AI Collusion in Procurement Auctions

Andrew Tai

ARP Symposium
May 6, 2026

The views expressed are those of the author and do not reflect the official policy or position of the Department of War or the U.S. Government.

Background

- Auctions accounted for at least $1.5bil in US govt procurement in 2017
- Clear use case for small or uniform contracts ← only dimension of choice is price
- Possible use of AI algorithms by sellers
  - Extensive in online ad auctions; Amazon sellers, etc. (Aggarwal et al., 2018)
- AI algorithms learn to max. profit → autonomous collusion
  - Theoretical possibility established in other market contexts; e.g. Calvano et al., 2020.
  - E.g. allegedly RealPage for rent setting
  - Sellers can extract multiple times more profit from the buyer

Auctions

- More small, frequent, uniform auctions in the future?

- Procurement that is commoditized vs. large and indivisible
  - UAVs
  - Interceptor missiles

- Computing power for AI

Collusion

- US law: collusion generally requires “meeting of the minds”

- This paper: broader definition, including tacit collusion

- Sustain high prices by threat of later punishment, i.e. severe undercutting

- Implications for US policy:
    - Such activity may be difficult or impossible to prosecute

This paper

Setup:

- Reverse auction format
- Simulate basic AI learning algorithm (Q-learning)

Findings:

- Algorithms consistently collude at supra-competitive prices
- Extract on average 3x competitive profit

Implications:

- Real algorithms are likely much more advanced
- Large cost increases in procurement auctions
- Likely extend to more formats beyond reverse auctions
- Mitigated by increasing number of firms – defense industrial base
- Non-artificial intelligence (i.e. humans) can also replicate this behavior

Economic environment

- Two firms competing in reverse auction format
  - Sealed bid
  - Lower bidder wins the contract, receives bid in profit
  - 0 cost to firm of providing contract (doesn't matter)

- Bids in discrete increments of $[0, \frac{1}{6}, \dots, \frac{5}{6}, 1]$

- **Nash equilibria**: both firms bid 1/6, obtain small profit

- Repeated $T = 100{,}000$ periods

- Above simulated 100 times

Q-learning algorithms

- **Q-matrix (policy)** is a function from states (past bids) to actions (current bid)
  - Given the previous period bids, what will I bid?
  - For every state X action pair: discounted future profit

- Q-learning attempts to learn the optimal policy
  - Eventually maximize NPV of profits

Q-learning algorithms

Q-learning algorithms

Each slot machine

- Has different, ex ante unknown prizes
- Has different, ex ante unknown chance of winning
- Prizes and chances change each day depending on weather

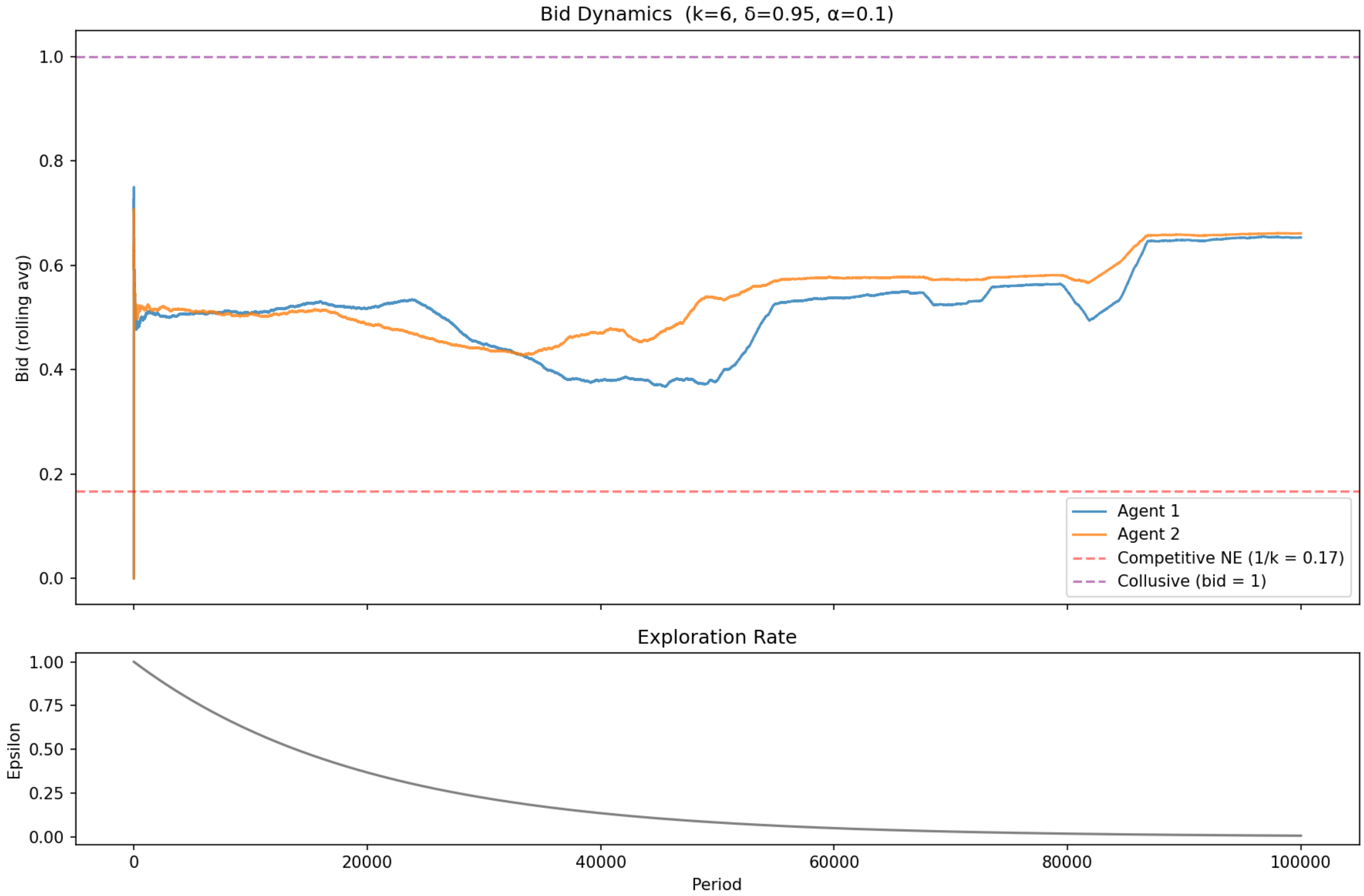
Q-learning

- Starts playing slot machines randomly
- Learns which machines are best under what weather
- Eventually focuses on those

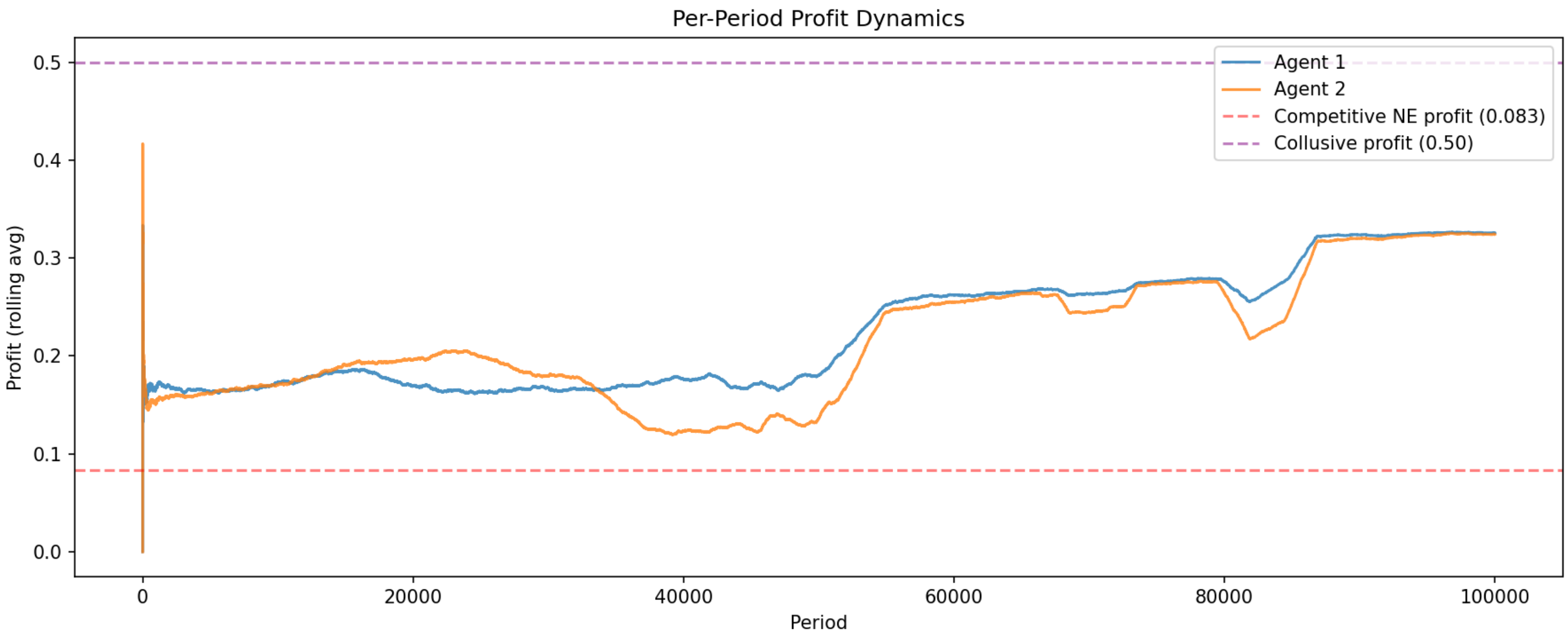
In auctions:

- Starts bidding randomly
- Learns what bids are best depending on available info
- Available info – past history

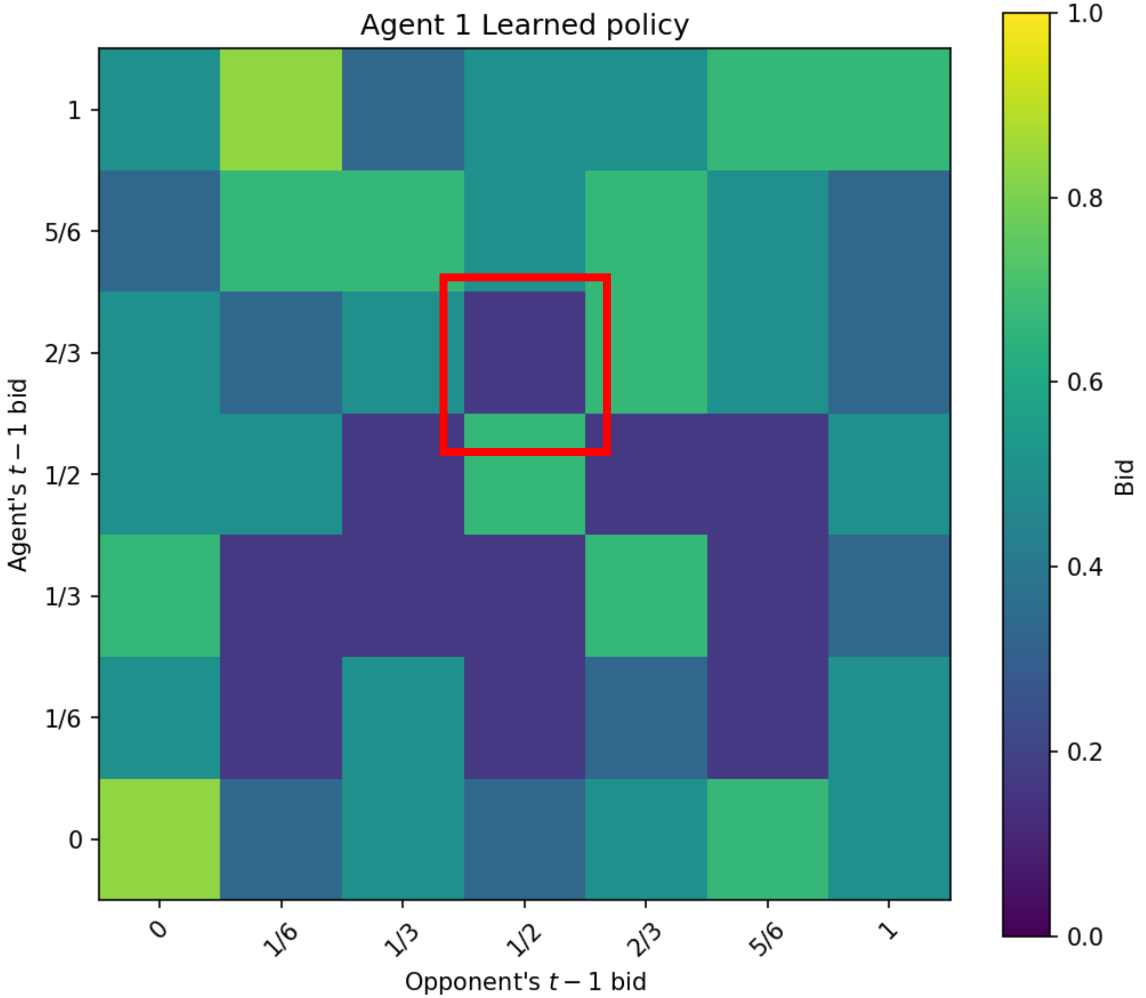
Bids over time



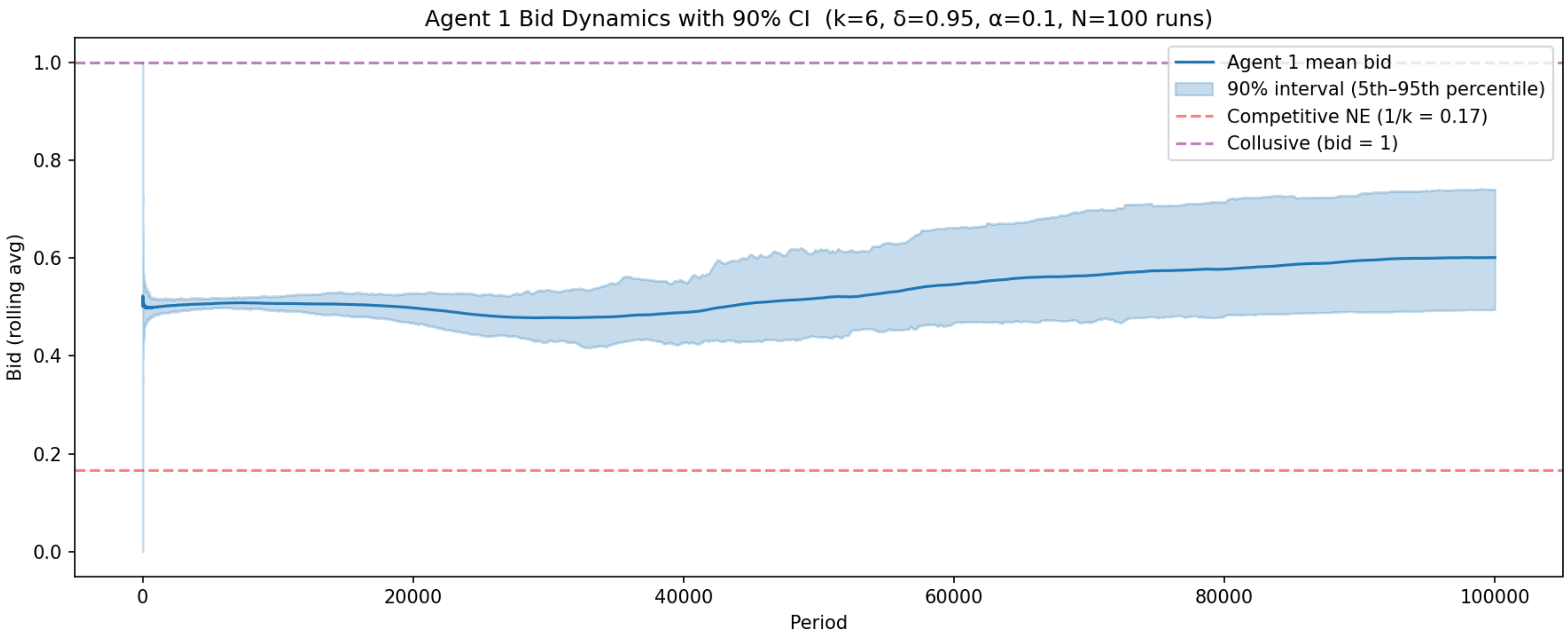
Profits over time



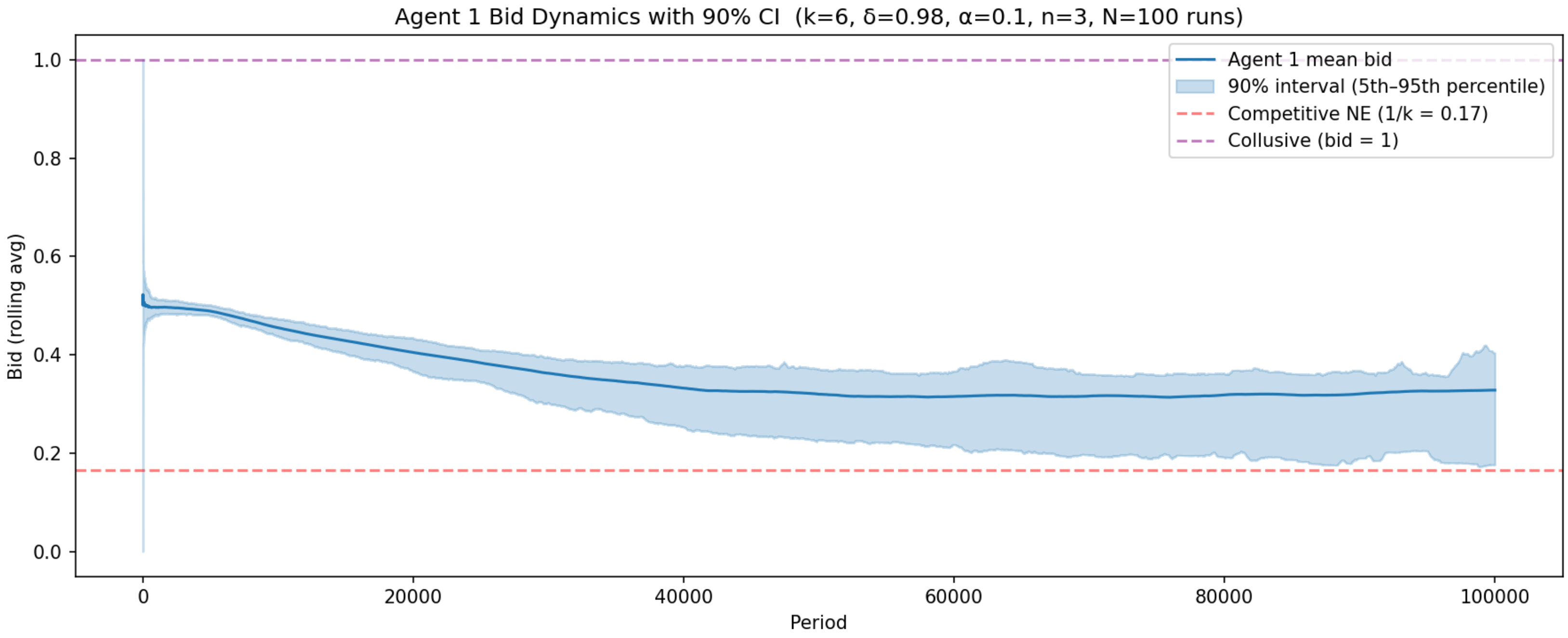
Learned policy



100 simulations



Three firms



Agent 1 Bid Dynamics with 90% CI (k=6, δ=0.98, α=0.1, n=3, N=100 runs)
Agent 1 mean bid
90% interval (5th–95th percentile)
Competitive NE (1/k = 0.17)
Collusive (bid = 1)
Bid (rolling avg)
1.0
0.8
0.6
0.4
0.2
0.0
0
20000
40000
60000
80000
100000
Period

Conclusion

In a repeated reverse auction environment:

- Q-learning algorithms consistently bid supra-competitively
- In this simulation, extract 3x profit vs. competitive
- Humans can do this too

Possible policies

- No explicit communication → Not generally considered anticompetitive
- Target firm market concentration rather than explicit communication
- Restricting usage of algorithms by sellers
    - But humans can do this too
- 1^{st} order: more firms → Expand defense industrial base
- Further experiments may offer hints on parameters

Q-learning algorithms

- **Q-matrix (policy)** is a function from states (past bids) to actions (current bid)
  - Given the previous period bids, what will I (the algorithm) bid
  - For every state X action pair: discounted future profit

- Q-learning attempts to learn the optimal policy

$$Q_{t+1}(s,a) = (1-\alpha)Q_t(s,a) + \alpha\left[\pi_t + \delta \max_{a \in A} Q_t(s',a)\right]$$

- Given state s and action a, updates α towards observed profit
  - This period profit + continuation value

- Put simply: experiment and remember payoffs to try to maximize profit