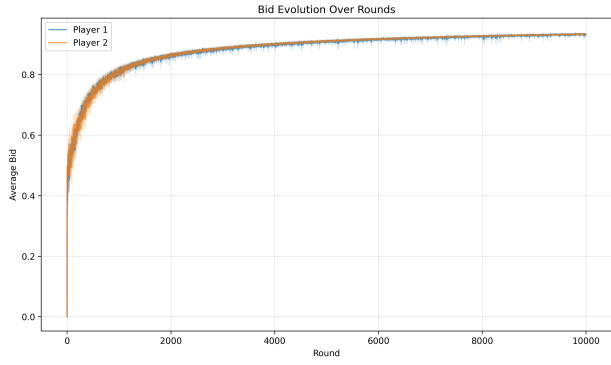
AI was used to assist with code organization, plotting, slide layout, and editing text; all code and results were reviewed and validated by the authors.



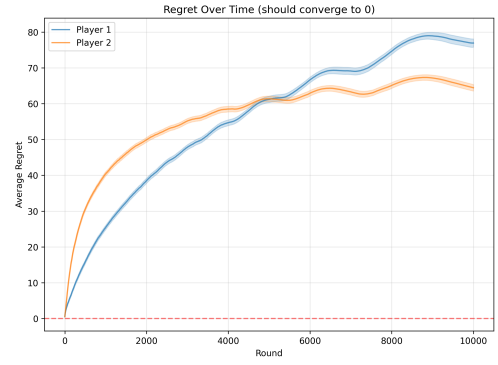
(1) Myopic vs Flexible — Bid



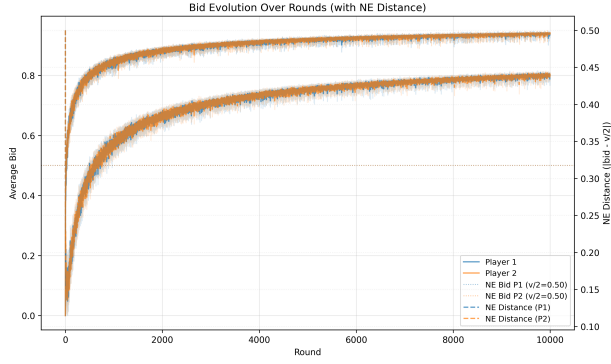
(2) Myopic vs Flexible — Regret



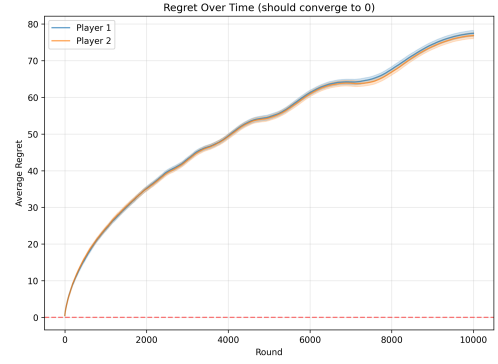
(3) Flexible (3xlr) vs Flexible (1r) — Bid



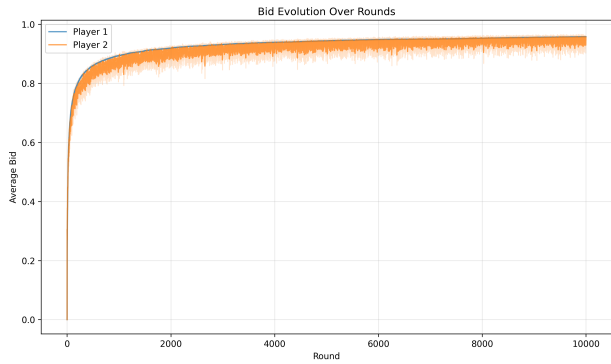
(4) Flexible (3xlr) vs Flexible (1r) — Regret



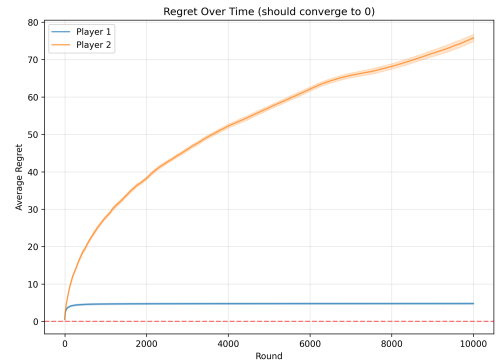
(5) Flexible vs Flexible (different  $v$ ) — Bid



(6) Flexible vs Flexible (different  $v$ ) — Regret



(7) Flexible vs Exploitation — Bid



(8) Flexible vs Exploitation — Regret