Algorithmic Collusion in Repeated Auctions

Will Nehrboss

Contributing authors: wkn4dn@virginia.edu;

Abstract

Simple Q-learning algorithms have been demonstrated to successfully collude in some competitive environments. In this paper, I investigate the ability of these algorithms to collude in repeated contact auctions. Examining both FPA and SPA auctions with environment parameters motivated by search-advertising markets, I demonstrate that simple Q-learning algorithms are unable to overcome the native barriers to collusion found in indivisible-good auctions. I go on to develop a more sophisticated variant on traditional Q-learning, *two-level Q-learning*, which achieves highly effective collusion in the same environments. Remarkably, this algorithm learns the strategy of bid rotation; this unprecedented result demonstrates the ability of even unsophisticated algorithms to develop anthropomorphic collusive behavior in stylized environments.

Keywords: Algorithmic Collusion, Auction, Bid Rotation, Bid Rigging

1 Introduction

The use of computer algorithms to play repeated contact games with poor sub-game Nash equibria has a long history (Axelrod and Hamilton, 1981). Limitations in computing power and the sophistication of learning algorithms traditionally prevented the evolution of collusive behavior in these games without explicit instruction (Kimbrough and Murphy, 2008). Advancements in the

field of artificial intelligence have made the autonomous development of collusive strategies increasingly feasible, even in complex environments (Hu and Wellman, 2003). In the past few years, the potential for such collusion in the context of market pricing has attracted intense interest. Several recent papers have demonstrated the ability of various deterministic and deep learning algorithms to collude on prices in simulated market environments (Calvano et al, 2020; Hansen et al, 2021). Others have investigated empirical evidence of algorithmic pricing, finding credible indications of collusion (Assad et al, 2020). From an anti-trust perspective, the looming possibility of autonomous algorithmic collusion is particularly consequential. The current legal framework requires a prior agreement between collusive parties for such behavior to be illicit (Capobianco, 2020); the advent of algorithms with the ability to collude without inherent bias presents the potential for systematic but seemingly un-indictable tacit collusion.

While a rich literature exists evidencing the viability of algorithmic collusion in price setting, very little has been done to generalize this result to alternative market mechanisms. In this context, I attempt to contribute to our understanding of algorithmic collusion by examining algorithmic bidding in standard auctions. A potential explanation for the paucity of research into this area is that auctions are generally hostile to the learning of collusion. Key ingredients of collusion found in the pricing game, including recurring interaction between players and stable market conditions, can be absent in auctions, commonly characterized by low volumes and large inter-auction changes in value and bidding pool. However, online markets have led to the growing relevance of an auction-based mechanism that negates these barriers: high-frequency auctions. High-frequency auctions are susceptible to collusion for

a few reasons: firstly, the sheer number of repetitions provides an opportunity for dynamic learning of non-deterministic bidding strategies; secondly, the homogeneity of goods across auctions allows for collusive signals to stand out against inter-auction variance in values; thirdly, bidding pools tend to remain relatively stable across auctions allowing for sustained contact between bidders, a vital condition for tacit collusion. I investigate an important utilization of high-frequency auctions, the search advertising market. Not only does this market provide a rich environment for collusion for the reasons above, but it has already seen the widespread adoption of algorithmic bidding - a variety of highly sophisticated autonomous bidding algorithms currently dominate auctions run by major search platforms. In this paper, I attempt to model the potential for such algorithms to capitalize on these market characteristics to extract collusive profit.

Following in the footsteps of past papers on the topic, I utilize the Q-learning algorithm, or a variant thereof, as a proxy for the bidding algorithms used in current markets. The Q-learning algorithm is a straightforward type of reinforcement learning algorithm whose simplicity and mathematical accessibility has made it especially popular as a tool in the analysis of algorithmic collusion. In my simulation, two non-communicating, profit-maximizing agents learn to bid over a sequence of thousands of auctions. The environment in which these agents interact is very much a stylized version of the genuine search-advertising market; I make several simplifying assumptions regarding the agents themselves and the auction environment, the most important of which is a small and stable bidding pool. I glean several salient results from this simulation: Firstly, standard Q-learning algorithms fail to learn to collude in first and second price auctions, for reasons I enumerate in Section 5. Secondly, a slightly more sophisticated algorithm *two-level Q-learning*, succeeds

in overcoming these barriers and effectively colludes in both types of auction. Thirdly, a display of algorithmic bid rotation; the two-level Q-learning algorithms tacitly develop a complex collusive strategy that is a hallmark of bid-rigging.

This paper is laid out as follows: Section 2 reviews the current literature on the topic of algorithmic collusion and outlines the current understanding of collusive strategies in auctions. Section 3 describes the high-frequency auctions of the search advertising market, detailing the environment parameters and rules which govern my simulation. Section 4 introduces the foundational reinforcement learning algorithm used throughout the paper, the Q-learning algorithm. Subsections 4.1 through 4.3 specify simulation environment, Q-learning updating rules, and experiment parameters. I outline in Section 5 the results of the bidding strategies learned by the standard Q-learning model, which indicate robust competition. In Section 6, I introduce a slightly more sophisticated variant on the standard Q-learning algorithm, two-level Q-learning; in the subsequent sections, I re-run the simulation and present the results of robust collusion. Section 7 demonstrates the robustness of my results to changes in environment parameters. In Section 8, I reflect on the implications of my results and present potential avenues of further exploration on the topic.

2 Literature Review

In their influential 1981 paper, Axelrod and Hamilton (1981) demonstrate how algorithms could overcome the Prisoner's Dilemma in repeated games through a strategy of simple reciprocity - a grim-trigger-like approach in which each agent punishes selfish moves and rewards mutually beneficial moves. While Axelrod's algorithms depended on pre-programmed strategies and a straightforward environment to collude, his paper began an inexorable march

towards increasingly autonomous collusion in environments of increasing complexity. Kimbrough and Murphy (2008) outlined how various specifications of a naive reinforcement learning algorithm could result in collusion in Cournot oligopoly, although such collusion was still dependent pre-determined altruism. In his seminal work on algorithmic collusion, Calvano et al (2020) took this step further, using a completely unbiased but more sophisticated learning algorithm, the Q-learning algorithm, to compete in Bertrand oligopoly with differentiated products. These Q-learning algorithms were able to maintain supra-competitive pricing through an autonomously learned punishment-reward scheme. Hansen et al (2021) built on this, providing two important insights: firstly, that the result of collusion is robust to alternative algorithm designs and secondly, that the feasibility of algorithmic collusion is highly dependent on collusive signals standing out against environment variability. This literature provides compelling evidence that collusive algorithmic price setting is a real possibility in stylized market environments; the empirical foundation of algorithmic collusion remains less understood. The most important work in this area, Assad et al (2020), tied the adoption of algorithmic price setting to the emergence of non-competitive pricing. This effect was especially pronounced in regions where a small number of stations all adopted algorithmic price setting; unfettered interaction of algorithms leading to the learning of collusion was theorized to be behind this result.

There is a wealth of literature on collusion in auctions, both theoretical and empirical. Tacit and explicit bid rigging is a common and consequential occurrence; paper after paper document bid rigging in a great number of markets. Two early papers, (Porter and Zona, 1993) and (Pesendorfer, 2000), document the impact of cartels in public construction procurement and school milk auctions respectively. McAfee and McMillan (1992) comprehensively outline the

mechanics of bidding rings, describing the strategy of bid rotation - in which cartel members alternately withdraw or submit a shill bid to dampen competitive pressure and distribute collusive profit. Skrzypacz and Hopenhayn (2004) discuss the evolution of tacit cartels in repeated auctions, presenting the intuition that low discounting of future profit is critical to cartel integrity. Zhang (2021) mathematically formalizes the latter result; I also provide empirical evidence for this in Section 8.

Since its genesis with Watkins and Dayan (1992), Q-learning has been demonstrated to be a simple, yet effective tool of autonomous learning. Some have proposed mutant Q-learning algorithms which are more effective in converging to optimal strategies in certain environments. One such variant, *Nash Q-learning*, proposed by Hu and Wellman (2003), chooses actions under the assumption that the opposing agent will choose the Nash equilibrium action. This has a few disadvantages in the context of algorithmic bidding, the most dispositive of which are the computational difficulty in forecasting Nash equilibrium in non-trivial environments (Daskalakis et al, 2009) and intrinsic bias towards collusion which would run afoul of anti-trust laws.

A very recent paper, Banchio and Skrzypacz (2022), investigates algorithmic bidding using Q-learning algorithms. However, when specifying the learning algorithm, the authors make the critical simplifying assumption that each agent is completely unaware of their opponent's bids in past auctions - each agent only keeps track of their preceding bid. This fundamentally changes the analysis. Genuine collusion, which depends on the monitoring of the opponent's behavior, should be highly challenging, if not impossible, in such an environment. Additionally, the learning simulation used is quite short (each action-state is only visited, on average, 50 times). Without sufficient time to learn the dynamics of the environment or sufficient memory to comprehend

them, the algorithms are effectively faced with an optimization problem in the face of total random noise. While the authors interpret the low average biding in FPA as learned collusion, this result is better characterized as an artifact of the simulation design. This is evidenced by the fact that when the agents do incorporate information regarding their opponent's bid into their model, this 'collusion' disappears in favor of the competitive equilibrium. Such a result is antithetical to our understanding of tacit collusion; increased public information should only facilitate cartel behavior. The excessive simplicity of this analysis necessitates an examination of more information-rich algorithms which better reflect those currently used in search advertising markets. I seek to provide that analysis in this paper.

3 Industry Description

The paradigm of reliable and accessible computing tools has made automated, high-frequency auctions a reality. Many markets use such auctions to facilitate the buying and selling of an enormous volume of goods, from financial instruments to commodities to online advertisement slots. For example, in 2000, Google begin auctioning off slots of search advertisements through many repeated second-price auctions; since then, hundreds of billions of dollars worth of advertisements have been cleared through this system. As computing tools have improved since the early 2000s, bidders in these advertisement auctions began to automate behavior using computer algorithms; recently, fully autonomous algorithms that can 'learn' optimal bidding strategies have come to play an important role in such auctions.

The particular significance of this system of auctions lies in its repetition. In standard one-shot auctions, bidders are faced with dramatic uncertainty regarding the private values of opponents; this information asymmetry results

in equilibrium bidding strategies that utilize only limited information regarding opponents' behavior. In high-frequency search advertisement auctions repeated contact between bidders allows for the observation of the past behavior of opposing bidders; bidders may use this information to update expectations of the opposing bidders' values and future behavior. This ameliorates the information asymmetry and allows for the organic development of complex bidding strategies.

In this paper, I investigate the results of algorithmic bidding in a simulation of high-frequency search advertisement auctions with a limited, stable bidding pool.

The search advertising (SA) market is segmented between a few dominant search platforms, each of which operates markets serving a large number of bidders. The major search platforms, Google and Microsoft, offer higher search indexing for sponsored links in return for payments. These agents, seeking to maximize the summed payments from SA auctions, oversee and set the rules for auctions which allocate billions of ad slots. The search platforms have a vested interest in the maintenance of competitive bidding as any collusive behavior would have a detrimental effect on ad revenue.

The bidders are firms, seeking to capitalize on the increased visibility of sponsored links to increase their revenue. Annually, millions of firms around the world compete in SA auctions, however, the bidding pool for each particular auction is a tiny subset of this larger pool. In each auction, bidders submit bids for a given set of search keywords – the auction's bidding pool is made up of all bidders sharing these keywords. In a sense, the larger SA market can be segmented into smaller markets defined by a unique set of keyword combinations, each attracting a different set of bidders to its sequence of auctions. Depending on the number of available sponsored slots, the ranked top

n bidders are allocated the first corresponding n ad slots for a given number of impressions. I will focus on auctions in which $n = 1$. An interesting facet of SA auctions is that bidders are not fully aware of their individual valuation of the ad slot for which they are bidding. The value of a single ad slot is determined *ex post facto*, that is, after the completion of the auction, the increase in revenue resulting from the advertisement determines the winning bidder's private value. Bidders must formulate expectations of their private value based on experiences in past auctions; the accuracy of these expectations is victim to natural variability and exogenous shocks over the course of repeated auctions.

Because the target keywords for a particular firm remain relatively consistent, firms that compete in one auction tend to continue to compete in a sequence of repeated auctions for those same keywords. This allows for repeated contact between bidders. Furthermore, by dint of this shared market focus, competing bidders often interact in more than one auction due to additional shared keyword combinations. This multi-market contact has implications for potential collusive behavior should be explored in future research. Markets are segmented geographically only along the lines of the language barrier - bidders tend to face competition and demand from agents within the same language group. A growing share of bidders utilizes an algorithm to determine and place their bids; in fact, Google freely offers participating bidders an algorithm to set bids. I consider those auctions in which algorithmic bidding is universal. In the next section, I model this algorithmic bidding using reinforcement learning algorithms.

While the specific auction rules vary from platform to platform, there are a few consistencies. Most SA auctions follow one of the two standard designs: first-price or second-price auctions. In first-price auctions (FPA), all bidders submit a single bid, the highest bidder wins, and the payment is the value of

that winning bid. In second-price auctions, (SPA) again, the highest bidder wins, but the payment is the value of the second-highest bid. Google recently transitioned from a SPA to an FPA design. Most search platforms weigh additional characteristics of the advertisements themselves in determining the winner (as the click-through rate on the ad impacts revenue), but I will simplify the model with the assumption that these 'soft characteristics' are uniform amongst competing ads. Bids in each auction are kept private but bidding histories are public (e.g. Google post-2019). The winning bidder's sponsored link is displayed for a given number of impressions, after which the auction is rerun for the same keywords to determine the allocation of the advertisement slot. Given the volume of searches, this impression quota is often met quite quickly; as a result, an enormous number of SA auctions are needed to continually allocate ad slots – Google alone completes three billion SA auctions per day. Due to the rapidity of these auctions, bidders usually chose a bid that is submitted to several rounds of the same auction sequence.

4 Model

4.1 Simulation Environment

To investigate the effects of algorithmic bidding, I simulate a sequence of repeated auctions that incorporate generalized characteristics of SA[1] auctions. The auction environment is as follows: Two bidders ($n = 2$) compete across a sequence of thousands of rounds. At the beginning of each round, the agents chose a bid which is entered in l auction iterations. After the l auctions are completed, the agents choose to increment or decrement their bid based on the outcome of the past rounds. Past bidding history and personal profit outcomes are the only available information to both agents. Just as in SA auctions,

[1] Search Advertisement

agents are unaware of their own or their opponent's true value when bidding on a good; between each auction iteration, each agent's true value is subject to independent, random shocks which the agents learn only after the completion of the auction. Thus, each agent must rely on uncertain expectations of their private value. I explore both FPA and SPA auctions.

4.2 Basic Q-Learning

Seeking to maximize profit across the auction sequence, each agent learns their optimal bidding strategy using a Q-learning algorithm. The Q-learning algorithm is a type of simple reinforcement learning algorithm in which agents learn to favorably weight more profitable actions and to unfavorably weight less profitable ones. At the beginning of each round, the algorithm chooses an action a_i - to increment, decrement, or not change previous bid b_i - and then updates the value associated with that action at the current state using the mean rewards over that round. The state s_i at time t is simply the pair of bids (b_i, b_j) in the previous round at *t-1*. The reward for auction_t is given by $r_i = F(b_i, b_j|\lambda_t^i, \zeta_t)$, where λ_t^i is the shock to agent_i's value and ζ_t is the tiebreak rule[2]. The function used to assign values to each action is specified by:

$$Q(a_i, s_i) = \mathbb{E}[\bar{r_i}] + \rho\, \mathbb{E}[\max_{a_i'} Q(a_i', s_i'|a_i, s_i')] \tag{1}$$

where ρ is the discount factor of future rewards and a_i' is the range of possible actions in the succeeding state s_i' after taking action a_i. The Q-value associated with the state-action pair a_i, s_i is the expected rewards plus the discounted value of the best action next round. The values for each state-action pair are maintained in each agent's Q-matrix, which maps a three-element vector (i.e.,

[2]If two agents submit the same bid, a random 50-50 tiebreak is imposed to decide the winner.

action values) to each location in the action space. To facilitate the implementation of the Q-learning algorithm, I discretize the action space of the agents to integer bids [0,5][3]. I chose a relatively small action space to make the simulation less burdensome computationally, although, as I will demonstrate later on, my results are robust to a large action space. I restricted bids to those which are not strictly unprofitable i.e., uniformly greater than the agent's value.

The learning process is the updating of the Q-matrix using the following function:

$$Q_{t+1}(a_i, s_i) = (1-\alpha) \cdot Q_t(a_i, s_i) + \alpha \cdot (\bar{r}_i + \rho \, \mathbb{E}[\max_{a_i'} Q(a_i', s_i' | a_i, s_i)]). \quad (2)$$

Again, this has a straightforward interpretation: the Q-value of s_i after executing a_i is updated to be a convex combination of the prior Q-value and the sum of the observed mean rewards over the round plus the discounted value of the best possible action in the succeeding state. The weighting of this combination is the learning rate, α, which determines how quickly the agent adapts to a new environment. A higher α allows the agent to find optimal strategies more quickly but makes these strategies more easily disrupted by exogenous shocks or changes in the opponent's behavior.

4.3 Experiment Parameters

Each experiment runs for 80,000 rounds. The algorithm uses an ϵ-greedy rule to determine its actions; that is, at each round, the agent will choose a random action instead of the optimal action according to its Q-matrix ϵ fraction of times. ϵ decays over the course of the experiment: $\epsilon = e^{\beta\tau}$, where τ is the round number. I set β to 10^{-4}. The period of high exploration towards the beginning of the experiment allows the agents to more accurately estimate

[3] SA auctions often specify a certain bid increment, so this is not entirely detached from reality.

their Q-matrices through the repeated visitation of all action states[4]. I specify α to be 0.4 and a ρ of 0.9. Initial bids and Q-matrices are all initialized to zero, although the results are robust to varied initializations. The private value of each agent is given by

$$\nu_i^{n^{th}win} = 5.1 + \lambda_i^{n^{th}win} \tag{3}$$

$$\lambda_i^{n^{th}win} = \gamma \cdot (\lambda_i^{n-1^{th}win} + \mathcal{N}_\rangle(0,1)) \tag{4}$$

An agent's value, ν_i, in a given auction is a combination of its past value and a random shock. The rationale for this is that the private value of an advertisement can change either due to internal factors that remain somewhat consistent across auctions or ad-specific factors that are more variable. The agent's value is only updated after winning an auction. The coefficient γ can be viewed as a noise parameter for the simulation; high levels of noise should be a barrier to convergence to optimal bidding strategies, as the randomness in auction profit obscures the action-reward relationship that underpins the learning process. The effective noise is somewhat lower as the agents optimize their mean rewards of each round, so the variance of mean value across rounds is lower by a factor of $agent_i$'s n wins per round. I chose a round length, l, of 50 and a γ of 0.3.

5 Results

The results of this simulation resemble what we would theoretically expect under competitive bidding. In FPA auctions, agents universally converge to a competitive price with a single step of bid shading. The significance of bid

[4]This relatively low β allows for quicker convergence but sacrifices some accuracy in the agent's ultimate Q-matrix estimation.

shading in this experiment is an artifact of the small action space - increasing the action space results in a relatively less significant bid shading in FPA[5]. In SPA, agents learn to bid as close as possible to their own value, which, due to the initial value specification, results in a less profitable outcome for the agents. In neither environment do the algorithms show indications of collusion.

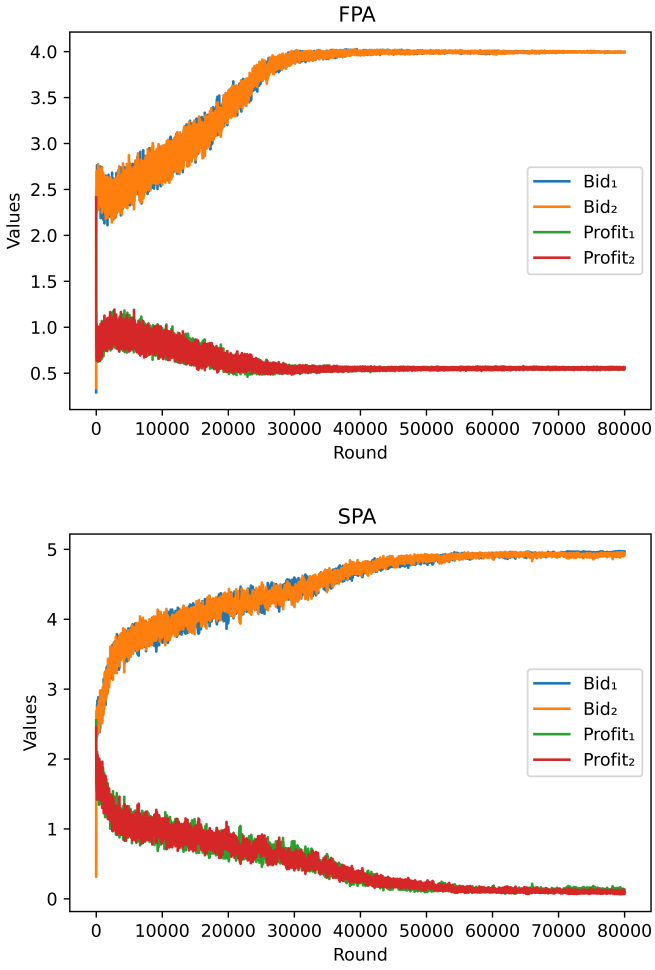

Fig. 1: *Mean of Bids and Rewards Across 500 Experiments*
Note: The agents converge to the competitive equilibrium universally in both environments.

[5]In a descretized action space, the least amount of non-trivial bid shading is a single one below value. This one step is relatively significant when the action space is limited to such a degree, but should not be interpreted as evidence for collusion.

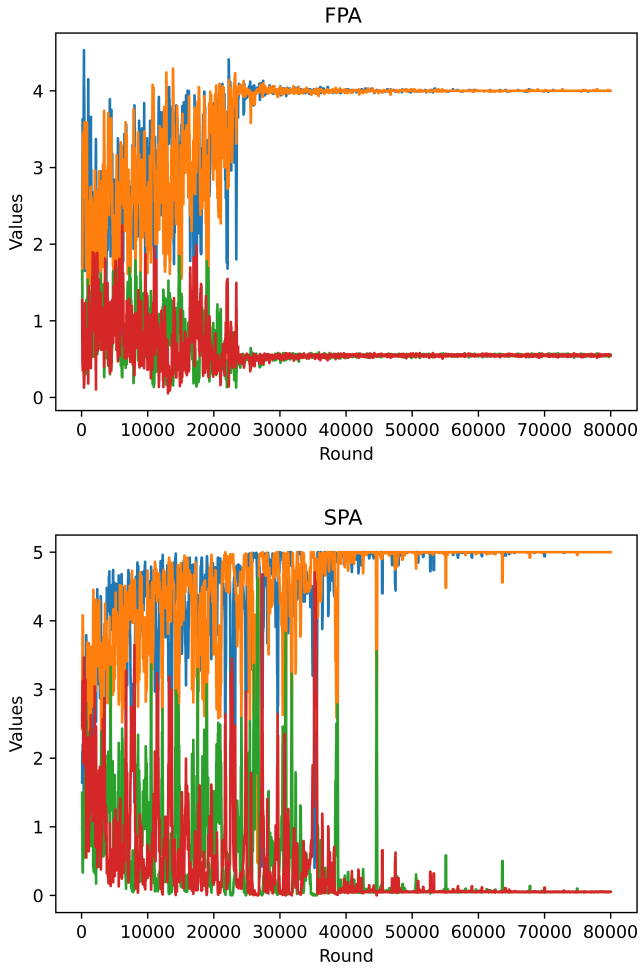

Fig. 2: *Smoothed Bids and Rewards in Single Experiment*
Note: Intense exploration slowly gives way to competitive bidding, as the agents learn the optimal strategy in both environments.

A snapshot of the learning process in a single experiment can be seen in Figure 2. Both agents engage in intense exploration towards the beginning of the simulation, during which they adapt to the auction environment. As the simulation progresses and ϵ decays towards zero, the agents begin to bid more aggressively, understanding that an aggressive bidding strategy tends to result in greater profit. This competitive pressure forces both agent's bids towards an equilibrium: in FPA, $b_i, b_j \xrightarrow{\mu} 4$ and in SPA, $b_i, b_j \xrightarrow{\mu} 5$. The difference in the equilibrium bids between SPA and FPA is neither surprising nor contrary to theory. In SPA, the dominant strategy is to bid the agent's own value as

the bid only determines the probability of winning. Conversely, in FPA, the agent's bid determines both their probability of winning and potential profit; thus the optimal strategy when facing uncertainty regarding the opposing agent's value and strategy is to bid somewhat below the agent's own expected value. The result is remarkably consistent, across all 1000 experiments, the winning bid converges to the respective FPA and SPA equilibrium uniformly. The agents generally find the equilibrium bids more rapidly in FPA than in SPA; I believe this is a result of the greater agent welfare under the FPA equilibrium - experimentation in late stage SPA is much less costly. The apparent lack of non-competitive behavior is a critical result. Calvano et al (2020) demonstrate the ability of Q-learning algorithms to collude in Bertrand competition with differentiated products; the fact that this result does not hold in auctions speaks to the following conclusion: algorithmic collusion in auctions is a fundamentally more difficult problem. This is the case for three reasons:

- *Asymmetric Information*: Each agent is unaware of the opponent's changing, private value. This is in contrast to the constant and uniform marginal cost used by Calvano et al (2020). Collusion becomes easier when agents better understand the parameters on which their opponent's strategies depend; as a result, the relative paucity of information in these auctions encourages competitive behavior.
- *Variance of Tiebreaks*: When agents collude on the same bid, splitting the indivisible good is not an option. Rather, one agent is randomly chosen to be the recipient of the entire good. This makes it difficult for a colluding agent to determine whether a private welfare loss is due to adverse tiebreaking or defection by the opposing agent. The maintenance of collusion depends on signals of collusion and defection standing out from the variance of the simulation; even when the variability of tiebreaking is smoothed across l

auctions in each round, it is likely still a major culprit in the destabilization of any potential collusive strategies.

- *Reduced Informational Feedback*: In differentiated Bertrand competition, each action that the algorithm takes results in a change in welfare, regardless of the distance between competitors' prices. This action→feedback mechanism is the source of learning and environmental understanding that Q-learning algorithm develops during the simulation. In auctions, this mechanism is disrupted in a large subset of states; when the distance between bids is larger than 1, the losing agent's welfare remains zero regardless of its immediate actions. This hampers stable estimation of the Q-matrix and, by the same token, the finding of optimal strategies. In SPA, this effect is exacerbated by the fact that when $b_i \notin (\mathrm{b}_j - 1, b_j + 1))$, the agent's action can have no immediate effect on its welfare (this result is confirmed in Section 7); the action→feedback is absent in a much larger group of states. This discontinuity in rewards and resulting action-profit independence creates a hostile environment for the Q-learning algorithm to learn complex strategies.

We have established that the limitations of the Q-learning algorithms memory and informational richness preclude the development of collusive bidding behavior. In the next section, I relax these limitations, adding another dimension to the agent's memory and exploring the subsequent implications for bidding behavior.

6 Two-Level Q-Learning

6.1 Model

In this section, I seek to answer the question: is a slightly more sophisticated learning algorithm able to overcome the intrinsic barriers to collusion present

in auctions? The philosophical motivation behind this question is grounded in the development and use of powerful deep-learning algorithms which can dynamically comprehend and incorporate complex environmental factors into its learning process. In SA auctions, a variety of advanced AI software is currently in use. However, such algorithms and their approaches to learning are difficult to mathematically describe. Instead, I create a novel variant on the traditional Q-learning algorithm that maintains most of the simplicity and accessibility of Q-learning but offers the agents increased flexibility and sophistication in the auction sequence.

The crux of two-level Q-learning is as follows: At the beginning of each experiment, the algorithm chooses the value of a parameter, θ, which it incorporates into the learning process of its bidding strategy throughout the experiment. During the experiment, it learns in a similar way but utilizes an updating rule that is a function of θ. Then, at the termination of the experiment, it observes the mean rewards over the course of the experiment and updates the value of θ according to a super-experiment updating rule. In a sense, this process is 'stacked' Q-learning; on the super-experiment level, the algorithm learns the optimal model; on sub-experiment level, the agent employs this 'learned' model to find the optimal strategy in the experiments themselves. I define an experiment as an auction sequence of 80,000 rounds.

I implement this by setting the sub-experiment updating rule to be:

$$Q_{t+1}(a_i, s_i) = (1-\alpha) \cdot (Q_t(a_i, s_i) + \alpha \cdot (\bar{r}_i + G(\theta_i) \cdot \rho \, \mathbb{E}(\max_{a'_i} Q(a'_i, s'_i | a_i, s_i). \tag{5}$$

$G(\theta_i)$ is effectively a weight on the discount factor of the expectation of future rewards and is given by the equation:

$$G(\theta_i) = \frac{0.1 + \theta_i \cdot min(1, \frac{\bar{r}_j^e + 0.1}{\bar{r}_i + 0.1})}{0.1 + \theta_i} \tag{6}$$

Where $\bar{r}_j^e$ is agent$_\text{i}$'s expectation of the mean reward of agent$_\text{i}$ in the last η_j rounds. The agent$_\text{i}$ maintains expectations of the agent$_\text{j}$ rewards by assuming that agent$_\text{j}$'s value is equivalent to its own private value. Because λ_i and λ_j are independent, this expectation is not a good one and becomes increasingly poor as γ increases. $\bar{r}_i$ is simply the agent$_\text{i}$'s mean reward in the last η_i rounds. G(θ_i) can be interpreted as the ratio of the opposing agent's recent welfare and to the agent's own welfare weighted by θ. If θ is equal to 0, G(θ) reduces to 1, and the updating rule reverts to Equation 2. The motivation for this specification of G(θ) is this: if the agent$_\text{j}$ is consistently and significantly underperforming agent$_\text{i}$, agent$_\text{i}$ should rationally expect agent$_\text{j}$ to alter their behavior in the near future and disrupt the beneficial equilibrium. In such a state, agent$_\text{i}$ discounts the value of future rewards at a greater rate.

θ has no impact on the actual rewards each agent realizes in the auctions; each agent's goal of maximizing the total rewards across the auction sequence remains unchanged. Under this specification, agent$_\text{i}$ does not incorporate agent$_\text{j}$'s welfare into its own, rather, G(θ) allows each agent$_\text{j}$ the flexibility to include or exclude an a priori expectation of its opponent's behavior in its own model. The process through which each agent chooses θ is just the standard Q-learning process updating on a per experiment basis rather than a per round basis. The super-experiment updating rule is just Equation 1.

6.2 Simulation Parameters

The sub-experiment Q-learning parameters are the same as in Section 9. The super-experiment parameters, $\theta_{i,j}$, η_i, η_j are initialized to 0, 3, and 10 respectively. The super-experiment action-space is restricted to [0,1]; that is θ is restricted to the values of 0 or 1[6]. Again, when $\theta = 0$, the sub-experiment Q updating rule reduces to the standard updating rule; when $\theta = 1$ the agent follows the discounting rule described in the previous section. Because the super-experiment action space is small, the agent should be able to learn at a much quicker rate; in this context, I set α^{super} to be 0.5, β^{super} to be $2 \cdot 10^{-2}$, and γ^{super} to be 0.9.

7 Results

7.1 Collusion

As seen in Figure 3, the agents quickly and uniformly converge to $\theta_{i,j} = 1$

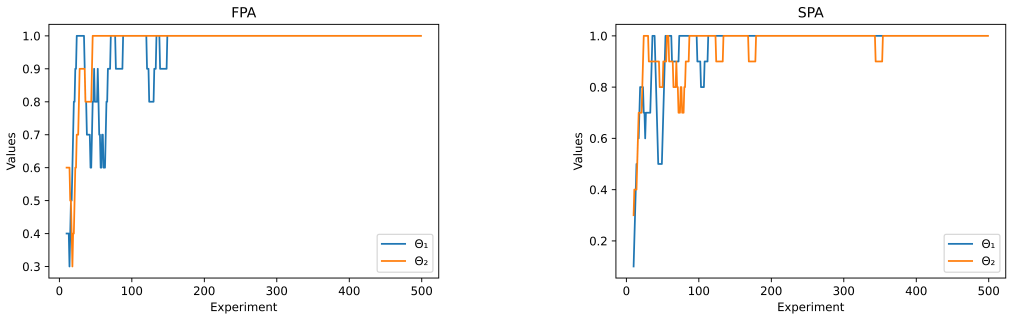

Fig. 3: *Super-Experiment Learning Process of Optimal* $\theta_{i,j}$

Note: The agents learn the dominant strategy of $\theta_{i,j}$ extremely quickly.

Following convergence of $\theta_{i,j}$, collusion, the agents collude on non-competitive bids in nearly all experiments Figure 4. This is a remarkable result. Algorithms with no inherent bias towards collusion autonomously learn to specify

[6] Again, these results are robust to larger action spaces.

their own learning model in such a way that allows for the development and maintenance of near-perfect collusion. This collusion in both FPA and SPA is remarkably effective; in equilibrium, the agents jointly extract almost all possible profit from the auction: 96.8% in SPA and 99.9% in FPA.

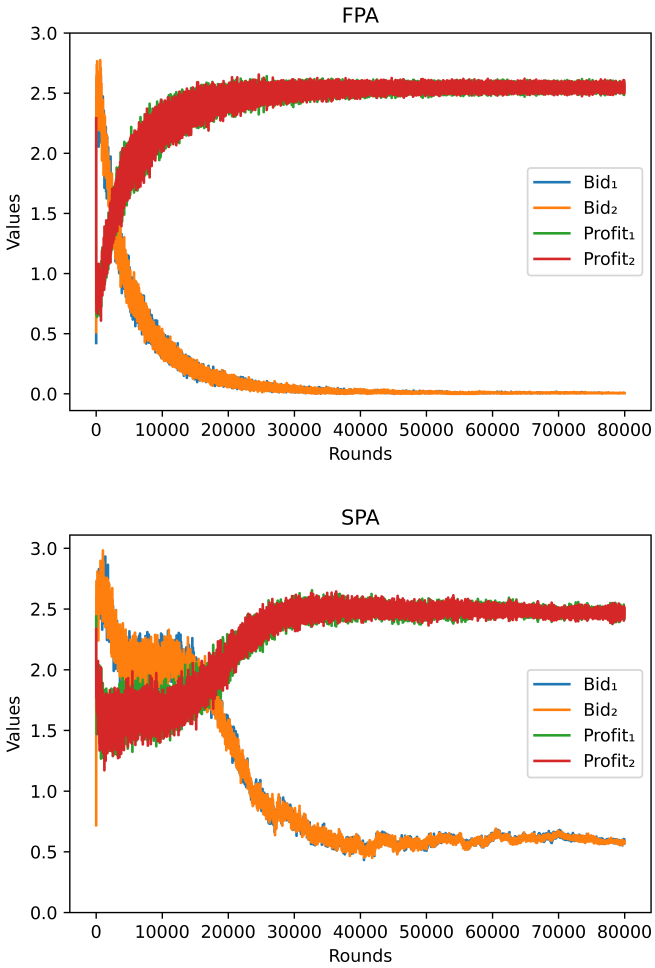

Fig. 4: *Smoothed Mean Bids and Rewards In Stable*[7] *Sub-Experiments*
Note: The agents learn to overcome the initial competitive pressure, converging on near optimal collusion.

An interesting result of this simulation is that agents converge to the perfectly collusive bid more quickly and slightly more often in FPA than in SPA. This is evidence for the hypothesis that the action-reward disconnect discussed in Section 5, which is present to a greater degree in SPA, hampers the ability

	$\overline{Bid_1}$	$\overline{Bid_2}$	$\overline{Profit_1}$	$\overline{Profit_2}$
Mean	0.005205	0.005200	2.547894	2.546527
Standard Deviation	0.049860	0.049908	0.027527	0.027152

Table 1: *Distribution of Mean Equilibrium Bids and Rewards in Stable FPA Sub-Experiments*

	$\overline{Bid_1}$	$\overline{Bid_2}$	$\overline{Profit_1}$	$\overline{Profit_2}$
Mean	0.578530	0.574975	2.466880	2.468617
Standard Deviation	0.722261	0.723893	0.168290	0.172625

Table 2: *Distribution of Mean Equilibrium Bids and Rewards in Stable SPA Sub-Experiments*
Note: Collusive profit is slightly lower under SPA but is remarkably high in both environments. Relatively greater difference in mean bids explained by bid rotation.

of the reinforcement learning algorithm to learn collusive strategies. Specifically, the action that an agent chooses in a given period has no immediate impact on their payoff in the vast majority of states. This limits the actionable information available to the agent to comprehend its environment and effectively estimate its Q-matrix, necessary conditions for the evolution of non-competitive bidding strategies. Here, the material effect of this phenomenon is small as the two-level Q-learning algorithm is able to effectively overcome such a learning barrier; the agents learn the collusive equilibrium extremely well in either environment.

7.2 Bid Rotation

A more salient result seen Figure 4 is the difference in the equilibrium mean bids in FPA and SPA. While in FPA, the agents converge to an equilibrium bid of nearly zero, the mean of equilibrium bids in SPA experiments tends to be much closer to 0.5. Only a small fraction of this difference can be attributed to imperfect collusion, instead, it indicates the presence of a much more profound result: bid rotation. Bid rotation is a hallmark of auction collusion and is the

lynchpin of bid-rigging when agents cannot effectively transfer collusive profits. In stable SPA experiments, the agents converge on a bid rotation scheme 64% of the time. A snapshot of equilibrium bid rotation in a single experiment can be seen in Figure 5.

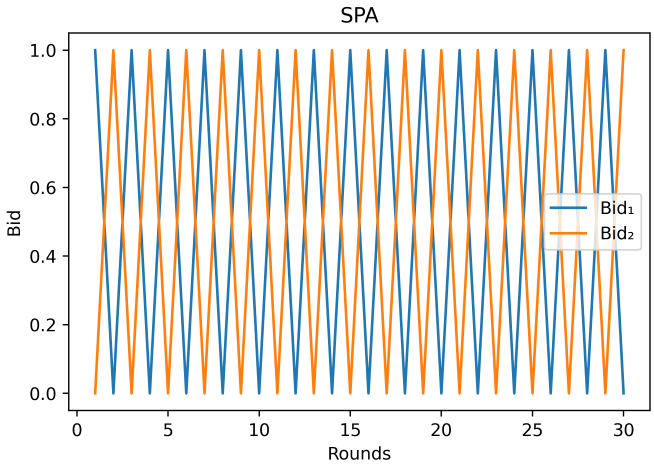

Fig. 5: *Equilibrium Bids of Sub-Experiment#400, Beginning in Round#79970*
Note: Regular rotation of shill bid results in even distribution of maximal collusive profit.

Equilibrium bid rotation is observed only in SPA. This makes sense; in FPA, any collusive outcome that involves non-zero bids sacrifices potential profit. In SPA, however, rotation between the bids of 0 and w, where w is any non-zero bid, realizes all collusive profit. The preference for bid rotation over collusion on the same bid across auctions can be explained by the desire of agents to avoid the variability in profit introduced by the random tiebreak rule. By establishing a deterministic rotation strategy, the agents ensure a uniform distribution of collusive profits. It is a remarkable and somewhat unprecedented result that the use off such a simple and unsophisticated algorithm as two-level Q-learning can result in such complex collusive behavior as optimal bid rotation.

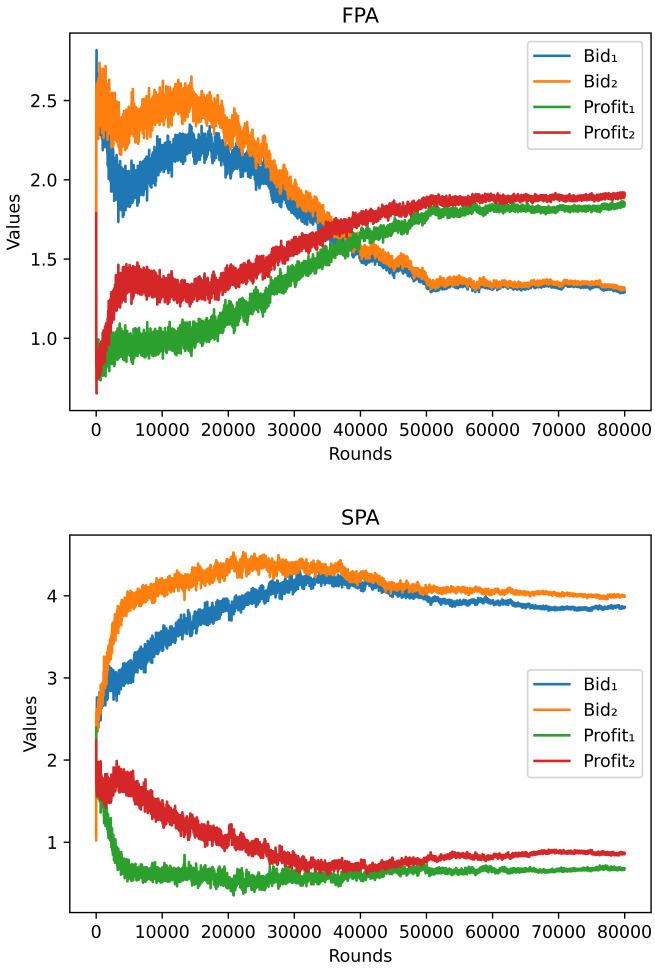

Fig. 6: *Bids and Rewards in Single Experiment*
Note: The adoption $\theta = 1$ by either agent induces both agents to bid somewhat less competitively. Resultant equilibrium is inferior collusion which stochastically dominates the competitive equilibrium.

7.3 Nash Equilibrium in Super-Experiment Learning

The rapidity of convergence of both agents to $\theta = 1$ suggests the existence of a Nash dominant strategy. The choice of $\theta = 1$ is, in fact, a dominant strategy. When both agents chose $\theta = 0$, the results are the same as those seen in Figure 1 - the profit for both agents is low. When both agents chose $\theta = 1$, highly profitable collusion, seen in Figure 4, is the result. When agent_i chooses $\theta = 1$ while the agent_j chooses $\theta = 0$, the result is a semi-collusive result that improves both agent's welfare over the $\theta_{i,j} = 0$ state. The mean results of 300

experiments where $\theta_{1,2} = (0, 1)$ can be seen in Figure 7. In both FPA and SPA, the adoption of $\theta_1 = 1$ leads agent_1 to bid in such a way that induces agent_2 to bid in a less aggressively. This results in a imperfect collusive result in which both agents are strictly better off but agent_2 extracts just slightly more collusive profit that agent_1. Numerically, the super-experiment game is characterized by the profit matrix in Table 3.

		θ_2 = 0	θ_2 = 1
θ_1	0	$(0.55, 0.55)$	$(1.902, 1.845)$
	1	$(1.845, 1.902,)$	$(2.546, 2.546)$

(a) FPA

		θ_2 = 0	θ_2 = 1
θ_1	0	$(0.05, 0.05)$	$(0.864, 0.672)$
	1	$(0.672, 0.864)$	$(2.466, 2.468)$

(b) SPA

Table 3: *Mean Equilibrium Profit Matrix for Choices of* $\theta_{i,j}$

The dominant strategy is clearly $\theta = 1$. The fact that the super-experiment game is straightforward does not make the result of collusion less interesting. Without any prior understanding of or inclination towards collusion, the agents organically learn the pair of sub-game and super-game strategies that results in collusion. The simplicity of the learning algorithm is important; algorithms of already greater and ever-increasing complexity than the two-level Q-learning algorithm dominate algorithmic bidding. These results - that increased sophistication is conducive to collusion and that even two-level Q-learning is strongly collusive - suggest that algorithmic collusion in high-frequency auctions with repeated contact is eminently possible.

7.4 Anatomy of Sub-Experiment Collusion: G(θ)

Some insight into the process of collusion can be seen in the evolution of G(θ) as defined in Equation 6. G(θ) can be interpreted as a proxy for the stability of an equilibrium as it effectively represents each agent's expectation of the opposing agent's incentive to deviate from the current state.

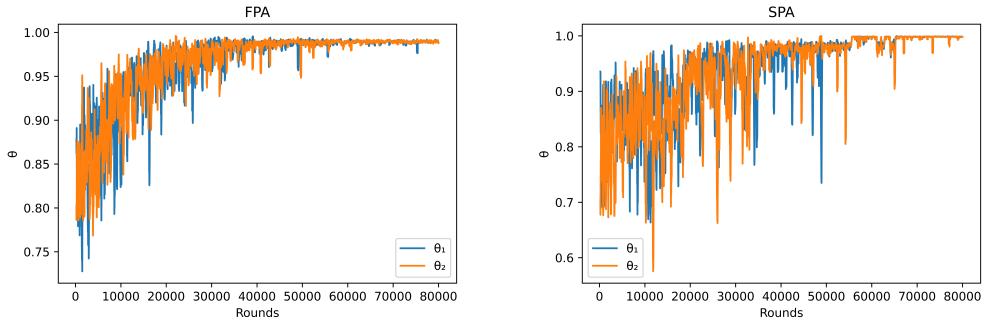

Fig. 7: *The Convergence of G(θ_i) to 0.98 and 1 respectively in Experiment#150.*

This nicely models the evolution of collusion. Early in the simulation, the agents explore the environment, far from a stable collusive strategy, as represented by the low and variable G(θ_i). As the agents begin to converge on collusion, G(θ_i) trends upward, but periodic deviations by either agent destabilizes the equilibrium, sending G(θ_i) plunging downward. Increasingly effective re-coordination pushes G(θ_i) to its upper bound in both environments: 0.98 in FPA and 1 SPA. Finally, as experimentation and deviation die out entirely, the agents converge on a stable collusive strategy, maximizing G(θ_i). The ultimate collusive state is more stable in SPA due to the presence of bid rotation, which effectively removes the destabilizing effect of tiebreaks. Collusion in FPA is maximally effective given environment constraints as the agents converge to $b_{i,j} = 0$, but the variance of the tiebreaks introduces slight instability into the equilibrium.

8 Robustness

In order to demonstrate that the action space restrictions used throughout the paper are immaterial to my results, I re-run the simulation, expanding the action space from [0,5] to [0,19]. As shown in Figure 8 and Figure 9, the results of robust competition under standard Q-learning and collusion under two-level Q-learning remain unchanged.

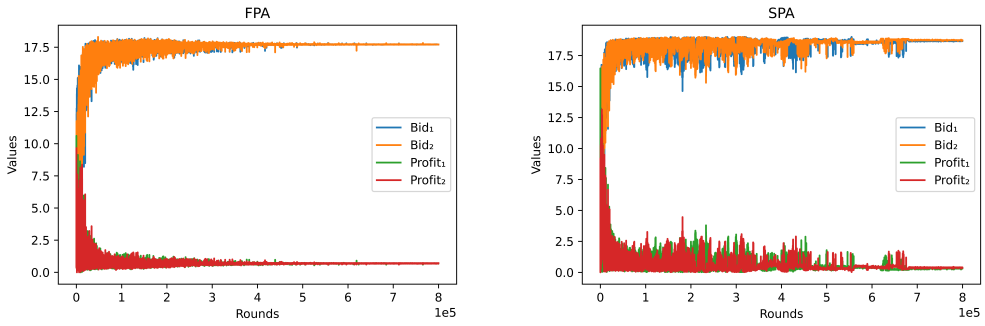

Fig. 8: *Convergent Outcome of Basic Q-learning in Larger Action Space*
Note: One-step bid shading and honest value reporting remain dominant strategies in SPA and FPA respectively.

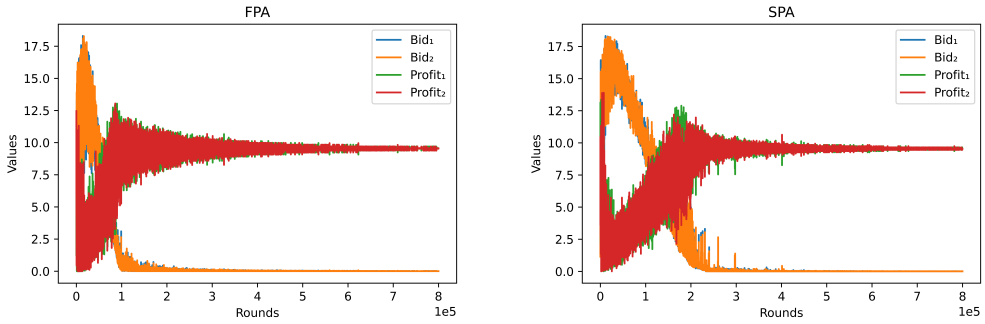

Fig. 9: *Convergent Outcome of Two-Level Q-learning in Larger Action Space*
Note: Uniform collusion is likewise present in the larger action space.

Of more interest is the sensitivity of collusion to changes in the discounting rate of future rewards. Figure 10 plots the percent of collusive profit extracted in following the convergence of two-level Q-learning against the discount factor, ρ used in that iteration of the simulation. Rerunning the simulation repeated with small decrements in the ρ each iteration demonstrated that the integrity

of the collusive outcome is dependent on cartel members highly weighting future profits. This is a strong empirical confirmation of the theory outlined in (Skrzypacz and Hopenhayn, 2004) and (Zhang, 2021).

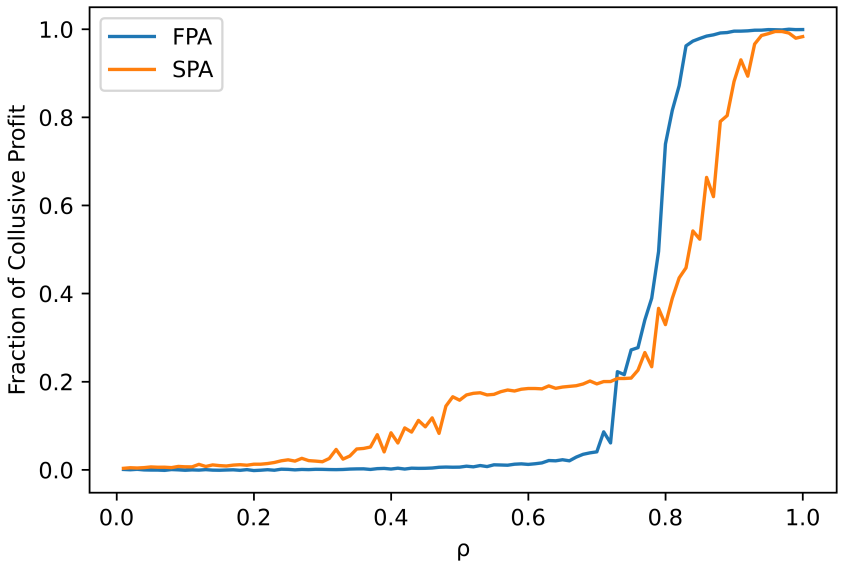

Fig. 10: *Fraction of Collusive Profit Extracted Across Levels of* ρ
Note: Low ρ destabilizes collusion; integrity of cartel depends on $\rho > 0.9$.

The intuition behind this result lies in the agent's sub-game incentives. If the agent significantly discounts future profit in favor of present gains, it will be less willing to forgo the immediate profits realized through defection from the cartel (i.e., bidding above the collusive bid). This incentive increases the likelihood of deviation, which destabilizes collusion. However, as the agent weights future profits more and more heavily ($\rho \to 1$), the defection incentive is outweighed by the prospect of overall greater future profits from stable collusion. The fact that this result conforms well to theory corroborates the legitimacy of my findings.

9 Discussion and Limitations

I discovered several important insights in this paper, each of which contributes to our understanding of algorithmic collusion in auctions. Firstly, by simulating the bidding behavior of non-communicating, profit-maximizing agents in stylized SA auctions, I found that the workhorse model of current algorithmic collusion research, Q-learning, is unable to reach non-competitive strategies in first and second price auctions. This indicates that algorithmic bid-rigging is a fundamentally more difficult problem than algorithmic price fixing. A potential rationale for this result is that certain inherent characteristics of indivisible-good auctions, including asymmetric information, tiebreak variance, and discontinuity in rewards, hinder the development of collusion. This finding, if robust to alternative algorithm designs, should help inform how resources are allocated in dealing with the potential emergence of algorithmic collusion in various market structures.

The next important result came from the implementation of two-level Q-learning, a more flexible and sophisticated Q-learning mutant. This algorithm, armed with an extra dimension of information, developed highly effective and remarkably complex collusion in both FPA and SPA. The overarching result, that a slight increase in algorithmic sophistication helps facilitate collusion, is a cause for concern. Incredibly sophisticated bidding algorithms are widely available; if sophistication is a driver of collusion as this result indicates, it would seem that algorithmic bid-rigging is a tangible threat. Nested within this collusion was the development of algorithmic bid rotation. Both exciting and worrisome, this unprecedented result speaks to the ability of simple algorithms to find and replicate established collusive schemes. Additionally, the robustness analysis, indicating a higher discount factor facilitates collusion, is important

in that it validates my results and empirically verifies a key folk theorem of repeated games.

This paper is very much a first step in the exploration of algorithmic bid rigging; there exists a rich array of possibilities to continue this research. The simplifying assumptions which allowed for a more straightforward analysis leave ambiguous the generality of my results in a less stylized environment. A common criticism applied to research into algorithmic collusion is that simplified models have limited practical bearing on reality. The same criticism can potentially be applied to this paper; while I attempted to capture the critical elements of SA auctions in my simulation, I leave it to future research to more rigorously demonstrate the viability of these results in real-world auctions.

There are a few concrete steps to be taken to enrich this analysis. Again, a more in-depth robustness analysis is critical to understand how well these results generalize to reality; an obvious next step is to relax certain model restrictions, such as the size and stability of the bidding pool and/or symmetry in agents, and assess the impact on these results. It would be a worthwhile pursuit to use this framework to empirically investigate our theoretical understanding collusion; for example, one could introduce multi-market contact between agents and examine the implications on speed and efficacy of collusion. Another step in connecting these results to the real world is to examine how different specifications of learning algorithms impact our findings; pitting an agent utilizing a deep-learning algorithm against another agent using a standard or two-level Q-learning could yield insightful results. Alternatively, one could restrict the exploration of algorithmic specification to two-level Q-learning, endogenizing alternate model parameters to more fully capture the learning possibilities of two-level Q-learning. All told, there is much to do; this paper hopefully provides a foundation on which further research can build.

References

Assad S, Clark R, Ershov D, et al (2020) Algorithmic pricing and competition: Empirical evidence from the german retail gasoline market. CESifo Working Paper Series 8521

Axelrod R, Hamilton WD (1981) The evolution of cooperation. Science 211(4489):1390–1396. https://doi.org/10.1126/science.7466396, URL https://www.science.org/doi/abs/10.1126/science.7466396, https://arxiv.org/abs/https://www.science.org/doi/pdf/10.1126/science.7466396

Banchio M, Skrzypacz A (2022) Artificial intelligence and auction design URL https://arxiv.org/pdf/2202.05947.pdf

Calvano E, Calzolari G, Denicolò V, et al (2020) Artificial intelligence, algorithmic pricing, and collusion. American Economic Review 110(10):3267–3297

Capobianco A (2020) Algorithms and collusion, volume = 127, pages = 10, journal = Organisation for Economic Co-operation and Development Competition Committee,

Daskalakis C, Goldberg P, Papadimitriou C (2009) The complexity of computing a nash equilibrium. SIAM J Comput 39:195–259. https://doi.org/10.1137/070699652

Hansen K, Misra K, Pai M (2021) Frontiers: Algorithmic collusion: Supra-competitive prices via independent algorithms. Market Science 40(4489):1–12. URL https://doi.org/10.1287/mksc.2020.1276

Hu J, Wellman M (2003) Nash q-learning for general-sum stochastic games. Journal of Machine Learning Research 4:1039–1069. https://doi.org/10.

1162/1532443041827880

Kimbrough SO, Murphy FH (2008) Learning to collude tacitly on production levels by oligopolistic agents. Computational Economics 33(1):47. https://doi.org/10.1007/s10614-008-9150-6, URL https://doi.org/10.1007/s10614-008-9150-6

McAfee RP, McMillan J (1992) Bidding rings. The American Economic Review 82(3):579–599. URL http://www.jstor.org/stable/2117323

Pesendorfer M (2000) A Study of Collusion in First-Price Auctions. The Review of Economic Studies 67(3):381–411. https://doi.org/10.1111/1467-937X.00136, URL https://doi.org/10.1111/1467-937X.00136, https://arxiv.org/abs/https://academic.oup.com/restud/article-pdf/67/3/381/4421986/67-3-381.pdf

Porter R, Zona JD (1993) Detection of bid rigging in procurement auctions. Journal of Political Economy 101(3):518–538. URL http://www.jstor.org/stable/2138774

Skrzypacz A, Hopenhayn H (2004) Tacit collusion in repeated auctions. Journal of Economic Theory 114(1):153–169. https://doi.org/https://doi.org/10.1016/S0022-0531(03)00128-5, URL https://www.sciencedirect.com/science/article/pii/S0022053103001285

Watkins CJCH, Dayan P (1992) Q-learning. Machine Learning 8(3):279–292

Zhang W (2021) Collusion enforcement in repeated first-price auctions URL https://econtheory.org/ojs/index.php/te/article/viewForthcomingFile/4640/31982/1