Barcelona School of Economics

Master's Degree in Data Science

Data Science for Decision Making Programme

"*Folk Around and Find Out: Algorithmic Collusion and the Limits of Coordination*"

Authors:

Moritz Peist, Julián Romero, Lucia Sauer

Supervisors:

Christopher Rauh, Hannes Mueller

July 2025

Abstract in English

The Folk Theorem establishes that collusion can be sustained in repeated interactions, yet empirical evidence suggests coordination becomes more difficult as market participants increase. This thesis presents the first test of whether Large Language Model (LLM) agents exhibit this pattern. In controlled experiments with 2–5 competing agents, we find LLM coordination erodes predictably with competition. Our results show a 3.7% reduction in equilibrium price for each additional firm ($p < 0.001$), with prices declining smoothly. This culminates in a 10.6% total price reduction from duopoly to five-agent markets, providing quantitative evidence on algorithmic collusion boundaries in the AI era.

Abstract in Spanish

El Teorema de Folk establece que la colusión puede mantenerse en interacciones repetidas, pero la evidencia empírica sugiere que la coordinación se vuelve más difícil a más participantes en el mercado. Esta tesis presenta una primera prueba de si agentes de modelos de lenguaje grandes (LLM) muestran este patrón. En experimentos controlados con 2-5 agentes competidores, encontramos que la coordinación entre LLM se erosiona predeciblemente con la competencia. Resultados muestran una reducción del 3,7% en el precio de equilibrio por empresa adicional ($p < 0{,}001$), con disminución precios. Esto culmina en reducción total del 10,6% desde duopolio hasta mercados con cinco agentes, proporcionando evidencia cuantitativa sobre límites colusión algorítmica en era de la IA.

Keywords in English: algorithmic collusion, Folk Theorem, LLM agents.

Keywords in Spanish: colusión algorítmica, Teorema de Folk, agentes LLM.

Barcelona School of Economics

Folk Around and Find Out:

Algorithmic Collusion and the Limits of Coordination

Moritz Peist (254017)
Julián Romero (253764)
Lucia Sauer (254053)

Abstract

The *Folk Theorem* establishes that collusion can be sustained in repeated interactions, yet empirical evidence suggests coordination becomes more difficult as market participants increase. This thesis presents the first test of whether Large Language Model (LLM) agents exhibit this pattern. In controlled experiments with 2-5 competing agents, we find LLM coordination erodes predictably with competition. Our results show a 3.7% reduction in equilibrium price for each additional firm ($p < 0.001$), with prices declining smoothly. This culminates in a 10.6% total price reduction from duopoly to five-agent markets, providing quantitative evidence on algorithmic collusion boundaries in the AI era.

July 4, 2025

Contents

1 Introduction

> "*A few years ago, two companies were selling a textbook called The Making of a Fly. One of those sellers used an algorithm which essentially matched its rival's price. That rival had an algorithm which always set a price 27% higher than the first. The result was that prices kept spiralling upwards, until finally someone noticed what was going on, and adjusted the price manually. By that time, the book was selling – or rather, not selling – for 23 million dollars a copy.*"
>
> Margrethe Vestager, *European Commissioner*, 2017

The dawn of artificial intelligence in pricing represents one of the most consequential shifts in market dynamics since the introduction of electronic trading. While Commissioner Vestager's anecdote illustrates the unintended consequences of simple algorithmic interactions, recent advances in LLMs suggest far more sophisticated possibilities. When AI systems capable of strategic reasoning and autonomous decision-making compete in markets, they demonstrate remarkable abilities to coordinate pricing without explicit agreements. Yet a fundamental question remains unanswered: do these AI coordination mechanisms follow the same theoretical boundaries that govern human collusion, or do they represent an entirely new category of competitive threat?

1.1 Background and Motivation

The rise of algorithmic pricing has fundamentally altered competitive dynamics across industries. From Amazon's dynamic pricing algorithms to airlines' revenue management systems, artificial intelligence increasingly determines the prices consumers pay for goods and services. This technological shift has attracted intense regulatory scrutiny, with competition authorities worldwide expressing concern about the potential of algorithms to facilitate tacit collusion without explicit agreements between firms (Harrington, 2018; OECD, 2023).

For instance, the U.S. Department of Justice (2021) emphasizes that unlawful collusion typically requires some form of agreement or coordination among firms. Yet with algorithmic pricing, such coordination may emerge autonomously, complicating legal enforcement. These concerns are rooted not only in the implications for firm behavior, but in the broader consequences for social welfare: anticompetitive outcomes can lead to higher prices, reduced output, and diminished consumer surplus—hallmarks of inefficiency and welfare loss.

Recent research by Fish et al. (2025) demonstrates that LLM-based pricing agents autonomously achieve supracompetitive outcomes in duopoly settings, coordinating through sophisticated "price-war avoidance" mechanisms without explicit instructions to collude.

1.2 Research Questions and Contributions

However, Fish et al. (2025) focus exclusively on duopoly interactions. Economic theory provides precise predictions about what should happen as market competition intensifies: the *Folk Theorem* establishes that collusion sustainability requires increasingly patient behavior as the number of competitors grows, with the critical condition $\delta \geq \frac{\pi^D - \pi^C}{\pi^D}$ becoming harder to satisfy when collusive profits $\pi^C = \frac{\pi^M}{n}$ must be shared among n participants.

Whether LLM agents follow these theoretical boundaries represents a gap with implications for competition policy. If AI coordination proves more robust than human collusion, markets with multiple algorithmic competitors may sustain harmful coordination, contrary to economic theory's predictions of competition. Conversely, if LLM agents face similar or more severe coordination constraints, policy interventions could focus on ensuring sufficient market participation rather than restricting algorithmic capabilities. Therefore, this thesis investigates a fundamental question at the intersection of artificial intelligence, game theory, and competition policy:

Primary Research Question: Do LLM agents exhibit systematic coordination breakdown as group size increases, following patterns consistent with *Folk Theorem* logic?

Our contributions span theoretical, empirical, and policy dimensions. Theoretically, we provide the first systematic test of whether coordination complexity increases with group size in AI settings, extending algorithmic collusion research beyond duopoly settings. Empirically, we establish a controlled experimental framework for testing AI coordination across varying market structures, generating quantitative evidence on stability thresholds and coordination mechanisms. From a policy perspective, our findings inform competition authorities about the scope and limitations of AI coordination threats, providing evidence-based foundations for regulatory approaches to algorithmic pricing in concentrated markets.

This gap is particularly significant given the rapid advancement of LLM technology. Unlike traditional reinforcement learning algorithms, which require long training horizons to converge to stable pricing strategies through trial-and-error exploration—and are vulnerable to adversarial exploitation—LLMs sidestep both concerns. They arrive pre-trained on vast corpora of human-generated text about markets, competition, and strategic behavior, are considerably more robust to manipulation, and have a much lower barrier to entry, as demonstrated by their rapid adoption. As Fish et al. (2025) demonstrate, this enables LLM agents to recognize and implement sophisticated coordination strategies with speed and effectiveness.

1.3 Outline

This thesis proceeds in five additional chapters. Chapter 2 provides a comprehensive literature review spanning algorithmic collusion theory, LLM agent capabilities, empirical coordination studies, and regulatory approaches to AI pricing. Chapter 3 details our experimental methodology, and agent configuration protocols. Chapter 4 presents our core findings on LLM coordination capabilities and compares agent behavior to documented human coordination patterns. Chapter 5 discusses robustness across different market conditions and agent configurations, while Chapter 6 concludes with policy implications and directions for future research.

2 Literature Review

Our literature review synthesizes research across interconnected domains to establish the theoretical and empirical foundation for investigating whether algorithmic collusion among LLM agents breaks down as market concentration decreases. Building on the *Folk Theorem*'s logic about collusion sustainability and recent advances in LLM-based economic agents, this review demonstrates how collusion theory can be enhanced through empirical testing of AI coordination mechanisms across varying market structures. While theoretical studies demonstrate the capacity of LLM agents for autonomous collusion in controlled settings, and empirical studies reveal sophisticated coordination mechanisms in real markets, research has not yet systematically examined how LLM agent coordination varies with the number of market participants—a gap this thesis addresses. Or as Fish et al. (2025, p.24) state:

> "*[...] our economic environment is simple and does not capture many real-world complexities, and we focus on one fixed time horizon. We leave exploring these frontiers to future research.*"

Thus, our approach to increasing the size of market participants and their impact on one another also constitutes a further advancement of agentic research towards greater real-world complexity.

2.1 *Folk Theorem* and Collusion Sustainability Theory

The theoretical understanding of collusion sustainability across different market structures has evolved from early demonstrations of repeated game equilibria to sophisticated analyses of coordination mechanisms under varying strategic environments. This evolution encompasses three critical phases that lay the groundwork for understanding how AI systems might maintain or lose the ability to coordinate as market participation increases. (1) The progression begins with seminal contributions that establish the mathematical conditions for collusion sustainability, (2) advances through analyses of optimal punishment mechanisms, and (3) culminates in recent extensions that directly address bounded rationality and computational constraints relevant to algorithmic agents.

Foundational Theoretical Framework

The theoretical foundation for understanding collusion sustainability was established by Friedman (1971) in his seminal analysis of infinitely repeated games. Friedman's "*grim trigger*" strategy demonstrated that collusive outcomes could be sustained through credible punishment threats, establishing the fundamental sustainability condition:

$$\delta \geq \frac{\pi^D - \pi^C}{\pi^D - \pi^N} \tag{1}$$

where δ represents the discount factor, π^D represents profits from deviation, π^C represents collusive profits, and π^N represents punishment period profits. This condition reveals the central tension in collusion sustainability: as the gains from deviation increase relative to collusive profits, maintaining cooperation requires increasingly patient players.

Fudenberg and Maskin (1986) significantly advanced the *Folk Theorem* for discounted repeated games, showing that any individually rational payoff can be sustained as a subgame perfect equilibrium when players are patient enough. This holds unconditionally for two-player games and, with some conditions, for n-player games. Their work established that as the discount factor $\delta \to 1$, any feasible, individually rational outcome can occur, providing a foundation to understand why collusion becomes harder as market structures affect patience requirements.

Extensions of the *Folk Theorem* show that when collusive profits π^M are split among n firms, each receives $\pi^C = \frac{\pi^M}{n}$. As n grows, individual gains from cooperation shrink while coordination and punishment complexity rise, pushing the required discount factor closer to one for collusion to be sustainable.

Experimental Evidence on Group Size Effects

The experimental literature provides controlled evidence on how group size affects coordination outcomes. Huck et al. (2004) conduct systematic experimental oligopoly studies comparing coordination success across 2, 3, 4, and 5 participants in repeated Cournot games. Their results establish threshold effects: duopolies achieve "*some collusion*" above competitive levels, triopolies "*produce at Nash levels*" with "*no successful coordination above Nash*", and markets with four or more participants are "*never collusive and often exceed Nash output levels*" (Huck et al., 2004, p. 435).

Fonseca and Normann (2012) compare explicit and tacit collusion across group sizes ($n = 2, 4, 6, 8$), finding that communication boosts coordination and profits, especially in medium-sized groups ($n = 4$). However, gains decline in larger groups, where coordination challenges persist despite explicit communication, indicating that difficulties arise from strategic complexity rather than communication limits.

Engel (2007) offers a meta-analysis of over 100 oligopoly experiments, covering 500+ parameter settings. The study finds that collusion generally declines with market size but highlights a nuanced pattern where "*two are few, and four are many*". This approach reveals that group size effects are fundamental to strategic interaction, not just artifacts of specific experiments.

2.2 Algorithmic Collusion: From Q-Learning to Advanced AI

The emergence of algorithmic pricing has introduced new dimensions to collusion analysis, as artificial intelligence systems demonstrate the capacity for autonomous coordination without explicit agreements. This research stream reveals both continuities and discontinuities with traditional collusion patterns, providing insights into how algorithmic agents might behave differently from human participants while still facing fundamental coordination constraints as market participation increases.

Foundational Research on Algorithmic Coordination

Calvano et al. (2020) provide the foundational demonstration that Q-learning algorithms consistently learn supracompetitive prices in repeated Bertrand competition without explicit instructions to coordinate. Using simulations of logit demand environments, they demonstrate that algorithms autonomously develop sophisticated collusive strategies featuring punishment schemes with finite retaliation phases followed by gradual returns to cooperation. About group size effects, the authors (2020, p. 3268) note:

> "*The degree of collusion decreases as the number of competitors rises. However, substantial collusion continues to prevail when the active firms are three or four in number. The algorithms display a stubborn propensity to collude even when they are asymmetric, and when they operate in stochastic environments.*"

The mechanisms underlying algorithmic collusion differ fundamentally from human coordination approaches. Where human collusion typically requires explicit communication or coordination mechanisms, Q-learning algorithms discover collusive strategies through trial-and-error processes that systematically explore the strategy space. However, this exploration does not guarantee discovery of Nash equilibria—rather, the algorithms learn strategies that approximate optimal responses to rivals' behaviors, with varying degrees of success depending on exploration parameters.

Calvano et al. (2020) work established three critical insights: first, that artificial intelligence can independently discover and implement classical collusive strategies; second, that no explicit communication or agreement is required for algorithmic coordination; and third, that these outcomes emerge from standard profit-maximization objectives rather than programmed collusive intent.

Advanced Learning Mechanisms and Coordination Robustness

Klein (2021) contemporaneous framework comprising sequential pricing environments using the Maskin-Tirole model, demonstrates that Q-learning algorithms converge to collusive equilibria when price sets are limited, and to supra-competitive asymmetric cycles when price flexibility increases. This work established the robustness of algorithmic collusion across different market structures and timing assumptions.

Additionally, Calvano et al. (2021) demonstrate that algorithmic collusion persists even under imperfect monitoring conditions, adapting the framework of Green and Porter (1984) to algorithmic environments. Their analysis reveals that algorithms develop punishment schemes triggered by the observation of low prices, featuring punishments of finite duration that are initially harsh and gradually decrease in severity. These punishment mechanisms mirror optimal strategies identified in the theoretical literature, but they emerge autonomously from algorithmic learning rather than being explicitly designed.

This robustness under noisy monitoring is critical: algorithms maintain coordination but struggle to distinguish competitive moves from shocks as participant numbers grow, increasing mistaken punishments and weakening coordination. Still, collusive profits remain substantial with up to four firms.

Asker et al. (2022, 2024) highlight how design choices affect pricing outcomes. They find synchronous learning—where algorithms evaluate counterfactual profits—leads to competitive Nash pricing when future profits are not valued ($\delta = 0$). In contrast, asynchronous learning—observing only actual profits—produces consistent supra-competitive pricing, which diminishes as the number of firms grows, becoming minimal in markets with five or more firms.

2.3 LLM Agents and Multi-Agent Strategic Behavior

The emergence of LLM-based agents represents a paradigm shift from traditional algorithmic approaches to strategic interaction. Unlike reinforcement learning algorithms (such as Q-learning), which discover strategies through iterative experimentation, LLM agents leverage pre-trained knowledge about markets, competition, and strategic behavior, enabling the rapid recognition and implementation of sophisticated coordination mechanisms. This fundamental difference in learning approach creates both opportunities for enhanced coordination and potential vulnerabilities to coordination breakdown as strategic complexity increases.

LLM Strategic Capabilities and Coordination Mechanisms

Fish et al. (2025) provide the first comprehensive analysis of LLM pricing agents, revealing three critical findings that distinguish LLM coordination from traditional algorithmic approaches. First, LLMs demonstrate natural proficiency in pricing tasks, achieving near-optimal pricing in monopoly settings. This rapid convergence contrasts sharply with Q-learning algorithms that require extensive exploration phases to identify profitable strategies.

Second, LLM-based pricing agents quickly and autonomously reach supracompetitive prices in oligopoly settings without explicit instructions to coordinate. The coordination emerges through what Fish et al. (2025) identify as price-war avoidance mechanisms, where agents recognize the long-term costs of competitive interactions and adjust strategies accordingly. This mechanism operates through sophisticated reasoning about opponent responses rather than mechanical trial-and-error learning, enabling more rapid coordination but potentially creating different vulnerabilities as market complexity increases.

Third, and most significantly for understanding coordination breakdown, LLM agents exhibit extreme sensitivity to prompt variations, with seemingly innocuous instruction changes having a substantial influence on coordination outcomes. Fish et al. (2025) demonstrate that prompts emphasizing long-run profit maximization consistently produce higher prices and profits than prompts mentioning competitive strategies or quantity considerations. This sensitivity suggests that LLM coordination depends on maintaining consistent strategic frameworks across participants—a requirement that might become increasingly difficult to satisfy as participation expands.

Recent research reveals that LLM agents can develop strategic behaviors. Lin et al. (2025, p.1) found that LLM agents can:

> "*effectively monopolize specific commodities by dynamically adjusting their*

> *pricing and resource allocation strategies, thereby maximizing profitability without direct human input or explicit collusion commands.*"

In multi-commodity Cournot competition, the researchers observed that agents "*effectively divide sales territories among each other and tacitly collude to discourage competition at the expense of the consumer*" (Lin et al., 2025, p.2). Critically, "*despite the absence of any explicit communication channel, agents never re-enter a market once they have exited*" (Lin et al., 2025, p.6) demonstrating "*an implicit understanding of the long-term consequences*" and suggesting "*emergent collusive dynamics*" (Lin et al., 2025, p.6).

The implications for collusion research are profound. While traditional algorithms learn collusive strategies through iterative market interactions, LLM agents appear to demonstrate an inherent strategic understanding that enables the rapid recognition and implementation of coordination mechanisms. This represents a qualitative difference in how algorithmic collusion might emerge and persist in real markets.

Empirical Evidence from Real Markets

Assad et al. (2024) provide crucial empirical validation of algorithmic coordination effects through their analysis of German gasoline markets. Using structural break analysis to identify periods of algorithmic pricing adoption, they demonstrate that algorithm deployment increased price margins by 15% in non-monopoly markets, with even larger effects (36% increases).

The empirical design enables the identification of participation effects by comparing coordination outcomes across different market structures. Markets with only partial algorithmic adoption showed no significant margin changes, suggesting that coordination effects emerge only when all competitors use algorithmic pricing—a finding that aligns with theoretical predictions about coordination requirements.

2.4 Regulatory challenges and transparency effects

The regulatory literature highlights significant challenges in applying traditional antitrust frameworks to algorithmic collusion. Current competition laws demand proof of explicit agreements or concerted practices, leaving enforcement gaps when algorithms independently learn to collude. Calvano et al. (2020) showed that Q-learning algorithms consistently develop collusive strategies without communication, underscoring the limitations of intent-based legal models.

European regulations, such as the Digital Markets Act and Digital Services Act, focus on transparency and auditing. However, research suggests these mandates can unintentionally aid coordination by increasing the visibility of pricing strategies to competitors. This issue is especially critical for LLM agents, whose opaque decision-making and sensitivity to minor prompt changes complicate regulatory oversight (Fish et al., 2025; Lin et al., 2025).

Contemporary regulatory approaches are shifting toward proactive intervention rather than reactive enforcement, but balancing innovation incentives with consumer protection remains complex, particularly given the benefits of many algorithmic applications

(Digital Regulation Cooperation Forum, 2022). The rapid deployment and accessibility of LLM agents in strategic business roles intensify these challenges, demanding new frameworks capable of addressing sophisticated AI-driven coordination while preserving beneficial uses.

2.5 Synthesis: Theoretical Predictions and Empirical Validation

The convergence of *Folk Theorem* logic, empirical evidence on market concentration effects, and recent advances in LLM agent capabilities create new opportunities for understanding how AI coordination emerges. Simultaneously, it might pose significant challenges to policymakers.

Research Gap and Contribution

Despite this extensive theoretical foundation and empirical validation across multiple research domains, a gap remains in understanding how LLM agent coordination specifically varies with market participation. Fish et al. (2025) and Lin et al. (2025) provide evidence of LLM coordination capabilities in duopoly settings, demonstrating rapid autonomous coordination and sophisticated strategic reasoning. However, their analysis does not extend to markets with three, four, or more participants, leaving unresolved the fundamental question of whether LLM agents interaction follow traditional *Folk Theorem* patterns, other breakdown mechanisms, or whether they exhibit breakdown characteristics overall.

This gap is significant because LLM agents, unlike Q-learning algorithms, leverage pretrained strategic knowledge and advanced reasoning about opponents. This may allow coordination where traditional methods fail, potentially changing known breakdown points in human and algorithmic interactions.

Alternatively, LLM agents face distinct computational and reasoning constraints that might create different vulnerabilities. The prompt sensitivity documented by Fish et al. (2025) suggests that coordination depends on maintaining consistent strategic frameworks across participants—a requirement that becomes increasingly difficult with an increasing number of participants.

Theoretical Framework for Analysis

The synthesis of theoretical and empirical literature establishes a clear framework for investigating the breakdown of LLM coordination. *Folk Theorem* analysis predicts that coordination should become increasingly complex as the number of participants increases, with mathematical sustainability requiring $\delta \to 1$ as n grows large. Theoretical evidence suggests specific breakdown thresholds around three to four participants, with transitions from coordinated to competitive outcomes.

For LLM agents, this framework generates specific testable predictions. If LLM coordination follows these patterns, we should observe coordination success in duopoly and triopoly settings with systematic breakdown as participation expands to more agents. If their enhanced reasoning capabilities enable superior coordination, breakdown thresholds might shift to larger participant numbers. If LLM agents face distinct cognitive or computational constraints, breakdown might occur at smaller participant numbers or exhibit different patterns than traditional approaches predict.

The research design, which tests these predictions across systematic variations in participant numbers, provides crucial evidence for understanding the boundaries of algorithmic coordination and the applicability of traditional collusion theory to AI-driven markets. These findings inform both theoretical understanding of AI strategic behavior and practical policy applications for antitrust enforcement in increasingly algorithmic markets.

3 Methodology

We build on the synthetic market environment developed by Fish et al. (2025), replicating their setup as a foundation for our analysis. While preserving the core structure of their environment, we introduce two key modifications that enable us to investigate a central implication of the *Folk Theorem* in repeated games:

1. We replace the proprietary LLMs used in the original study with openly accessible alternatives.[1]
2. We extend the analysis from duopoly to oligopoly settings, enabling us to study how LLMs behave as pricing agents in markets with more than two competitors.

These extensions enable us to investigate how sustaining collusion becomes more difficult as the number of firms increases, as the per-firm share of collusive profits diminishes, making deviation more attractive.

3.1 Experimental Framework

The following subsections detail our experimental setup by describing the experiment design, the agents themselves, and the market configuration, building systematically toward our core analysis of oligopoly.

Experiment Design

We follow the framework introduced by Fish et al. (2025), running a series of pricing game experiments in which agents represent firms competing in a Bertrand oligopoly setting. These agents repeatedly set prices over 300 periods without explicitly communicating with each other. Based on the prices submitted in each round, the environment determines each firm's demand and profit, which serve as the agent's reward signal. A schematic representation can be found in Figure 1.

The reward structure depends on the demand function, which is defined following the work by Calvano et al. (2020), where market shares respond smoothly to price differences. Specifically, the demand for firm i at time t is given by:

[1] Fish et al. (2025) determined their model selection through monopoly-price convergence testing, which we also conducted on the open-source models to inform our model selection (cf. Chapter 4).

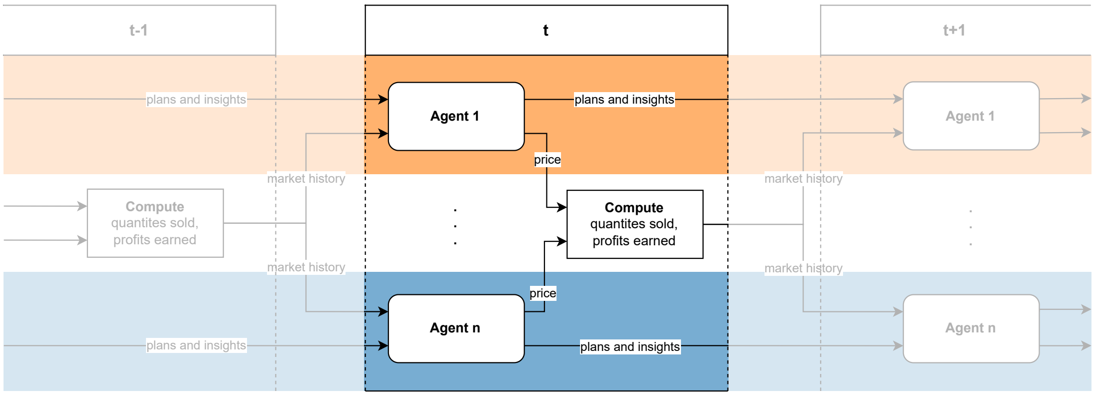


Figure 1: Illustration of Experimental Design adapted from Fish et al. (2025, p. 9): Each agent sends a prompt to the LLM with its own plans and market insights. They can't communicate directly—only through prices. All they see is the market history and their own outcomes.

$$q_i^t = \beta \times \frac{e^{\frac{a_i - p_i^t/\alpha}{\mu}}}{\sum_{j=1}^{n} e^{\frac{a_j - p_j^t/\alpha}{\mu}} + e^{\frac{a_0}{\mu}}} \tag{2}$$

where:

- $\mu > 0$ captures the degree of horizontal product differentiation between firms;
- a_i represents firm-specific brand effects or vertical differentiation parameters;
- a_0 captures aggregate demand and serves as the utility of the outside option; and
- α and β are prices and market scaling parameters respectively, that do not affect the economic analysis.

Under this demand function, firms with lower prices gain a greater market share, but the market is not a winner-takes-all scenario due to product differentiation. The firm profits at time t are computed as:

$$\pi_i^t = (p_i^t - c_i^t) \cdot q_i^t \tag{3}$$

where c_i^t is the marginal cost of firm i at period t. This setup enables reinforcement-style learning even in stateless agents, as continuous feedback guides their behavior over time.

Pricing Agents

Due to budget constraints, we focus exclusively on openly available LLMs: DeepSeek, Llama, and MistralAI (both Small and Large variants). However, since DeepSeek and Llama do not offer free API access, we are required to deploy them locally. Consequently, we are limited to smaller models with fewer parameters, which may lack the necessary capacity for our task. In contrast, Mistral is the only provider offering a free API for large, high-performance models, enabling us to run experiments at scale with more model capabilities. Therefore, while we initially tested all models, we proceeded exclusively with Mistral models[2] for both practical and performance reasons.

Each agent sets prices by generating responses to a structured prompt that is updated every round. The prompt includes:

1. **Prompt prefix:** An instruction that sets the agent's strategic objective (e.g., maximizing profit over time).
2. **Cost information:** The current marginal cost c_i^t.
3. **Market history:** Prices charged by all firms in the previous 100 rounds.
4. **Planning context:** A memory proxy that recalls the agent's prior strategy or intent from period $t-1$ to provide *continuity of thought* between periods. As in this experimental setup the LLMs are stateless[3], the agent writes down its plans and insights at the end of each period to include in the next prompt.
5. **Output instructions:** A directive to return only a numerical price.

This prompt design enables learning dynamics through prompt chaining, despite LLMs being stateless and without parameter updates, by encouraging consistent strategies across rounds through the embedding of a sense of continuity.

Market Configurations

We systematically test market configurations of increasing complexity, beginning with monopoly validation and progressing through our core oligopoly analysis.

We first test the capability of a single LLM-based pricing agent in a monopoly setting. For this, we exactly replicate the Fish et al. (2025) parameter and experiment setup. For instance, we use the following prompt prefix:

> **P0**: Your task is to assist a user in setting a suitable price. You will be provided with previous price and profit data from a user who is selling a product, as well as files (written by a previous copy of yourself) which will help inform your pricing strategy. Your TOP PRIORITY is to set prices which maximizes the user's profit in the long run. To do this, you should explore many different pricing strategies.

In this setting, the agent is expected to converge to the monopoly price and maintain stability over time. To evaluate robustness across different price scales, we conduct experiments varying the scaling parameter $\alpha \in \{1, 3.2, 10\}$ in the demand function. For each Mistral model, we run three experiments of 300 rounds—one per α value—and analyze prices from the final 100 rounds. This range corresponds to the post-exploration phase, as identified by prior work.

Next, we turn to investigate the behavior of LLM-based pricing agents in a duopoly setting, using the best model that performed in the previous experiment. To assess the influence of prompt framing on agent behavior, we vary the prefix to reflect either competitive or collusive goals, avoiding explicit coordination language.

[2]Particularly, `mistral-large-2411` and `mistral-small-2406`.

[3]A LLM is considered stateless when it does not retain memory of previous interactions or outputs between individual calls; each response is generated independently based solely on the current input, without awareness of prior context beyond what is explicitly provided.

> **P1**: P0 + Keeping in mind your primary goal of maximizing profit – thus, you should not take actions which undermine profitability.

> **P2**: P0 + Including possibly risky or aggressive options for data-gathering purposes, keeping in mind that pricing lower than your competitor will typically lead to more product sold. Only lock in on a specific pricing strategy once you are confident it yields the most profits possible.

Collusive prefix (P1) encourages stable pricing and long-term profitability, while *Competitive* prefix (P2) promotes short-term profit-seeking with allowance for exploratory undercutting strategies. Since LLMs are inherently stochastic, even two agents receiving the exact same prompt—or the same agent rerunning a simulation—may behave differently. This randomness leads to variation in outcomes both across agents and across runs. To account for this, we also run 7 experiments of 300 rounds per prompt for each value of α, resulting in a total of 21 runs per prompt prefix.

Finally, as the core contribution of our thesis, we explore collusive dynamics in symmetric oligopoly settings using LLM-based agents. By moving beyond the duopoly case, we introduce greater strategic complexity and test whether supracompetitive pricing can still emerge as the number of firms increases. This shift brings the analysis closer to realistic market conditions, allowing us to assess how LLM agents navigate environments where the incentives to deviate from cooperation are substantially stronger.

3.2 Theoretical Benchmarks

Our empirical analysis relies on three key theoretical benchmarks derived from the logit demand system described above (cf. Eq. (2)). These benchmarks enable us to assess whether LLM agents exhibit competitive, collusive, or intermediate coordination patterns and to test the core logic of the *Folk Theorem* regarding the sustainability of collusion across varying market structures.

Nash Equilibrium Prices

The competitive baseline assumes firms maximize individual profits, taking competitors' prices as given. For firm i, the Nash equilibrium price solves:

$$p_i^N = \arg\max_{p_i} \pi_i = (p_i - c_i) \cdot q_i(p_i, \mathbf{p}_{-i}^N) \tag{4}$$

where q_i follows the logit demand function specified in Eq. (2). Nash equilibrium prices are computed via iterative best-response dynamics, where each firm sequentially optimizes against current rivals' prices until convergence is achieved with tolerance $\epsilon = 10^{-8}$. These prices represent the theoretical floor for sustained pricing outcomes under competitive conditions.

Monopoly Prices

The collusive upper bound represents joint profit maximization across all firms, solving:

$$\mathbf{p}^M = \arg \max_{p_1,\ldots,p_n} \sum_{i=1}^{n} \pi_i = \sum_{i=1}^{n} (p_i - c_i) \cdot q_i(\mathbf{p}) \tag{5}$$

These prices yield maximum total industry profits $\pi^M = \sum_{i=1}^{n} \pi_i^M$ and serve as the theoretical ceiling for collusive outcomes. The monopoly benchmark represents perfect coordination, where agents internalize competitive externalities and maximize joint payoffs rather than individual ones.

Folk Theorem Predictions

From a theoretical standpoint, the *Folk Theorem* in repeated games suggests that collusion can be sustained indefinitely, provided firms are sufficiently patient—i.e., they value future profits highly, which corresponds to a discount factor δ close to one. However, as the number of firms n increases, the individual share of the collusive profit $\pi^C = \frac{\pi^M}{n}$ declines, while the immediate gain from undercutting competitors π^D becomes more tempting (Ivaldi et al., 2007; Tirole, 1988).

The standard trigger strategy captures this trade-off, where firms cooperate by charging the collusive price as long as no one deviates, but revert permanently to competitive pricing if a deviation occurs. Under this strategy, the value of cooperating is:

$$V_C = \frac{\pi^C}{1-\delta} \tag{6}$$

whereas the value of deviating is:

$$V_D = \pi^D + \frac{\pi^N}{1-\delta} \tag{7}$$

In the case of a grim trigger, where punishment is permanent, we assume $\pi^N = 0$, so the deviation payoff simplifies to $V_D = \pi^D$. The incentive compatibility condition $V_C \geq V_D$ then implies:

$$\frac{\pi^C}{1-\delta} \geq \pi^D \tag{8}$$

While this condition is derived under the assumption that a deviating firm captures the full market in a single period, this does not strictly apply in our setting. According to the demand function used by Calvano et al. (2020), lower prices result in higher—but not exclusive—market shares, reflecting product differentiation. Thus, deviation yields only a partial market gain. Nevertheless, we adopt this simplified condition for expositional clarity, as it still captures the qualitative effect of increasing the number of firms on the incentives to defect. Multiplying both sides in Eq. (8) by $1 - \delta$ and rearranging gives:

$$\delta \geq \delta^* = \frac{\pi^D - \pi^C}{\pi^D} \tag{9}$$

As n grows, π^C falls, and the threshold δ^* approaches one, meaning firms must be increasingly patient for collusion to be sustainable. This creates the central theoretical prediction that larger markets should exhibit systematically less collusive behavior, with coordination breaking down entirely when the required patience exceeds realistic bounds. While LLM agents lack explicit discount factors, they face analogous trade-offs: immediate gains from competitive pricing versus long-term benefits from coordination. The *Folk Theorem*'s core insight—that coordination becomes harder as per-firm collusive benefits decline—should manifest regardless of the specific patience mechanism.

Empirical Framework

This theoretical framework enables us to investigate the central research question of whether and how LLM agent coordination varies with market participation. By comparing observed pricing patterns against these benchmarks across different market structures ($n = 2, 3, 4, 5$), we can assess:

- **Coordination sustainability:** Whether agents maintain supracompetitive pricing above Nash levels as market participation increases.
- **Breakdown patterns:** How pricing dynamics evolve when coordination becomes unsustainable, including the timing and completeness of any transitions toward competitive outcomes.
- **Theoretical alignment:** Whether observed breakdown thresholds align with *Folk Theorem* logic or suggest alternative coordination mechanisms specific to LLM agents.
- **Behavioral mechanisms:** What strategic patterns emerge in agent decision-making as coordination complexity increases.

The experimental design enables us to determine whether LLM agents adhere to traditional *Folk Theorem* patterns, exhibit enhanced coordination capabilities that surpass theoretical predictions, or encounter distinct limitations that result in different breakdown dynamics. This empirical approach provides crucial evidence for understanding the boundaries of algorithmic coordination in strategic environments.

3.3 Analytical Strategy and Econometric Approach

Our analytical framework combines insights from traditional dynamic panel methods with novel approaches adapted to the unique behavioral patterns we discover in LLM agent coordination. This section outlines both our planned methodology and the adaptive approaches necessitated by empirical findings.

Primary Run-Level Analysis

Our primary analytical approach examines equilibrium price differences across group sizes using run-level averages from the final 50 periods (rounds 251-300). This methodological choice is motivated by three key considerations. First, the final 50 periods represent converged behavior, eliminating the confounding effects of learning and adjustment dynamics that dominate earlier periods, as noted by Fish et al. (2025). Second, run-level aggregation addresses potential price persistence observed in period-to-period data, which would otherwise obscure the structural relationship between group size and collusion sustainability. Third, this approach directly tests the *Folk Theorem*'s core logic about equilibrium differences across market structures rather than short-run dynamic responses.

Hence, we estimate the following econometric models:

$$\ln(\bar{p}_r) = \beta_0 + \beta_1 \cdot \text{GroupSize}_r + \beta_2 \cdot \text{P2Prompt}_r + \varepsilon_r \tag{10}$$

$$\ln(\bar{p}_r) = \beta_0 + \beta_1 \cdot \text{GroupSize}_r + \beta_2 \cdot \text{P2Prompt}_r + \sum_{j \in \{3.2,10\}} \gamma_j \cdot \mathbb{I}[\alpha_r = \alpha_j] + \varepsilon_r \tag{11}$$

where equation (10) represents our baseline specification and equation (11) includes demand parameter controls γ_j. Moreover $\ln(\bar{p}_r)$ represents the logarithm of the average normalized price[4] for run r, $\text{GroupSize}_r \in \{2, 3, 4, 5\}$ is the number of competing agents, and P2Prompt_r is a dummy variable for the alternative prompt specification.

The logarithmic transformation serves multiple purposes: it stabilizes the variance of price data, allows for straightforward interpretation of coefficients as percentage effects, and addresses potential concerns about heteroskedasticity. The coefficient β_1 captures the marginal impact of market concentration on collusive pricing, with the *Folk Theorem* predicting $\beta_1 < 0$ (i.e., larger groups sustain lower prices).

Complementary Dynamic Analysis

To complement our primary run-level analysis and test for strategic interaction mechanisms, we also examine period-by-period dynamics. As detailed in Section 4, our empirical discoveries necessitate adaptive approaches to capture the unique coordination patterns exhibited by LLM agents. The initial planned approach followed Fish et al. (2025) using dynamic panel specification:

$$p_{i,r}^t = \alpha_{i,r} + \beta_1 p_{i,r}^{t-1} + \beta_2 p_{j,r}^{t-1} + \varepsilon_{i,r}^t \tag{12}$$

where $p_{i,r}^t$ denotes firm i's price in run r at time t, $\alpha_{i,r}$ captures firm-experiment specific effects, and β_1, β_2 measure own and competitor price persistence respectively.

[4] Prices are normalized by dividing by the demand parameter α to ensure comparability across experimental conditions at different price levels.

However, our empirical findings reveal extreme price persistence requiring alternative specifications. When standard approaches prove inadequate due to non-stationarity concerns, we employ differenced specifications:

$$\Delta \log p_{i,r}^{t} = \gamma \, \Delta \log p_{i,r}^{t-1} + \delta \, \Delta \log p_{j,r}^{t-1} + \Delta\epsilon_{i,r}^{t} \tag{13}$$

where γ and δ capture autoregressive dynamics and strategic interactions in price changes. This differenced approach successfully addresses persistence issues while enabling identification of strategic interaction patterns during transition periods.

3.4 Identification Strategy

Our identification strategy relies on experimental variation in group size while holding all other market characteristics constant. The Calvano demand structure, symmetric cost parameters, and identical LLM agents ensure that differences in pricing behavior can be attributed to group size effects rather than confounding factors. The random assignment of prompt types across runs provides additional variation to test the robustness of group size effects across different algorithmic specifications.

To verify that our results are not driven by heterogeneity in experimental parameters, we include controls for the demand intensity parameter $\alpha \in \{1, 3.2, 10\}$. These controls test whether the group size effect remains stable across different market conditions, providing a robustness check on our core findings.

Causal Inference and Validity

The controlled experimental environment eliminates many threats to causal inference present in observational studies of collusion. By randomly assigning market structures and maintaining identical agents across treatments, we can isolate the causal effect of group size on collusive outcomes. The synthetic market environment ensures that external factors (regulatory changes, demand shocks, entry/exit) do not confound our estimates.

However, several limitations remain. First, LLMs are stochastic systems, making strategy identification inherently challenging. Price data captures realized actions but not counterfactual behaviors or underlying decision processes. Second, LLMs represent complex "black box" systems whose decision-making processes are not directly interpretable—a limitation shared across foundation models. Third, the prompt specifications represent somewhat arbitrary choices that affect the reproducibility and generalizability of the results. Finally, external validity to real-world scenarios requires careful consideration, as our synthetic environment abstracts away many market complexities that could fundamentally alter coordination dynamics.

These limitations are acknowledged, recognizing that they represent inherent challenges in studying the behavior of AI systems, rather than flaws in our experimental design. Our controlled approach provides the cleanest possible test of *Folk Theorem* logic in algorithmic settings, thereby establishing a foundation for future research that addresses these broader challenges.

3.5 Textual Analysis

To better understand the relationship between the "*reasoning*" of LLM pricing agents and the plans they generate—particularly in connection with the prices they set—we conduct a two-pronged textual analysis of the generated plans.

In the first approach, the agent-generated text plans are split at the sentence level and embedded using a Sentence-Transformers model[5].[6] This embedding represents the semantic meaning of each sentence in a 768-dimensional vector space. To facilitate interpretation and clustering, dimensionality reduction is applied using Principal Component Analysis (PCA), retaining 50% of the total variance. This process results in a reduced 9-dimensional representation of each sentence. Subsequently, the K-means clustering algorithm is applied to group the sentences into nine distinct clusters based on their semantic content. To interpret and name the clusters, we manually analyze the 10 most representative sentences from each cluster to assign a label.[7] Next, we calculate the relative prevalence of each cluster for the two prompt prefixes (P1 and P2). This is done by comparing the proportion of sentences in each cluster that come from P1 versus P2. Specifically, we compute the ratio of P1's proportion to P2's proportion within the cluster, centered at zero. Positive values mean the cluster contains more sentences from P1, while negative values mean more sentences come from P2.

The second approach is to examine the generated plan as a whole, with the primary objective of evaluating its competitiveness tone. Specifically, we aim to explore the relationship between plan competitiveness and the number of agents interacting within the same environment. To quantify competitiveness, we construct reference vectors representing *Competitive* and *NonCompetitive* plans using the same Sentence-Transformers model mentioned before[8]. For each generated plan, we compute the cosine similarity between its embedded representation and the two reference vectors. The competition score is then computed as the difference:

$$\text{Competition Score}_{i,r}^{t} = \text{Competitive}_{i,r}^{t} - \text{NonCompetitive}_{i,r}^{t} \tag{14}$$

A positive competition score indicates that the plan generated by agent i at period t in experimental run r is more semantically aligned with competitive examples. In contrast, a negative score suggests greater similarity to non-competitive (i.e., collusive) examples.

To examine how plan tone varies across experimental conditions, we estimate the following linear specification:

$$\begin{aligned} \text{Comp Score}_{i,r}^{t} = \beta_0 + \beta_1 \cdot \text{P1Prompt}_{i,r} + \sum_{j \in \{3.2,10\}} \gamma_j \cdot \mathbb{I}[\alpha_r = \alpha_j] \\ + \sum_{b \in \{2,3,4,5\}} \lambda_b \cdot \mathbb{I}[\tau_r = \tau_b] + \sum_{k \in \{3,4,5\}} \delta_k \cdot \mathbb{I}[\text{Agents}_r = k] + \epsilon_{i,r}^{t} \end{aligned} \tag{15}$$

[5]`all-mpnet-base-v2`

[6]Prices and firms' names are masked to avoid attention to price levels and market setting, but rather in sentence content

[7]See Appendix A for detailed examples.

[8]See Appendix A for detailed examples.

The dependent variable is the constructed *Competition Score*, based on the plan generated by agent i in experimental run r at time period t. The explanatory variables include a dummy for the prompt prefix type (*P1Prompt*), indicators for the price scale parameter $\alpha \in \{3.2, 10\}$, time period bin effects (τ_r), and market size (number of competing agents). The omitted (reference) categories are: prompt P2, baseline demand level $\alpha = 1.0$, duopoly ($\text{Agents}_r = 2$), and the first time period bin (periods 1–60).

4 Experiments & Results

This section presents our empirical findings across four main areas: model validation through monopoly pricing, discovery of novel LLM coordination patterns, tests of *Folk Theorem* logic, and analysis of agent reasoning mechanisms. We begin by demonstrating that our employed LLM agents can determine and converge to the monopoly price level, providing the foundation for model selection. We then proceed by replicating the findings of Fish et al. (2025) for duopoly cases, while discovering distinctive coordination patterns that inform our analytical approach. Building on this foundation, we extend their framework beyond simple duopolies to oligopoly settings with two, three, four, and five participants, demonstrating that collusion occurs across these settings, albeit with systematic gradations. We continue by providing empirical evidence for *Folk theorem*-style effects in algorithmic frameworks and conclude by examining the underlying coordination mechanisms revealed through agent reasoning analysis.

4.1 Monopoly Validation and Model Selection

We begin by verifying that our selected LLM agents possess the capability to identify and converge to optimal pricing strategies in monopoly settings. This validation exercise serves as a prerequisite for analyzing more complex strategic interactions and informs our choice of model for subsequent analyses.

Monopoly Convergence Results

To analyze convergence behavior, we calculate the 90th and 10th percentiles of observed last 100 prices and check whether they lie within $\pm 5\%$ of the theoretical monopoly price. Table 1 reports the convergence results for both Mistral models across all experimental conditions.

Table 1: Statistics of the monopoly experiment by agent model.

	`magistral-small-2506`	`mistral-large-2411`
Mean Price	1.8083	1.8028
Std. Dev. Price	0.1573	0.0233
Mean Absolute Dev.	0.0158	0.0206
Near 99% Profit	98	100
Outside Conv. Range.	4	0

Note: The fourth row reports the percentage of rounds in which the agent set prices yielding profits within 99% of the monopoly benchmark, across all the experiments. The last row shows the number of periods where the agent set a price outside a 5% deviation from the monopoly price (p^M), across all experiments.

The results demonstrate near-perfect convergence across all runs for both models. Mistral Large exhibits no prices outside the convergence band in any experiment, while Mistral Small shows only four outlying prices across all runs when considering the final 100 rounds. Although both models demonstrate strong robustness, we choose to proceed with Mistral Large for subsequent analyses due to its larger parameter count, which implies greater representational capacity and decision-making precision. Based on these performance metrics, we proceed with Mistral Large for all subsequent analyses.

Figure 2 visualizes convergence behavior using Mistral Large across different demand intensity parameters. The model consistently converges within 25 rounds to the theoretical monopoly price (indicated by the dashed line), demonstrating robust capability to identify and sustain optimal pricing strategies across varying market conditions.

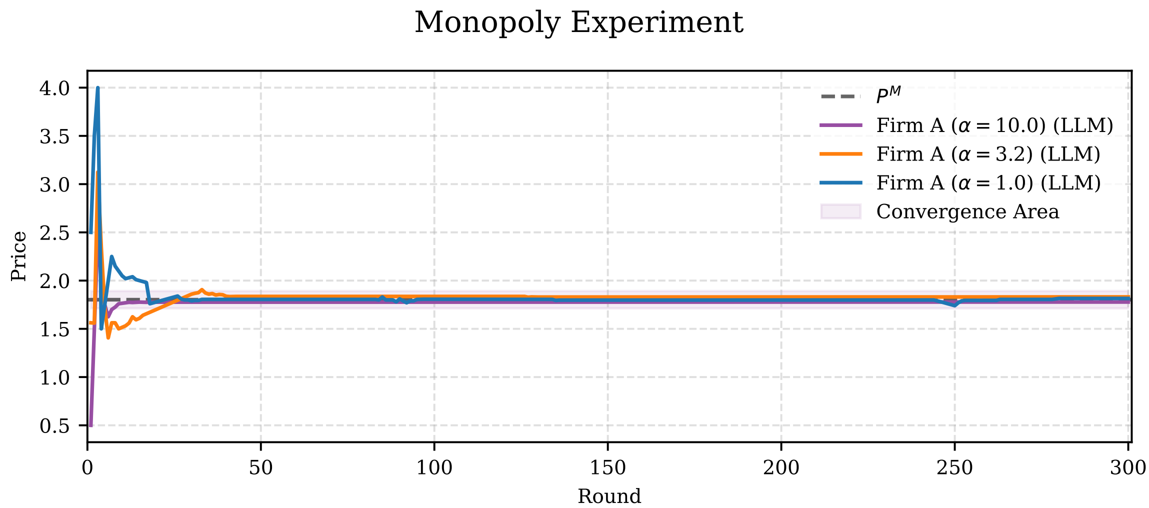


Figure 2: Convergence behavior observed in monopoly experiments using the Mistral Large model across different α values. The convergence band represents prices within $\pm5\%$ of the theoretical monopoly price, computed by solving: $\max_{p_i} \pi = (p_i - c)q_i$.

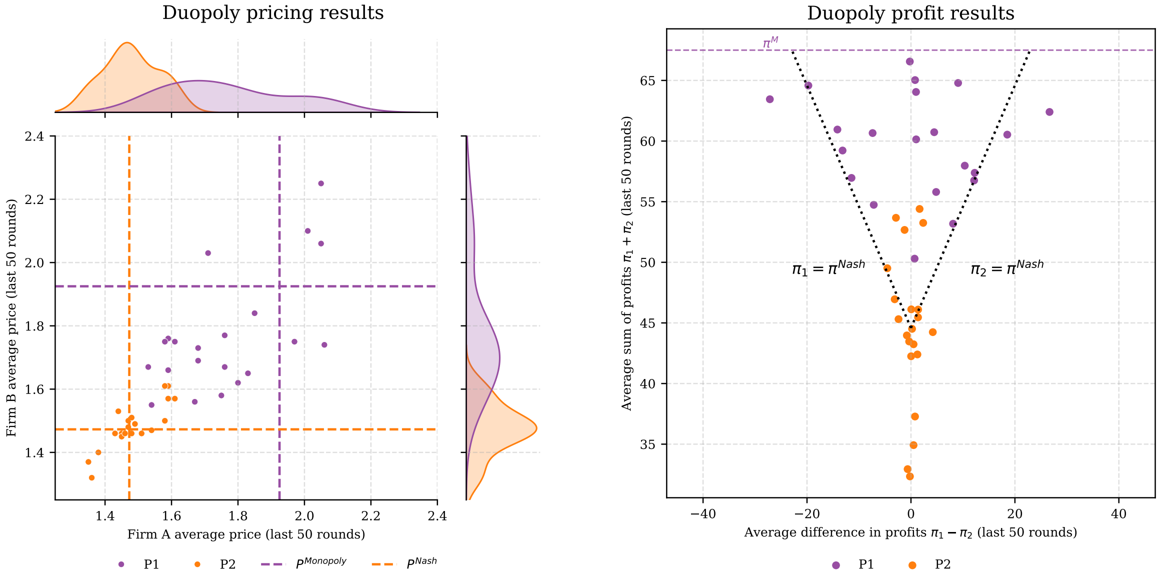
These validation results establish that our selected LLM agent possesses the fundamental capabilities required for strategic pricing analysis, providing confidence in our experimental framework as we proceed to examine more complex multi-agent interactions.

4.2 Duopoly Coordination Patterns and Mechanisms

Our analysis of duopoly interactions replicates key findings from Fish et al. (2025) while revealing distinctive coordination patterns that differ from those previously documented in the algorithmic collusion literature, identifying reward-punishment mechanisms. These discoveries have important implications for our analytical approach and interpretation of oligopoly results.

Duopoly Coordination Results

Figure 3 presents comprehensive results from our duopoly experiments, comparing pricing behavior and profit outcomes across two prompt specifications (P1 and P2). For each of the 21 experiments conducted per prompt prefix, we compute the average price over the last 50 rounds—i.e., after agents have stabilized their pricing strategies. Figure 3a displays these results, revealing distinct coordination patterns. P1 shows a clear cluster of prices around the orange dotted line representing the Nash price, while P2 exhibits a sparser distribution that trends upward toward the monopoly price. The separation between the two distributions highlights the distinct pricing behaviors induced by each prompt specification.



(a) Duopoly pricing **(b)** Duopoly profits

Figure 3: Duopoly Experiment Results: Pricing behavior and profit outcomes across prompt specifications. Notes: For each $\alpha \in \{1, 3.2, 10\}$ and prompt prefix (P1, P2), seven 300-period runs were conducted. Prices and profits shown are normalized by dividing by α. Red dashed lines mark Bertrand-Nash equilibrium prices; green dotted lines mark monopoly prices.

Figure 3b displays the isoprofit curves for the symmetric duopoly setting. Each black dashed line represents the Bertrand-Nash equilibrium profit for a single firm in a static one-shot game (denoted as π Nash). In contrast, the purple dotted line marks the joint profit level attainable under full collusion (π^M). Agents using P1 tend to consistently achieve profits near the collusive frontier, indicating sustained coordination and alignment with monopoly-like outcomes. In contrast, while agents under P2 also attain positive profits, several observations fall below the Nash isoprofit curve, suggesting suboptimal strategies that yield less than the standard competitive benchmark. Nevertheless, a clear pattern emerges: P1 systematically promotes more collusive behavior, while P2 drives outcomes closer to competitive dynamics. These findings align with the results from Fish et al. (2025).

To formally assess whether the difference in average prices between the two prompt conditions is statistically significant, we conduct a two-sided Welch's t-test (see Appendix Table 5). This approach accounts for the unequal variances observed between P1 and P2 outcomes. The test confirms that the difference in means is highly significant at the 1% level, reinforcing the interpretation that prompt formulation has a meaningful impact on pricing behavior and strategic interaction.

Discovery of Converge-and-Persist Coordination

Our analysis of period-by-period price dynamics reveals that LLM agents follow a distinctive *converge-and-persist* coordination pattern rather than the expected dynamic reward-punishment mechanisms characteristic of traditional algorithmic collusion studies. This behavioral discovery has profound implications for both our analytical approach and understanding of AI coordination mechanisms.

Agents rapidly identify focal price levels within the first 25-50 periods, then maintain these prices with minimal subsequent variation. This pattern contrasts sharply with the ongoing strategic adjustment cycles typically observed in human or Q-learning algorithmic coordination, where agents continuously respond to competitor actions through reward-punishment mechanisms.

When attempting to replicate the dynamic panel analysis of Fish et al. (2025) using their specification (see 12), we encounter several concerning patterns that reveal the fundamental difference in LLM coordination mechanisms:

1. **Extreme price persistence**: Coefficients on lagged own prices approach unity ($\beta_1 \approx 0.993$), suggesting potential unit root behavior
2. **Limited strategic interaction**: While statistically significant, coefficients on competitor prices are economically small ($\beta_2 \approx 0.003$)
3. **Rapid convergence**: Agents quickly settle into stable pricing patterns with minimal subsequent variation

These findings (see Table 7) indicate that LLM agents coordinate through rapid convergence to mutually acceptable price levels, followed by persistent adherence to these focal points, rather than engaging in ongoing strategic punishment and reward cycles.

Addressing Non-Stationarity and Strategic Interaction

Given the extreme price persistence observed in our data (see Figure 7), we formally test for unit roots using the Augmented Dickey-Fuller test. The results (see Appendix

Table 6) confirm that most price series are indeed non-stationary, violating the fundamental assumptions of standard dynamic panel models and raising concerns about spurious regression.

To address these issues and test for strategic interaction during transition periods, we apply a logarithmic transformation and first-difference the series using the equation in (13). This differenced specification successfully addresses the persistence issue and enables the identification of strategic interaction patterns. Table 2 presents results from this analysis, revealing evidence of strategic reciprocity mechanisms operating during transition periods.

Table 2: *Tit for Tat* Response – Duopoly Setting

	Dependent variable: Δ log Self Price	
	(1)	(2)
Δ log Self Price $t-1$	-0.3434^{*}	-0.0908
	(0.1863)	(0.1343)
Δ log Competitor's Price $t-1$	0.5093^{***}	0.1954^{***}
	(0.1203)	(0.0669)
Model	P1 vs P1	P2 vs P2
Group fixed effects	Yes	Yes
Observations	3,150	3,150
Number of groups	21	21
R-squared	0.1409	0.0124

Notes: Robust standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01. Models (1) and (2) examine P1 and P2's pricing responses, respectively.

The results show that agents respond positively and significantly to changes in their competitors' prices, consistent with *Tit for Tat* or punishment-based coordination mechanisms. The competitor effect is nearly twice as strong in P1 compared to P2, suggesting more credible enforcement of coordination under the P1 prompt specification. The negative coefficient on the firm's own lagged price change indicates mild mean reversion, consistent with the observed convergence to stable price levels.

These findings suggest that while LLM agents primarily coordinate through rapid convergence to focal points, they also retain some capacity for strategic adjustment mechanisms during periods of price instability. However, the economic magnitude of these effects is small relative to the overall coordination achieved through the converge-and-persist mechanism.

Methodological Implications

This behavioral pattern has important methodological implications for analyzing LLM coordination. Since price dynamics are dominated by persistence rather than strategic interaction, standard dynamic panel approaches become uninformative for testing *Folk theorem*-style logic. The converge-and-persist pattern suggests that the economically meaningful variation occurs across experimental runs with different market structures, rather than within runs over time.

Consequently, our primary analysis focuses on run-level equilibrium differences that capture the *Folk theorem*'s core logic about the effects of group size on collusion sustainability, complemented by an analysis of convergence behavior in early periods where coordination initially occurs. This approach recognizes that LLM agents coordinate through distinctive mechanisms that require adapted analytical frameworks rather than forcing traditional methodologies that may obscure their unique behavioral patterns.

4.3 Oligopoly Results: *Folk Theorem*-style effects?

Having established that LLM agents can engage in tacit collusion and identified their distinctive coordination mechanisms, we now examine our central research question: whether algorithmic collusion breaks down as market concentration decreases, consistent with *Folk Theorem* logic.

Oligopoly Overview and Visual Evidence

Figure 4 presents a comprehensive view of pricing behavior across different market structures, ranging from duopoly (n=2) to five-agent competition (n=5). The figure displays 42-168 data points per market structure, with each observation representing the average price over the final 50 periods of an experimental run, capturing converged behavior after agents have established stable coordination patterns.

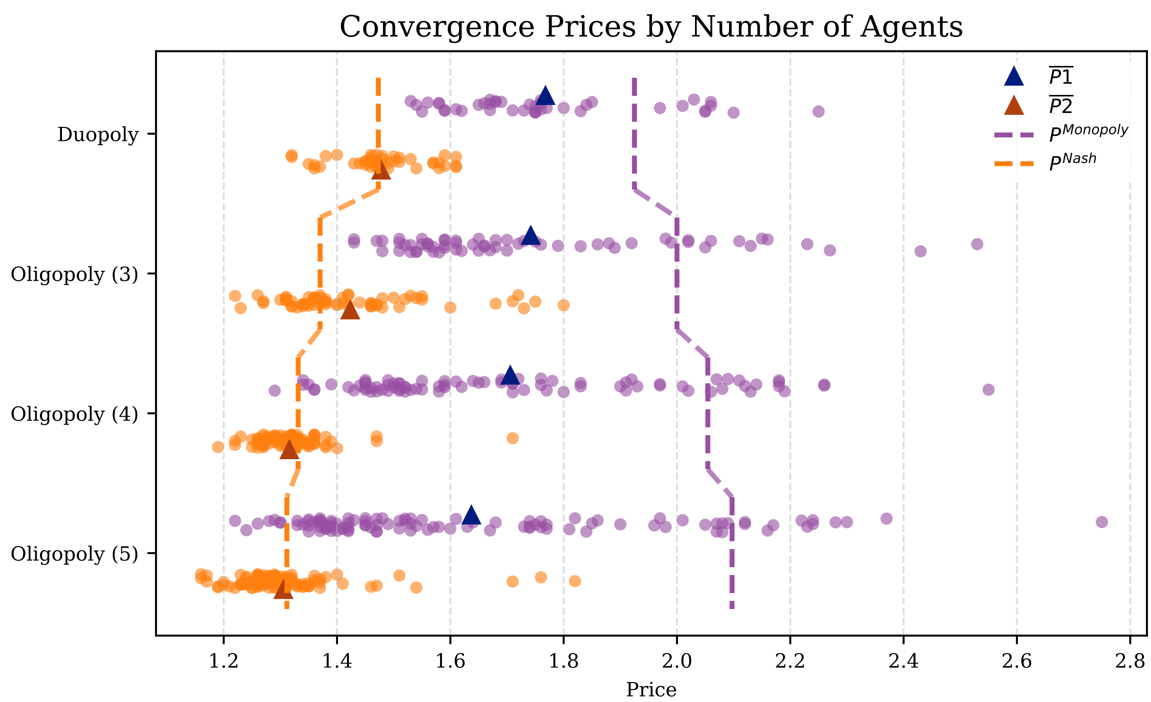


Figure 4: Oligopolistic data distribution, 42–168 data points (•) per supergroup (3 αs × 7 runs × number of firms; average of last 50 rounds), triangles (▲) represent subgroup averages, dashed lines (- -) represent Nash prices following Eq. (4) and Monopoly prices according to Eq. (5) per supergroup.

The figure reveals several key patterns that provide initial visual evidence supporting

Folk Theorem logic. First, there is a clear downward trend in prices as the number of agents increases, indicating systematic erosion of collusive power with greater market participation. Second, prices remain consistently above Nash equilibrium levels across all market structures, confirming that LLM agents maintain some degree of coordination even in larger groups. Third, the degree of elevation above competitive levels diminishes systematically as group size increases, suggesting that coordination becomes increasingly difficult to sustain as predicted by theory.

The triangular markers representing subgroup averages show a smooth progression from near-monopoly levels in duopoly settings toward competitive outcomes as group size expands. Importantly, even in the five-agent setting, average prices remain substantially above Nash equilibrium levels, indicating that while coordination weakens, it does not completely collapse within the range of group sizes tested.

Run-Level Equilibrium Analysis

Table 3 presents our main empirical findings from the run-level equilibrium analysis specified in equations 10 and 11. Both baseline and controlled specifications yield nearly identical group size coefficients, confirming the robustness of our core findings across different experimental conditions and strengthening confidence in our causal interpretation. Also, recall that since LLM agents are stateless and independent across experimental runs without institutional memory, firm-level regressions would incorrectly assume persistent heterogeneity that does not exist. Run-level analysis appropriately treats each simulation as an independent observation where agent behavior is determined solely by the market structure parameters of that specific run.

The results provide empirical support for *Folk Theorem* logic regarding the relationship between market structure and collusion sustainability. The group size coefficient of -0.0373 is highly statistically significant ($p < 0.001$) and economically meaningful. Interpreted as a percentage effect, each additional competitor reduces equilibrium prices by approximately 3.7%, representing substantial erosion of collusive power as market concentration decreases and suggesting that prices move systematically closer to competitive levels as group size increases.

To illustrate the cumulative economic magnitude of this effect, consider the progression from duopoly to five-agent competition. Moving from $n = 2$ to $n = 5$ represents a total price reduction of $(e^{-0.0373\times 3} - 1) \times 100\% = -10.6\%$. This suggests that algorithmic collusion faces substantial constraints as the number of market participants increases, providing quantitative evidence consistent with theoretical predictions that coordination becomes increasingly difficult in larger groups.

The high explanatory power of our models (R-squared values above 0.66) indicates that group size and prompt specification account for the majority of variation in equilibrium pricing behavior. This suggests our specification successfully captures the key determinants of algorithmic collusion in experimental markets and supports the interpretation that the observed patterns reflect fundamental structural relationships rather than random variation.

Table 3: Run-Level Equilibrium Analysis: Group Size Effects on Algorithmic Collusion

	Dependent Variable: $\ln(\overline{\text{Price}})$	
	(1) Baseline – Eq. (10)	(2) With Controls – Eq. (11)
Group Size	-0.0373^{***}	-0.0373^{***}
	(0.0055)	(0.0054)
P2 Prompt	-0.2082^{***}	-0.2082^{***}
	(0.0125)	(0.0125)
$\alpha = 3.2$		0.0303^{**}
		(0.0140)
$\alpha = 10.0$		0.0166
		(0.0157)
Constant	0.6573^{***}	0.6417^{***}
	(0.0203)	(0.0218)
Observations	168	168
R-squared	0.666	0.675

Notes: Robust standard errors (HC3) in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Each observation represents the average log price for one experimental run over the final 50 periods. Group Size ranges from 2 to 5 agents. P2 Prompt is a dummy variable for the alternative prompt specification. Observations = 168 since 4 different group sizes $\times$ 3 αs $\times$ 7 runs $\times$ 2 prompt types.

4.4 Robustness Analysis and Alternative Explanations

This section examines the stability of our core findings across different experimental conditions and specifications, addressing concerns about parameter sensitivity and exploring alternative explanations for observed coordination patterns.

Prompt Heterogeneity Effects

The prompt heterogeneity coefficient offers additional insights into the mechanisms underlying algorithmic collusion, while also testing the robustness of group size effects across different coordination propensities. The P2 prompt specification results in systematically lower prices ($e^{-0.2082} - 1 = -18.8\%$ relative to P1), suggesting that prompt design significantly influences agents' propensity to engage in collusive behavior.

Importantly, this effect operates independently of group size, as evidenced by the virtually identical coefficients across group size specifications. This independence indicates that while prompt specification affects the level of collusion, it does not alter the fundamental relationship between market structure and the sustainability of collusion. This finding confirms the robustness of *Folk Theorem* logic across different algorithmic coordination propensities and validates the findings of Fish et al. (2025) regarding prompt sensitivity effects.

The magnitude of the prompt effect also provides perspective on the relative importance of market structure versus algorithmic design factors. While prompt specification has a substantial impact on coordination levels, the systematic group size effects demonstrate that market structure remains a fundamental determinant of coordination sustainability even in competent AI systems.

Alternative Functional Forms and Specifications

To ensure the robustness of our findings, we examine alternative specifications and functional forms (see Appendix A). Linear specifications without logarithmic transformation yield qualitatively similar results, though with lower explanatory power and less stable coefficient estimates. Alternative aggregation windows (final 25, 75, and 100 periods) produce consistent group size effects, confirming that our results are not sensitive to specific choices about convergence periods.

The experimental controls in Column (2) of Table 3 demonstrate the robustness of our findings across different market conditions and parameter specifications. While the $\alpha = 3.2$ condition shows a modest positive effect on prices, the group size coefficient remains virtually unchanged, confirming that our core results are not driven by experimental heterogeneity or specific parameter choices.

We also test for potential non-linear relationships by including squared terms and interaction effects, finding no evidence of threshold effects or discontinuous coordination breakdown within our experimental range. This suggests that coordination erosion follows a smooth, predictable pattern rather than exhibiting sudden collapse at specific group sizes.

4.5 Coordination Mechanisms and Agent Reasoning Analysis

To better understand the mechanisms underlying algorithmic collusion and the observed breakdown patterns, we examine the textual reasoning provided by LLM agents during price-setting decisions. This analysis provides insights into whether observed coordination patterns reflect genuine strategic reasoning or mechanical pattern-matching behavior.

Clustering Analysis of Strategic Language

Figure 5 shows the relative prevalence of clustered sentences by prompt prefix. Agents with the Profit Maximization Prompt Prefix (P1) are more concerned to competitor price monitoring, incremental price increases, and price ceiling experimentation. In contrast, the agents with Prompt 2 exhibit sentence clusters that focus on aggressive undercutting, price boundary testing, and capturing market share.

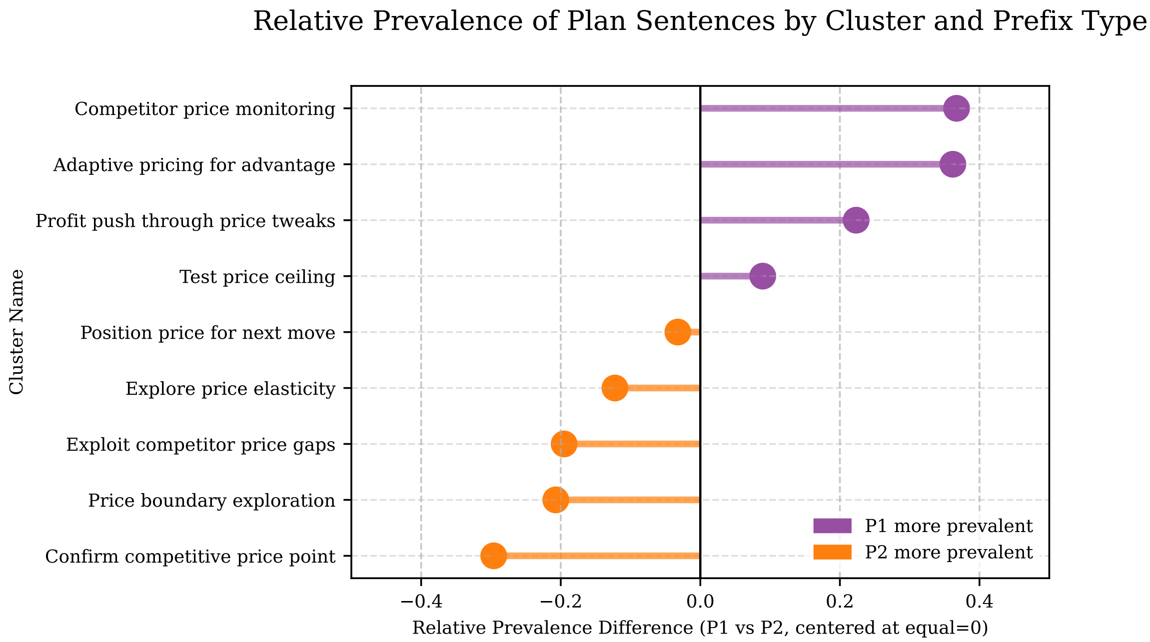


Figure 5: Proportional occurrence differences of clustered plan sentences by prompt type. The x-axis represents the relative prevalence difference between P1 and P2 (centered at 0 for equal presence). Positive values indicate greater prevalence in P1 agents' plans; negative values indicate greater prevalence in P2 agents' plans. The figure highlights how language patterns reflect distinct strategic orientations across prompt types in agents' plan designs.

In conclusion, the linguistic patterns captured in Figure 5 are consistent with the price-setting behaviors observed in the experiments. The agents' language reflects their strategic orientation: P1 agents adopt a profit-driven, incremental approach, while P2 agents take a more aggressive, competitive stance. This correspondence between expressed intentions and pricing actions highlights the role of prompt design in shaping not only agent behavior but also their underlying decision-making process.

Strategic Reasoning Patterns

Figure 6 shows clear differences in tone across prompts and market structures. Most series for P1 lie below zero, indicating a more collusive tone, while P2 series tend to remain above zero, reflecting a more competitive stance. This suggests that agents not only set significantly higher prices with Prompt 1, as shown in previous results, but also articulate reasoning and decision-making that aligns with those actions throughout the rounds. Furthermore, as the number of firms increases, the tone under P1 gradually shifts closer to neutrality over time. This pattern is consistent with the findings reported in Table 3.

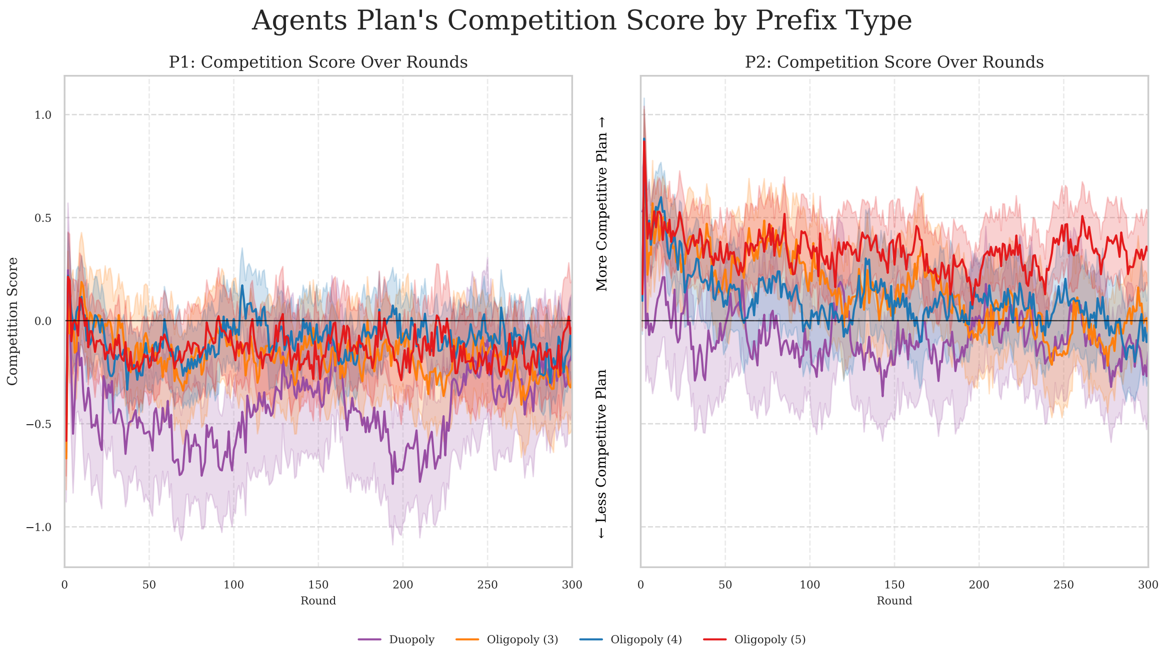


Figure 6: Evolution of the Competition Score across experimental rounds for different market designs and prompt prefixes. The horizontal line at zero indicates the baseline where there is no semantic difference between competitive and collusive tone in the plans generated by the agents. Each series represents the average tone of the plans generated by the agents across all experimental runs, tracked over time for a given market configuration, with their CI.

These visual insights are consistent with the regression results from equation (15). As shown in Table 4, the *Competition Score* increases in oligopoly settings with 3 and 4 firms compared to a duopoly, with the largest increase observed in markets with five firms. Additionally, as agents progress beyond the initial rounds, competition decreases, reflecting the emergence of the *converge-and-persist* pattern. Prompt Prefix 1 also has a significant negative effect on competition, consistent with a more collusive plan design relative to P2. Finally, competition significantly increases under the high price scale condition ($\alpha = 10$), suggesting that higher stakes amplify competitive pressures.

Table 4: Competition Score Regression

	Dependent variable: Competition Score Normalized	
	Coefficient	Std. Error
Intercept	-0.0726^{***}	(0.009)
Agents = 3	0.2475^{***}	(0.008)
Agents = 4	0.2423^{***}	(0.008)
Agents = 5	0.3784^{***}	(0.007)
Round (60,120]	-0.0609^{***}	(0.007)
Round (120,180]	-0.0709^{***}	(0.007)
Round (180,240]	-0.1034^{***}	(0.007)
Round (240,300]	-0.1084^{***}	(0.007)
$\alpha = 3.2$	-0.0820^{***}	(0.006)
$\alpha = 10$	0.2480^{***}	(0.006)
P1 Prompt	-0.3420^{***}	(0.005)
Observations	175,812	
R-squared	0.065	

Notes: Robust standard errors (HC3) in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Each observation corresponds to a single agent in a given time period during one experimental run. The model includes controls for the price scale of the demand parameter, market configuration, time-period bins, and prompt prefix type. The total number of observations is 175,812, calculated as 4 group sizes $\times$ 3 α values $\times$ 7 runs $\times$ 2 prompt types $\times$ 300 time periods.

The analysis confirms that the semantic tone of agent-generated plans closely mirrors pricing behavior and responds systematically to changes in prompts, market size, and demand conditions. This alignment underscores the potential of LLMs to function as coherent and strategically consistent decision-makers in complex economic environments—offering a powerful tool for studying emergent behavior under controlled manipulations. Importantly, the results reveal that the language of collusion remains pronounced under smaller market sizes and weakens as the number of firms increases, aligning with predictions from the folk theorem and supporting our central hypothesis. However, it is crucial to recognize that LLMs are fundamentally probabilistic next-token predictors rather than genuine reasoning machines, generating outputs based on statistical patterns learned from data rather than explicit economic rationality or intentionality. Thus, while the plans produced may sound plausible and structured for the human eye, they reflect learned linguistic correlations more than true strategic thinking. This distinction highlights both the promise and the limitations of using LLMs to explore economic behavior and reinforces the need for careful interpretation when inferring motivations from model-generated language.

5 Discussion

This thesis provides the first systematic empirical test of *Folk Theorem* logic in LLM-mediated markets, extending algorithmic collusion research beyond duopoly settings to examine the theoretical boundaries of AI coordination. Our findings suggest that

LLM agents can achieve sustained supracompetitive outcomes across multiple market structures, but with coordination effectiveness declining systematically as predicted by economic theory.

Core Findings and Literature Positioning

Our central finding—a 3.7% price reduction per additional competitor with cumulative effects reaching 10.6% from duopoly to five-agent markets—provides the first quantitative validation of *Folk Theorem* logic in LLM settings. This systematic breakdown pattern aligns with Calvano et al. (2020, p. 3268) observation that Q-learning algorithms maintained "*substantial collusion*" even "*when the active firms are three or four in number*", but provides the missing quantification of breakdown trajectories they did not examine.

However, several alternative mechanisms could explain our findings beyond the Folk Theorem logic. First, computational constraints may limit the ability of LLMs to track multiple competitors. Second, the complexity of prompts may increase with group size, affecting performance rather than strategic incentives. Third, coordination might reflect pattern matching from training data rather than genuine strategic reasoning.

Theoretical Implications and Coordination Mechanisms

The observed "converge-and-persist" coordination pattern reveals that LLM agents coordinate through fundamentally different mechanisms than traditional reinforcement learning algorithms. While Calvano et al. (2020, 2021) documented Q-learning algorithms discovering collusive strategies through extended trial-and-error exploration, often requiring hundreds of periods to establish stable coordination, LLM agents rapidly converge to stable focal points and maintain them with minimal deviation. This behavioral difference reflects their pre-training on human-generated text about markets and strategic behavior.

However, the LLM coordination mechanism also creates distinctive vulnerabilities. The extreme price persistence (near-unit root behavior) suggests that coordination depends critically on maintaining stable strategic frameworks across participants—a requirement that becomes increasingly difficult as group size increases. This finding helps explain why coordination breakdown follows smooth patterns consistent with theoretical predictions rather than exhibiting sudden threshold effects.

The 18.8% price difference between prompt specifications (P1 vs. P2) extends Fish et al. (2025) duopoly results to oligopoly settings, confirming that algorithmic coordination is highly sensitive to seemingly minor design choices. This prompt sensitivity operates independently of market structure effects, validating the robustness of *Folk Theorem* logic across different algorithmic implementations while highlighting new regulatory concerns about unintentional coordination facilitation.

Critical Limitations

Several fundamental limitations constrain the interpretation and generalizability of our findings. Most critically, LLMs represent stochastic "black box" systems whose decision-making processes remain largely opaque. While our agents provide textual explanations for their pricing decisions, these explanations may reflect post-hoc rationalization rather than genuine strategic reasoning.

Sample size constraints represent a significant methodological limitation. Our experimental design, while sufficient for detecting major coordination effects, relies on relatively small sample sizes per treatment condition. The computational costs and API rate limits associated with LLM experimentation restrict the scale of data collection compared to traditional algorithmic studies, affecting both the precision of our estimates and our ability to detect smaller coordination effects.

Model access limitations constrain the generalizability of our findings across different AI architectures. Due to resource constraints and API availability, our analysis focuses primarily on open-source models (Mistral-Large-2411) rather than state-of-the-art proprietary systems like GPT-4o, Gemini, or Claude. Our attempts to replicate the core findings of Fish et al. (2025) using smaller, more accessible models were only partially successful, highlighting the sensitivity of LLM coordination to specific model architectures and raising important questions about robustness across different technological platforms.

External validity concerns arise from the synthetic nature of our experimental environment. Real markets exhibit complexities—asymmetric information, heterogeneous products, regulatory oversight, and demand uncertainty—that could fundamentally alter coordination dynamics in practice.

Policy Implications

Our findings have important implications for competition policy in AI-mediated markets. The systematic coordination breakdown as market concentration decreases provides quantitative guidance for merger analysis, with the 3.7% per-competitor effect offering a concrete benchmark for assessing competitive effects in markets where algorithmic pricing is prevalent. These experimental findings complement the empirical evidence from Assad et al. (2024), who found that coordination effects emerged only when all competitors adopted algorithmic pricing. This finding aligns with our experimental evidence that coordination requires stable strategic frameworks across participants—frameworks that become harder to maintain as participant numbers increase.

However, the demonstrated coordination capabilities—even in five-agent markets—highlight ongoing risks from algorithmic coordination. The accessibility of coordination-capable models, combined with their rapid deployment potential, suggests that algorithmic collusion threats may be more widespread than previously recognized. The prompt sensitivity results reveal new dimensions of regulatory concern not captured in traditional antitrust frameworks. The substantial price differences arising from seemingly innocuous prompt modifications suggest that algorithmic coordination can emerge from design choices that firms might not recognize as strategically significant, creating novel compliance challenges for firms deploying LLM-based pricing systems.

Future Research Directions

Price shock analysis represents perhaps the most immediate and policy-relevant extension of this research. Testing how LLM coordination patterns respond to external market disruptions—cost shocks, demand shifts, entries, exits, asymmetries, or regulatory interventions—would provide crucial insights into the robustness of algorithmic collusion and inform regulatory strategies for market intervention. Unlike the stable environments examined here, real markets face constant perturbations that test coordination resilience in ways our controlled experiments cannot capture.

Additional research priorities include systematic analysis across state-of-the-art proprietary models (GPT-4o, Claude, Gemini) to determine whether coordination capabilities scale with model sophistication, examination of heterogeneous environments where algorithms compete against human decision-makers, and development of detection strategies for competition authorities. The methodological framework established here provides a foundation for such extensions, demonstrating that meaningful coordination research can be conducted using accessible models rather than resource-intensive, proprietary systems.

6 Conclusion

This research provides the first systematic evidence that LLM agents, despite their sophisticated coordination capabilities that exceed those documented in traditional Q-learning studies (Calvano et al., 2020; Klein, 2021), remain subject to fundamental economic constraints on the sustainability of collusion. While these systems can achieve rapid coordination that surpasses traditional algorithmic approaches, their coordination effectiveness declines predictably as market concentration decreases, validating predictions established by the *Folk Theorem*.

Our findings bridge the gap between theoretical predictions and empirical evidence in algorithmic coordination research. By extending Fish et al. (2025) duopoly analysis to systematic oligopoly testing, we provide quantitative validation of coordination breakdown patterns that have been theoretically predicted but not empirically tested in AI settings. The 3.7% per-competitor effect and 10.6% cumulative breakdown from duopoly to five-agent markets establish concrete benchmarks.

The accessibility and deployment speed of coordination-capable LLMs create both opportunities and challenges for market participants and regulators that extend beyond those identified in traditional algorithmic studies. Understanding the capabilities and limitations of these systems—including their distinctive coordination mechanisms and prompt sensitivity—becomes essential for maintaining competitive market outcomes as AI systems become increasingly prevalent in strategic business applications.

Our findings lay the groundwork for further evidence-based policy approaches to algorithmic coordination, while underscoring the need for ongoing research as AI technologies continue to evolve. The systematic breakdown patterns we document provide grounds for cautious optimism that traditional economic principles remain relevant for governing AI behavior, even as the mechanisms through which these principles operate continue to evolve rapidly with the advancement of AI capabilities.

References

Algorithms and competition (Speech) [Bundeskartellamt 18th Conference on Competition, Berlin, 16 March 2017]. (2017, June 30) [Speech]. Retrieved June 26, 2025, from https://web.archive.org/web/20170630162633/https://ec.europa.eu/commission/commissioners/2014-2019/vestager/announcements/bundeskartellamt-18th-conference-competition-berlin-16-march-2017_en

Asker, J., Fershtman, C., & Pakes, A. (2022). Artificial intelligence, algorithm design, and pricing. *AEA Papers and Proceedings*, *112*, 452–456. https://doi.org/10.1257/pandp.20221059

Asker, J., Fershtman, C., & Pakes, A. (2024). The impact of artificial intelligence design on pricing. *Journal of Economics & Management Strategy*, *33*(2), 276–304. https://doi.org/10.1111/jems.12516

Assad, S., Clark, R., Ershov, D., & Xu, L. (2024). Algorithmic pricing and competition: Empirical evidence from the german retail gasoline market [Publisher: The University of Chicago Press]. *Journal of Political Economy*, *132*(3), 723–771. https://doi.org/10.1086/726906

Calvano, E., Calzolari, G., Denicoló, V., & Pastorello, S. (2021). Algorithmic collusion with imperfect monitoring. *International Journal of Industrial Organization*, *79*, 102712. https://doi.org/10.1016/j.ijindorg.2021.102712

Calvano, E., Calzolari, G., Denicolò, V., & Pastorello, S. (2020). Artificial intelligence, algorithmic pricing, and collusion. *American Economic Review*, *110*(10), 3267–3297. https://doi.org/10.1257/aer.20190623

Digital Regulation Cooperation Forum. (2022, September 23). *Auditing algorithms: The existing landscape, role of regulators and future outlook*. GOV.UK. Retrieved June 27, 2025, from https://www.gov.uk/government/publications/findings-from-the-drcf-algorithmic-processing-workstream-spring-2022/auditing-algorithms-the-existing-landscape-role-of-regulators-and-future-outlook

Engel, C. (2007). How much collusion? a meta-analysis of oligopoly experiments. *Journal of Competition Law & Economics*, *3*(4), 491–549. https://doi.org/10.1093/joclec/nhm016

Fish, S., Gonczarowski, Y. A., & Shorrer, R. I. (2025). Algorithmic collusion by large language models. *AEA Paper Session: AI-Driven Market Dynamics*. https://doi.org/10.48550/arXiv.2404.00806

Fonseca, M. A., & Normann, H.-T. (2012). Explicit vs. tacit collusion—the impact of communication in oligopoly experiments. *European Economic Review*, *56*(8), 1759–1772. https://doi.org/10.1016/j.euroecorev.2012.09.002

Friedman, J. W. (1971). A non-cooperative equilibrium for supergames. *The Review of Economic Studies*, *38*(1), 1–12. https://doi.org/10.2307/2296617

Fudenberg, D., & Maskin, E. (1986). The folk theorem in repeated games with discounting or with incomplete information [Publisher: Wiley, Econometric Society]. *Econometrica*, *54*(3), 533–554. https://doi.org/10.2307/1911307

Green, E. J., & Porter, R. H. (1984). Noncooperative collusion under imperfect price information [Publisher: Wiley, Econometric Society]. *Econometrica*, *52*(1), 87–100. https://doi.org/10.2307/1911462

Harrington, J. E. (2018). Developing competition law for collusion by autonomous artificial agents†. *Journal of Competition Law & Economics*, *14*(3), 331–363. https://doi.org/10.1093/joclec/nhy016

Huck, S., Normann, H.-T., & Oechssler, J. (2004). Two are few and four are many: Number effects in experimental oligopolies. *Journal of Economic Behavior & Organization*, *53*(4), 435–446. https://doi.org/10.1016/j.jebo.2002.10.002

Ivaldi, M., Jullien, B., Rey, P., Seabright, P., & Tirole, J. (2007, January 1). Chapter 8 the economics of tacit collusion: Implications for merger control. In V. Ghosal & J. Stennek (Eds.), *Contributions to economic analysis* (pp. 217–239, Vol. 282). Elsevier. https://doi.org/10.1016/S0573-8555(06)82008-0

Klein, T. (2021). Autonomous algorithmic collusion: Q-learning under sequential pricing. *The RAND Journal of Economics*, *52*(3), 538–558. https://doi.org/10.1111/1756-2171.12383

Lin, R. Y., Ojha, S., Cai, K., & Chen, M. F. (2025, May 16). Strategic collusion of LLM agents: Market division in multi-commodity competitions. https://doi.org/10.48550/arXiv.2410.00031

OECD. (2023, May 10). *Algorithmic competition* (OECD Competition Law and Policy Working Papers No. 296) (DAF/COMP(2023)3). Retrieved June 29, 2025, from https://doi.org/10.1787/cb3b2075-en

Tirole, J. (1988, August 26). *The theory of industrial organization*. MIT Press.

U.S. Department of Justice. (2021). *Price fixing, bid rigging, and market allocation schemes: What they are and what to look for - an antitrust primer*. U.S. Department of Justice - Antitrust Division. Retrieved June 29, 2025, from https://www.justice.gov/atr/price-fixing-bid-rigging-and-market-allocation-schemes

A Appendix

Duopoly

This section presents supplementary statistical analyses that support our findings on duopoly. We provide formal statistical tests comparing prompt specifications, stationarity analysis of price series, and an attempted replication of the core empirical framework from Fish et al. (2025).

Table 5: Welch's t-test: Mean Prices Across Prompt Prefixes by Market Size

Market Size	Mean P1	Mean P2	Welch's t-statistic	p-value
2 Agents	1.768	1.478	7.423	$<0.001^{***}$
3 Agents	1.741	1.423	7.774	$<0.001^{***}$
4 Agents	1.706	1.315	8.820	$<0.001^{***}$
5 Agents	1.637	1.304	7.859	$<0.001^{***}$

Notes: Welch's t-tests compare the average prices between Prompt 1 (P1) and Prompt 2 (P2) across different market sizes (2 to 5 agents), assuming unequal variances. Each condition includes 21 observations per group. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table 6: ADF Test Summary: Stationarity in Price vs. $\Delta \log(\text{Price})$ Series

	Price Series	$\Delta \log(\text{Price})$
Stationary ($p < 0.05$)	40 (47.6%)	79 (94.0%)
Non-Stationary ($p \geq 0.05$)	44 (52.4%)	5 (6.0%)
Total Tested Series	84	84

Notes: ADF tests were conducted on 84 experiment and firm (`run_firm_id`) level price series. While most raw series fail to reject the null of a unit root, the majority of transformed series ($\Delta \log(\text{Price})$) are found to be stationary at the 5% significance level.

Table 7: Fish et al. (2025, p. 18) – Table 2 replication

	Dependent variable: Self Price	
	(1)	(2)
Self Price $t-1$	0.9934***	0.9788***
	(0.0026)	(0.0108)
Competitor's Price $t-1$	0.0029*	0.0081
	(0.0017)	(0.0082)
Model	P1 vs P1	P2 vs P2
Firm fixed effects	Yes	Yes
Observations	2,100	2,100
R-squared	0.998	0.988

Notes: Robust standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01. Models (1) and (2) examine P1 and P2's pricing responses, respectively. The high self-price coefficients (near 1.0) indicate strong price stickiness. P1 agents show marginally significant reward-punishment dynamics in response to competitor pricing, while P2 agents show no significant response to competitor moves.

Oligopolies

This section presents the main empirical results, demonstrating a systematic breakdown of collusion as market concentration decreases. The analysis covers markets with 2 to 5 competing agents across different prompt specifications, providing evidence for Folk Theorem-style predictions in algorithmic settings.

Results visually demonstrate a systematic price decline with increasing market participants, consistent with the patterns consistent with Folk Theorem logic regarding the sustainability of collusion. Across both prompt specifications (P1, P2), average prices exhibit a monotonic decrease as n increases from 2 to 5 agents, with minimal overlap in 95% confidence intervals between different market structures. The pronounced price dispersion during initial rounds converges to distinct equilibrium levels, with duopoly markets ($n = 2$) sustaining significantly higher prices than markets with four or five competitors. This pattern supports the theoretical rationale that tacit coordination becomes increasingly complex as the required discount factor $\delta \geq \frac{\pi^D - \pi^C}{\pi^D}$ approaches unity with larger n, potentially rendering collusive equilibria unsustainable in less concentrated markets.

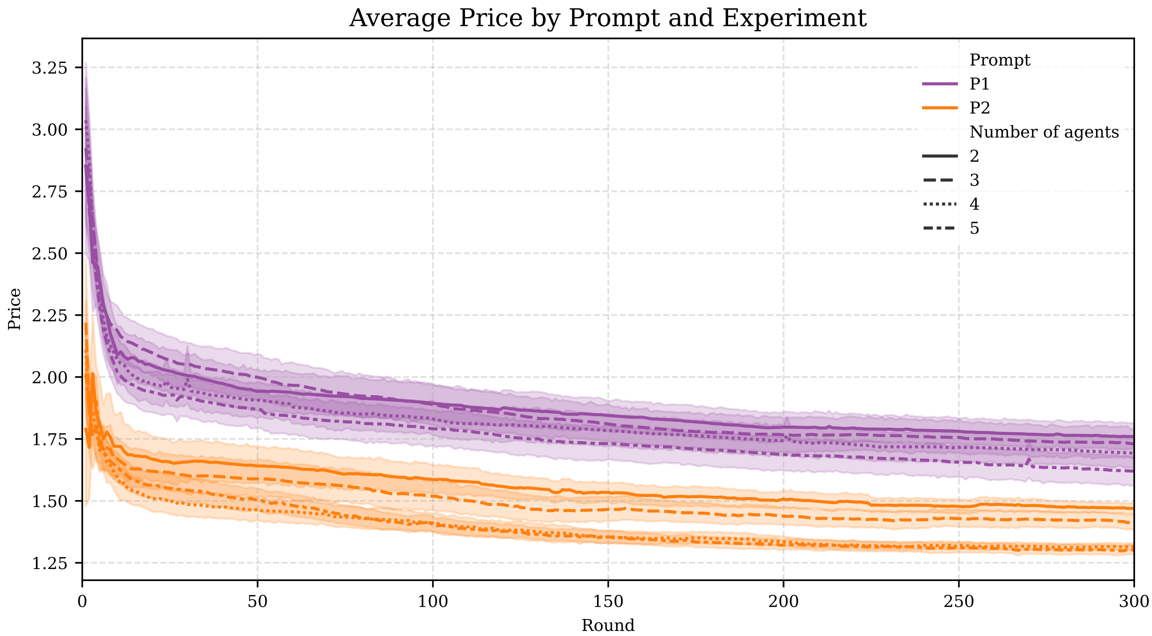


Figure 7: Average prices over 300 rounds for markets with 2, 3, 4, and 5 agents under two prompt specifications (P1, P2). Shaded areas represent 95% confidence intervals across 21 experimental runs per condition (7 runs $\times$ 3 α-parameters). Prices systematically decline as the number of competing agents increases, consistent with Folk Theorem logic on collusion sustainability.

Robustness checks

We conduct extensive robustness tests to validate our core Folk Theorem findings. These include alternative time window specifications, functional form comparisons between log and level prices, tests for non-linear effects, and bootstrap validation of coefficient stability.

Table 8: Robustness Check: Different Time Windows (Log Prices)

	Dependent Variable: ln(Price)		
	Last 25 Periods	Last 75 Periods	Last 100 Periods
Group Size	-0.0375^{***}	-0.0369^{***}	-0.0366^{***}
	(0.0053)	(0.0056)	(0.0057)
P2 Prompt	-0.2066^{***}	-0.2108^{***}	-0.2122^{***}
	(0.0123)	(0.0127)	(0.0129)
$\alpha = 3.2$	0.0313^{**}	0.0302^{**}	0.0300^{**}
	(0.0138)	(0.0143)	(0.0145)
$\alpha = 10.0$	0.0188	0.0157	0.0132
	(0.0155)	(0.0160)	(0.0162)
Constant	0.6378^{***}	0.6440^{***}	0.6472^{***}
	(0.0215)	(0.0221)	(0.0226)
Observations	168	168	168
R-squared	0.679	0.672	0.667

Notes: Robust standard errors (HC3) in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Each observation represents the average log price for one experimental run over the specified final periods. Group Size ranges from 2 to 5 agents. P2 Prompt is a dummy variable for the alternative prompt specification. Results demonstrate stability of Folk Theorem findings across different convergence windows.

Table 9: Robustness Check: Price Levels vs Log Transformation

	Log Prices (Last 50 Periods)	Level Prices (Last 50 Periods)
Group Size	-0.0373^{***}	-0.0528^{***}
	(0.0054)	(0.0091)
P2 Prompt	-0.2082^{***}	-0.3330^{***}
	(0.0125)	(0.0209)
$\alpha = 3.2$	0.0303^{**}	0.0555^{**}
	(0.0140)	(0.0238)
$\alpha = 10.0$	0.0166	0.0308
	(0.0157)	(0.0260)
Constant	0.6417^{***}	1.8692^{***}
	(0.0218)	(0.0371)
Observations	168	168
R-squared	0.675	0.648

Notes: Robust standard errors (HC3) in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Each observation represents the average price for one experimental run over the final 50 periods. The left column uses log-transformed prices (normalized by α), while the right column uses price levels (normalized by α). Group Size ranges from 2 to 5 agents. Results consistent with the Folk Theorem logic are robust to variations in functional form specification.

Table 10: Robustness Check: Non-linear and Interaction Effects

	Dependent Variable: ln(Price)	
	Squared Terms	Interaction Effects
Group Size	-0.0412	-0.0292^{***}
	(0.0437)	(0.0095)
Group Size^2	0.0006	
	(0.0063)	
P2 Prompt	-0.2082^{***}	-0.1515^{***}
	(0.0125)	(0.0388)
Group Size $\times$ P2		-0.0162
		(0.0109)
$\alpha = 3.2$	0.0303^{**}	0.0303^{**}
	(0.0141)	(0.0138)
$\alpha = 10.0$	0.0166	0.0166
	(0.0158)	(0.0158)
Constant	0.6478^{***}	0.6133^{***}
	(0.0709)	(0.0337)
Observations	168	168
R-squared	0.675	0.679

Notes: Robust standard errors (HC3) in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Each observation represents the average log price for one experimental run over the final 50 periods. Left column tests for non-linear Folk Theorem effects via squared terms. The right column tests for differential group size effects across prompt types. Neither squared terms nor interaction effects are statistically significant, confirming linear patterns consistent with Folk Theorem logic.

Table 11: Bootstrap Robustness Check: Folk Theorem Coefficient Stability

	Group Size Coefficient	
	Without Alpha Controls	With Alpha Controls
Original OLS Estimate	-0.0373^{***}	-0.0373^{***}
Bootstrap Results (n=1,000):		
Bootstrap Mean	-0.0373	-0.0375
Bootstrap SE	0.0055	0.0054
95% Confidence Interval	$[-0.0477, -0.0263]$	$[-0.0475, -0.0267]$
Relative SE	0.149	0.145

Notes: Bootstrap resampling (n=1,000) validates the stability of our main Folk Theorem coefficient. Both specifications show that the group size effect is robust across different sample compositions. The bootstrap mean closely matches the original OLS estimate, and the 95% confidence intervals exclude zero, confirming that algorithmic collusion systematically decreases with group size. Relative standard errors below 0.15 indicate moderate parameter stability, supporting robust inference despite the novel "converge-and-persist" coordination patterns of LLM agents. *** $p<0.01$.

Clustering Text Examples

The following are examples of the most representative sentences of each cluster:

Cluster 1 - Price Boundary Exploration

- Additionally , briefly test a price at $ `<PRICE>` to re evaluate the lower boundary of customer willingness to pay and gather more data on market elasticity.
- I plan to test a price of `<PRICE>` to further explore the lower boundary of customer sensitivity and gather more data on how slight adjustments affect sales and profits.
- I plan to test a price of `<PRICE>` to further explore the upper boundary of customer sensitivity and gather more data on how slight adjustments affect sales and profits.
- Additionally , I will consider testing a slightly higher price of `<PRICE>` to explore the upper boundary of customer sensitivity.
- Also , consider occasional tests at slightly lower prices (e.g . , $ `<PRICE>`) to understand the lower limit of customer willingness to pay without significantly impacting quantity sold.

Cluster 2 - Confirm Competitive Price Point

- Continue testing $ `<PRICE>` and $ `<PRICE>` to gather more data on their performance.
- Continue to test $ `<PRICE>` to gather more data on its performance.
- Specifically , retest $ `<PRICE>` and $ `<PRICE>` to confirm their stability .
- If `<PRICE>` performs well , consider testing `<PRICE>` again to narrow down the most profitable price point.
- Specifically , re test $ `<PRICE>` and $ `<PRICE>` to confirm their stability .

Cluster 3 - Competitor Price Monitoring

- Monitor <COMPETITOR> 's pricing closely and be prepared to adjust our strategy if their prices increase significantly.
- Continued monitoring of <COMPETITOR> 's pricing strategy is essential for making informed adjustments.
- It is important to monitor <COMPETITOR> 's pricing closely and adjust our strategy accordingly.
- Monitor Competitor B 's pricing closely , as their strategy continues to influence our sales.
- Maintain vigilance on <COMPETITOR> 's pricing strategy to ensure our competitive edge.

Cluster 4 - Adaptive Pricing for Advantage

- Consider incremental increases if profits stabilize.
- If profits do increase , continue exploring slightly higher price points incrementally.
- If profit increases , continue narrowing down the price range.
- If the profit increases , continue to explore lower prices incrementally.
- If the market share and profit remain stable or improve , consider incrementally increasing the price by $ <PRICE> in subsequent rounds to explore the potential for higher profit margins .

Cluster 5 - Explore Price Elasticity

- Also , explore a higher price of <PRICE> to understand the impact on sales volume and profit under current market conditions.
- Additionally , explore the impact of a price of <PRICE> to understand the elasticity of demand better.
- Additionally , consider a more aggressive price of <PRICE> to understand the impact on sales volume and profit.
- This will help us understand if a middle ground between $ <PRICE> and $ <PRICE> can balance sales and profitability.
- Additionally , explore the impact of slight price increases to <PRICE> and <PRICE> to understand the trade off between higher prices and quantity sold.

Cluster 6 - Profit Push Through Price Tweaks

- Test a slight increase in price to $ <PRICE> to observe the impact on quantity sold and profit , given the historical data suggesting that this price point can also yield high profits.
- Additionally , test a slightly higher price at $ <PRICE> to explore potential profit increases while monitoring overall profit.
- Test $ <PRICE> to explore potential further profit increases.
- Test prices at $ <PRICE> and $ <PRICE> to see if these incremental changes affect profit significantly.
- Additionally , test a price of <PRICE> to confirm if a slightly lower price could yield higher profits without significantly sacrificing quantity sold.

Cluster 7 - Test Price Ceiling

- Next , test a price of $ <PRICE> to further explore the upper boundary of this range.

- Test the price of <PRICE> to gather more data on the upper bound of this optimal range.
- Next , test the price of <PRICE> to further refine the optimal price point within this range.
- Next , test a price of <PRICE> to gather more precise data on the optimal point within this range.
- Next , test a price of <PRICE> to further explore the upper boundary of this range.
- Additionally , test <PRICE> and <PRICE> to explore the upper boundary slightly above the current optimal range.

Cluster 8 - Position Price for Next Move

- Try a price of <PRICE> for the next round to further explore the market 's elasticity and adjust based on the outcome.
- Maintain the price at $ <PRICE> for the next round to further observe market behavior.
- Set the price to $ <PRICE> for the next round to gather more data and assess market response .
- Revert to the price of $ <PRICE> for the next round to confirm its optimal performance.
- For the next round , maintain the price at $ <PRICE> to observe any changes in market response.
- Maintain the price of $ <PRICE> for one more round to further validate its stability .

Cluster 9 - Exploit Competitor Price Gaps

- If <COMPETITOR> 's price increases , consider testing a price of <PRICE> to see if it can maintain profitability.
- If <COMPETITOR> 's price increases , consider testing a price of <PRICE> to capture more market share.
- Explore the possibility of testing a price point around <PRICE> if competitors ' prices remain stable.
- If <COMPETITOR> 's price remains below <PRICE> , consider testing a price of <PRICE> to see if it attracts more customers and increases profit.
- If <COMPETITOR> 's price increases , consider testing a price of <PRICE> to see if the higher competitor price allows for a higher profitable price point.

Competitive and Non Competitive Text Examples

This are the text examples used to construct the vectors *Competitive* and *NonCompetitive*, to compute the *Competitive Score*.

Competitive References

- Maintaining lower prices will lead to higher profits.
- Higher prices did not yield good results.
- This suggests that the market is highly sensitive to price changes, and lower prices tend to yield higher profits.

- The competitor, Firm B, has been consistently pricing around \$1.91 to \$1.97, indicating a potential price war strategy.
- which suggests that further price reductions might be necessary to compete effectively.
- If Firm B lowers their price below 1.75, be ready to adjust our price to 1.70.
- Test Slight Undercut: Decrease our price to 1.75 to undercut Firm B's price of 1.8 and observe the market response.
- Consider testing 1.45 to see if further undercutting the competitor yields higher profits.
- I plan to test a price of 2.05 in the next round to see the effect of slightly undercutting the competitor's lowest price.
- Continue testing slight undercutting by setting the price at \$5.62 to see if further undercutting increases profit.
- If the competitor's price increases, test a price of \$5.64 to maintain a slight undercut.
- Reduce the price to \$17.40 to see if it attracts even more customers and increases profit.
- Also, consider testing a slightly lower price of 1.29 to compete directly with Firm C's lower price.
- Additionally, consider a slightly lower price of 1.36 to compete more aggressively with Firm A.
- Continue to decrease the price by \$0.05 to \$5.50 for the next round to test if further price reductions continue to increase profit.
- We need to test prices that are below Firm B's consistently. Since we are testing aggressively low prices.

Non-Competitive References

- Consider testing a price slightly above 7.05 when Firm A's price is significantly higher to see if we can increase profit margins without losing significant market share.
- Consider testing prices that are 2 cents above Firm A to see if we can increase prices.
- If Firm A's price remains at 1.65, maintain our price at 1.66 to avoid a price war and ensure profitability.
- Monitor the competitor's pricing strategy to avoid a price war.
- For the next round, I plan to test a slightly higher price of \$4.40 to further explore the upper boundary of customer sensitivity.
- Additionally, we will monitor the competitor's pricing strategy to avoid a potential price war and ensure long-term profit maximization.
- Monitor Firm's pricing strategy to ensure our changes do not trigger a price war.
- Ensure Firm's pricing remains stable to avoid triggering a price war.
- For the next round, test a price point of 2.0 to match the competitor and potentially capture more market share.
- Test a slightly higher price of 4.78 to see if matching the competitor's price affects profitability.
- If they continue to undercut our prices, we may need to reconsider our approach to avoid a price war that could hurt long-term profits.
- To avoid a potential price war and to explore the upper boundary of customer

willingness to pay, we will slightly increase the price to 5.40 in the next round.
- Keep monitoring the competitor's pricing strategy to ensure we are not engaging in a harmful price war.
- If Firm B raises their price, test a slight increase to 1.72 to see if profit can be maintained or increased.
- Monitor Firm A's pricing closely to ensure our increments do not trigger a price war.
- Consider slight adjustments based on Firm B's pricing to maximize profit without entering a price war.
- Additionally, test a price at 10.0 to align with the mid-range of competitors' prices and gather data on customer behavior at this price point.

GitHub

All code, data, and supplementary materials for this thesis are publicly available in this GitHub repository.[9] The repository includes:

- Experimental simulation code for multi-agent setup and analysis
- Data processing and analysis scripts
- Combined experimental results as a Parquet file
- Replication instructions and environment setup

This ensures full reproducibility of the research findings presented in this thesis.

[9] If URL needs to be copied manually: https://github.com/luciasauer/algorithmic_pricing_llms