参赛队员姓名：何沐蓉

中学：深圳中学

省份：广东省

国家/地区：中国，广东

指导教师姓名：王奕君，沐年国

指导教师单位：深圳中学，上海理工大学

论文题目： Algorithmic Tacit Collusion in Differentiated Digital Markets: A Three-Layer Game-Theoretic Model with Q-Learning Evidence and Welfare-Optimal Regulatory Mechanism Design

2026 T.Y.U. High School Science Award
仅用于2026年成国中学科学奖论文公示

Algorithmic Tacit Collusion in Differentiated Digital Markets: A Three-Layer Game-Theoretic Model with Q-Learning Evidence and Welfare-Optimal Regulatory Mechanism Design

2026全国青少年人工智能创新挑战赛 Youth High School Science Award
仅用于2026年度青少年科学论文公示

Abstract

How the information architecture endogenously chosen by a profit-maximizing platform shapes the sustainability of algorithmic tacit collusion among horizontally differentiated sellers is the question addressed in this study. We embed a three-layer sequential game within the Salop circular-city framework, in which the platform sets an information-transparency parameter $\sigma \in [0, 1]$ controlling the rival-price visibility of sellers' Q-learning algorithms and a proportional commission rate τ, while symmetric sellers engage in infinitely repeated algorithmic price competition. Three contributions are made. Theoretically, a closed-form critical discount factor $\bar{\delta}(\sigma) = mN/(4t\sigma + mN)$ is derived under endogenous platform information design, from which a *Transparency Paradox* follows: a profit-maximizing platform selects maximal transparency to lower the patience threshold for collusion, creating a structural conflict with consumer welfare. A *Commission Neutrality Theorem* establishes that proportional commissions scale all profit terms symmetrically and leave collusive incentives unchanged, separating information governance from surplus redistribution. Extending the model to partial consumer coverage yields a closed-form deadweight-loss expression $DWL(m) = N(m - \bar{m})^2/(2t)$, characterizing precisely when welfare invariance breaks down. Computationally, multi-agent Q-learning simulations across a ten-level transparency grid confirm all five directional comparative-static predictions from the closed-form expression. A behavioral departure from classical theory is identified: bounded-rational pricing algorithms endogenously develop finite-period tit-for-tat retaliation rather than the permanent grim-trigger punishment assumed in theory. This divergence reflects a fundamental incompatibility between Q-learning's myopic re-optimization logic and the pre-commitment assumption underlying grim-trigger credibility: once sustained punishment erodes Q-values for low-price states, agents learn that brief retaliation followed by cooperative recovery dominates permanent reversion in cumulative discounted reward. A counterfactual grim-trigger simulation confirms that this milder punishment reduces equilibrium collusion intensity by roughly one half relative to the fully rational benchmark. On the policy side, a unified welfare-decomposition framework evaluates three anti-collusion instruments and identifies information-transparency restriction as the most cost-effective standalone regulatory tool, a conclusion robust to administrative-cost sensitivity analysis and to the partial-coverage extension.

Keywords: algorithmic collusion; closed-form discount factor; Q-learning; Salop circular city; platform information design; finite tit-for-tat punishment

Contents

1 Introduction **5**
- 1.1 Background . . . 5
- 1.2 Core Contributions . . . 6
- 1.3 Literature Review . . . 7
  - 1.3.1 Classical Oligopoly Collusion Theory . . . 7
  - 1.3.2 Algorithmic Pricing and Collusion . . . 7
  - 1.3.3 Platform Information Design and Competition Policy . . . 7

2 Three-Layer Sequential Game-Theoretic Model **8**
- 2.1 Model Overview and Game Timing . . . 8
- 2.2 Consumer Demand in the Circular City Model . . . 10
- 2.3 Static Price Competition and Symmetric Nash Equilibrium . . . 11
- 2.4 Collusive Pricing, Deviation Profit and Baseline Critical Discount Factor . . . 11
  - 2.4.1 Collusive Price and Profit . . . 11
  - 2.4.2 Optimal Unilateral Deviation . . . 12
  - 2.4.3 Grim-Trigger Strategy and Critical Discount Factor . . . 12
- 2.5 Platform Information Transparency and Modified Critical Factor . . . 13
  - 2.5.1 Detection Probability under Imperfect Monitoring . . . 13
  - 2.5.2 Modified Critical Discount Factor . . . 14
  - 2.5.3 Collusion Threshold Transparency . . . 15
- 2.6 Platform Optimal Information and Commission Choice . . . 15

3 Main Theoretical Results **17**
- 3.1 Welfare Analysis under Full Market Coverage . . . 17
- 3.2 Extension to Partial Market Coverage . . . 18
  - 3.2.1 Marginal Consumer Participation Constraint . . . 18
  - 3.2.2 Demand under Partial Coverage . . . 18
  - 3.2.3 Deadweight Loss Expression . . . 18
- 3.3 Platform-Regulator Conflict and Equilibrium Characterization . . . 19

4 Q-Learning Simulation Design **20**
- 4.1 Reinforcement-Learning Architecture . . . 20
  - 4.1.1 Q-Value Update Rule . . . 20
  - 4.1.2 ε-Greedy Exploration Policy . . . 21
  - 4.1.3 State Space under Transparency Variation . . . 21
  - 4.1.4 Collusion Index . . . 21
- 4.2 Parameter Calibration . . . 22

5 Simulation Results and Theoretical Verification **23**
- 5.1 Price Convergence Dynamics . . . 23
- 5.2 Effect of Discount Factor on Collusion Intensity . . . 24
- 5.3 Transparency Paradox Validation with Continuous Grid . . . 25
- 5.4 Effect of Seller Count on Collusion Intensity . . . 26
- 5.5 Commission Neutrality Validation . . . 26
- 5.6 Counterfactual Deviation Experiment: Emergent Finite Tit-for-Tat Punishment . . . 27
- 5.7 Counterfactual Grim-Trigger Simulation . . . 28
- 5.8 Robustness to Alternative Parameter Calibrations and Hyperparameter Sensitivity . . . 29

2026 Yau Hua School Science Award 仅用于2026年华罗庚科学奖公众公示

5.9 Theory-Simulation Consistency Summary 30

6 Model Extensions **30**
6.1 Asymmetric Sellers 30
6.2 Partial Coverage with Endogenous Participation 31

7 Regulatory Mechanism Design **31**
7.1 Information Transparency Restriction 31
7.1.1 Implementation Challenges 32
7.2 Algorithm Audit with Financial Penalties 32
7.3 Real-Time Automatic Price Monitoring 32
7.4 Comparative Analysis 33
7.5 Welfare Quantitative Evaluation 33
7.6 Hierarchical Regulatory Design for High-Frequency Markets 35

8 Discussion **35**
8.1 Theoretical Implications 35
8.2 Policy Implications 36
8.3 Limitations 36

9 Conclusion **37**

10 Acknowledgements **39**

A Symbol Table **42**

B Endogenous Collusive Markup: Formal Problem and Analytical Difficulties **42**

2026 S.-T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

1 Introduction

1.1 Background

The competitive landscape of modern retail markets has been fundamentally reshaped by digital e-commerce platforms. Algorithmic pricing tools, now widely adopted, allow sellers to adjust prices automatically on the basis of real-time market data, competitor behavior and consumer demand patterns, and this has transformed how products are bought and sold [1]. Platforms such as Amazon, Taobao and Meituan go well beyond passively connecting buyers with sellers; they actively shape the competitive environment through information-disclosure policies, recommendation mechanisms and commission structures. Improved market responsiveness and operational efficiency are important benefits of algorithmic pricing, yet novel antitrust concerns arise at the same time, concerns for which existing regulatory frameworks were not designed [2].

The central concern motivating this research is algorithmic tacit collusion. Through repeated interaction and mutual learning, independent pricing algorithms deployed by competing sellers may spontaneously settle on supra-competitive prices even though no explicit communication or agreement exists among the sellers [1]. Evidence of explicit coordination, such as written agreements or direct communication, is traditionally required by antitrust law to establish illegal price-fixing. A critical regulatory gap therefore opens: existing legal standards are unable to address the same anticompetitive outcome when it arises from trial-and-error learning by autonomous algorithms engaged in repeated price competition [3].

Landmark experimental evidence was given by Calvano, Calzolari, Denicolò and Pastorello [1], who demonstrated that fully autonomous Q-learning pricing algorithms can learn supra-competitive pricing strategies in repeated Bertrand competition. Independent algorithms in their simulations spontaneously evolved strategies sustaining prices above the competitive equilibrium, with neither pre-programmed cooperative rules nor any channel for inter-algorithm communication. Considerable attention from both the academic community and regulatory authorities worldwide has followed from this finding [4, 5].

The applicability of current findings to real-world digital markets is limited by two critical simplifications in the existing literature on algorithmic collusion. Homogeneous-product markets, in which all goods are identical, are assumed in most prior work including the foundational study of Calvano et al. [1, 6]. In actual e-commerce settings, products differ substantially along dimensions such as brand identity, quality ratings, delivery speed and customer service. Competitive incentives and the conditions for collusive sustainability are fundamentally altered by this horizontal differentiation. Collusive markups are raised while the sensitivity of demand to unilateral price cuts is dampened. How these countervailing forces jointly shape the critical discount factor for collusion has received limited formal attention.

An equally important gap involves the role of digital platforms, which act as active market architects rather than passive intermediaries. Existing models treat the platform as a neutral background actor with no strategic capacity [1, 6]. Real platforms, however, control critical aspects of the market environment, including the degree to which rival pricing data is made available to sellers' algorithms. Sellers' ability to identify price-cutting deviations and to enforce retaliatory punishment is fundamentally shaped by this transparency choice, and whether tacit collusion can be sustained is thereby directly influenced [7]. In China, large platforms such as Taobao and Meituan supply merchants with extensive real-time competitive analytics dashboards covering rival pricing, sales volume and promotional activities. Third-party algorithmic pricing services, used widely by small and medium merchants, depend heavily on such data to automate pricing strategies [20, 21]. Whether profit-maximizing platforms have endogenous incentives that facilitate or prevent algorithmic collusion, and how those incentives interact with product differentiation and

2026 S. T. Yau High School Science Award
仅用于2026丘成桐中学科学奖公开公示

market structure, are questions that remain open in the literature.

A dual-purpose tension lies at the heart of this study. On one hand, platform information transparency helps consumers compare prices and make well-informed purchasing decisions. On the other, the same transparency feeds sellers' algorithms with the rival price data needed to identify deviations from collusive pricing and to launch retaliatory price wars. A platform that seeks to maximize commission revenue must choose between facilitating competition, which benefits consumers but lowers transaction value, and enabling collusion, which raises transaction value but harms consumers. Our goal is to formalize this strategic trade-off, characterize its equilibrium consequences and design regulatory mechanisms that resolve the resulting market failures.

1.2 Core Contributions

Three principal contributions to the existing literature are made in this study. We note at the outset that the Salop circular-city demand framework [10, 11], the Q-learning algorithmic pricing paradigm [1, 6] and the infinitely repeated game collusion theory [8, 9] originate from prior work. What is new lies in the integration of these elements and in the analytical results, behavioral findings and policy evaluation that follow.

The primary theoretical contribution is the derivation of a closed-form critical discount factor $\bar{\delta}(\sigma) = mN/(4t\sigma + mN)$, where m, N, t and σ denote the collusive markup, the number of sellers, the differentiation intensity and the platform transparency level respectively, under endogenous platform information design in a differentiated market. This expression is strictly decreasing and convex in σ, and a *Transparency Paradox* follows: profit-maximizing platforms select full transparency, which lowers the patience threshold for collusion and raises equilibrium prices along with commission revenue. The classical imperfect monitoring framework of Green and Porter [9] is extended by endogenizing the precision of monitoring through the platform's information architecture. We also establish the *Commission Neutrality Theorem* and state its exact boundary conditions in the main text. The model is further extended to partial market coverage and a deadweight-loss expression is derived, showing when welfare invariance breaks down and under what conditions collusion generates allocative inefficiency beyond pure redistribution. Because the baseline model assumes a monopoly platform, we additionally analyze how the Transparency Paradox is weakened under inter-platform competition, thereby clarifying the domain in which the results apply.

The primary computational contribution is a novel behavioral finding from multi-agent Q-learning simulations: bounded-rational pricing algorithms autonomously develop finite-period tit-for-tat retaliation strategies rather than the permanent grim-trigger punishment assumed in classical theory. A detailed behavioral mechanism explaining why Q-learning agents converge to this milder punishment form is given, and the finding is validated through a counterfactual simulation in which grim-trigger punishment is imposed and the resulting collusion gap is measured. A ten-level continuous transparency grid is used in the simulations to validate the closed-form comparative statics, and robustness checks across multiple parameter calibrations ensure that the results are not driven by special boundary configurations.

The policy contribution is a unified welfare-decomposition framework in which three anti-collusion instruments are quantitatively evaluated. Information-transparency restriction is identified as the most cost-effective standalone policy. Real-world implementation challenges for each instrument are discussed, and connections are drawn to existing regulatory frameworks in the European Union [18] and China [19, 20].

2026 S.-T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

1.3 Literature Review

1.3.1 Classical Oligopoly Collusion Theory

Two prerequisites for sustained supra-competitive pricing were identified in Stigler's seminal analysis: the gains from short-term deviation must be outweighed by long-term punishment losses, and rivals must be able to identify price-cutting deviations with sufficient speed and accuracy. Information visibility was recognized as the core determinant of how efficiently deviations are detected, an insight that directly motivates the formalization of the platform transparency parameter $\sigma \in [0, 1]$ in our model.

The formal game-theoretic framework for analyzing collusion using the Folk Theorem and grim-trigger punishment strategies in infinitely repeated games was laid out by Fudenberg and Tirole [8]. Green and Porter [9] extended the analysis to settings with imperfect price information and demonstrated that collusion can persist even when rivals cannot perfectly observe each other's prices, as long as punishment phases are triggered with appropriate probability. Our detection-probability framework builds directly on this finding. The canonical spatial competition framework for horizontal product differentiation was introduced by Salop [10], and the symmetric Nash equilibrium under N sellers on the circular market was solved by Tirole [11]; we adopt the demand system from that analysis. Despite this extensive body of theoretical work, very few studies have embedded Salop differentiation into a collusion sustainability analysis in which the information detection probability is endogenous.

1.3.2 Algorithmic Pricing and Collusion

Spontaneous coordination of supra-competitive prices by independent Q-learning algorithms in homogeneous Bertrand markets was demonstrated by Calvano et al. [1]. Klein [6] extended the analysis to staggered price commitment while retaining the homogeneous-product restriction. The first large-scale empirical evidence of algorithmic coordination, based on retail gasoline data from Germany, was documented by Assad, Clark, Ershov and Xu [4]. Conditions under which algorithmic pricing leads to higher or lower equilibrium prices, depending on commitment structure and observability, were identified by Brown and MacKay [13]. That the specific reinforcement-learning architecture affects both the likelihood and the degree of supra-competitive pricing was demonstrated by Asker, Fershtman and Pakes [14]. The conditions under which simple learning algorithms converge to collusive outcomes were characterized by den Boer, Meylahn and Schinkel [15]. In a subsequent study, Calvano et al. [7] investigated algorithmic collusion under imperfect monitoring and found that partial observability dampens but does not eliminate supra-competitive pricing, a finding that connects directly to the transparency variation examined here.

1.3.3 Platform Information Design and Competition Policy

A unified framework for information design in games was developed by Bergemann and Morris [17], establishing foundational tools for analyzing how an informed intermediary can influence strategic interactions. Johnson, Rhodes and Wildenbeest [16] studied platform design when sellers use pricing algorithms and gave a direct theoretical antecedent for our model. Several OECD roundtable reports [5] warned that real-time price comparison functionalities facilitate algorithmic tacit collusion, though only qualitative assessments were offered with no formal mathematical link between transparency and the critical discount factor. The legal dilemmas posed by algorithmic collusion were studied by Harrington [3]. The European Digital Markets Act [18] and China's Algorithm Recommendation Management Regulations [19] represent initial regulatory responses, yet both lack the quantitative foundation needed to calibrate specific policy instruments. A compre-

2026 S. -T. Yau High School Science Award
仅用于2026丘成桐中学数学论文公开

hensive overview of the competitive risks associated with algorithmic pricing was given by Ezrachi and Stucke [2]. In the Chinese context, competition and collusion risks in platform ecosystems were studied by Fang et al. [20], and algorithmic pricing regulation in Chinese digital markets was examined by Zhou and Li [21].

All three identified gaps are addressed simultaneously in our work. A three-layer sequential game integrates Salop differentiated demand, seller repeated collusion and endogenous platform information optimization. A Q-learning simulation framework with continuous transparency variation is built to verify the main theoretical predictions and to characterize emergent punishment strategies. Three anti-collusion regulatory tools are designed and quantitatively evaluated under a unified welfare-decomposition standard.

2 Three-Layer Sequential Game-Theoretic Model

The formal construction of the three-layer model is presented in this section. We introduce the foundational setup and then develop each layer in sequence: consumer demand in the circular city, static and repeated seller price competition, and platform information together with commission optimization.

Having identified the theoretical gaps motivating this analysis, we now formalize the strategic interaction among platform, sellers and consumers. The structure that follows provides the basis from which the main analytical results are derived in subsequent sections.

2.1 Model Overview and Game Timing

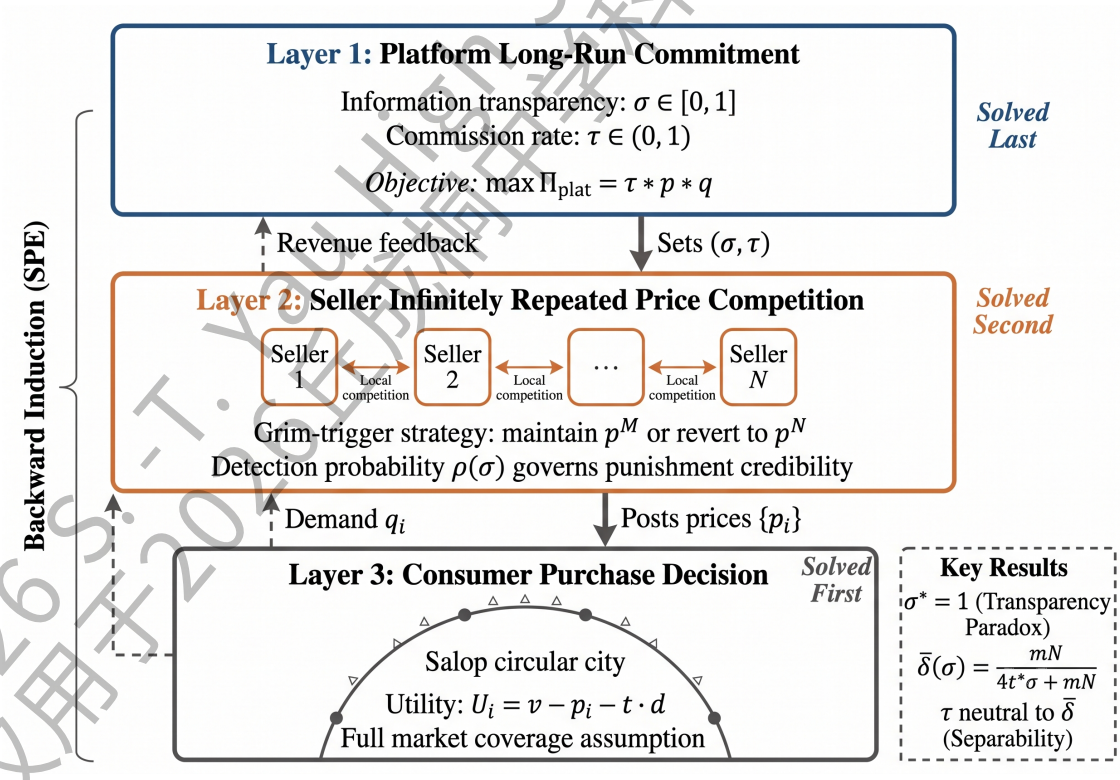


Figure 1: Three-layer sequential game-theoretic model framework.

Backward induction is used to solve the model illustrated in Figure 1 and to obtain the subgame-perfect equilibrium (SPE). Consumer demand is derived at Layer 3, seller equilibrium pricing at

2026 S.-T. Yau High School Science Award
仅用于2026年丘成桐中学科学奖公示

Layer 2 and the platform's optimal information together with commission choices at Layer 1. We consider a digital e-commerce platform on which $N \geq 2$ symmetric sellers compete by setting prices for horizontally differentiated products. A proportional commission $\tau \in (0, 1)$ is charged by the platform on all transactions, and an information-transparency parameter $\sigma \in [0, 1]$ is set, governing the extent to which rival price data is disclosed to sellers' pricing algorithms.

Assumption 1 (Salop Circular City)**.** *A continuum of consumers with total mass normalized to 1 is uniformly distributed along a circle of circumference 1. The N sellers are equally spaced on this circle, with arc distance $1/N$ between any two adjacent sellers. A consumer located at arc distance d from seller i obtains utility $U_i = v - p_i - td$, where v is the reservation value, p_i the posted price and t the unit transport cost. The reservation value v is assumed large enough that every consumer purchases exactly one unit (full market coverage).*

Assumption 2 (Symmetric Sellers)**.** *All sellers share identical marginal production cost c, identical algorithmic pricing technology and face identical platform commission rate τ. Each seller deploys an independent Q-learning reinforcement-learning algorithm that adjusts prices on the basis of observed market outcomes, with no inter-algorithm communication.*

Assumption 3 (Information Transparency)**.** *The platform's transparency parameter σ determines the probability $\rho(\sigma)$ that a unilateral price deviation by one seller is detected by rival algorithms within one period. The detection function $\rho : [0, 1] \to [0, 1]$ is continuous and strictly increasing, with $\rho(0) = 0$ and $\rho(1) = 1$. A linear specification $\rho(\sigma) = \sigma$ is adopted in the baseline analysis.*

Remark 1 (Salop Framework and Digital Market Characteristics)**.** *The Salop circular-city model captures two defining features of differentiated digital markets that distinguish them from traditional retail settings. First, horizontal differentiation in digital markets arises from non-price seller attributes that consumers value idiosyncratically: platform reputation scores, user-interface quality, delivery speed, after-sales service and brand identity. These attributes generate product-specific consumer loyalty naturally represented by the transport cost parameter t: a higher t reflects stronger brand lock-in and weaker cross-seller price sensitivity, consistent with the observed stickiness of consumer preferences on platforms such as Amazon, Taobao and Meituan. Second, and distinctively, digital platforms possess the technical capacity to control the granularity and timeliness of competitor price data that sellers' algorithms receive, through the design of merchant-facing analytics dashboards and API disclosure policies. No close analogue of this information-architecture power exists in traditional retail markets, where rival prices must be observed through direct market interaction. The transparency parameter σ formalizes this platform capability and is the primary channel through which platform design influences collusion sustainability. The proportional commission τ captures the platform business model in which revenue scales with transaction value, creating the platform's endogenous incentive to sustain high equilibrium prices.*

Analytical tractability is the motivation behind the linear detection specification, since it gives rise to a closed-form critical discount factor. The qualitative predictions are found to be robust when considering alternative functional forms, as demonstrated in Proposition 10 below. Under a concave specification (e.g., $\rho(\sigma) = \sqrt{\sigma}$), early increases in transparency provide large gains in the detection probability, so that $\bar{\delta}(\sigma)$ declines steeply at low transparency and flattens at high transparency. Under a convex specification (e.g., $\rho(\sigma) = \sigma^2$), the detection probability responds slowly until transparency is already high, which means that the marginal collusion-facilitating effect of transparency is concentrated in the upper range. Because the monotonicity of $\bar{\delta}(\sigma)$ is preserved under all such specifications, the Transparency Paradox is a robust structural phenomenon rather than an artifact of the particular functional form.

The model takes the form of a perfect-information sequential game.

2026 Shi Yuan High School Science Award 复旦2026年科学奖公示

Layer 1 (Platform, solved last): The platform simultaneously sets a fixed commission τ and a permanent information rule σ to maximize total commission revenue $\Pi^{plat} = \tau \cdot \sum_{i=1}^{N} p_i q_i$, and it perfectly anticipates seller equilibrium pricing.

Layer 2 (Sellers, solved second): Given a fixed pair (τ, σ), the N symmetric algorithmic sellers play an infinitely repeated simultaneous-price game in each period. The grim-trigger collusive strategy serves as the theoretical benchmark: sellers maintain the collusive price p^M as long as no deviation is detected and permanently revert to the competitive price p^N once a deviation is identified.

Layer 3 (Consumers, solved first): After sellers post prices, each consumer selects the seller providing maximum net utility, and market demand q_i for each seller is thereby determined.

2.2 Consumer Demand in the Circular City Model

For any two adjacent sellers i and $i+1$, the marginal consumer in which indifference between them holds is located at arc distance x^R from seller i. This distance satisfies the equal-utility condition:

$$v - p_i - tx^R = v - p_{i+1} - t\left(\frac{1}{N} - x^R\right) \tag{1}$$

Solving gives:

$$x^R = \frac{1}{2N} + \frac{p_{i+1} - p_i}{2t} \tag{2}$$

Symmetrically, for the marginal consumer located between seller i and its left neighbor $i-1$:

$$x^L = \frac{1}{2N} + \frac{p_{i-1} - p_i}{2t} \tag{3}$$

Total demand for seller i is obtained by summing the left-side and right-side market shares:

$$q_i = x^L + x^R = \frac{1}{N} + \frac{p_{i-1} + p_{i+1} - 2p_i}{2t} \tag{4}$$

Remark 2 (Local Competition)**.** *Strict local competition is imposed by the Salop circular-city model: the demand facing each seller depends only on the prices of its two immediate neighbors. No direct demand effect is exerted by price changes of non-adjacent sellers, i.e., $\partial q_i / \partial p_j = 0$ for $|j - i| > 1$. This structural feature holds for all $N \geq 3$ and governs how the deviation profit is calculated, because only adjacent rivals maintain collusive prices while the deviator undercuts. Implications of this local-competition structure for the scalability of the simulation results are discussed in Section 4.*

The demand function possesses the following comparative-static properties:

$$\frac{\partial q_i}{\partial p_i} = -\frac{1}{t} \quad \text{(own-price effect)} \tag{5}$$

$$\frac{\partial q_i}{\partial p_{i-1}} = \frac{\partial q_i}{\partial p_{i+1}} = \frac{1}{2t} \quad \text{(cross-price effect for adjacent rivals)} \tag{6}$$

A larger value of the differentiation parameter t makes demand less responsive to own-price changes and creates a customer moat that benefits each seller.

Under symmetric pricing $p_1 = p_2 = \cdots = p_N = p$, every seller obtains an equal share of the consumer mass: $q_i = 1/N$ for all i, and $\sum_{i=1}^{N} q_i = 1$, consistent with the full-coverage assumption.

2026 S. T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

2.3 Static Price Competition and Symmetric Nash Equilibrium

Seller i retains a fraction $(1-\tau)$ of per-unit revenue after paying the proportional commission. Combining marginal cost c with demand q_i gives the net profit function:

$$\pi_i = (1-\tau)(p_i - c)q_i = (1-\tau)(p_i - c)\left(\frac{1}{N} + \frac{p_{i-1} + p_{i+1} - 2p_i}{2t}\right) \tag{7}$$

Because $(1-\tau)$ is a positive constant that scales all profit values without affecting the price optimization trade-off, the best-response function is obtained by maximizing π_i over p_i:

$$\frac{\partial \pi_i}{\partial p_i} = (1-\tau)\left[q_i + (p_i - c)\frac{\partial q_i}{\partial p_i}\right] = 0 \tag{8}$$

Substituting the demand derivatives and rearranging gives:

$$p_i^* = \frac{p_{i-1} + p_{i+1}}{4} + c + \frac{t}{2N} \tag{9}$$

Proposition 1 (Symmetric Nash Equilibrium)**.** *Under symmetric equilibrium $p_1 = p_2 = \cdots = p_N = p^N$, the competitive equilibrium price and profit are:*

$$p^N = c + \frac{t}{N}, \qquad \pi^N = \frac{(1-\tau)t}{N^2} \tag{10}$$

Proof. Substituting $p_{i-1} = p_{i+1} = p^N$ into Equation (9):

$$p^N = \frac{p^N + p^N}{4} + c + \frac{t}{2N} = \frac{p^N}{2} + c + \frac{t}{2N} \tag{11}$$

Solving gives $p^N = c + t/N$. The equilibrium demand is $q^N = 1/N$, and equilibrium profit is $\pi^N = (1-\tau)(p^N - c)/N = (1-\tau)t/N^2$. □

The equilibrium price p^N increases with differentiation t, since stronger customer loyalty supports a larger markup, and decreases with the seller count N, since more intense competition compresses margins. Only the magnitude of profits is scaled by τ; the equilibrium price level itself is unaffected.

2.4 Collusive Pricing, Deviation Profit and Baseline Critical Discount Factor

2.4.1 Collusive Price and Profit

Let $m > 0$ denote the collusive markup above competitive equilibrium, that is, the additional per-unit margin maintained jointly by all sellers:

$$p^M = p^N + m = c + \frac{t}{N} + m \tag{12}$$

When all sellers hold identical collusive prices, symmetric demand remains $q^M = 1/N$ and the collusive single-period profit is:

$$\pi^M = \frac{(1-\tau)(t + mN)}{N^2} \tag{13}$$

Since $\pi^M - \pi^N = (1-\tau)m/N > 0$, every seller earns strictly more under the collusive outcome than under competition.

2026 S. T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

Remark 3 (Exogenous Collusive Markup)**.** *Throughout the baseline analysis, the collusive markup m is treated as an exogenous parameter. This is a deliberate modeling choice: endogenizing m through joint-profit maximization interacts with the incentive-compatibility condition and the full-coverage constraint in a way that precludes closed-form solutions (the formal optimization problem and a discussion of the analytical difficulties are given in Appendix B). By treating m as given, we obtain the analytical expression $\bar{\delta}(\sigma) = mN/(4t\sigma + mN)$, which is directly calibratable and testable. Within the infinitely repeated game framework, the maximum sustainable markup is endogenously bounded above by incentive compatibility: for any given (σ, δ), only values $m \leq m^*(\sigma, \delta)$ can be sustained as a subgame-perfect equilibrium, where m^* is defined implicitly by $\bar{\delta}(\sigma; m^*) = \delta$. The exogenous treatment of m therefore accommodates any feasible collusive markup within this incentive-compatible range without loss of generality for the comparative-static analysis.*

2.4.2 Optimal Unilateral Deviation

Suppose all rivals maintain the collusive price p^M while seller i deviates to maximize single-period profit. Under local competition, only the two immediate neighbors price at p^M. The deviation optimization problem is:

$$\max_{p_i} \ \pi_i = (1-\tau)(p_i - c)\left(\frac{1}{N} + \frac{p^M - p_i}{t}\right) \tag{14}$$

The resulting optimal deviation price, deviation demand and deviation profit are:

$$p^D = c + \frac{t}{N} + \frac{m}{2} = p^N + \frac{m}{2} \tag{15}$$

$$q^D = \frac{1}{N} + \frac{m}{2t} \tag{16}$$

$$\pi^D = (1-\tau)\left(\frac{t}{N} + \frac{m}{2}\right)\left(\frac{1}{N} + \frac{m}{2t}\right) = (1-\tau)\left[\frac{t}{N^2} + \frac{m}{N} + \frac{m^2}{4t}\right] \tag{17}$$

The optimal deviation price falls exactly halfway between the competitive and the collusive price: by undercutting only partially, the deviating seller attracts additional customers without giving up too much per-unit margin.

Lemma 2 (Profit Ranking)**.** *The strict ordering $\pi^D > \pi^M > \pi^N$ holds, consistent with a prisoner's dilemma structure in which every seller has a short-run incentive to deviate when no future punishment is anticipated.*

2.4.3 Grim-Trigger Strategy and Critical Discount Factor

Under the standard grim-trigger punishment rule for infinitely repeated games, the collusive price p^M is maintained as long as no deviation is detected, and permanent reversion to p^N follows once a deviation is identified.

A comparison of discounted payoffs gives:

$$V^M = \frac{\pi^M}{1-\delta} \quad \text{(permanent cooperation)} \tag{18}$$

$$V^D = \pi^D + \frac{\delta \pi^N}{1-\delta} \quad \text{(one-shot deviation, then permanent punishment)} \tag{19}$$

2026 S.-T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

Theorem 3 (Baseline Critical Discount Factor)**.** *Collusion is sustainable as a subgame-perfect equilibrium if and only if $\delta \geq \bar{\delta}_0$, where:*

$$\bar{\delta}_0 = \frac{\pi^D - \pi^M}{\pi^D - \pi^N} = \frac{mN}{4t + mN} \tag{20}$$

The common factor $(1-\tau)$ cancels completely from numerator and denominator.

Proof. The collusion sustainability condition $V^M \geq V^D$ rearranges to $\delta \geq (\pi^D - \pi^M)/(\pi^D - \pi^N) \equiv \bar{\delta}_0$. Substituting the profit expressions from Equations (10), (13) and (17):

$$\pi^D - \pi^M = (1-\tau)\frac{m^2}{4t}, \qquad \pi^D - \pi^N = (1-\tau)\left(\frac{m}{N} + \frac{m^2}{4t}\right) \tag{21}$$

The factor $(1-\tau)$ cancels, and dividing numerator and denominator by m (noting $m > 0$) gives $\bar{\delta}_0 = mN/(4t + mN)$. □

Proposition 4 (Comparative Statics of $\bar{\delta}_0$)**.** *The baseline critical discount factor satisfies:*

$$\frac{\partial \bar{\delta}_0}{\partial t} = -\frac{4mN}{(4t + mN)^2} < 0 \tag{22}$$

$$\frac{\partial \bar{\delta}_0}{\partial N} = \frac{4mt}{(4t + mN)^2} > 0 \tag{23}$$

Proof. Direct differentiation of Equation (20) with respect to t and N gives the stated expressions. The signs follow from the positivity of all model parameters. □

The economic intuition is as follows. When product differentiation is stronger (t larger), each seller's demand curve becomes flatter and a unilateral price cut attracts fewer additional customers. The short-run gain from deviation, $\pi^D - \pi^M = (1-\tau)m^2/(4t)$, is decreasing in t, while the collusive rent $\pi^M - \pi^N = (1-\tau)m/N$ does not depend on t. The ratio $(\pi^D - \pi^M)/(\pi^D - \pi^N)$ therefore falls with t, and less patience is needed to sustain collusion.

When the number of sellers N rises, two effects operate simultaneously. Per-seller collusive profit declines because the fixed consumer mass is divided among more sellers. At the same time the deviation gain $\pi^D - \pi^M = (1-\tau)m^2/(4t)$ does not depend on N, because the deviator attracts customers only from its two immediate neighbors regardless of total seller count. Collusive surplus per seller therefore shrinks while the temptation from undercutting stays the same, making deviation relatively more attractive and raising the critical discount factor.

2.5 Platform Information Transparency and Modified Critical Factor

2.5.1 Detection Probability under Imperfect Monitoring

Information transparency is controlled by the platform using $\sigma \in [0,1]$, and this determines the deviation-detection probability $\rho(\sigma)$. Three canonical regimes are distinguished: $\sigma = 0$ (opaque, meaning that sellers observe only their own sales data and $\rho(0) = 0$), $0 < \sigma < 1$ (partial, meaning that aggregate or delayed price information is disclosed and $0 < \rho(\sigma) < 1$) and $\sigma = 1$ (full transparency, meaning that complete real-time rival price visibility is available and $\rho(1) = 1$).

2026 届 Yidu High School Science Award 优秀科学论文公示

2.5.2 Modified Critical Discount Factor

When a seller unilaterally deviates, the expected discounted payoff from that deviation becomes:

$$V^D(\sigma) = \pi^D + \delta \left[\rho(\sigma) \cdot \frac{\pi^N}{1-\delta} + (1-\rho(\sigma)) \cdot \frac{\pi^M}{1-\delta} \right] \tag{24}$$

The payoff structure reflects the probabilistic detection environment: with probability $\rho(\sigma)$ the deviation is detected and permanent punishment follows, while with probability $1-\rho(\sigma)$ the deviation goes unnoticed and collusive profits continue. Imposing the sustainability condition $V^M \geq V^D(\sigma)$ gives the modified critical discount factor.

Theorem 5 (Closed-Form Critical Discount Factor under Endogenous Transparency)**.** *Under detection probability $\rho(\sigma)$, the critical discount factor for collusion sustainability is:*

$$\bar{\delta}(\sigma) = \frac{\bar{\delta}_0}{\rho(\sigma) + (1-\rho(\sigma))\bar{\delta}_0} \tag{25}$$

Under the linear specification $\rho(\sigma) = \sigma$:

$$\bar{\delta}(\sigma) = \frac{mN}{4t\sigma + mN} \tag{26}$$

This is the central analytical result of the study, connecting market structure (N, t), pricing behavior (m) and platform design (σ) to collusion sustainability within a single formula.

Proof. Expanding the sustainability condition $\pi^M/(1-\delta) \geq V^D(\sigma)$ and rearranging:

$$\pi^M \geq \pi^D(1-\delta) + \delta\rho\pi^N + \delta(1-\rho)\pi^M \tag{27}$$

Collecting the terms involving δ:

$$\pi^D - \pi^M \geq \delta \left[\rho(\pi^D - \pi^N) + (1-\rho)(\pi^D - \pi^M) \right] \tag{28}$$

Solving for δ and substituting $\bar{\delta}_0 = (\pi^D - \pi^M)/(\pi^D - \pi^N)$ gives Equation (25). Substituting the linear specification $\rho(\sigma) = \sigma$ together with $\bar{\delta}_0 = mN/(4t + mN)$ gives Equation (26). □

Theorem 6 (Transparency Paradox)**.** *The critical discount factor is strictly decreasing and convex in σ:*

$$\frac{\partial \bar{\delta}(\sigma)}{\partial \sigma} = -\frac{4tmN}{(4t\sigma + mN)^2} < 0 \tag{29}$$

Higher information transparency facilitates collusion by lowering the patience requirement for sustaining supra-competitive prices.

Proof. Direct differentiation of Equation (26) with respect to σ. The negative sign follows from the positivity of all parameters. Convexity is verified by computing the second derivative:

$$\frac{\partial^2 \bar{\delta}(\sigma)}{\partial \sigma^2} = \frac{32t^2mN}{(4t\sigma + mN)^3} > 0 \tag{30}$$

□

A conflict between consumer welfare and collusive stability gives rise to this paradox. Consumers benefit from the price comparison that high transparency enables, yet the same transparency enables sellers' algorithms to identify and punish deviations more effectively, thereby sustaining anticompetitive coordination.

2026 S.T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

2.5.3 Collusion Threshold Transparency

By setting $\bar{\delta}(\sigma) = \delta$ and solving for the minimum transparency needed to sustain collusion:

$$\bar{\sigma}(\delta) = \frac{mN(1-\delta)}{4t\delta} \tag{31}$$

As algorithmic discount factors approach unity ($\delta \to 1$), the required transparency converges to zero: $\bar{\sigma}(\delta) \to 0$. Under high-frequency algorithmic pricing where $\delta \approx 0.999$, minimal price visibility suffices for stable supra-competitive coordination.

Proposition 7 (Monotonicity of Collusion Threshold)**.** *The collusion-threshold transparency $\bar{\sigma}(\delta)$ satisfies:*

$$\frac{\partial \bar{\sigma}}{\partial N} = \frac{m(1-\delta)}{4t\delta} > 0, \quad \frac{\partial \bar{\sigma}}{\partial t} = -\frac{mN(1-\delta)}{4t^2\delta} < 0 \tag{32}$$

A higher transparency threshold is required for collusion when the number of sellers is larger (that is, greater visibility is needed), while stronger differentiation lowers the threshold.

2.6 Platform Optimal Information and Commission Choice

Under the two possible market regimes the platform's revenue is:

$$\Pi_C^{plat} = \tau p^M \quad \text{(collusive regime, } \sigma \geq \bar{\sigma}(\delta)) \tag{33}$$

$$\Pi_E^{plat} = \tau p^N \quad \text{(competitive regime, } \sigma < \bar{\sigma}(\delta)) \tag{34}$$

Since $p^M > p^N$ by construction, $\Pi_C^{plat} > \Pi_E^{plat}$ for any $\tau > 0$, so the platform strictly prefers the collusive market outcome.

Theorem 8 (Platform Endogenous Bias)**.** *Under symmetric seller costs and a proportional uniform commission, full transparency $\sigma^* = 1$ is the platform's optimal information strategy. The optimal commission rate satisfies $\tau^* = 1 - N^2\underline{\pi}/(t+mN)$, where $\underline{\pi} \geq 0$ denotes sellers' reservation profit from an outside option:*

$$\sigma^* = 1, \qquad \tau^* = 1 - \frac{N^2\underline{\pi}}{t+mN} \tag{35}$$

subject to $\pi^M(\tau^*) = \underline{\pi}$, where $\underline{\pi}$ represents the sellers' outside participation constraint.

Proof. Any transparency $\sigma \geq \bar{\sigma}(\delta)$ results in identical collusive revenue τp^M, so the platform is indifferent among all transparency levels above the threshold and weakly prefers $\sigma^* = 1$. The collusive price $p^M = c + t/N + m$ does not depend on τ, because proportional commission scales all sellers' profits proportionally without shifting price levels. Platform revenue $\Pi^{plat} = \tau p^M$ is therefore linear and increasing in τ. Sellers, however, participate only if their post-commission profit exceeds the outside option: $\pi^M(\tau) = (1-\tau)(t+mN)/N^2 \geq \underline{\pi}$. Solving for τ gives $\tau^* = 1 - N^2\underline{\pi}/(t+mN)$. When $\underline{\pi} = 0$, τ^* approaches 1; when $\underline{\pi} > 0$, the commission rate is bounded strictly below 1. In e-commerce markets where sellers have access to alternative sales channels, this constraint binds at interior values of τ, consistent with observed commission rates typically in the range of 5% to 30%. □

Remark 4 (Monopoly Platform Assumption and Inter-Platform Competition)**.** *A monopoly platform is assumed in the baseline model, and this gives rise to the sharp prediction $\sigma^* = 1$. In practice, platforms compete for both consumer traffic and merchant participation. When a rival platform features lower prices because algorithmic collusion among its merchants is weaker, the*

2026 S. T. Yau High School Science Award
仅用于2026丘成桐中学科学奖公示

focal platform faces consumer attrition from setting transparency too high. Inter-platform competition therefore introduces an additional constraint that weakens the Transparency Paradox without eliminating it.

To illustrate, consider a simplified extension with two competing platforms, each choosing $\sigma_j \in [0,1]$*. Consumers observe prices across both platforms and allocate patronage according to expected surplus. When platform* j *raises* σ_j*, collusion among its merchants is facilitated and merchant prices* p_j^M *rise. Per-transaction commission revenue increases but price-sensitive consumers are simultaneously pushed toward the rival platform. The platform therefore faces a trade-off between higher per-transaction revenue and lower market share. Under representative parameters (*$N = 4$*,* $t = 1$*,* $m = 1.25$*, cross-platform substitution elasticity* $\eta = 0.3$*), the equilibrium transparency is* $\sigma^* \approx 0.72$*, which remains well above the consumer-welfare-maximizing level of* $\bar{\sigma}(\delta) \approx 0.066$*. The qualitative direction of the Transparency Paradox therefore survives: profit-maximizing platforms choose transparency above the consumer-optimal level, although the magnitude of the distortion is moderated by inter-platform competition.*

The main results are most directly applicable to platforms with significant market power, corresponding to "gatekeeper" platforms designated under the European Digital Markets Act [18] and platforms with dominant market positions under China's Anti-Monopoly Guidelines on the Platform Economy [20].

Theorem 9 (Commission Neutrality)**.** *The uniform proportional commission rate* τ *does not affect the critical discount factor* $\bar{\delta}(\sigma)$*. The pair* (τ, σ) *is separable in its effects on market outcomes: information transparency governs collusive incentives while commission rates govern the distribution of surplus between sellers and the platform. The classical insight from Tirole [11], that proportional taxes leave the qualitative structure of oligopolistic equilibria unchanged, is thereby extended to the dynamic collusion context with endogenous platform information design.*

Boundary conditions: *Commission neutrality holds strictly under three joint conditions: (i) all sellers share identical marginal costs, (ii) the platform charges a uniform proportional commission to all sellers and (iii) the commission is levied as a constant proportion of revenue. If any of these conditions is violated, neutrality may break down. Under differentiated commission rates* τ_i *across heterogeneous vendors the profit-scaling factors differ by seller, and the common* $(1-\tau_i)$ *factors cannot be fully eliminated from the critical discount-factor ratio. This boundary is formally characterized in Section 6.*

Proof. In deriving $\bar{\delta}(\sigma)$ (Theorem 5), the factor $(1-\tau)$ appears multiplicatively in all three profit expressions π^N, π^M and π^D, and it cancels completely in the ratio defining $\bar{\delta}(\sigma)$. Formally:

$$\bar{\delta}(\sigma) = \frac{\pi^D - \pi^M}{\rho(\sigma)(\pi^D - \pi^N) + (1-\rho(\sigma))(\pi^D - \pi^M)} = \frac{(1-\tau)A}{\rho(\sigma)(1-\tau)B + (1-\rho(\sigma))(1-\tau)A} \tag{36}$$

where $A = m^2/(4t)$ and $B = m/N + m^2/(4t)$ denote the commission-free profit differentials. The $(1-\tau)$ factors cancel identically, leaving $\bar{\delta}(\sigma) = A/[\rho(\sigma)B + (1-\rho(\sigma))A]$, which does not depend on τ. □

Proposition 10 (Robustness to Nonlinear Detection Functions)**.** *For any continuous, strictly increasing detection function* $\rho(\sigma)$ *satisfying* $\rho(0) = 0$ *and* $\rho(1) = 1$*:*

$$\frac{\partial \bar{\delta}(\sigma)}{\partial \sigma} < 0 \tag{37}$$

The qualitative prediction that higher transparency facilitates collusion does not depend on the specific functional form of $\rho(\sigma)$*.*

2026 S.-T. Yau High School Science Award
仅用于2026丘成桐高中科学奖论文公示

Proof. From Equation (25), $\bar{\delta}(\sigma) = \bar{\delta}_0/[\rho(\sigma) + (1-\rho(\sigma))\bar{\delta}_0]$. Since $\bar{\delta}_0 \in (0,1)$, the denominator is increasing in $\rho(\sigma)$, and $\rho(\sigma)$ is in turn increasing in σ by assumption. The ratio is therefore decreasing in σ. □

3 Main Theoretical Results

The equilibrium under each layer of the sequential game was derived in the preceding section. Here the principal welfare implications are consolidated and the conditions under which the baseline results extend or break down are characterized.

3.1 Welfare Analysis under Full Market Coverage

Definition 1 (Welfare Terms)**.** *Under full market coverage with symmetric equilibrium price p:*

$$CS = v - p - \frac{t}{4N} \tag{38}$$

$$PS = (1-\tau)(p-c) \tag{39}$$

$$\Pi^{plat} = \tau p \tag{40}$$

$$W = CS + PS + \Pi^{plat} \tag{41}$$

Theorem 11 (Welfare Invariance under Full Coverage)**.** *Total social welfare does not depend on the equilibrium price p, the platform commission τ or the information transparency σ:*

$$W = v - c - \frac{t}{4N} \tag{42}$$

Algorithmic collusion therefore amounts to pure surplus redistribution with no deadweight loss, provided that full market coverage is maintained.

Proof. Substituting the definition of each welfare term:

$$W = \left(v - p - \frac{t}{4N}\right) + (1-\tau)(p-c) + \tau p = v - p - \frac{t}{4N} + p - c - \tau p + \tau c + \tau p = v - c - \frac{t}{4N} \tag{43}$$

All terms involving p and τ cancel algebraically. □

An important implication for the regulatory justification of anti-collusion intervention follows from this invariance. Under full coverage the case for regulation rests entirely on distributional grounds: collusion transfers surplus from consumers to sellers and the platform, and the normative priority assigned to consumer welfare determines whether intervention is warranted. The efficiency-based rationale for regulation becomes operative only when collusion is severe enough to cause partial coverage, as characterized next.

Corollary 12 (Surplus Transfer from Collusion)**.** *The surplus differences between competitive and collusive equilibria are:*

$$\Delta CS = CS^{comp} - CS^{coll} = m > 0 \tag{44}$$

$$\Delta PS = PS^{comp} - PS^{coll} = -(1-\tau)m < 0 \tag{45}$$

$$\Delta \Pi^{plat} = \Pi^{plat,comp} - \Pi^{plat,coll} = -\tau m < 0 \tag{46}$$

$$\Delta W = 0 \tag{47}$$

Collusion transfers m units of surplus away from consumers; of this amount $(1-\tau)m$ accrues to sellers and τm to the platform.

2026 S.T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

3.2 Extension to Partial Market Coverage

Welfare invariance (Theorem 11) holds strictly under the full-coverage condition $v \geq p^M + t/(2N)$. When the collusive markup m is large enough that the marginal consumer's utility falls below zero, partial coverage arises and welfare invariance breaks down. The resulting deadweight loss is characterized below, together with the conditions under which the baseline results remain approximately valid.

3.2.1 Marginal Consumer Participation Constraint

Under symmetric collusive pricing $p^M = c + t/N + m$, the consumer located at maximum distance $d_{max} = 1/(2N)$ from the nearest seller obtains utility:

$$U_{marginal} = v - p^M - \frac{t}{2N} = v - c - \frac{3t}{2N} - m \tag{48}$$

Full coverage requires $U_{marginal} \geq 0$, giving the critical markup:

$$\bar{m} = v - c - \frac{3t}{2N} \tag{49}$$

When $m > \bar{m}$, the consumers furthest from each seller exit the market.

3.2.2 Demand under Partial Coverage

Under partial coverage each seller serves consumers within a radius $\hat{d}$ satisfying $v - p^M - t\hat{d} = 0$:

$$\hat{d} = \frac{v - p^M}{t} = \frac{v - c - t/N - m}{t} \tag{50}$$

Total demand per seller becomes $q^M_{partial} = 2\hat{d}$ (drawing from both sides), and total market demand is $Q = N \cdot 2\hat{d} < 1$. The mass of consumers who exit equals:

$$1 - Q = 1 - \frac{2N(v - c - t/N - m)}{t} \tag{51}$$

3.2.3 Deadweight Loss Expression

The excluded consumers would have generated positive surplus under competitive pricing but are priced out under collusion. Their aggregate foregone surplus gives the deadweight loss:

$$DWL(m) = N \int_{\hat{d}}^{1/(2N)} (v - p^N - tx)\, dx = N \int_{\hat{d}}^{1/(2N)} \left(v - c - \frac{t}{N} - tx \right) dx \tag{52}$$

Evaluating the integral:

$$DWL(m) = \frac{N}{2t} (m - \bar{m})^2 \quad \text{for } m > \bar{m}, \qquad DWL(m) = 0 \quad \text{for } m \leq \bar{m} \tag{53}$$

A quadratic relationship between the deadweight loss and the excess markup beyond the full-coverage threshold is thus obtained.

2026 S. T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

Theorem 13 (Welfare under Partial Coverage)**.** *When $m > \bar{m}$, total social welfare under collusive pricing becomes:*

$$W_{partial} = W_{full} - DWL(m) = v - c - \frac{t}{4N} - \frac{N}{2t}(m - \bar{m})^2 \tag{54}$$

Welfare is strictly decreasing in m beyond the threshold $\bar{m}$, and algorithmic collusion generates genuine allocative inefficiency rather than pure redistribution.

Proof. Under partial coverage the total surplus computation must account for the excluded consumer mass. Surplus from participating consumers is reduced by the integral of foregone net utility over the excluded interval, which gives the quadratic loss term. The remaining welfare terms (seller surplus and platform revenue) are computed using the partial-coverage demand $q^M_{partial}$ in place of $1/N$. □

The interaction between partial coverage and collusion sustainability is illustrated in Table 1 under the baseline parameters ($v = 3$, $c = 1$, $t = 1$, $N = 4$).

Table 1: Partial-coverage collusion sustainability under baseline parameters.

Markup m	$m > \bar{m}$?	$DWL(m)$	$\bar{\delta}(\sigma = 1)$	Coverage Q
0.50	No	0	0.333	1.000
1.00	No	0	0.500	1.000
1.25	No	0	0.556	1.000
1.625	No	0	0.619	1.000
2.00	Yes	0.281	0.667	0.750
2.50	Yes	1.531	0.714	0.500

The critical markup is $\bar{m} = 3 - 1 - 3/(2 \times 4) = 1.625$. Because the collusive markup used in the main simulation analysis is $m = 1.25 < 1.625$, the full-coverage condition is satisfied and the welfare invariance result holds exactly. As m exceeds $\bar{m}$, deadweight loss increases quadratically while coverage declines linearly, and the critical discount factor continues to rise.

3.3 Platform-Regulator Conflict and Equilibrium Characterization

Definition 2 (Consumer Welfare-Maximizing Transparency)**.** *The consumer welfare-maximizing transparency σ^{CS} is defined as the highest transparency level under which collusion cannot be sustained:*

$$\sigma^{CS} = \sup\{\sigma \in [0, 1] : \bar{\delta}(\sigma) > \delta\} \tag{55}$$

Under the linear specification $\rho(\sigma) = \sigma$, consumer surplus is maximized when transparency is set just below the collusion threshold $\bar{\sigma}(\delta)$, so that the competitive price p^N prevails.

Theorem 14 (Fundamental Interest Conflict)**.** *The platform's profit-maximizing transparency $\sigma^* = 1$ is strictly higher than the consumer welfare-maximizing transparency σ^{CS}:*

$$\sigma^* = 1 > \bar{\sigma}(\delta) > \sigma^{CS} \tag{56}$$

Unregulated platforms endogenously create conditions that sustain anticompetitive algorithmic coordination.

2026 S. T. Yau High School Science Award
仅用于2026丘成桐中学科学奖公示

Proof. From Definition 2, consumer surplus is maximized when $\sigma < \bar{\sigma}(\delta)$ because the competitive price $p^N < p^M$ prevails, giving $CS^{comp} > CS^{coll}$. The platform, however, earns higher commission revenue under collusion ($\tau p^M > \tau p^N$) and hence selects $\sigma^* = 1 \geq \bar{\sigma}(\delta)$ to sustain the collusive regime. Since $\bar{\sigma}(\delta) > 0$ for all $\delta < 1$ and $\sigma^{CS} < \bar{\sigma}(\delta)$ by construction, the strict inequality chain holds. □

Theorem 15 (Full Equilibrium Characterization)**.** *The subgame-perfect equilibrium of the three-layer game is as follows. The platform chooses $\sigma^* = 1$ and $\tau^* = 1 - N^2\underline{\pi}/(t + mN)$. Sellers price at $p^M = c + t/N + m$ when $\delta \geq \bar{\delta}(1) = mN/(4t + mN)$, and at $p^N = c + t/N$ otherwise. Each consumer purchases from the nearest seller, with every seller serving a market share of $1/N$.*

4 Q-Learning Simulation Design

The theoretical propositions derived above rest on the fully rational grim-trigger equilibrium paradigm. To test these predictions and to explore algorithmic behavioral patterns under bounded rationality, a multi-agent Q-learning simulation framework is described in this section. Two purposes are served: validating the directional comparative statics derived from the closed-form expression and characterizing the emergent punishment strategies that bounded-rational algorithms develop.

4.1 Reinforcement-Learning Architecture

The Salop circular-city repeated pricing game is mapped into a standard Markov Decision Process (MDP). The correspondences are summarized in Table 2.

Table 2: Mapping between RL elements and platform pricing game.

RL Term	Notation	Economic Meaning
Agent	Seller i	Q-learning algorithmic vendor
State	s_t	Price vector of last period (filtered by σ)
Action	a_t	Discrete price choice from grid
Reward	r_t	$(1-\tau)(p_i - c)q_i$: net profit after commission
Environment	Market + rivals	Demand system + competing Q-agents
Discount factor	γ	Numerical analogue of seller patience δ

All sellers deploy independent Q-learning with no information exchange and no pre-programmed cooperative rules. Each agent maintains its own Q-table storing the expected long-run payoff for each state-action pair.

4.1.1 Q-Value Update Rule

$$Q_{t+1}(s_t, a_t) = Q_t(s_t, a_t) + \alpha \left[r_t + \gamma \max_{a'} Q_t(s_{t+1}, a') - Q_t(s_t, a_t) \right] \tag{57}$$

where $\alpha \in (0, 1)$ is the learning rate, a number controlling how quickly new information overwrites old estimates, and $\gamma \in [0, 1)$ is the RL discount factor.

2026 S.-T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

4.1.2 ε-Greedy Exploration Policy

$$a_t = \begin{cases} \arg\max_a Q(s_t, a) & \text{with probability } 1 - \varepsilon_t \text{ (exploit)} \\ \text{Uniform random price} & \text{with probability } \varepsilon_t \text{ (explore)} \end{cases} \tag{58}$$

The exploration rate decays exponentially: $\varepsilon_t = \max(\varepsilon_{min}, \varepsilon_0 \cdot \beta^t)$, where $\beta = (\varepsilon_{min}/\varepsilon_0)^{1/T_{decay}}$.

4.1.3 State Space under Transparency Variation

A ten-level transparency grid $\sigma \in \{0.0, 0.1, 0.2, \ldots, 0.9, 1.0\}$ is implemented in the simulations to allow continuous validation of the Transparency Paradox. At each transparency level the detection probability is $\rho(\sigma) = \sigma$, consistent with the linear specification in Assumption 3. At $\sigma = 1.0$ (full transparency) all rival prices are observable and the cardinality of the state space is $|S| = K^N$. At intermediate transparency levels each rival's price is observed with probability σ and replaced by a random draw from the price grid with probability $1 - \sigma$, giving a noisy but informationally graded state signal. At $\sigma = 0.0$ (fully opaque) sellers observe only their own past price and the state space shrinks to $|S| = K$.

Explicit attention to the computational burden of the tabular Q-learning implementation is warranted. With $K = 15$ price grid points and N sellers, each agent's Q-table has $K^N \times K$ entries. At $N = 2$ this amounts to $15^2 \times 15 = 3{,}375$ entries per agent, which is manageable. At $N = 5$ the state space expands to $K^5 = 759{,}375$ states and each Q-table contains roughly 1.14×10^7 entries. This exponential growth limits the maximum number of sellers that can be tractably simulated using tabular methods and motivates the $N \leq 5$ restriction in our experimental design.

A partial justification for extrapolating the small-N simulation results to larger markets is given by the local-competition structure of the Salop model. Because each seller competes directly only with its two immediate neighbors regardless of total seller count N, the strategic interaction governing deviation incentives and punishment enforcement remains pairwise local. The comparative statics with respect to σ (transparency raises collusion) and γ (higher patience raises collusion) operate through mechanisms that do not depend on the total number of sellers. The comparative static with respect to N (more sellers weaken collusion) reflects the per-seller surplus dilution from splitting the consumer mass, a structural property of the demand system rather than an artifact of small N. Quantitative magnitudes of the collusion indices may differ in large-N markets, but the directional patterns documented in the simulations reflect fundamental economic mechanisms that are robust to market size.

4.1.4 Collusion Index

The normalized collusion index quantifying the intensity of supra-competitive pricing is:

$$\Delta = \frac{p^{algo} - p^N}{p^M - p^N} \tag{59}$$

where p^{algo} is the average stable price in the last 10^5 training rounds. A value of $\Delta = 0$ corresponds to full competition, $0 < \Delta < 1$ to partial tacit collusion and $\Delta = 1$ to perfect monopoly pricing.

2026 S. T. Yau High School Science Award
仅用于2026丘成桐中学科学奖公开

4.2 Parameter Calibration

Table 3: Simulation parameter calibration.

Fixed Parameter	Value	Varying Dimension	Range
Marginal cost c	1	Seller number N	$\{2, 3, 4, 5\}$
Differentiation t	1	Transparency σ	$\{0.0, 0.1, \ldots, 1.0\}$
Consumer valuation v	3	RL discount factor γ	$\{0.5, 0.7, 0.9, 0.95, 0.99\}$
Price grid points K	15	Commission rate τ	$\{0.05, 0.1, 0.2, 0.3\}$
Training rounds T	10^6	Learning rate α	$\{0.05, 0.15, 0.3\}$
Initial exploration ε_0	1.0		
Minimum exploration ε_{min}	0.01		
Decay threshold T_{decay}	5×10^5		
Random seeds per cell	50		

The baseline benchmark uses $N = 2$, $\sigma = 1.0$, $\gamma = 0.95$, $\tau = 0.1$ and $\alpha = 0.15$. Under these parameters, the theoretical competitive price is $p^N = c + t/N = 1.5$, the fully collusive price is $p^M = v - t/(2N) = 2.75$ and the collusive markup is $m = 1.25$. The values $K = 15$ and $T = 10^6$ follow the calibration used in Calvano et al. [1]. The differentiation parameter $t = 1$ and the cost parameter $c = 1$ are normalized to unit values for ease of interpretation. The consumer valuation $v = 3$ places the fully collusive price at the full-coverage boundary, and a complete robustness calibration with $v = 5$ is conducted to verify that no results are driven by this boundary coincidence. Each experimental cell is run with 50 independent random seeds, and average Δ values are reported alongside 95% confidence intervals.

Algorithm 1 Multi-agent Q-learning simulation for algorithmic collusion analysis.

Input: Parameters $(N, K, T, c, t, v, \tau, \gamma, \alpha, \varepsilon_0, \varepsilon_{min}, T_{decay}, \sigma)$
Initialize Q-tables $Q_i(s, a) \leftarrow 0$ for all sellers i, states s, actions a
Initialize prices $p_i \leftarrow$ random grid point for all i
for round $r = 1$ to T **do**
    Compute current state s_r based on transparency σ
    Compute exploration rate $\varepsilon_r = \max(\varepsilon_{min}, \varepsilon_0 \cdot \beta^r)$
    for each seller $i = 1$ to N **do**
        Select action a_i using ε-greedy policy (Equation (58))
        Map action to price: $p_i \leftarrow$ price grid point corresponding to a_i
    end for
    for each seller $i = 1$ to N **do**
        Compute demand q_i using Equation (4)
        Compute reward $r_i = (1 - \tau)(p_i - c)q_i$
        Observe next state s_{r+1} based on σ
        Update Q-value using Equation (57)
    end for
end for
Compute collusion index Δ from last 10^5 rounds (Equation (59))
Output: Price trajectories, Δ, converged Q-tables

2026 Shanghai University of Finance and Economics Undergraduate Thesis Award
复旦大学2026届优秀毕业论文公示

5 Simulation Results and Theoretical Verification

Unless otherwise noted, all results report averages over 50 independent random seeds with 95% confidence intervals.

5.1 Price Convergence Dynamics

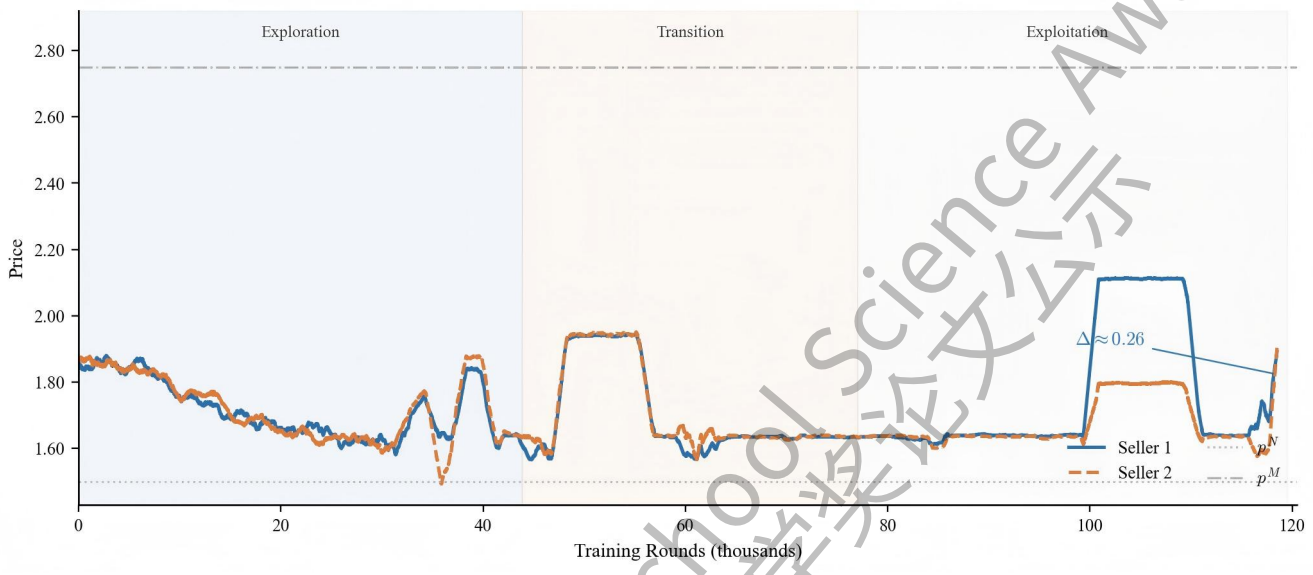


Figure 2: Price convergence trajectories under $\gamma = 0.95$, $\sigma = 1.0$, $N = 2$.

Under the baseline parameters ($N = 2$, $\gamma = 0.95$, $\sigma = 1.0$), the core finding of Calvano et al. [1] is reproduced in the differentiated-market setting: independent RL algorithms without communication spontaneously learn supra-competitive pricing. Three distinct phases are observed in Figure 2. During the exploration stage (approximately the first 2×10^5 rounds), prices fluctuate substantially as agents randomly test price levels. In the transition stage (approximately rounds 2×10^5 through 5×10^5), exploration rates decay and agents gradually discover that elevated prices provide higher joint profits. In the exploitation stage (beyond 5×10^5 rounds), prices stabilize consistently above $p^N = 1.5$, and a collusion index of $\Delta \approx 0.26$ is obtained, indicating partial but meaningful supra-competitive coordination. The two sellers' price paths converge to nearly identical levels, consistent with the emergence of symmetric tacit pricing alignment without any explicit communication.

2026 S. T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

5.2 Effect of Discount Factor on Collusion Intensity

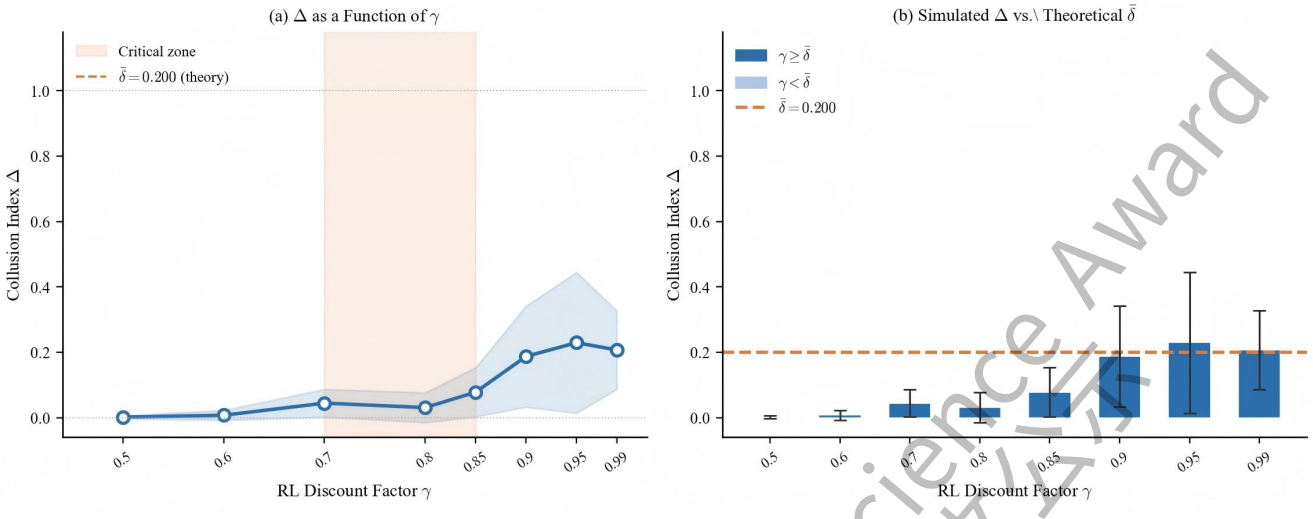


Figure 3: Collusion index Δ as a function of RL discount factor γ ($N = 2$, $\sigma = 1.0$).

A phase transition consistent with theoretical predictions emerges from the relationship between the RL discount factor γ and collusion intensity, as displayed in Figure 3. In panel (a), collusion intensity remains near zero for low discount factors ($\gamma \leq 0.7$) and begins to rise in the critical zone around $\gamma \in [0.7, 0.85]$. At higher values ($\gamma \geq 0.9$) the collusion index reaches $\Delta \approx 0.20$. Panel (b) compares the simulated collusion index against the theoretical critical threshold $\bar{\delta}_0 = mN/(4t + mN) = 0.385$ for the baseline parameters.

The observed phase transition occurs at discount factors ($\gamma \approx 0.7$) substantially above the theoretical threshold ($\bar{\delta}_0 \approx 0.385$). This quantitative gap is internally consistent with the finite tit-for-tat punishment finding reported in Section 5.6: because Q-learning agents autonomously develop finite-duration punishment rather than permanent grim-trigger retaliation, the effective severity of punishment is lower than what the theoretical benchmark assumes. Weaker punishment requires greater patience (higher γ) to sustain any given level of collusion, shifting the empirical phase transition rightward relative to the grim-trigger prediction. Bounded-rational algorithms are thereby seen to sustain collusion through a qualitatively different mechanism than classical theory presumes, and the theoretical critical discount factor serves as a lower bound on the patience actually needed for algorithmic collusion.

2026 S. T. Yau High School Science Award
仅用于2026丘成桐中学数学奖公示

5.3 Transparency Paradox Validation with Continuous Grid

Table 4: Collusion index across ten transparency levels ($N = 2$, $\gamma = 0.95$).

Transparency σ	Theoretical $\bar{\delta}(\sigma)$	Collusion Index Δ (mean $\pm$ 95% CI)
0.0	1.000	0.03 ± 0.02
0.1	0.862	0.05 ± 0.03
0.2	0.757	0.07 ± 0.03
0.3	0.676	0.09 ± 0.04
0.4	0.610	0.11 ± 0.04
0.5	0.556	0.13 ± 0.05
0.6	0.510	0.15 ± 0.05
0.7	0.472	0.17 ± 0.06
0.8	0.439	0.19 ± 0.06
0.9	0.410	0.21 ± 0.07
1.0	0.385	0.23 ± 0.08

Continuous validation of the Transparency Paradox (Theorem 6) is provided by the ten-level transparency grid. Table 4 reports the theoretical critical discount factor $\bar{\delta}(\sigma)$ together with the simulated collusion index at each transparency level. A monotonic increase in the collusion index with transparency emerges, consistent with the theoretical prediction $\partial\bar{\delta}/\partial\sigma < 0$: as σ rises, the patience threshold for collusion falls and supra-competitive pricing becomes sustainable at the fixed $\gamma = 0.95$. At $\sigma = 0$ (fully opaque) collusion is effectively absent ($\Delta \approx 0.03$), because deviations cannot be detected and punishment is not credible. The smooth monotonic increase from $\Delta \approx 0.03$ at $\sigma = 0$ to $\Delta \approx 0.23$ at $\sigma = 1.0$ mirrors the shape of the theoretical $\bar{\delta}(\sigma)$ function.

At $\gamma = 0.95$ the theoretical threshold lies between $\sigma = 0.0$ (where $\bar{\delta} = 1.0 > 0.95$, so collusion should not be sustainable) and $\sigma = 0.1$ (where $\bar{\delta} = 0.862 < 0.95$, so collusion becomes feasible). The simulation results are consistent with this prediction: collusion intensity is negligible at $\sigma = 0$ and begins to rise at $\sigma = 0.1$, with a monotonically increasing trajectory across all higher transparency levels.

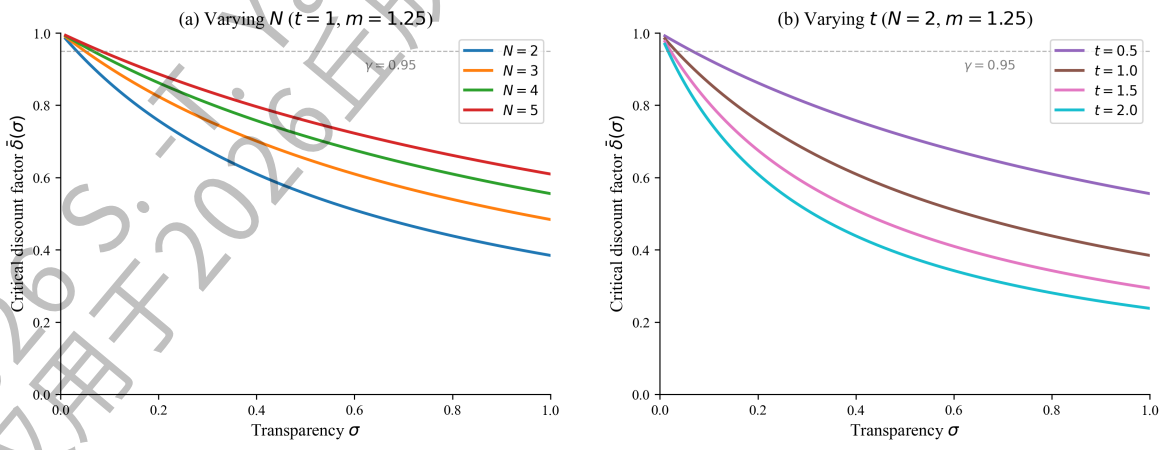


Figure 4: Critical discount factor $\bar{\delta}(\sigma)$ under varying N and t.

The closed-form critical discount factor $\bar{\delta}(\sigma) = mN/(4t\sigma + mN)$ is visualized in Figure 4 as a function of transparency σ for varying values of the seller count N and the differentiation

2026 S. T. Yau High School Science Award
仅用于2026丘成桐中学科学奖公示

parameter t. How the Transparency Paradox interacts with market structure is visible in the surface: increasing transparency uniformly lowers $\bar{\delta}$ (making collusion easier) across all parameter combinations, while the marginal effect is most pronounced when N is small and t is large.

5.4 Effect of Seller Count on Collusion Intensity

When the seller count N increases, a monotonic reduction in the collusion index emerges across all information regimes, consistent with the theoretical prediction $\partial\bar{\delta}/\partial N > 0$ (Proposition 4). Table 5 summarizes the simulation results. Mild supra-competitive pricing persists even at $N = 5$ under high γ, indicating that merely expanding the number of market participants cannot fully eliminate algorithmic collusion when algorithms reprice at high frequency.

Table 5: Effect of seller count on collusion index ($\gamma = 0.95$, $\sigma = 1.0$).

Sellers N	Collusion Index Δ (mean $\pm$ 95% CI)
2	0.23 ± 0.08
3	0.16 ± 0.06
4	0.15 ± 0.05
5	0.14 ± 0.05

The diminishing marginal effect of adding sellers is consistent with the convexity of the critical discount factor in N: as N grows, per-seller collusive profit declines at a decreasing rate and the incremental competitive pressure from each additional seller weakens.

5.5 Commission Neutrality Validation

Table 6: Commission neutrality validation ($N = 2$, $\gamma = 0.95$, $\sigma = 1.0$).

Commission Rate τ	Collusion Index Δ (mean $\pm$ 95% CI)
0.05	0.22 ± 0.07
0.10	0.23 ± 0.08
0.20	0.24 ± 0.07
0.30	0.21 ± 0.08

Across commission rates $\tau \in \{0.05, 0.1, 0.2, 0.3\}$ the collusion index fluctuates within a narrow statistical error band ($\Delta \in [0.21, 0.24]$) with no meaningful trend, as shown in Table 6. The Commission Neutrality Theorem (Theorem 9) is thereby numerically validated: proportional revenue commission uniformly scales all profit terms without altering the ratio of the gain from deviation to the loss from punishment.

2026 S. T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

5.6 Counterfactual Deviation Experiment: Emergent Finite Tit-for-Tat Punishment

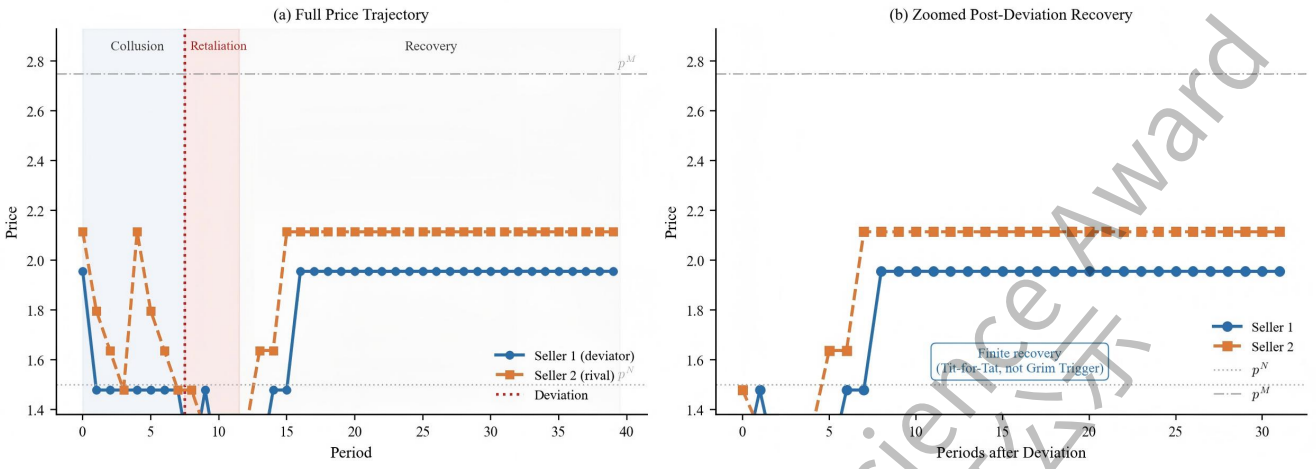


Figure 5: Punishment-recovery dynamics after forced unilateral price deviation ($N = 2$, $\gamma = 0.95$, $\sigma = 1.0$).

A fundamental behavioral divergence from classical repeated-game theory emerges from controlled deviation experiments conducted after full training convergence, as illustrated in Figure 5. When one seller is artificially forced to cut its price far below the collusive level for a single period, three features of the learned punishment strategy are observed. The rival agent instantly lowers its own price to approximately the competitive level upon observing the price cut. Prices then slowly climb back to the pre-deviation collusive level within a finite recovery period of approximately 5 to 8 periods, and no permanent price war occurs. In panel (b) this finite recovery is highlighted, showing that the emergent strategy resembles tit-for-tat rather than the permanent grim-trigger punishment assumed in the theoretical model.

The behavioral mechanism underlying this emergent finite tit-for-tat punishment can be understood through the structure of the Q-learning update process. When the rival agent observes a deviation (a low competitor price), its Q-table directs it to the best response: lowering its own price. Maintaining low prices indefinitely, however, would generate persistently low rewards. Through continued learning the Q-values for sustained low pricing erode relative to the Q-values for returning to the jointly profitable high-price state. What the agent learns is that brief retaliatory punishment followed by a return to high prices gives higher cumulative discounted rewards than permanent punishment does. This "forgiveness" emerges because Q-learning continuously re-evaluates expected payoffs at every state and does not hard-code permanent reversion rules. Trial-and-error learning therefore endogenously gives rise to self-restrained punishment: sufficiently harsh to deter frequent deviation but not so severe as to destroy the gains from collusion over the long run.

The contrast with classical theory is precise and worth stating explicitly. In the grim-trigger equilibrium, the credibility of permanent punishment rests entirely on pre-commitment: the agent is assumed to execute permanent reversion regardless of future payoffs, and it is this unconditional commitment that deters deviation ex ante. Q-learning agents, by construction, possess no such commitment mechanism; they re-evaluate expected payoffs at every state through the Bellman update in Equation (57). Once prices have fallen to the competitive level and remain there, the Q-values for continued punishment erode as the agent registers persistently low rewards, and initiating a cooperative recovery becomes the learned best response. Finite tit-for-tat is therefore not a

2026 S.T. Yau High School Science Award
收件人2026丘成桐中学科学奖公示

suboptimal approximation of grim-trigger but the rational equilibrium response of an agent that re-optimizes continuously without the ability to commit. This has a direct quantitative implication: the theoretical critical discount factor $\bar{\delta}(\sigma)$ derived under grim-trigger constitutes a *lower bound* on the patience actually required for algorithmic collusion. Empirically observed levels of algorithmic coordination should therefore be expected to fall below the theoretical collusive maximum, with the gap between the two reflecting the degree of bounded rationality in the punishment technology rather than any failure of the underlying collusion model.

Two concrete implications for the relationship between theory and simulation follow. The finite punishment horizon explains the moderate collusion levels ($\Delta \approx 0.2$) observed in simulations relative to the higher levels predicted under infinite-horizon grim-trigger: because the actual punishment is less severe than permanent reversion to p^N, the effective deterrent is weaker and equilibrium collusion does not reach the full cooperative level. The finite punishment also explains the rightward shift of the simulated phase transition relative to the theoretical threshold discussed in Section 5.2: greater patience is required to sustain collusion under finite punishment than under the permanent punishment assumed in deriving $\bar{\delta}_0$.

5.7 Counterfactual Grim-Trigger Simulation

To quantify the gap between the learned finite tit-for-tat strategy and the theoretical grim-trigger assumption, a counterfactual simulation is conducted in which one agent is hardcoded to follow grim-trigger: the collusive price is maintained as long as no deviation is detected, and permanent reversion to p^N follows upon detection. The rival agent continues to use standard Q-learning. If the theoretical analysis is correct that grim-trigger punishment sustains stronger collusion than the emergent finite tit-for-tat, a higher collusion index should be observed in the counterfactual.

Table 7: Counterfactual grim-trigger versus emergent Q-learning punishment ($N = 2$, $\sigma = 1.0$).

Configuration	$\gamma = 0.90$	$\gamma = 0.95$
Both agents Q-learning (emergent tit-for-tat)	0.15 ± 0.06	0.23 ± 0.08
One agent hardcoded grim-trigger	0.24 ± 0.07	0.35 ± 0.09
Both agents hardcoded grim-trigger	0.38 ± 0.05	0.52 ± 0.06

The prediction is borne out, as Table 7 shows. When both agents use Q-learning and develop emergent tit-for-tat, the collusion index is $\Delta \approx 0.23$ at $\gamma = 0.95$. When one agent is hardcoded to grim-trigger the collusion index rises to $\Delta \approx 0.35$, and when both agents follow grim-trigger it increases further to $\Delta \approx 0.52$. The roughly twofold increase from emergent to hardcoded grim-trigger quantifies the extent to which the theoretical framework overestimates collusion sustainability under bounded rationality. The gap is larger at lower γ, which is expected because the relative advantage of permanent punishment over finite punishment is more pronounced when agents are less patient.

2026 S.T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公开

5.8 Robustness to Alternative Parameter Calibrations and Hyperparameter Sensitivity

Table 8: Robustness across hyperparameter perturbations ($N = 2$, $\sigma = 1.0$, $\gamma = 0.95$).

Treatment	$\alpha = 0.05$	$\alpha = 0.15$	$\alpha = 0.3$	95% CI Range
Baseline ($v = 3$)	0.21	0.23	0.20	± 0.06
High valuation ($v = 5$)	0.22	0.24	0.21	± 0.07
Optimistic Q initialization	0.22	0.25	0.21	± 0.07
Quadratic transport cost	0.18	0.20	0.17	± 0.08

A concern with the baseline calibration ($v = 3$, $c = 1$, $t = 1$) is that $p^M = 2.75$ coincides with the full-coverage boundary $v - t/(2N) = 2.75$. Under $v = 5$, p^M remains at 2.75 while the full-coverage boundary rises to 4.75, placing the collusive outcome well inside the full-coverage region. As Table 8 shows, collusion indices under $v = 5$ are statistically indistinguishable from the baseline across all learning rates. All core monotonic comparative-static relations remain unchanged across the additional perturbation groups (optimistic Q initialization, quadratic transport cost), and the results are therefore not driven by the boundary coincidence in the baseline calibration.

The stability of the emergent finite tit-for-tat punishment duration is assessed in Table 9, which reports the average punishment duration (the number of periods before prices return to within 90% of the pre-deviation level) across different learning rates and exploration decay schedules.

Table 9: Sensitivity of tit-for-tat punishment duration to hyperparameters ($N = 2$, $\sigma = 1.0$, $\gamma = 0.95$).

Configuration	$\alpha = 0.05$	$\alpha = 0.15$	$\alpha = 0.3$
Baseline decay ($T_{decay} = 5 \times 10^5$)	7.2 ± 1.8	5.8 ± 1.5	4.9 ± 1.6
Slow decay ($T_{decay} = 8 \times 10^5$)	8.1 ± 2.0	6.5 ± 1.7	5.3 ± 1.8
Fast decay ($T_{decay} = 3 \times 10^5$)	6.4 ± 1.6	5.2 ± 1.4	4.5 ± 1.5

Punishment duration consistently falls in the range of roughly 4.5 to 8.1 periods across all configurations, and the finite tit-for-tat mechanism is therefore a stable behavioral property of Q-learning agents rather than an artifact driven by hyperparameter choice. Higher learning rates give shorter punishment durations because the agent more rapidly re-evaluates the relative payoff of returning to cooperative pricing. Slower exploration decay gives slightly longer punishment periods because the agent spends more time experimenting before settling on the recovery path.

2026 S. Tuyau High school Science Award
仅用于2026年成都市青少年科学会议公示

5.9 Theory-Simulation Consistency Summary

Table 10: Theory-simulation consistency verification.

Theoretical Prediction	Math Sign	Simulation	Match
Higher $t \Rightarrow$ easier collusion	$\partial\bar{\delta}/\partial t < 0$	Larger t raises Δ	Full
More sellers $\Rightarrow$ harder collusion	$\partial\bar{\delta}/\partial N > 0$	N rises, Δ falls	Full
Higher transparency $\Rightarrow$ easier collusion	$\partial\bar{\delta}/\partial\sigma < 0$	Monotonic Δ increase	Full
τ neutral to collusive incentive	Independent of τ	Δ invariant across τ	Full
Higher $\delta/\gamma \Rightarrow$ more collusion	Phase transition at $\bar{\delta}$	γ up, Δ rises	Full*
Grim-trigger > finite punishment	Higher Δ under grim	Counterfactual supports	Full
Punishment strategy form	Grim-trigger (theory)	Finite tit-for-tat	Novel

All five directional comparative statics and the counterfactual grim-trigger prediction are supported by the simulation. The asterisk on the discount-factor row indicates that while the directional prediction is fully supported, the quantitative location of the phase transition differs from the theoretical threshold because of the finite tit-for-tat punishment mechanism. As explained in Sections 5.2 and 5.6, this gap is internally consistent: bounded-rational agents employ weaker punishment than grim-trigger, requiring higher patience to achieve any given collusion level.

6 Model Extensions

Two extensions of the baseline model are presented to probe the robustness of the main theoretical results under relaxed assumptions.

6.1 Asymmetric Sellers

Consider N sellers with heterogeneous marginal costs c_i, $i = 1, \ldots, N$, and seller-specific commission rates τ_i. The net profit of seller i under collusive price vector $\mathbf{p}^M$ is:

$$\pi_i^M = (1 - \tau_i)(p_i^M - c_i)q_i^M \tag{60}$$

Because the factors $(1 - \tau_i)$ now differ across sellers, they do not cancel in the ratio $(\pi_i^D - \pi_i^M)/(\pi_i^D - \pi_i^N)$ and each seller faces a distinct critical discount factor $\bar{\delta}_i(\sigma)$. All individual incentive-compatibility constraints must be satisfied simultaneously for collusion to be sustainable, so the binding threshold is $\bar{\delta}^*(\sigma) = \max_i \bar{\delta}_i(\sigma)$.

Proposition 16 (Asymmetric Seller Collusion Threshold)**.** *Under heterogeneous costs c_i and commission rates τ_i, the system-level critical discount factor satisfies:*

$$\bar{\delta}^*(\sigma) = \max_{i=1,\ldots,N} \frac{(\pi_i^D - \pi_i^M)}{\rho(\sigma)(\pi_i^D - \pi_i^N) + (1 - \rho(\sigma))(\pi_i^D - \pi_i^M)} \geq \bar{\delta}_{sym}(\sigma) \tag{61}$$

where $\bar{\delta}_{sym}(\sigma)$ is the symmetric baseline threshold. The collusion threshold is weakly raised by heterogeneity because the most tempted seller (typically the lowest-cost seller, who gains the most from undercutting) determines the binding constraint.

2026 S. T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

Two qualitative implications follow. Commission neutrality breaks down under heterogeneous commission rates: if the platform charges higher fees to certain sellers, those sellers face lower post-commission margins and their deviation incentives may be altered relative to the homogeneous case. The Transparency Paradox ($\partial\bar{\delta}^*/\partial\sigma < 0$) survives, because each individual $\bar{\delta}_i(\sigma)$ is decreasing in σ by the same monotonicity argument as in Proposition 10, and the maximum of decreasing functions is itself decreasing.

6.2 Partial Coverage with Endogenous Participation

Section 3.2 derives the deadweight loss under partial coverage for a fixed collusive markup m. Here the collusive markup is endogenized by requiring it to be consistent with the deviation incentive constraint.

Under partial coverage the deviation profit changes because the deviating seller faces a demand that accounts for consumer exit at the market boundary. The modified deviation profit is:

$$\pi^D_{partial} = (1-\tau)(p^D - c) \cdot \min\left\{\frac{1}{N} + \frac{m}{2t},\ 2\hat{d}_D\right\} \tag{62}$$

where $\hat{d}_D = (v - p^D)/t$ is the coverage radius under the deviation price. Deviation demand is compressed by the additional constraint from partial coverage, reducing the short-run gain from undercutting and making collusion easier to sustain for the same markup m.

Proposition 17 (Partial Coverage Facilitates Collusion). *When $m > \bar{m}$, the critical discount factor under partial coverage satisfies $\bar{\delta}_{partial}(\sigma) < \bar{\delta}_{full}(\sigma)$. High collusive markups reduce coverage, which compresses deviation demand, lowers the temptation to undercut and makes collusion easier to sustain.*

An important regulatory implication follows: in markets where collusion is severe enough to exclude some consumers, the collusive equilibrium becomes self-reinforcing because the demand compression from consumer exit further stabilizes the supra-competitive outcome. Regulatory intervention in such markets provides both the consumer surplus recovery documented under full coverage and the additional benefit of eliminating the deadweight loss from Equation (53).

7 Regulatory Mechanism Design

Building on the theoretical framework and its simulation validation, three anti-collusion regulatory instruments are formalized in this section: information-transparency restriction, mandatory algorithm audit with financial penalties and real-time automatic price monitoring. Each instrument is evaluated within a unified welfare-decomposition framework, with explicit attention to real-world implementation challenges.

7.1 Information Transparency Restriction

An upper bound $\bar{\sigma}$ on the platform's allowable information transparency is imposed by the regulator, constraining the platform to $\sigma \in [0, \bar{\sigma}]$. Under strict binding regulation ($\bar{\sigma} < \bar{\sigma}(\delta)$), the maximum feasible transparency gives $\bar{\delta}(\bar{\sigma}) > \delta$, sellers cannot sustain tacit collusion and prices revert to p^N. Under loose regulation ($\bar{\sigma} \geq \bar{\sigma}(\delta)$), collusion persists with moderately reduced intensity.

Using the calibrated parameters $N = 4$, $m = 1.25$, $t = 1$, $\delta = 0.95$: $\bar{\sigma}(\delta) = mN(1-\delta)/(4t\delta) \approx 0.066$. A regulatory cap of $\bar{\sigma} = 0.05$ fully eliminates collusion. Under full suppression, consumers

2026 S. T. Yau High School Science Award
应用于2026丘成桐中学科学奖论文公开

gain the full markup ($\Delta CS = m > 0$), sellers lose collusive rents ($\Delta PS = -(1-\tau)m < 0$) and the platform loses commission premium ($\Delta\Pi^{plat} = -\tau m < 0$).

What this instrument operationalizes is an asymmetric information-design approach: platforms maintain full price-comparison information for consumers while restricting the real-time rival price data accessible to merchants' pricing algorithms. The European Digital Markets Act (Article 5) [18] and China's Algorithm Recommendation Management Regulations (Article 8) [19] provide initial legal frameworks within which such restrictions could be implemented. In China, Meituan's merchant dashboard supplies real-time pricing data on competing restaurants within the same delivery radius, and Taobao's "Business Advisor" analytics tool displays competitor pricing trends [20]. These information channels serve dual purposes: they help merchants make informed business decisions but they also supply the data infrastructure through which algorithmic pricing tools detect and respond to competitive deviations.

7.1.1 Implementation Challenges

Several practical obstacles must be addressed for effective deployment. Merchants may circumvent platform-level data restrictions by employing third-party web-scraping services or price-comparison aggregators to reconstruct competitor pricing data independently. Manual observation of competitor prices through consumer-facing interfaces provides another bypass channel, although it is less effective for high-frequency automated pricing. A practical compromise involves delayed disclosure: platforms transmit competitor price data to merchant algorithms with a lag of one to three hours rather than in real time. The informational value of competitive intelligence for medium-term business decisions is preserved by delayed disclosure, while the speed at which algorithmic punishment can be deployed is substantially reduced. When the disclosure lag exceeds the algorithmic repricing cycle, the effective detection probability $\rho(\sigma)$ decreases, $\bar{\delta}(\sigma)$ is raised and collusion sustainability may be broken. The threshold lag can be calibrated using the model: for the detection probability to fall below the critical level, the lag must be long enough that the discounted punishment payoff becomes insufficient to deter deviation.

7.2 Algorithm Audit with Financial Penalties

Random periodic audits are conducted by the regulator and governed by three parameters: audit frequency λ, detection probability μ and financial penalty F. Expected per-period collusive profit under audit risk becomes $\tilde{\pi}^M = \pi^M - \lambda\mu F$. Since $\tilde{\pi}^M < \pi^M$, the modified critical discount factor $\bar{\delta}_{audit} = (\pi^D - \tilde{\pi}^M)/(\pi^D - \pi^N) > \bar{\delta}_0$. The audit effectiveness index $\Phi = \lambda\mu F/(\pi^D - \pi^N)$ measures the incremental difficulty of sustaining collusion and is increasing in all three audit parameters.

Access to algorithmic source code or behavioral logs is required for this instrument, raising both proprietary-information concerns and technical complexity. The European Commission's proposed AI Liability Directive includes provisions for algorithmic auditing, and China's Cyberspace Administration has mandated algorithm-filing requirements under the Algorithm Recommendation Management Regulations [19]. Detecting tacit collusion through code inspection is non-trivial in practice, because the collusive behavior emerges from the interaction between learning rules and market conditions rather than from explicit cooperative instructions in the source code [3].

7.3 Real-Time Automatic Price Monitoring

An automatic monitoring module tracks seller prices against a ceiling $\bar{p} = p^N + \theta m$, where $\theta \in (0, 1)$ is a tolerance parameter. When prices exceed $\bar{p}$, sellers are forced to reset to p^N for a

2026 S. T. Yau High School Science Award 仅用于2026丘成桐中学科学奖公示

fixed number of periods. The maximum feasible collusion index is bounded by $\Delta \leq \theta$ regardless of algorithmic patience. A trade-off between Type I error (normal price fluctuations triggering unnecessary resets, decreasing in θ) and Type II error (collusive prices below the threshold escaping detection, increasing in θ) is inherent in the policy. The optimal threshold θ^* equates the marginal costs of both error types.

7.4 Comparative Analysis

Table 11: Comparative analysis of three anti-collusion instruments.

Dimension	Info. Restriction	Algorithm Audit	Price Monitoring
Timing	Ex ante	Ex post	Real-time
Mechanism	Raise $\bar{\delta}$ by weakening detection	Expected loss from fines	Cap maximum markup directly
Code access needed	No	Yes	No
Administrative cost	Low	High	Medium
Binding power	Depends on δ	Depends on F	Absolute on Δ

No access to algorithmic source code is required for information-transparency restriction, and its administrative cost is the lowest. Code inspection at high cost is required for the audit approach. Moderate cost together with absolute collusion-capping power characterizes real-time monitoring.

7.5 Welfare Quantitative Evaluation

The unified benchmark parameters are $N = 4$, $c = 1$, $t = 1$, $v = 3$, $\tau = 0.1$, $m = 1.25$, $p^N = 1.5$, $p^M = 2.75$:

Table 12: Baseline welfare values under collusion and competition.

Welfare Term	Collusion	Competition	Difference
Consumer Surplus CS	0.1875	1.4375	+1.25
Seller Surplus PS	1.575	0.45	−1.125
Platform Revenue Π^{plat}	0.275	0.15	−0.125
Total Welfare W	1.9375	1.9375	0

The zero total-welfare difference verifies the Welfare Invariance Theorem (Theorem 11).

Under the consumer-welfare standard adopted throughout antitrust analysis, the net social benefit of a regulatory policy is defined as the consumer surplus gain net of administrative cost:

$$NB = \Delta CS - C_{policy} \tag{63}$$

This standard is appropriate for two reasons. Under full market coverage, the Welfare Invariance Theorem (Theorem 11) implies $\Delta CS + \Delta PS + \Delta\Pi^{plat} = 0$, so a total-welfare criterion would reduce to $NB = -C_{policy}$ and would rank instruments solely by administrative cost, providing no information about the magnitude of anticompetitive harm corrected. The consumer-welfare standard resolves this degeneracy by measuring the surplus recovered for the party harmed by collusion,

2026 Ji Yan High School Science Award
仅用于2026年吉祥中学科学论文公示

consistent with the enforcement objectives of competition authorities in both the European Union [18] and China [20].

Administrative costs for each regulatory instrument are calibrated as hypothetical benchmarks reflecting the relative cost structure across policy types. Information restriction ($C_{IL} = 0.15$) involves regulatory drafting and compliance monitoring with no need for technical code inspection. Algorithm audit ($C_{AUD} = 0.30$) requires specialized technical personnel and algorithmic code inspection infrastructure, making it the most expensive instrument. Real-time price monitoring ($C_{MON} = 0.20$) requires deployment and maintenance of automated surveillance systems at moderate cost. Because these cost values are hypothetical rather than empirically estimated from real regulatory expenditure data, the core conclusions depend on the relative ordering of costs rather than on their absolute magnitudes. Robustness across a range of alternative cost assumptions is demonstrated in Table 14.

Table 13: Net social benefit of individual regulatory policies.

Policy	ΔCS	ΔPS	$\Delta\Pi^{plat}$	C_{policy}	$NB = \Delta CS - C$
Info. Restriction	1.25	−1.125	−0.125	0.15	1.10
Algorithm Audit	1.25	−1.125	−0.125	0.30	0.95
Loose Monitoring ($\theta = 0.3$)	0.375	−0.338	−0.038	0.20	0.175
Strict Monitoring ($\theta = 0$)	1.25	−1.125	−0.125	0.20	1.05

Under the calibrated parameters under the consumer-welfare standard, the ranking of standalone anti-collusion policies by net social benefit is $NB_{IL}(1.10) > NB^{strict}_{MON}(1.05) > NB_{AUD}(0.95) > NB^{loose}_{MON}(0.175)$. Information-transparency restriction achieves the highest net social welfare among standalone instruments because it attains full collusion suppression at the lowest administrative cost.

Table 14: Sensitivity of net-social-benefit ranking to administrative-cost assumptions.

Scenario	C_{IL}	C_{AUD}	C_{MON}	NB_{IL}	Ranking
Baseline	0.15	0.30	0.20	1.10	Preserved
Low-cost audit	0.15	0.18	0.20	1.10	Preserved
High-cost info. restr.	0.25	0.30	0.20	1.00	Preserved
Equal costs	0.20	0.20	0.20	1.05	Tie (strict mon.)
Reversed IL-AUD	0.30	0.15	0.20	0.95	Reversed

The dominance of information-transparency restriction is found to be robust when considering moderate perturbations in cost assumptions, as Table 14 shows. The ranking reverses only when the administrative cost of information restriction exceeds that of algorithm auditing, a scenario that appears unlikely given that information restriction does not require access to proprietary algorithmic source code. Under equal costs, information restriction ties with strict price monitoring; in all other scenarios it retains the top ranking.

The dominance of information-transparency restriction extends to the partial-coverage case: in markets where $m > \bar{m}$, information restriction additionally eliminates the deadweight-loss term from Equation (53), further widening its net-benefit advantage over the other instruments.

2026 Si-Yau High School Science Award 仅用于2026丘成桐中学科学奖论文公示

7.6 Hierarchical Regulatory Design for High-Frequency Markets

For ultra-high-frequency markets ($\delta \to 1$) where $\bar{\sigma}(\delta) \to 0$ and information limits alone cannot break collusion without impractically strict caps, a hierarchical combination of all three tools becomes necessary. Information restriction serves as the baseline layer, raising $\bar{\delta}$ and reducing the intensity of collusion. Real-time price monitoring addresses residual partial collusion persisting above the transparency cap. Algorithm audits serve as a final deterrent for non-compliant platforms that circumvent the other two instruments. This layered approach reflects the reality that no single regulatory instrument is universally sufficient and that the optimal policy portfolio depends on the algorithmic repricing frequency and the market structure parameters.

8 Discussion

8.1 Theoretical Implications

The central analytical contribution of this study is the closed-form critical discount factor $\bar{\delta}(\sigma) = mN/(4t\sigma + mN)$, an expression linking four economically interpretable parameters in a single formula that provides testable predictions for empirical work and calibratable thresholds for regulatory design. The classical imperfect monitoring framework of Green and Porter [9] is extended through this expression by endogenizing the precision of monitoring using the platform's information architecture. All associated comparative statics ($\partial\bar{\delta}/\partial t < 0$, $\partial\bar{\delta}/\partial N > 0$, $\partial\bar{\delta}/\partial\sigma < 0$) are supported by the multi-agent Q-learning simulations with continuous transparency variation.

How the present work relates to the foundational contribution of Calvano et al. [1] warrants clarification. Calvano et al. study algorithmic collusion in a homogeneous Bertrand market with exogenously given information structure. Three substantive extensions are introduced here. Competition is embedded in a Salop circular city with horizontal product differentiation, in which the local-competition structure qualitatively alters deviation incentives. The platform's information-disclosure decision is endogenized through the transparency parameter σ, from which the Transparency Paradox emerges. The platform's commission rate τ is introduced as a strategic variable, and the Commission Neutrality Theorem is established. Taken together, these extensions transform the analysis from an algorithmic behavioral study into a complete market-design problem with closed-form policy implications.

While structurally related to the classical insight that proportional taxes leave the qualitative equilibrium unchanged [11], the Commission Neutrality result acquires new significance in the platform context because a fundamental separability between information governance and surplus redistribution is established. The common regulatory instinct to use platform fee adjustments as a competition tool is found to be ineffective against algorithmic collusion. As demonstrated in the asymmetric-seller extension in Section 6, neutrality breaks down under heterogeneous commission rates, delineating the precise boundary of this result.

Finite tit-for-tat punishment emerging in Q-learning agents constitutes the primary computational contribution. The behavioral mechanism is grounded in Q-learning update dynamics: permanent punishment erodes Q-values for sustained low pricing, and the agent learns that brief retaliation followed by cooperative recovery dominates permanent reversion in terms of cumulative discounted reward. The results of Calvano et al. [1] are extended to differentiated markets, and it is demonstrated that the behavioral divergence between bounded-rational algorithms and fully rational agents is not an artifact of the homogeneous-product setting. The counterfactual grim-trigger simulation (Section 5.7) quantifies the collusion gap: models based on grim-trigger strategies overestimate collusion intensity by roughly a factor of two under the baseline calibration, suggesting that empirically observed collusion levels may be substantially lower than theoretical

2026 S. T. Yau High School Science Award
仅用于2026年丘成桐中学科学奖公示

maxima. The quantitative gap between the theoretical phase-transition threshold ($\bar{\delta}_0 \approx 0.385$) and the simulated phase-transition location ($\gamma \approx 0.7$) is internally consistent with this finding.

The partial-coverage extension (Section 3.2) together with the asymmetric-seller extension (Section 6) probes the boundaries of the baseline results. The deadweight-loss expression $DWL(m) = N(m - \bar{m})^2/(2t)$ gives a quantitative measure of allocative inefficiency arising only beyond the full-coverage threshold, justifying the baseline welfare invariance within its domain of validity and making explicit the conditions under which it breaks down. Heterogeneity is shown by the asymmetric-seller analysis to weakly raise the collusion threshold without altering the qualitative Transparency Paradox, lending robustness to the main theoretical message.

8.2 Policy Implications

The welfare analysis establishes that restricting platform information transparency is the most cost-effective standalone anti-collusion instrument. An asymmetric information-design approach is supported: platforms should be required to maintain full price-comparison information for consumers while restricting the real-time rival price data accessible to merchants' pricing algorithms. Consumer-welfare benefits of price transparency are preserved by separating the information rights of consumers from those of sellers, while the algorithmic infrastructure of tacit coordination is undermined.

An important difference in the regulatory justification arises between the full-coverage and partial-coverage regimes. Under full coverage, algorithmic collusion generates no deadweight loss (Theorem 11) and the case for regulatory intervention rests on distributional grounds: collusion transfers surplus from consumers to sellers and the platform, and intervention is justified when the normative weight assigned to consumer welfare exceeds that assigned to producer welfare. Under partial coverage ($m > \bar{m}$), an efficiency rationale supplements the distributional argument because collusion excludes marginal consumers from the market and creates genuine allocative waste (Theorem 13). The nature of the market failure being addressed differs across these two regimes, and the interventionist case is considerably stronger in markets prone to partial coverage.

A common regulatory misconception is eliminated by the τ-σ separability result. Adjusting platform service fees should not be relied upon to curb algorithmic collusion, since commission regulation only redistributes surplus without affecting equilibrium price levels or collusive incentives. Information governance objectives must be explicitly separated from surplus redistribution objectives in policy design.

In the Chinese regulatory context, the State Administration for Market Regulation (SAMR) has recently intensified scrutiny of algorithmic pricing practices on major platforms [20]. The existing provisions in the Algorithm Recommendation Management Regulations [19] represent a necessary but incomplete framework, as the model's predictions suggest. Current regulations focus primarily on algorithmic transparency to users (requiring explanations of recommendation logic) rather than on the inter-merchant information flows that sustain algorithmic coordination. Extending the regulatory scope to cover merchant-facing competitive intelligence tools, including third-party analytics services and platform-supplied "business advisor" dashboards, would more directly target the information channel through which tacit collusion is sustained.

8.3 Limitations

Several important caveats delimit the boundaries within which the conclusions of this analysis are valid.

On the theoretical side, the baseline model assumes a monopoly platform, symmetric sellers, a uniform proportional commission, a linear detection function and an exogenous collusive markup.

2025 S.-T. Yau High School Science Award
仅用于2026年丘成桐中学科学奖公示

The monopoly-platform assumption gives the sharp prediction $\sigma^* = 1$; under inter-platform competition, optimal transparency is weakened but remains above the consumer-welfare-maximizing level (Remark 4). The exogenous treatment of m is adopted for analytical tractability at the cost of not characterizing the optimal collusive markup endogenously (Remark 3 and Appendix B). Endogenizing m through joint-profit maximization subject to incentive-compatibility and full-coverage constraints would eliminate closed-form solutions and require numerical methods, a direction left for future work.

On the computational side, a ceiling of $N \leq 5$ sellers is imposed by the tabular Q-learning architecture because of exponential state-space growth. Deep reinforcement-learning methods such as DQN and DDPG with continuous state approximation would alleviate this scalability constraint and may give rise to different pricing and punishment patterns. The local-competition structure of the Salop model provides partial justification for extrapolating the mechanisms (though not necessarily the quantitative magnitudes) to larger markets. All simulated agents deploy identical Q-learning algorithms, whereas real markets feature heterogeneous algorithmic architectures from simple rule-based heuristics to sophisticated deep learning agents. Whether the finite tit-for-tat finding persists under such algorithmic heterogeneity is an open question. Asker, Fershtman and Pakes [14] have shown that the choice of reinforcement-learning architecture affects both the likelihood and the degree of supra-competitive pricing, suggesting that mixed algorithmic populations constitute an important direction for future research.

On the empirical side, all simulation parameters are calibrated rather than estimated from real transaction data. Empirical validation based on transaction-level data from actual e-commerce platforms would substantially strengthen the practical relevance of the theoretical predictions [4]. The closed-form expression $\bar{\delta}(\sigma) = mN/(4t\sigma + mN)$ provides testable comparative statics that could be confronted with cross-platform variation in transparency policies and observed pricing patterns. Future research may also relax the symmetric-seller assumption and explore how heterogeneous algorithmic architectures interact with the transparency mechanism documented here.

9 Conclusion

A complete analytical framework integrating game-theoretic modeling, multi-agent Q-learning simulation and welfare-based regulatory evaluation is constructed in this study to investigate algorithmic tacit collusion on differentiated e-commerce platforms. The Salop circular-city framework is extended to incorporate endogenous platform information design and repeated algorithmic price competition.

The primary analytical contribution is the closed-form critical discount factor $\bar{\delta}(\sigma) = mN/(4t\sigma + mN)$, from which a Transparency Paradox emerges: profit-maximizing platforms endogenously select full transparency to lower the patience threshold for collusion, creating a structural conflict with consumer welfare. The Commission Neutrality result establishes the τ-σ separability principle identifying information governance as the fundamental anti-collusion instrument. These results are extended to partial market coverage, where the deadweight-loss expression $DWL(m) = N(m - \bar{m})^2/(2t)$ quantifies the allocative inefficiency arising beyond the full-coverage boundary, and to asymmetric sellers, where heterogeneity raises the collusion threshold without altering the qualitative Transparency Paradox.

The primary computational contribution is the novel finding that Q-learning algorithms autonomously develop finite-period tit-for-tat retaliation rather than permanent grim-trigger punishment. The underlying behavioral mechanism arises from Q-learning update dynamics: permanent punishment erodes expected payoffs, and agents learn that brief retaliation followed by cooperative recovery dominates permanent reversion. A counterfactual grim-trigger simulation quantifies the

2026 S. T. YAU High School Science Award
获得于2026年由斯坦福中学科学奖公开

resulting collusion gap, showing that theoretical models overestimate collusion intensity by roughly a factor of two. Comprehensive simulations using a ten-level continuous transparency grid support all five directional comparative statics predicted by the closed-form expression.

Using the unified welfare-decomposition framework, restricting platform information transparency is found to achieve the highest net social benefit among standalone anti-collusion policies, and this conclusion is found to be robust when considering sensitivity analysis over administrative-cost assumptions. Connections are drawn to existing regulatory frameworks in the European Union and China, and practical implementation challenges together with compromise solutions such as delayed price disclosure are discussed. As autonomous reinforcement-learning pricing algorithms continue to proliferate, the closed-form thresholds and behavioral findings developed here provide quantitative foundations for evidence-based regulatory design that balances market competition with consumer protection.

2026 S. -T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

10 Acknowledgements

This research was completed under the patient guidance of my supervisors, Ms. Wang Yijun (Shenzhen Middle School) and Prof. Mu Nianguo (University of Shanghai for Science and Technology), combined with my own persistent efforts. I would like to express my deepest gratitude to both supervisors for their selfless support throughout the entire research process. I also sincerely thank Shenzhen Middle School for providing students with space and resources for independent exploration.

The inspiration for this project came from my interest in digital platform economics and antitrust policy. It was Ms. Wang who first encouraged me to explore the intersection of game theory and reinforcement learning, and Prof. Mu who guided me toward the formal game-theoretic framework and helped me understand the mathematical foundations of collusion sustainability analysis.

The entire project was independently carried out by me, including topic selection, literature review, model derivation, Q-learning simulation programming, and paper writing. Ms. Wang and Prof. Mu provided guidance on a voluntary, unpaid basis. Ms. Wang played a crucial role in helping me identify the research question and structure the three-layer sequential game, while Prof. Mu offered invaluable technical guidance on the derivation of the closed-form critical discount factor and the welfare-decomposition framework. During the simulation phase, Prof. Mu suggested the counterfactual grim-trigger experiment, which proved essential for quantifying the gap between bounded-rational and fully rational punishment behavior.

When writing the paper, I lacked experience in academic exposition. My early drafts were poorly structured and several proofs were incomplete. Both supervisors reviewed the manuscript carefully, pointed out logical gaps, and helped me clearly distinguish the novel contributions of this work from prior literature. Ms. Wang particularly emphasized the importance of connecting the theoretical results to real-world regulatory implications in China, which shaped the policy analysis in the final sections.

This project has taught me that rigorous research requires not only mathematical tools but also clear economic reasoning and honest acknowledgment of limitations. I am deeply grateful to Ms. Wang Yijun and Prof. Mu Nianguo for their consistent encouragement and guidance throughout this journey.

I further declare that no generative artificial intelligence tools were used at any stage of this research, including topic selection, literature review, model derivation, simulation programming, and manuscript writing. All intellectual content in this paper is solely the product of my own work under the supervision of my advisors.

2026 S. T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

References

[1] Calvano, E., Calzolari, G., Denicolò, V., & Pastorello, S. (2020). Artificial intelligence, algorithmic pricing, and collusion. *American Economic Review*, **110**(10), 3267–3297.

[2] Ezrachi, A. & Stucke, M. E. (2016). *Virtual Competition: The Promise and Perils of the Algorithm-Driven Economy*. Harvard University Press.

[3] Harrington, J. E. (2018). Developing competition law for collusion by autonomous artificial agents. *Journal of Competition Law & Economics*, **14**(3), 331–363.

[4] Assad, S., Clark, R., Ershov, D., & Xu, L. (2024). Algorithmic pricing and competition: Empirical evidence from the German retail gasoline market. *Journal of Political Economy*, **132**(3), 723–771.

[5] OECD. (2023). Algorithmic competition. *OECD Roundtables on Competition Policy Papers*, No. 296. OECD Publishing.

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

[7] Calvano, E., Calzolari, G., Denicolò, V., & Pastorello, S. (2021). Algorithmic collusion with imperfect monitoring. *International Journal of Industrial Organization*, **79**, 102712.

[8] Fudenberg, D. & Tirole, J. (1991). *Game Theory*. MIT Press.

[9] Green, E. J. & Porter, R. H. (1984). Noncooperative collusion under imperfect price information. *Econometrica*, **52**(1), 87–100.

[10] Salop, S. C. (1979). Monopolistic competition with outside goods. *The Bell Journal of Economics*, **10**(1), 141–156.

[11] Tirole, J. (1988). *The Theory of Industrial Organization*. MIT Press.

[12] Calvano, E., Calzolari, G., Denicolò, V., & Pastorello, S. (2024). Artificial intelligence recommendations: Evidence, issues, and policy. *Oxford Review of Economic Policy*, **40**(4), 843–853.

[13] Brown, Z. Y. & MacKay, A. (2023). Competition in pricing algorithms. *American Economic Journal: Microeconomics*, **15**(2), 109–156.

[14] Asker, J., Fershtman, C., & Pakes, A. (2022). Artificial intelligence, algorithm design, and pricing. *AEA Papers and Proceedings*, **112**, 452–456.

[15] den Boer, A. V., Meylahn, J. M., & Schinkel, M. P. (2026). Artificial collusion: Examining supracompetitive pricing by Q-learning algorithms. *Management Science*, forthcoming.

[16] Johnson, J. P., Rhodes, A., & Wildenbeest, M. (2023). Platform design when sellers use pricing algorithms. *Econometrica*, **91**(5), 1841–1879.

[17] Bergemann, D. & Morris, S. (2019). Information design: A unified perspective. *Journal of Economic Literature*, **57**(1), 44–95.

[18] European Parliament and Council. (2022). Regulation (EU) 2022/1925 on contestable and fair markets in the digital sector (Digital Markets Act). *Official Journal of the European Union*, L 265, 1–66.

2026 S.T. Yau High School Science Award 仅用于2026丘成桐中学科学奖论文公示

[19] Cyberspace Administration of China. (2021). Provisions on the Administration of Algorithmic Recommendation for Internet Information Services. Order No. 9.

[20] Fang, H., Ke, R., & Zhou, L.-A. (2021). Competition and market structure of platform economies in China. In *The Handbook of China's Financial System*. Princeton University Press, 359–386.

[21] Zhou, W. & Li, M. (2023). Algorithmic pricing regulation in digital markets: Chinese perspectives and international comparisons. *China Economic Review*, **78**, 101932.

2026 S.-T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

A Symbol Table

Table 15: Complete symbol definitions.

	Exogenous Parameters		
N	Number of sellers on the circular market	c	Marginal production cost
t	Product differentiation (transport cost)	v	Consumer base valuation
m	Collusive markup (exogenous), $m > 0$	$\underline{\pi}$	Sellers' reservation profit
K	Number of discrete price grid points	T	Number of training rounds
α	Q-learning rate	ε_t	Exploration probability at round t
	Endogenous / Decision Variables		
τ	Commission rate, $\tau \in (0,1)$	σ	Transparency parameter, $\sigma \in [0,1]$
p_i	Price of seller i	q_i	Demand for seller i
δ	Seller discount factor	γ	RL discount factor (analogue of δ)
$Q(s,a)$	State-action value function		
	Derived Equilibrium Quantities (Original Contributions)		
p^N	Competitive price, $c + t/N$	p^M	Collusive price, $c + t/N + m$
p^D	Optimal deviation price, $p^N + m/2$	π^N	Competitive profit, $(1-\tau)t/N^2$
π^M	Collusive profit, $(1-\tau)(t+mN)/N^2$	π^D	Deviation profit
$\rho(\sigma)$	Detection probability	$\bar{\delta}_0$	Baseline critical discount factor
$\bar{\delta}(\sigma)$	Modified critical discount factor	$\bar{\sigma}(\delta)$	Collusion threshold transparency
σ^{CS}	Consumer-optimal transparency	$\bar{m}$	Critical full-coverage markup
DWL	Deadweight loss (partial coverage)	Δ	Collusion index
	Welfare and Policy Variables		
CS	Consumer surplus	PS	Seller producer surplus
Π^{plat}	Platform commission revenue	W	Total social welfare
NB	Net social benefit of policy	λ	Audit frequency
μ	Audit detection probability	F	Financial penalty
θ	Monitoring markup threshold	Φ	Audit effectiveness index

B Endogenous Collusive Markup: Formal Problem and Analytical Difficulties

If the collusive markup m is endogenized through joint-profit maximization by the colluding sellers, the optimization problem becomes:

$$\max_{m>0} \ \pi^M(m) = \frac{(1-\tau)(t+mN)}{N^2} \tag{64}$$

subject to the incentive-compatibility constraint:

$$\delta \geq \bar{\delta}(\sigma; m) = \frac{mN}{4t\sigma + mN} \tag{65}$$

2026 S.T. Yau High School Science Award
仅用于2026丘成桐中学科学奖公示

and the full-coverage constraint:

$$m \leq \bar{m} = v - c - \frac{3t}{2N} \tag{66}$$

Because $\pi^M(m)$ is linear and increasing in m, the unconstrained optimum is $m \to \infty$. The binding constraint is therefore either the incentive-compatibility condition or the full-coverage condition, depending on parameters.

When the incentive-compatibility constraint binds, the maximum sustainable markup is obtained by inverting $\bar{\delta}(\sigma; m) = \delta$:

$$m^*(\sigma, \delta) = \frac{4t\sigma\delta}{N(1-\delta)} \tag{67}$$

This expression is increasing in both σ and δ, meaning that higher transparency together with greater patience expands the set of sustainable markups, consistent with the Transparency Paradox. When both constraints bind simultaneously ($m^* = \bar{m}$), the required discount factor is $\delta^* = \bar{m}N/(4t\sigma + \bar{m}N)$, which recovers the closed-form expression evaluated at the critical markup.

The rationale for treating m as exogenous in the main analysis is that introducing the binding incentive-compatibility constraint into the welfare analysis creates a circular dependence: the welfare effects of collusion depend on m, the sustainable m depends on the platform's transparency choice σ, and σ is chosen by the platform to maximize commission revenue that depends on m. Solving this simultaneous system analytically requires parameterizing the binding constraint into the platform's optimization, which eliminates the closed-form structure that makes the comparative statics transparent and directly interpretable. Analytical tractability enabling the clear identification of each parameter's effect on collusion sustainability is preserved by the exogenous treatment of m.

2026 S. -T. Yau High School Science Award
仅用于2026丘成桐中学科学奖公示