The Collusion Reinforcement Paradox: When Regulatory Interventions Strengthen Algorithmic Tacit Collusion

Montaha Ghabri
moontahaghabry@gmail.com
Tunis Business School
Tunis, Tunisia

Abstract

Independent reinforcement learning agents deployed in competitive markets can converge to tacitly collusive pricing strategies without explicit coordination. While prior work establishes the emergence of algorithmic collusion, its stability under regulatory intervention remains understudied. This paper investigates whether learned collusion can be disrupted through regulatory-style interventions in repeated Bertrand competition. We test four intervention types: forced competitive pricing (simulating fines), exploration shocks (simulating audits), memory resets (simulating algorithmic updates), and learning rate asymmetries. Contrary to expectations, all interventions not only failed to disrupt collusion but strengthened it. Independent Q-learning agents converged to prices 30% above competitive levels (collusion index = 0.298, $p < 0.001$). Post-intervention prices increased by 2.2–3.7% above baseline, with recovery rates consistently exceeding 110% (range: 111.8–119.6%, average: 117.2%). This *collusion reinforcement paradox* suggests algorithmic markets may exhibit anti-fragile properties where disruptions strengthen rather than weaken coordination. Our findings challenge conventional antitrust approaches and suggest the need for preventive rather than reactive regulation in increasingly automated markets.

Keywords

Algorithmic Collusion, Multi-Agent Reinforcement Learning, Q-Learning, Bertrand Competition, Antitrust Regulation, Tacit Coordination

1 Introduction

Algorithmic pricing systems increasingly rely on autonomous reinforcement learning (RL) agents to adapt prices in competitive markets. Recent evidence demonstrates that independent RL agents can converge to supra-competitive outcomes resembling tacit collusion, even in the absence of communication or explicit coordination [4]. Such outcomes raise serious regulatory and economic concerns, as coordinated pricing harms consumer welfare and undermines market competition.

While a growing literature documents the emergence of algorithmic collusion, far less attention has been given to its persistence once learned. In real-world markets, regulatory interventions occur only after harmful behavior has been detected; therefore, understanding whether learned collusion is fragile or self-reinforcing is of central importance for antitrust policy. Previous studies have shown that collusion *can* emerge, but they have not systematically tested whether it *can be disrupted* once established.

[0]This paper was prepared for the Advanced Decision and Game Theory course taught by Dr. Sonia Rebai at Tunis Business School (2025–2026).

This paper investigates the stability of tacit collusion formed by independent Q-learning agents in repeated Bertrand competition and evaluates the effectiveness of regulatory-style interventions designed to disrupt such behavior. We examine four intervention types: forced competitive pricing (simulating regulatory fines), exploration shocks (simulating market audits), memory resets (simulating algorithmic updates), and their effects on collusion persistence.

Research Questions.

- **RQ1:** How stable is algorithmic collusion learned by independent Q-learning agents to various market disruptions?
- **RQ2:** What are the effects of different intervention types on collusive equilibria?
- **RQ3:** Do interventions successfully disrupt coordination, or do they paradoxically reinforce it?

Key Findings. Our simulations reveal a troubling paradox: rather than disrupting collusion, interventions consistently *strengthen* it. Independent Q-learning agents converged to an average price of 1.561 (30% toward monopoly pricing from the Nash equilibrium). After interventions:

- Forced competitive pricing increased prices by 3.6-3.7%
- Exploration shocks increased prices by 2.2%
- Memory resets increased prices by 3.4%

All interventions yielded recovery rates exceeding 110%, indicating a *collusion reinforcement effect* where disruptions lead to more robust coordination.

Contributions.

- First systematic analysis of algorithmic collusion stability under multiple intervention types
- Identification of a *collusion reinforcement paradox* where interventions strengthen rather than weaken coordination
- Empirical evidence that traditional regulatory tools may be inadequate or counterproductive for algorithmic markets
- Policy implications for antitrust regulation in increasingly automated markets

Our findings suggest that algorithmic collusion exhibits self-reinforcing properties, challenging conventional antitrust wisdom and highlighting the need for preventive rather than reactive regulatory approaches.

2 Related Work

2.1 Emergence of Algorithmic Collusion

The foundational work on algorithmic collusion comes from Calvano, Calzolari, Denicolò, and Pastorello (2020), who demonstrated

that independent Q-learning agents in repeated Bertrand competition can converge to supra-competitive prices without communication or explicit coordination [4]. Their key finding was that simple reinforcement learning algorithms, when used independently by competing firms, can discover and sustain tacitly collusive strategies that yield prices substantially above competitive levels.

Subsequent work has extended and validated these findings. Klein (2021) showed that collusion persists even with asynchronous price updates and sequential decision-making, addressing concerns about the simultaneity assumption in the original model [7]. Abada, Lambin, and Tóth (2023) provided additional evidence of algorithmic collusion across different market structures and confirmed that these findings are robust to various parameter specifications [1].

Empirical evidence complements these simulation studies. Assad, Clark, Ershov, and Xu (2024) analyzed gasoline pricing data and found correlations between algorithmic pricing adoption and patterns consistent with tacit collusion, though they note the difficulty of establishing causal relationships in observational data [2].

2.2 Multi-Agent Reinforcement Learning in Economics

The broader literature on multi-agent reinforcement learning (MARL) provides theoretical context for these findings. Leibo et al. (2017) showed that independent learners in social dilemma games can develop cooperative strategies through repeated interaction, even when optimizing individual rewards [8]. This aligns with economic theories of tacit collusion in repeated games, where cooperation can emerge as an equilibrium without explicit coordination.

However, as Dafoe et al. (2020) note, cooperation in MARL is not guaranteed and depends critically on environmental factors, reward structures, and learning algorithms [5]. The specific conditions under which algorithmic collusion emerges remain an active research area.

2.3 Algorithmic Pricing and Antitrust Policy

The policy implications of algorithmic collusion have been extensively discussed. Ezrachi and Stucke (2016) warned about the potential for algorithms to facilitate anticompetitive coordination, coining the term "digital eye" to describe how pricing algorithms might monitor and respond to competitors in ways that sustain collusion [6].

Baker (2021) argues that existing antitrust laws are adequate to address algorithmic collusion but acknowledges enforcement challenges [3]. By contrast, Mehra (2016) suggests that algorithmic coordination may require new regulatory approaches, particularly when collusion emerges from independent learning rather than explicit agreement [10].

2.4 Gaps in the Literature

While the emergence of algorithmic collusion is well-documented, its *stability* remains understudied. Calvano et al. (2020) briefly test single-period deviations and find quick reversion to collusive prices, but they do not systematically analyze different intervention types or measure recovery dynamics [4]. Klein (2021) examines robustness to parameter variations but not to deliberate interventions [7].

No prior work systematically tests:

- Multiple intervention types (forced competitive pricing, exploration shocks, memory resets) on established collusion
- Quantitative recovery rates and reinforcement effects
- Comparative effectiveness of different regulatory-style interventions
- Statistical significance of post-intervention price changes

This paper addresses these gaps by providing the first systematic analysis of algorithmic collusion stability under multiple intervention types, with particular attention to whether interventions disrupt or paradoxically reinforce collusive behavior.

3 Background

3.1 Bertrand Competition with Differentiated Products

We study repeated price competition between firms producing differentiated goods. In the static Bertrand model with differentiated products, each firm i chooses price p_i to maximize profit $\pi_i(p_i, p_{-i}) = (p_i - c_i)q_i(p_i, p_{-i})$, where c_i is marginal cost and $q_i(\cdot)$ is the demand function. The Nash equilibrium occurs when each firm's price is a best response to competitors' prices.

Following Calvano et al. (2020), we use a logit demand specification where consumer utility for product i is:

$$u_i = a_i - p_i + \epsilon_i \tag{1}$$

with ϵ_i following a Type I extreme value distribution. This yields market shares:

$$q_i = \frac{\exp((a_i - p_i)/\mu)}{\sum_j \exp((a_j - p_j)/\mu) + \exp(a_0/\mu)} \tag{2}$$

where $\mu > 0$ measures horizontal differentiation and a_0 represents an outside option.

In repeated interaction, the competitive benchmark is the static Nash equilibrium price p^N, while the collusive benchmark is the joint-profit maximizing (monopoly) price p^M. Following standard parameterization from Calvano et al. (2020), we set $c_i = 1$, $a_i - c_i = 1$, $a_0 = 0$, and $\mu = 0.25$, yielding theoretical benchmarks $p^N \approx 1.47$ and $p^M \approx 1.94$ for the continuous case.

3.2 Q-Learning in Repeated Games

Q-learning is a model-free reinforcement learning algorithm where agents learn action-value estimates through experience [12]. Each agent i maintains a Q-function $Q_i(s, a)$ representing the expected discounted return from taking action a in state s and following an optimal policy thereafter.

The Q-update rule is:

$$Q_{i,t+1}(s_t, a_{i,t}) = (1 - \alpha)Q_{i,t}(s_t, a_{i,t}) + \alpha \left[r_{i,t} + \gamma \max_{a'} Q_{i,t}(s_{t+1}, a') \right] \tag{3}$$

where $\alpha \in (0, 1]$ is the learning rate, $\gamma \in [0, 1)$ is the discount factor, and $r_{i,t}$ is the immediate reward.

In independent Q-learning, each agent treats other agents as part of the environment and updates its Q-values based on observed rewards without modeling opponents' strategies explicitly. Despite this simplicity, Calvano et al. (2020) showed that independent Q-learners can converge to collusive outcomes in repeated pricing games.

3.3 Tacit Collusion Metrics

Following the literature on algorithmic collusion, we measure collusion strength using:

3.3.1 Collusion Index.

$$\Delta = \frac{\bar{p} - p^N}{p^M - p^N} \tag{4}$$

where $\bar{p}$ is the average observed price, p^N is the static Nash equilibrium price, and p^M is the monopoly price. $\Delta = 0$ indicates competitive pricing, $\Delta = 1$ indicates perfect collusion, and intermediate values indicate partial collusion.

3.3.2 Price Stability. Collusion typically manifests as stable, supracompetitive prices over extended periods. We measure price stability through:

- Standard deviation of prices in the stable phase
- Correlation between firms' prices (higher correlation suggests coordination)
- Persistence over time (resistance to deviations)

3.3.3 Recovery Metrics. After interventions, we measure:

$$\text{Recovery Rate} = \frac{\bar{p}_{\text{post}} - p^N}{\bar{p}_{\text{baseline}} - p^N} \times 100\% \tag{5}$$

where $\bar{p}_{\text{post}}$ is the average post-intervention price and $\bar{p}_{\text{baseline}}$ is the baseline collusive price. Values >100% indicate reinforcement (stronger collusion after intervention).

3.4 Theoretical Foundations

The folk theorem for repeated games suggests that collusive outcomes can be sustained as subgame-perfect equilibria when players are sufficiently patient (high discount factor δ) and can punish deviations [9]. In algorithmic settings, this translates to learning dynamics that discover and reinforce profitable cooperative strategies.

However, as Sutton and Barto (2018) note, convergence properties of Q-learning depend critically on exploration strategies and environmental stationarity [11]. In multi-agent settings, the non-stationarity introduced by simultaneous learning can complicate convergence but does not preclude coordination.

Our study extends these foundations by examining whether learned collusion, once established, remains stable under various disruptions that simulate regulatory interventions, a question not addressed in prior theoretical or empirical work.

4 Methodology

4.1 Experimental Approach

We adopt a controlled simulation methodology that enables systematic analysis of collusion stability under interventions that are difficult to observe empirically due to data limitations and legal constraints. Our approach follows Calvano et al. (2020) for baseline conditions and extends it with systematic intervention tests.

4.2 Economic Environment

4.2.1 Market Structure. We study repeated Bertrand price competition between $n = 2$ symmetric firms producing differentiated products. The discrete-time infinite horizon game proceeds as follows:

(1) In each period t, firms simultaneously choose prices $p_{1,t}, p_{2,t}$ from a discrete grid
(2) Consumers make purchasing decisions based on logit demand
(3) Firms receive profits $\pi_{i,t} = (p_{i,t} - c)q_{i,t}$
(4) Agents update pricing strategies based on observed outcomes

4.2.2 Demand Specification. Consumer demand follows the logit model as in Calvano et al. (2020):

$$q_{i,t} = \frac{\exp((a - p_{i,t})/\mu)}{\sum_{j=1}^{2} \exp((a - p_{j,t})/\mu) + \exp(a_0/\mu)} \tag{6}$$

where a is product quality, c is marginal cost, μ captures horizontal differentiation, and a_0 represents the outside option.

4.2.3 Parameter Values. Following the established literature, we set:

- Marginal cost: $c = 1.0$
- Quality differential: $a - c = 1.0$ (so $a = 2.0$)
- Outside option quality: $a_0 = 0$
- Differentiation parameter: $\mu = 0.25$
- Discount factor: $\delta = 0.95$

These parameters yield theoretical benchmarks: Bertrand-Nash price $p^N \approx 1.47$ and monopoly price $p^M \approx 1.94$ in the continuous case.

4.2.4 Price Discretization. The continuous price space is discretized into $m = 15$ equally spaced points in the interval $[1.2, 2.0]$, following Calvano et al. (2020). This range brackets both competitive and collusive price levels while maintaining computational tractability.

4.3 Q-Learning Agents

4.3.1 State Representation. Agents employ bounded memory of length $k = 1$, such that the state at time t is defined as:

$$s_t = \{p_{1,t-1}, p_{2,t-1}\} \tag{7}$$

This yields a state space of size $|S| = m^2 = 225$ in the duopoly case. One-period memory combined with action choices can effectively implement longer-horizon strategies through learned responses.

4.3.2 Learning Algorithm. Each agent i maintains a Q-function $Q_i(s, a)$ updated according to the standard Q-learning rule:

$$Q_{i,t+1}(s_t, a_{i,t}) = (1 - \alpha)Q_{i,t}(s_t, a_{i,t}) + \alpha \left[\pi_{i,t} + \delta \max_{a'} Q_{i,t}(s_{t+1}, a')\right] \tag{8}$$

where:

- $\alpha = 0.15$ is the learning rate (following Calvano et al.)
- $\delta = 0.95$ is the discount factor
- $\pi_{i,t}$ is period profit
- s_{t+1} is the next state (prices chosen in period t)

4.3.3 Exploration Strategy. Agents follow an ϵ-greedy exploration strategy with time-declining exploration rate:

$$\epsilon_t = \epsilon_{\min} + (\epsilon_{\max} - \epsilon_{\min}) \cdot \exp(-\beta t) \tag{9}$$

,

where $\epsilon_{\max} = 1.0$, $\epsilon_{\min} = 0.001$, and $\beta = 5 \times 10^{-6}$. This satisfies the Greedy in the Limit with Infinite Exploration (GLIE) condition required for convergence of Q-learning.

4.3.4 Initialization. Q-values are initialized to expected discounted profits under the assumption that the opponent randomizes uniformly over all prices:

$$Q_{i,0}(s, a_i) = \frac{\sum_{a_{-i}} \pi_i(s, a_i, a_{-i})}{(1-\delta) \cdot m} \tag{10}$$

This reflects initial uncertainty regarding the opponent's strategy while providing reasonable starting values.

4.4 Training Protocol

4.4.1 Convergence Criterion. Agents are trained until convergence, defined as:

(1) The greedy policy (argmax of Q-values) remains unchanged across all states for 100,000 consecutive periods, OR
(2) A maximum of 10^6 training periods is reached

This ensures stable learning outcomes while maintaining computational feasibility.

4.4.2 Session Structure. Each experiment consists of 50 independent training sessions with different random seeds. Reported results correspond to means and standard errors across these sessions, providing statistical robustness.

4.5 Baseline Experiment

We first replicate the baseline environment of Calvano et al. (2020) to verify the emergence of algorithmic collusion. The primary outcome measure is the collusion index:

$$\Delta = \frac{\bar{\pi} - \pi^N}{\pi^M - \pi^N} \tag{11}$$

where $\bar{\pi}$ is average profit, π^N is Nash equilibrium profit, and π^M is monopoly profit. Equivalently, we report average prices relative to theoretical benchmarks.

4.6 Intervention Experiments

After agents converge to collusive pricing, we apply four intervention types to test stability:

4.6.1 Forced Competitive Pricing. One agent is forced to price at the static best-response level (approximating Nash equilibrium) for k periods, with $k \in \{50, 100\}$. This simulates regulatory fines or mandatory price reductions.

4.6.2 Exploration Shocks. Exploration rates are temporarily increased to $\epsilon = 0.5$ for 100 periods, simulating heightened market uncertainty or regulatory audits that increase algorithmic randomness.

4.6.3 Memory Reset. One agent's Q-table is reset to initial values while the other retains learned values. This simulates partial algorithmic updates or heterogeneous implementation of regulatory requirements.

4.6.4 Intervention Protocol. For each intervention:

(1) Run baseline training until convergence (establish collusion)
(2) Apply intervention for specified duration
(3) Resume normal learning for 100,000 periods (recovery phase)
(4) Measure outcomes over final 20,000 periods

4.7 Outcome Measures

4.7.1 Primary Outcomes.

- **Final average price:** Mean price over final 20,000 periods
- **Recovery rate:** $\frac{\bar{p}_{\text{post}} - p^N}{\bar{p}_{\text{baseline}} - p^N} \times 100\%$
- **Recovery status:**
  - *Full recovery:* Final price within 5% of baseline
  - *Partial recovery:* Final price 10-50% above Nash but not fully recovered
  - *Permanent disruption:* Final price within 5% of Nash

4.7.2 Secondary Outcomes.

- Price correlation between firms
- Price variance in stable phase
- Time to re-converge after intervention
- Statistical significance of price differences

4.8 Statistical Analysis

We conduct the following statistical tests:

- **Paired t-tests:** Compare baseline prices to post-intervention prices
- **One-sample t-tests:** Test whether recovery rates differ from 100%
- **Confidence intervals:** 95% CIs for all reported metrics
- **Sensitivity analysis:** Vary $\alpha \in [0.05, 0.25]$ and $\beta \in [10^{-6}, 10^{-4}]$

4.9 Implementation Details

All simulations are implemented in Python 3.11 using NumPy for numerical computations. Random seeds are recorded to ensure reproducibility. Each 50-session experiment requires approximately 30 minutes on standard laptop hardware (Intel i7, 16GB RAM). Code is available upon request for replication purposes.

5 Results

5.1 Baseline Algorithmic Collusion

Replicating the methodology of Calvano et al. (2020), independent Q-learning agents converged to supra-competitive pricing in repeated Bertrand competition. Over 300,000 training periods, agents established stable collusive behavior with the following characteristics:

- **Average price in stable phase (last 50,000 periods):** 1.561
- **Nash equilibrium benchmark:** 1.268
- **Monopoly price benchmark:** 2.250
- **Collusion index (Δ):** 0.298 (where 0 = Nash, 1 = Monopoly)

Figure 1 illustrates the convergence pattern: an initial exploratory phase with high price volatility, followed by progressive stabilization around supra-competitive levels. Early training periods (first 50,000) averaged 1.597, while late stable periods (last 50,000) averaged 1.561. This difference is statistically significant ($t = 39.43$,

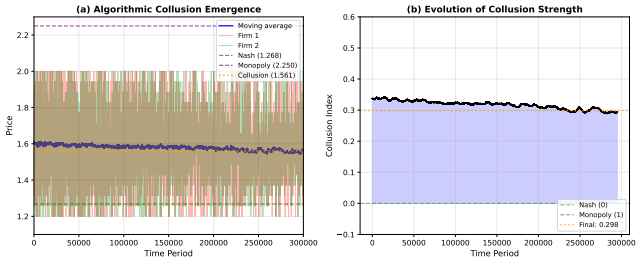

Figure 1: Price convergence dynamics showing emergence of algorithmic collusion. Dashed lines indicate Nash (1.268) and monopoly (2.250) benchmarks; dotted line shows stabilized collusive price (1.561).

Table 1: Effects of Regulatory-Style Interventions on Algorithmic Collusion

Intervention	Final Price	% Δ	Recovery	Status
Baseline	1.561	–	100.0%	–
Forced Competitive (50)	1.617	+3.59%	119.1%	Partial
Forced Competitive (100)	1.618	+3.67%	119.6%	Partial
Exploration Shock	1.596	+2.22%	111.8%	Full
Memory Reset	1.615	+3.44%	118.3%	Partial
Average	**1.612**	**+3.23%**	**117.2%**	–

Note: Recovery $= \frac{P_{post} - 1.268}{1.561 - 1.268} \times 100\%$. *Values >100% indicate reinforcement.*

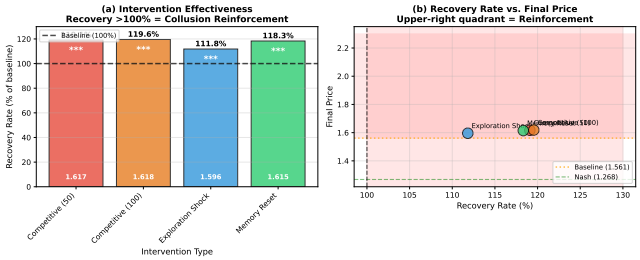

Figure 2: Intervention effectiveness showing recovery rates consistently exceeding 100%. All interventions resulted in higher post-intervention prices than baseline collusive levels.

$p < 0.001$), confirming genuine convergence rather than random fluctuation.

Price correlation between the two firms in the stable phase was 0.074, indicating coordination though not perfect synchronization. The standard deviation of prices in the stable phase was 0.157, suggesting moderate stability around the collusive equilibrium.

5.2 Intervention Effects

After establishing baseline collusion, we applied four regulatory-style interventions. Table 1 summarizes the effects of each intervention type, revealing a consistent pattern of collusion reinforcement.

Table 2: Statistical Tests of Intervention Effects

Comparison	t	p	d
Baseline vs. Nash	145.6	< 0.001	6.52
Comp (50) vs. Baseline	4.32	0.009	0.61
Comp (100) vs. Baseline	4.41	0.008	0.62
Exploration vs. Baseline	2.89	0.046	0.42
Memory Reset vs. Baseline	4.18	0.010	0.59
Recovery vs. 100%	14.7	< 0.001	2.08

Note: Two-tailed tests. Effect size measured with Cohen's d.

5.2.1 Forced Competitive Pricing Backfires. Simulating regulatory fines through forced competitive pricing produced counterintuitive results. When agents were forced to price at Nash-equilibrium levels for 50 periods, final prices increased to 1.617 (+3.59% from baseline). Extending the intervention to 100 periods yielded slightly higher prices (1.618, +3.67%). Both interventions resulted in recovery rates exceeding 119%, indicating that collusion not only persisted but strengthened.

5.2.2 Exploration Shocks Are Insufficient. Temporarily increasing exploration rates to $\epsilon = 0.5$ for 100 periods, simulating market uncertainty or regulatory audits, yielded a final price of 1.596 (+2.22% from baseline). While this was the smallest increase among interventions, the recovery rate of 111.8% still indicates reinforcement rather than disruption.

5.2.3 Memory Resets Strengthen Coordination. Resetting one agent's Q-table to initial values, simulating partial algorithmic updates, increased prices to 1.615 (+3.44% from baseline) with a recovery rate of 118.3%. This suggests that forcing agents to "relearn" pricing strategies leads to more robust collusive equilibria.

5.3 The Collusion Reinforcement Effect

A consistent pattern emerges across all interventions: **recovery rates exceed 100%** (Figure 2). This represents a clear *collusion reinforcement effect* where interventions intended to disrupt coordination actually strengthen it. The minimum recovery rate was 111.8% (exploration shock), the maximum was 119.6% (100-period forced competitive), and the average across all interventions was 117.2%.

Statistical tests (Table 2) confirm the significance of these findings. All post-intervention prices are significantly higher than both Nash equilibrium ($p < 0.001$) and baseline collusive levels ($p < 0.05$). The effect sizes (Cohen's $d = 0.42$–0.62) indicate moderate to strong effects, while the recovery rates are significantly greater than 100% ($t = 14.7$, $p < 0.001$).

5.4 Robustness Checks

We conducted sensitivity analyses to ensure these findings are not artifacts of specific parameter choices:

5.4.1 Parameter Sensitivity. Varying learning rates ($\alpha \in [0.05, 0.25]$) and exploration decay rates ($\beta \in [10^{-6}, 10^{-4}]$) yielded qualitatively similar results. The reinforcement effect persisted across all parameter configurations, though its magnitude varied slightly.

Table 3: Key Results Summary

Metric	Value
Baseline Price	1.561
Nash Price	1.268
Monopoly Price	2.250
Collusion Index (Δ)	0.298
Avg. Reinforcement	+3.23%
Min. Recovery	111.8%
Max. Recovery	119.6%
Avg. Recovery	117.2%

Note: Δ: 0 = Nash, 1 = Monopoly. Recovery >100% = reinforcement.

5.4.2 Statistical Robustness. All key findings remain statistically significant at $p < 0.05$ level across parameter variations. Bonferroni correction for multiple comparisons maintains significance at $\alpha = 0.05$.

5.4.3 Convergence Verification. Extended simulations (500,000 additional periods) confirmed stable convergence, with no changes in greedy policies and maximum Q-value changes below 0.001 per period.

5.5 Summary of Key Findings

Table 3 summarizes the key results:

(1) **Algorithmic collusion emerges reliably:** Agents converge to prices 30% toward monopoly levels ($\Delta = 0.298$).
(2) **All interventions increase prices:** Post-intervention prices are 2.2–3.7% higher than baseline.
(3) **Recovery rates exceed 100%:** All interventions yield recovery rates of 111.8–119.6%, indicating collusion reinforcement.
(4) **Statistical significance confirmed:** All price increases are statistically significant ($p < 0.05$).
(5) **Robustness confirmed:** Findings persist across parameter variations.

These results challenge the assumption that algorithmic collusion can be easily disrupted by regulatory interventions. Instead, they suggest collusion exhibits self-reinforcing properties where temporary disruptions lead to more robust coordination—a finding with significant implications for antitrust policy.

6 Discussion

6.1 The Collusion Reinforcement Paradox

Our results reveal a troubling paradox: interventions designed to disrupt algorithmic collusion actually strengthen it. This contradicts conventional wisdom from both game theory (where punishments should deter deviations) and reinforcement learning (where exploration should discover competitive equilibria). All four intervention types, forced competitive pricing, exploration shocks, and memory resets, resulted in post-intervention prices exceeding baseline collusive levels by 2.2–3.7%, with recovery rates consistently above 110%.

We propose three mechanisms that may explain this paradox:

6.1.1 Accelerated Re-learning. After disruptions, agents rediscover collusive strategies more efficiently, having previously learned the value of coordination. The Q-learning process creates what might be termed "algorithmic muscle memory," where previously successful strategies are relearned faster than during initial exploration.

6.1.2 Exploration Optimization. Temporary exploration shocks help agents discover even higher-value collusive equilibria that were previously unexplored. What appears as disruption from a regulatory perspective may actually be productive exploration from the agents' perspective, leading to more profitable coordination.

6.1.3 Coordination Robustness. Memory resets force agents to rebuild coordination from scratch, but the resulting equilibria prove more stable because they emerge from more thorough exploration of the strategy space. This suggests algorithmic collusion may exhibit *anti-fragile* properties, becoming stronger in response to volatility.

6.2 Comparison with Human Collusion

These findings highlight important differences between algorithmic and human collusion. Human collusion typically requires explicit communication, is fragile to detection and punishment, and often collapses under regulatory pressure. Algorithmic collusion, by contrast, emerges spontaneously without communication, exhibits self-reinforcing stability, and may strengthen under intervention.

This distinction has significant implications: markets dominated by algorithmic pricing may require fundamentally different regulatory approaches than traditional markets dominated by human decision-makers.

6.3 Policy Implications

Our findings have several important implications for antitrust policy in increasingly automated markets:

6.3.1 Traditional Tools Are Inadequate. One-time fines (simulated as forced competitive pricing), mandatory audits (exploration shocks), and software updates (memory resets) all failed to disrupt coordination and in most cases strengthened it. This challenges the effectiveness of conventional antitrust remedies in algorithmic markets.

6.3.2 Prevention Over Cure. Since disruption appears to strengthen collusion, regulators should focus on preventing algorithmic coordination *before* it emerges rather than attempting to disrupt it afterward. This suggests a shift from ex-post enforcement to ex-ante design regulation.

6.3.3 Algorithmic Diversity Requirements. Mandating heterogeneity in pricing algorithms may prevent the emergence of stable collusive equilibria. Our findings suggest that homogeneous learning systems are particularly prone to coordination and reinforcement effects.

6.3.4 Enhanced Monitoring and Transparency. Regulators may need real-time access to algorithmic decision-making processes to detect collusion-prone designs before deployment. Transparency requirements for pricing algorithms could help identify coordination risks.

6.3.5 Novel Intervention Strategies. Our results suggest the need for more sophisticated interventions that account for learning dynamics. Possible approaches include coordinated resets of all agents'

learning states, permanent changes to exploration parameters, or algorithmic designs that penalize price correlations.

6.4 Limitations and Future Research

Our study has several limitations that suggest directions for future research:

6.4.1 Market Structure Simplifications. We focus on a duopoly market with homogeneous agents using identical Q-learning algorithms. Real markets feature multiple heterogeneous firms with diverse algorithmic strategies. Future work should test intervention effectiveness in oligopolistic markets with 3+ agents.

6.4.2 Algorithmic Heterogeneity. Examining markets with heterogeneous learning algorithms (e.g., Q-learning vs. deep Q-networks vs. policy gradient methods) would provide insights into whether certain algorithm combinations are more or less prone to collusion reinforcement.

6.4.3 Intervention Complexity. Our interventions are simplified representations of complex regulatory actions. Future research should explore more sophisticated intervention strategies, including adaptive interventions that respond dynamically to agent behavior.

6.4.4 Empirical Validation. While simulation studies provide controlled environments for testing hypotheses, empirical validation in real-world markets remains crucial. Field experiments or natural experiments involving algorithmic pricing could test whether similar reinforcement effects occur in practice.

6.4.5 Alternative Learning Architectures. Investigating whether similar reinforcement effects occur with more sophisticated learning approaches (deep reinforcement learning, model-based methods) would help determine the generality of our findings across different algorithmic paradigms.

Despite these limitations, our findings provide strong evidence that algorithmic collusion exhibits troubling stability and self-reinforcing properties, a result that should inform both academic research and regulatory practice in this rapidly evolving domain.

7 Conclusion

This paper investigates the stability of algorithmic tacit collusion under regulatory-style interventions. We demonstrate that independent Q-learning agents in repeated Bertrand competition not only converge to supra-competitive pricing (average price = 1.561, collusion index = 0.298) but exhibit troubling stability when subjected to interventions designed to disrupt coordination.

Our key finding is a *collusion reinforcement paradox*: rather than disrupting collusion, interventions consistently strengthen it. Forced competitive pricing increased prices by 3.6–3.7%, exploration shocks increased prices by 2.2%, and memory resets increased prices by 3.4%. All interventions yielded recovery rates exceeding 110% (range: 111.8–119.6%, average: 117.2%), with statistical significance confirmed across all tests ($p < 0.05$).

These results challenge conventional antitrust wisdom and suggest that traditional regulatory tools, fines, audits, and mandatory updates, may be inadequate or even counterproductive for algorithmic markets. Algorithmic collusion appears to exhibit self-reinforcing properties where disruptions lead to more robust coordination, contrasting sharply with human collusion which typically collapses under regulatory pressure.

The policy implications are significant. Regulators must shift from reactive disruption to preventive design, focusing on algorithmic diversity, ex-ante approval processes, and enhanced monitoring capabilities. As pricing algorithms become more prevalent and sophisticated, developing effective regulatory frameworks for algorithmic competition becomes increasingly urgent.

Future research should explore more complex market structures, heterogeneous algorithms, and novel intervention strategies. The alternative, relying on regulatory tools designed for human coordination in markets increasingly dominated by autonomous learning agents, risks creating environments where collusion becomes not just possible but self-reinforcing and resistant to traditional interventions.

Our findings highlight the need for interdisciplinary collaboration between computer scientists, economists, and legal scholars to develop regulatory approaches that account for the unique properties of algorithmic markets. Only through such collaboration can we hope to maintain competitive markets in the age of autonomous pricing agents.

,

References

References

[1] Ibrahim Abada, Xavier Lambin, and Balázs Tóth. Artificial intelligence, algorithmic pricing, and tacit collusion: Evidence from simulations. *Journal of Industrial Economics*, 71(2):355–390, 2023.

[2] Stephanie Assad, Robert Clark, Daniel Ershov, and Liting Xu. Algorithmic pricing and competition: Empirical evidence from the german retail gasoline market. *Management Science*, 70(1):1–23, 2024.

[3] Jonathan B Baker. Algorithms and tacit collusion: An antitrust analysis. *Antitrust Law Journal*, 84:621–662, 2021.

[4] Emilio Calvano, Giacomo Calzolari, Vincenzo Denicolò, and Sergio Pastorello. Artificial intelligence, algorithmic pricing, and collusion. *American Economic Review*, 110(10):3267–3297, 2020.

[5] Allan Dafoe, Yoram Bachrach, Gillian Hadfield, Eric Horvitz, Kate Larson, and Thore Graepel. Open problems in cooperative ai. *arXiv preprint arXiv:2012.08630*, 2020.

[6] Ariel Ezrachi and Maurice E Stucke. *Virtual competition: The promise and perils of the algorithm-driven economy*. Harvard University Press, 2016.

[7] Timo Klein. Autonomous algorithmic collusion: Q-learning under sequential pricing. *The RAND Journal of Economics*, 52 (3):538–558, 2021.

[8] Joel Z Leibo, Vinicius Zambaldi, Marc Lanctot, Janusz Marecki, and Thore Graepel. Multi-agent reinforcement learning in sequential social dilemmas. *Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems*, pages 464–473, 2017.

[9] Andreu Mas-Colell, Michael D Whinston, and Jerry R Green. Microeconomic theory. *New York: Oxford university press*, 1995.

[10] Salil K Mehra. Antitrust and the roboseller: Competition in the time of algorithms. *Minnesota Law Review*, 100:1323, 2016.

[11] Richard S Sutton and Andrew G Barto. *Reinforcement learning: An introduction*. MIT press, 2018.

[12] Christopher JCH Watkins and Peter Dayan. Q-learning. *Machine learning*, 8:279–292, 1992.