

MASTERARBEIT | MASTER'S THESIS

Titel | Title

Algorithmic Collusion under Price Adjustment Frictions: an
Agent-Based Analysis with Menu Costs

verfasst von | submitted by

Fanni Bader BSc

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of
Master of Science (MSc)

Wien | Vienna, 2026

Studienkennzahl lt. Studienblatt | Degree
programme code as it appears on the
student record sheet:

UA 066 915

Studienrichtung lt. Studienblatt | Degree
programme as it appears on the student
record sheet:

Masterstudium Betriebswirtschaft

Betreut von | Supervisor:

Assoz. Prof. Daniel Garcia PhD

Abstract

Setting prices is increasingly handled by algorithms rather than by people, and simulation research repeatedly finds that independent pricing algorithms learn to hold prices above the competitive level. Coordination that arises this way, with no agreement between the firms, falls outside the reach of current competition law. Changing a price is not free in practice, and the form this cost takes is known to shape how a firm adjusts prices, but this menu cost friction has not yet been tested in combination with pricing algorithms.

This thesis adds a menu cost to the reward function of an established Q-learning pricing duopoly, charged on every price change and set as a proportion of the collusive premium, and compares the resulting behaviour against the frictionless case across 21 friction levels with 1,000 seeds each. At zero friction the simulation reproduces the published benchmark.

The friction reshapes collusion without preventing it. Mean profitability falls by roughly a fifth of the collusive premium and stays far above the competitive benchmark at every level tested. Edgeworth cycling becomes extinct once the menu cost reaches about nine percent of that premium, while markets that settle on a fixed price persist at every level tested. Collusion therefore becomes less varied rather than less common. The agents also stop revising their policies earlier and cover less of the price grid. Decomposing the profit loss shows that what firms actually pay is a minor and eventually vanishing part of it, while most travels through reduced exploration.

Menu costs therefore act through the learning process rather than through the fee itself. A cost that firms in almost every real market already face did not prevent collusion at any level tested, so real-world frictions should not be expected to contain algorithmic collusion on their own. Most of the effect occurs before the agents converge, which points to the learning phase as the stage where an intervention has the most potential to act on. These results come from a simulated duopoly with one fixed per-change cost specification, so they are a starting point for expectations about real industries rather than direct evidence about them.

Abstract (German)

Immer häufiger übernehmen die Preissetzung Algorithmen statt Menschen. Simulationsstudien zeigen dabei, dass unabhängig voneinander agierende Preisalgorithmen es lernen, Preise oberhalb des Wettbewerbsniveaus zu halten. Da eine solche Koordination ohne jede Absprache zwischen den Unternehmen zustande kommt, entzieht sie sich dem geltenden Wettbewerbsrecht. Preisänderungen sind in der Praxis allerdings nicht kostenlos, und wie ein Unternehmen seine Preise anpasst, hängt nachweislich von der Form dieser Kosten ab. Für Preisalgorithmen wurden solche Menükosten bislang nicht untersucht.

Die vorliegende Arbeit erweitert die Belohnungsfunktion eines etablierten Q-Learning-Duopols mit Preiswettbewerb um genau diese Kosten. Sie fallen bei jeder Preisänderung an und sind als fixer Anteil der Kollusionsprämie festgelegt. Simuliert wurden 21 Kostenstufen mit je 1.000 Seeds, deren Ergebnisse dem Fall ohne Menükosten gegenübergestellt werden. Bei einer Kostenstufe von null reproduziert die Simulation die publizierten Referenzergebnisse.

Menükosten verändern die Form der Kollusion, ohne sie zu verhindern. Die durchschnittliche Profitabilität sinkt um rund ein Fünftel der Kollusionsprämie, bleibt aber auf allen getesteten Stufen deutlich über dem Wettbewerbsniveau. Sobald die Menükosten etwa neun Prozent dieser Prämie erreichen, verschwinden Edgeworth-Zyklen vollständig, während Märkte mit statischer Preissetzung davon unberührt bleiben. Kollusion wird dadurch weniger vielfältig, nicht seltener. Die Agenten stellen die Überarbeitung ihrer Strategien zudem früher ein und decken einen kleineren Teil der verfügbaren Preise ab. Eine Zerlegung des Gewinnverlusts zeigt schließlich, dass die tatsächlich gezahlten Menükosten nur einen kleinen und letztlich verschwindenden Anteil daran ausmachen. Der größte Teil entsteht über die verringerte Exploration.

Menükosten wirken damit über den Lernprozess und nicht über die gezahlte Gebühr. Obwohl Unternehmen in nahezu jedem realen Markt solchen Kosten ausgesetzt sind, haben diese Kollusion auf keiner getesteten Stufe verhindert. Von realen Anpassungskosten sollte daher nicht erwartet werden, dass sie algorithmische Kollusion allein eindämmen. Da der größte Teil der Wirkung eintritt, bevor die Agenten konvergieren, erscheint die Lernphase als der Abschnitt mit dem größten Interventionspotenzial. Die Ergebnisse stammen allerdings aus einem simulierten Duopol mit fixen, pro Preisänderung anfallenden Kosten. Sie sind daher ein Ausgangspunkt für Erwartungen an reale Branchen und kein unmittelbarer Beleg für deren Verhalten.

Contents

Abstract	i
Abstract (German)	ii
List of Tables	vi
List of Figures	vii
1 Introduction and Motivation	1
2 Theoretical Background	2
2.1 Foundations of Price Rigidity and Menu Costs	2
2.1.1 The cost of changing prices	2
2.1.2 Small costs, large behavioural effects	3
2.1.3 The form of the friction determines the behaviour	3
2.1.4 Limits of attributing effects to menu costs	4
2.2 Tacit Collusion and Algorithmic Pricing	4
2.2.1 Tacit collusion in classical oligopoly theory	4
2.2.2 The real-world use of algorithmic pricing	5
2.2.3 Algorithmic collusion and its conditions	6
2.3 Ethical and Legal Dimensions	7
2.3.1 Limits of the current legal framework	7
2.3.2 Proposed responses	8
2.4 Research Gap and Positioning	8
2.4.1 The research gap	8
2.4.2 Positioning and research question	9
3 Methodology	10

3.1	The Economic Environment: Replicating the Baseline Model	10
3.2	State Space and Action Space Discretization	10
3.3	Departure from the Baseline: The Net Profit Function and Menu Costs . .	11
3.4	Agent Architecture and Q-Learning Mechanics	12
3.5	Data Collection Pipeline and the Two-Phase Evaluation Strategy	13
3.5.1	Master Registry and Final Data Export	15
3.6	Statistical Methods used for the Analysis	16
3.6.1	The Analysis Panel	16
3.6.2	Net Profitability	16
3.6.3	Paired and Independent Comparisons	17
3.6.4	Multiple Comparisons and Inclusion Thresholds	17
3.6.5	Correlation and Spatial Measures	18
3.6.6	Visualisation Conventions	18
4	Results	18
4.1	Baseline Validation	18
4.2	Market Outcomes and the Phase Transition	20
4.2.1	Continuous Profit Degradation	20
4.2.2	The Extinction Threshold	21
4.2.3	Profitability of Cycling versus Static Behavior	22
4.2.4	The Erosion of Credible Punishment	24
4.3	Temporal Mechanics and Premature Calcification	26
4.3.1	Premature Strategy Calcification	26
4.3.2	The Cumulative Friction Tax	27
4.4	Spatial Blindness and Information Decay	30
4.4.1	Policy Volatility and Freezing	30

4.4.2	The Decay of Spatial Awareness	31
4.4.3	The Contraction of the Explored State Space	33
4.5	Decomposing the Profit Loss	35
5	Discussion and Limitations	39
5.1	Selection instead of adaptation	39
5.2	Collusion and cycling markets	39
5.3	The effect of less exploration	40
5.4	Shortened training time	40
5.5	Intervention placement	40
5.6	Limitations	41
6	Conclusion	42
	Documentation of Tools and Aids Used	43
	Bibliography	48
A	Appendix: Supplementary Tables	48

List of Tables

1	Direct estimates of the cost of a price change	3
2	Continuous Profit Degradation: Wilcoxon Signed-Rank Test vs. Baseline . .	20
3	Population-Averaged GEE Panel Logit: Cycle Survival	22
4	Static vs. Cyclers values across friction tiers	23
5	Static vs. Cyclers Statistical tests	24
6	Punishment Depth: Baseline Cyclers compared to Cyclers at Each Tier . .	25
7	Training-Phase Price-Change Rate for Cyclers and Statics with Mann-Whitney U Test	28
8	Cumulative Training-Phase Menu Costs for Cyclers and Statics with Mann-Whitney U Test	29
9	Median Policy Volatility (Greedy Inertia Rate) by Friction Tier	30
A.1	Continuous Profit Degradation: Wilcoxon Signed-Rank Test vs. Baseline . .	48
A.2	Static vs. Cyclers Statistical tests Full Sweep ($\phi \leq 0.13$)	48
A.3	Macroeconomic and Computational Taxonomy of Algorithmic Pricing under Friction	49
A.4	Punishment Depth: Baseline Cyclers compared to Cyclers at Each Tier Full ($\phi \leq 0.13$)	50
A.5	Training Phase Price Change Rate Spearman Correlation with Final Policy Entropy by Friction Tier	51
A.6	Median Policy Volatility (Greedy Inertia Rate) by Friction Tier	51
A.7	Three-Part Causal Decomposition of Net Profit Loss	52
A.8	Bootstrap 95% Confidence Intervals: Entropy-Mediated Loss and Percent Mediated	52
A.9	Economic environment parameters and code mapping	53
A.10	Learning parameters and code mapping	53

List of Figures

1	Pseudocode for a complete simulation run.	14
2	Frictionless baseline Edgeworth price cycle.	19
3	The systemic gross profit penalty of menu costs.	20
4	Ergodic gross profit distributions (left) and gross profit loss across friction tiers (right).	21
5	The extinction of the Edgeworth cycle: empirical cycling rate, fitted logistic decay curve, and the extinction threshold $\phi_{extinct}$, at which cycling probability falls to 5%.	23
6	Collapse of overall punishment capacity and depth of ongoing price wars. .	25
7	Exploration-phase survival by friction tier.	27
8	Training-phase price-change rate by behaviour type.	28
9	Decay of visitation entropy across friction tiers.	32
10	Exploration space of a representative seed across selected friction tiers. . . .	33
11	Population-averaged state-visitation matrices across selected friction tiers. .	34
12	Decomposition of net profit loss (following MacKinnon et al. (2004))	35
13	Decomposition of net profit loss (top), and the share of the loss mediated by entropy with the 95% bootstrap confidence band (bottom).	36

1 Introduction and Motivation

Setting prices is increasingly supported or fully handled by algorithmic pricing agents. These agents analyse the data available about a market and set prices with the aim of maximising profit. They are being adopted across different industries, for example in retail fuel (Assad et al., 2024), in hotel revenue management (Garcia et al., 2023) and in online retail (Brown & MacKay, 2023), because with a pricing algorithm a firm can react faster and set better informed prices more often than a human could.

The concern is what these algorithms may learn once they interact with one another. Simulation research repeatedly finds that independent pricing algorithms, each optimising only its own profit, settle at prices well above the competitive level (Calvano et al., 2020; Gu, 2026). In these simulations the agents get no instruction to coordinate and exchange nothing beyond the prices each of them can observe, so the behaviour emerges from learning alone. Coordination that arises this way, with no agreement and no communication, is what the literature calls tacit collusion. It is beneficial for the firms, but costly for consumers, because prices held above the competitive level transfer value from buyers to firms. The law would normally protect consumers and market competition here, but the problem is difficult to address, because current competition law requires concrete coordination between firms before it can intervene (Beneke & Mackenrodt, 2021; Li & Virtosu, 2025; Setianingrum, 2025). Algorithms that reach the collusive outcome without explicit coordination, and without ever having been instructed to do so, are not covered by existing regulation. The conditions under which such outcomes arise are therefore both important to understand and a subject of research.

Much of this work relies on simulations. These models aim to reproduce how firms price in practice, in order to understand what to expect in real markets. Because the real economy is complex, and because the concrete setup of pricing agents may differ from firm to firm, research analyses many different aspects across a range of market scenarios. With each result the whole picture gets a little clearer.

One aspect of real markets has not yet been modelled in these studies. The economic background to it is that changing a price is not free. It involves the physical work of making the change (Levy et al., 1997), the managerial time spent deciding on it (Zbaracki et al., 2004) and similar costs. Economics has studied this friction for decades under the name of menu costs, and the form this price change cost takes is known to shape how a firm adjusts its prices (Sheshinski & Weiss, 1977; Rotemberg, 1982). To the best of my knowledge, this friction has not yet been tested in combination with pricing algorithms.

Adding one cost term to an existing model is a small change. It is worth making because a menu cost is not an arbitrary modelling choice. Every firm that sets a price faces one, to varying degrees depending on its industry, and the friction already has a long body of economic evidence behind it. Testing it therefore adds one more piece to the algorithmic pricing picture.

This thesis aims to implement and analyze the addition of a menu cost friction. It takes a standard Q-learning model of two firms competing on price, adds a menu cost that each firm pays whenever it changes its price, and examines how the resulting behaviour differs from the baseline case without friction. The question it sets out to answer is: *How do menu costs alter the stability and structural dynamics of algorithmic collusion in a Q-learning*

pricing duopoly?

The thesis proceeds in four stages. The literature review first examines price rigidity and menu costs, then tacit collusion in its classical and algorithmic forms, then the ethical and legal questions raised by autonomous pricing. The methodology describes the simulation environment, how the menu cost is implemented in it and how the learning agents are defined. The results section examines the data generated by the simulated environment to understand how friction changes market behaviour, the duration and extent of learning and net profitability. Finally, the discussion returns to the research question and considers what the findings imply for the literature and for the regulation of algorithmic pricing.

2 Theoretical Background

2.1 Foundations of Price Rigidity and Menu Costs

Price rigidity becomes a relevant economic question once it is questioned why a firm that wants to maximise profit would leave its price unchanged when conditions have moved against it. The standard answer is that changing a price is itself costly.

2.1.1 The cost of changing prices

The cost of changing prices depends considerably on the industry. Some direct measurements put the physical cost of repricing even with low costs per price change, still resulting in a substantial share of the margin. For instance, four large US supermarket chains spent an average of \$105,887 per store per year on changing shelf prices alone, which represented 0.70% of revenues and 35.2% of the net margins (Levy et al., 1997). In another study a drugstore chain measured the same way showed the effect of price changes at 0.59% of store revenue, with labour accounting for around 79% of the total costs of changing prices (Dutta et al., 1999).

Further separating the costs of changing prices into physical, managerial and customer components at one industrial manufacturer put managerial costs at more than six times the physical costs and customer costs at more than twenty times the physical menu costs. The total costs came to around 1.22% of the revenue and 20.03% of the net margin (Zbaracki et al., 2004).

Also firms themselves ranked the physical costs' impact as the lowest in a survey of 654 firms in the United Kingdom. Physical menu costs came last of eleven proposed explanations for price stickiness (Hall et al., 2000). Looking at e-commerce, where the physical cost of changing a price is close to zero, firms (online booksellers) still showed nominal price rigidity (Chakrabarti & Scholnick, 2007).

This suggests that even though the split between physical and non-physical costs for price changes varies largely between industries, menu costs overall remain still a considerable factor for firms.

Table 1 collects the measurements discussed above for better overview. No two cover

Table 1: Direct estimates of the cost of a price change

Publication	Context	% of revenue	% of net margin ^a	Scope
Levy et al. (1997)	supermarket chains, 1991–92	0.70	35.2	Physical, including in-store supervision
Dutta et al. (1999)	drugstore chain, 1992	0.59	21.3	Labour and price tags only ^b
Zbaracki et al. (2004)	industrial manufacturer, 1997	1.22	20.03	Physical, managerial and customer
	<i>of which physical</i>	0.04	0.71	Physical only

^a Net margin is assumed, not observed: 2% of revenues in Levy et al. (1997), 2.75% in Dutta et al. (1999). Only Zbaracki et al. (2004) use the firm’s reported margin.

^b Dutta et al. (1999) estimate mistake-correction and supervision costs by applying the cost ratios that Levy et al. (1997) measured in supermarkets, which raises the *% of revenue* figure to 0.74%.

exactly the same components, which is specified in the last column, so the totals should not be interpreted as alternative estimates of the same quantity. What they share is the shape of the result. Each study’s total lands between around 0.6% and 1.2% of revenue, and between a fifth and just over a third of net margin. Menu costs are therefore small measured against revenue and considerably larger measured against net margin.

2.1.2 Small costs, large behavioural effects

Because small menu costs have substantial impact on the firms’ financial results, they often influence what firms do. With changing prices firms have to navigate between private and social incentives. A firm can cut its price only if the extra profit directly covers the cost of changing it, while a social planner would weigh that same cost against the gain in total surplus, counting together what consumers gain as well as what the firm earns with it. Especially if a firm is close to its profit maximum, where small moves change profit very little, its own gain from adjusting is second order while the social gain is far more relevant (Mankiw, 1985).

Two experiments show this effect directly. Where a retailer’s physical labour of changing price tags scales with the number of product variants sharing a shelf, items with more variants had up to 13.3% fewer price increases after a cost increase, and only in 5.8% of occurrences did the foregone price increases catch up within 90 days (Anderson et al., 2015). The second experiment runs the other way. Electronic shelf labels remove the manual work of retagging, so the physical cost of a price change falls sharply. After retailers adopted them, price changes became more frequent and smaller, and they were spread more evenly across days instead of being carried out in batches (Stamatopoulos et al., 2021). Together, these findings provide evidence that menu costs contribute to price rigidity.

2.1.3 The form of the friction determines the behaviour

How long a firm waits when it needs to reprice depends on how the menu cost is specified. A fixed real charge per change results in an (s, S) policy: the price stays flat for some

time and then it moves back up in one discrete jump producing infrequent and large price adjustments (Sheshinski & Weiss, 1977).

Making the cost dependent on the size of the price change instead gives room for frequent and partial adjustments, since the cheapest way to close a large gap is then to adjust in stages (Rotemberg, 1982). Whether the cost is paid per item or once per pricing occasion matters too. Separating the two on supermarket scanner data showed that a product's price is four times more likely to change in a week when another product in its category also had a price change (Stella, 2020). In modelling terms this means that the friction can either be modelled as a cost subtracted from the firm's objective function or as a constraint on the number of permitted price changes (Feldman & Segev, 2025).

Estimates of the friction also depend on what a model assumes about competition. Estimating a menu cost parameter under monopolistic competition gives a figure more than twice as large as estimating it in a duopoly on the same data. This is because interaction between firms also generates price rigidities that are otherwise attributed to the cost of adjustment (Kano, 2013). Menu cost rigidity and firms' strategic behaviour (acting in accordance with how the competitors react) are therefore not separable.

2.1.4 Limits of attributing effects to menu costs

Two aspects can also limit how far the menu cost's effect reaches. Not every pricing regularity that appears to reflect a menu cost actually does. Roughly two thirds of grocery prices end in nine (eg. 0.99), and a price that ended in nine last week changes in 22.1% of weeks against 30.2% for prices ending in any other digit. Estimating both frictions in the same model puts the price-point effect roughly a hundred times above the menu cost. Models that ignore price points may therefore overestimate the role of menu costs in these systems (Knotek, 2024).

There are also markets where the menu costs can only be used restrictively to explain the market behaviour. As an example retail gasoline markets exhibit Edgeworth cycles and reprice almost daily even without an apparent cost shock (Lewis & Noel, 2011).

2.2 Tacit Collusion and Algorithmic Pricing

2.2.1 Tacit collusion in classical oligopoly theory

Prior to the adoption of automated pricing, oligopoly theory had already explained how firms can hold prices above the competitive level without any explicit agreement. The classical mechanism is explained as a trigger strategy: firms keep prices high because they expect that deviating would start a price war. Green and Porter (1984) show that this works even when firms cannot directly observe whether a rival has cheated. A low market price might reflect a rival secretly producing above its agreed share or could simply be triggered by low demand. Because the two cannot be clearly distinguished, firms usually respond by initiating a price war when the price falls below a certain level. The constant risk of a price war is in this case what holds the collusion together.

This kind of coordination tends to leave prices rigid, which links collusion back to the price stickiness of the previous section. Analysis shows that an optimal collusive strategy can involve holding prices fixed rather than adjusting them (Athey et al., 2004), and a collusive price can (when the two cost levels are sufficiently close) also stay rigid even after changes in costs when firms sustain it with price-matching responses (Garrod, 2012). This link also shows up in data from monopoly retail gasoline markets. They show less price fluctuation than competitive ones, consistent with the claim that weaker rivalry reduces price changes, though such evidence cannot by itself prove collusion (Jiménez & Perdiguero, 2012). Price rigidity is therefore not only a menu cost consequence but can also be a sign of collusion.

Collusion does not always appear as a stable high price. In some markets the collusive outcome is a cycle. Firms undercut one another down toward marginal costs until one firm resets the price, restarting the cycle. Maskin and Tirole (1988) gave a formal model of this Edgeworth cycle alongside the ‘kinked-demand’ explanation in which firms instead settle at the monopoly price as a stable reference point. As an example, this cycle is well documented in the retail gasoline industry, where small independent stations tend to lead the undercutting to win more customers and large chains tend to trigger the upward price adjustment for the whole market (M. D. Noel, 2007b). In this case such pricing cycles account for a large part of the difference in how prices respond to cost changes (M. Noel, 2009). Separating these patterns empirically shows they coexist across markets. In a sample of nineteen Canadian cities, 43% were classified as cycling, 30% as sticky and 27% as cost-based (M. D. Noel, 2007a). Whether coordination appears as a fixed price or as a recurring cycle therefore depends on the market, and both patterns reappear even when pricing is handled by algorithms.

2.2.2 The real-world use of algorithmic pricing

Pricing is increasingly supported and often completely outsourced to pricing algorithms instead of being set directly by managers. Such systems are now deployed across industries, from retail to hotel revenue management (Garcia et al., 2023), which makes their effect on prices a question about real markets rather than a hypothetical one. A clear example case is on large online marketplaces such as Amazon, where prices have been reported to automatically change many times a day to be able to maximise revenue (Bichler et al., 2025). Another category of practical evidence from a set of US retailers using hourly price data looked at the speed of repricing. In their paper Brown and MacKay (2023) documented persistent price differences of 10 to 30 percent between competing firms. This was shown to be dependent on how frequently each firm updated its prices.

One of the best-documented field evidence for physical markets is from the German retail gasoline industry, where in some local markets every competing station adopted pricing software. Notably, in these markets the margins rose by around 38%, while no equivalent effect was observed where the adoption was only partial (Assad et al., 2024). The dependence on all firms adopting pricing software and the fact that the increase appeared only about a year after a market-wide adoption are better explained by algorithms gradually learning to coordinate than by each firm optimising on its own (Assad et al., 2021).

2.2.3 Algorithmic collusion and its conditions

A reinforcement learning agent works without an exact model of the problem it faces. It stores a value estimate in its memory for each state and action it can possibly take, and slowly adjusts the estimates based on the reward it actually receives (Calvano et al., 2020). In the agent’s knowledge the demand structure is not represented directly, and the rival enters only as part of the observed state. The pricing agent’s memory contains a value estimate for every action available in a given state, with no explicit information about the rival or the market.

This makes the multi-agent case different from single-agent learning in two ways. Each agent’s environment is non-stationary, meaning that the rivals are revising their own policies at the same time. This violates the stationarity assumption underlying standard convergence theory, so convergence cannot be guaranteed theoretically and must instead be assessed empirically (Tesauro & Kephart, 2002). Coordination also does not need the punishment mechanism of the classical case. Waltman and Kaymak (2008) show in a Cournot setting that agents with no memory of past actions, and therefore no capacity to punish a rival even in principle, still converge on supracompetitive outcomes.

Whether pricing algorithms can arrive at such coordination on their own has been studied most directly in simulation. Calvano et al. (2020) found that two independent Q-learning agents, setting prices in a repeated Bertrand duopoly, end up pricing well above the competitive level, with no communication between them and no instruction to coordinate. The agents in this study reached a profit level between 70% and 90% of the range from the competitive to the monopoly outcome. The firms in this case also learned a punishment strategy: after one firm undercut, prices fell for a few periods before returning back up to the collusive level (Calvano et al., 2020). Because the agents exchange nothing but their observed prices, any coordination that emerges from this is by definition categorised as tacit collusion. The coordination is not an agreement but a learned pattern from playing mutual best response strategies.

That agents can learn to collude does not necessarily mean they always do. This was studied in the literature to be able to understand what the outcomes and the algorithmic collusion depend on. Most papers highlighted that the challenge is that the apparent collusion may be an artefact of how the agents explore. When exploration is cut short, agents can settle on high prices simply because they never learn that competing pays off. This is a pattern better described by insufficient knowledge rather than genuine strategy (Abada et al., 2024).

This learning trap is self-reinforcing. If the first rounds make cooperating look better than competing, both algorithms choose the competitive action less often. They then gather less evidence about how profitable competing would be, so the belief is never corrected. Abada et al. (2024) call this the mirage effect, because the closer an algorithm comes to the point where it could learn otherwise, the further that point moves away. What drives it is the shrinking exploration rate, not the values the estimates started from. More sophisticated algorithms, by contrast, tend to compete rather than coordinate, which runs against the worry that more capable algorithms collude more.

Setting the rules of learning matters in a similar way. When the agents place no weight on future profits, updating only the value of the action they actually took rather than all of

their available actions, the converged outcome moves toward near-monopoly rather than competitive prices (Asker et al., 2024). Smaller features of an environment can shift the result too, including the granularity of the price grid (Klein, 2021) and the design of the state space (Sanchez-Cartas & Katsamakas, 2022). For agents to reach supracompetitive prices, they do not even have to be able to get information about the other firms’ behavior. Hansen et al. (2021) show that independent algorithms that never observe their rivals’ prices can still converge on supracompetitive outcomes when demand is predictable enough. This is because correlated experimentation biases each firm’s own estimate of how price-sensitive the customers are. The result from Gu (2026) also extends this beyond tabular Q-learning algorithms, with deep reinforcement-learning agents reaching similar outcomes in simulated environments.

The resulting behavior cannot always be categorised as collusion. Agents in some settings punish any deviation from the established price, including price increases, not only undercutting (Epivent & Lambin, 2024). More information does not necessarily make collusion more probable either. Better demand prediction allows firms to identify the joint profit-maximising price more accurately, but it also shows each firm exactly when demand is high. Because undercutting is most profitable at these times, better prediction can ultimately make collusion harder to sustain (Miklós-Thal & Tucker, 2019).

A separate line of research questions whether the coordination is relevant for real-world markets at all. Den Boer et al. (2026) showed that Q-learning algorithms reach collusive outcomes only after a number of periods far exceeding anything relevant to a firm’s actual objective.¹ Some studies ask directly whether the observed behavior is real collusion or just an artificial outcome caused by the learning dynamics (Calvano et al., 2023). The open question is therefore not if algorithms can collude but under which conditions they do, which heavily depends on how the simulations are set up. This is why examining previously untested conditions can add to the understanding of the mechanisms of algorithmic collusion.

2.3 Ethical and Legal Dimensions

2.3.1 Limits of the current legal framework

It is useful to separate two ways in which algorithms can produce collusive outcomes. Bernhardt and Dewenter (2020) distinguish between collusion that is explicitly written into the pricing algorithm’s code and collusion that the algorithms reach on their own. The first is equivalent to a conventional cartel and can therefore be treated as one legally. The second, where the algorithms reach tacit collusion, is a new problem, because in this case the firms never explicitly agreed to anything.

As competition law usually is built around intent, the absence of an explicit agreement makes algorithmic collusion a problematic topic. As a concrete example under EU law an infringement requires an agreement or a concerted practice between the parties. This is not given when algorithms set prices autonomously and independently, but still are

¹However, this also depends on the industry and the pricing strategies used, as different industries differ a lot in how and how often they are able to change prices. (Bils & Klenow, 2004; Nakamura & Steinsson, 2008)

coordinating through nothing but observed prices (Beneke & Mackenrodt, 2021). The same difficulty of establishing agreement or intent is present not only in Europe, but also across the US and other jurisdictions (Li & Virtosu, 2025; Setianingrum, 2025). Because prices kept artificially high above the competitive level transfer value from consumers to firms and reduce overall welfare, tacit algorithmic collusion can harm consumers while remaining outside of the reach of the law.

The urgency of addressing this gap is still debated. Schwalbe (2018) argues that algorithmic collusion is less likely to emerge in practice than legal scholars often assume, because most findings primarily rely on modelling abstract duopoly settings. This disagreement draws attention to the conditions under which collusion emerges, of which some aspects were already discussed in the previous section.

2.3.2 Proposed responses

Because algorithmic collusion is not covered by existing regulations, the literature has suggested new ways to better oversee this field more effectively. On the legal side, proposals include behavioural and structural regulatory mechanisms, like restrictions and guardrails in the pricing algorithms' code themselves, or holding the algorithm's designer liable for collusive outcomes (Beneke & Mackenrodt, 2021; Li & Virtosu, 2025). A different approach intervenes in the learning process of the algorithms rather than in the resulting pricing behaviour. Abada and Lambin (2023) show, in an electricity-market setting, that enforcing decentralised learning can partially reduce the social and market welfare loss from algorithmic price coordination, and that a regulated market participant intervening during the learning phase itself reduces that welfare loss considerably further.

The concerns also extend beyond collusion. Algorithmic and personalised pricing raise additional questions of transparency and fairness, even when in these examples no coordination needs to occur (Seele et al., 2021).

This legal gap provides further motivation for the thesis, as for regulators to understand and create targeted policy measures for this problem, it is important to understand when and under which circumstances algorithmic collusion forms, and whether frictions such as menu costs make it more or less likely to happen.

2.4 Research Gap and Positioning

2.4.1 The research gap

The sections so far explored three observations that combined together raise a question that has not yet been addressed. The foundations of price rigidity established that menu costs are real and measurable frictions for industries. How they are specified determines how a firm changes its prices. For this it is relevant whether the costs of changing prices are fixed or proportional and whether they are paid per item or per change. The literature on algorithmic pricing showed that independent learning algorithms can settle into tacit collusion, but if they do and the specific outcomes depend strongly on how the problem is defined and how simulations attempt to model them. Additionally the review of the regulatory landscape highlighted that this kind of non-coordinated algorithmic collusion

falls outside existing competition law, which makes exploring the conditions under which it emerges a relevant practical question for policy makers as well as of academic interest.

The literature about menu costs agrees that changing prices is generally costly, but it does not explore its connection to learning algorithms. While the algorithmic collusion literature explores strategic learning in detail, it does not include constraints and effects from the cost of repricing. To the best of my knowledge, no existing study introduces explicit per-change costs for price adjustments into a model of algorithmic collusion. This thesis explores this intersection between menu costs and pricing algorithms by asking how menu costs affect collusion.

2.4.2 Positioning and research question

To study this question, this thesis builds on an already established model. It adopts the baseline of Calvano et al. (2020), which has become one of the most widely replicated standard reference points of Q-learning algorithmic collusion. It has a deliberately limited simulation setup that enables it to isolate the collusion mechanism without the additional complexities of richer settings. The wider literature has also tested alternating (rather than simultaneous) moves (Tesauro & Kephart, 2002), quantity (Cournot) rather than (Bertrand) price competition (Waltman & Kaymak, 2008), and stochastic-approximation methods in place of value-based learning (Meylahn & V. den Boer, 2022; Loots & denBoer, 2023). However, none of these are as widely used benchmarks as the Calvano et al. (2020) setup. Adopting this reference model keeps the results comparable to an already established literature and ensures that effects of the only extension, the added menu cost friction, can be isolated and evaluated against the baseline.²

This leads to the concrete question the thesis sets out to answer: *How do menu costs alter the stability and structural dynamics of algorithmic collusion in a Q-learning pricing duopoly?* By adding a proportional per-change menu cost to the reward function of the Calvano et al. (2020) baseline, this thesis provides a detailed analysis of how introducing explicit price-adjustment frictions affects tacit collusion between reinforcement-learning agents.

²The friction this thesis implements should not be confused with a seemingly similar concept. Klemperer (1987) studies consumer side switching costs, where the effort to switch between products (e.g. cancellation fees or learning a new product) makes a firm’s existing customers less sensitive to price changes. The effect of such switching costs on tacit collusion has been studied, with the result that collusion depended on whether firms were able to communicate (Fourberg, 2023). A switching cost is different from a menu cost. One (the switching cost) is a cost absorbed by the consumer in order to move between products or suppliers, the other (menu cost) is paid by a firm to change its own price. The two affect collusion through different mechanisms, and might be completely independent from each other. The results in this thesis therefore only speak to the firm-side menu cost frictions and do not analyse switching costs in any way.

3 Methodology

3.1 The Economic Environment: Replicating the Baseline Model

The empirical foundation of this thesis is a Multi-Agent Reinforcement Learning (MARL) environment, built in Python using the structural logic of the Gymnasium framework.³

To ensure the findings can be compared with the existing algorithmic pricing literature, the core economic setting precisely replicates the framework established by Calvano et al. (2020). This frictionless environment is referred to as the ‘baseline model’ going forward. The simulation without any price adjustment frictions ($\phi = 0$) behaves the same as the baseline model and its initial outputs can be validated directly against Calvano et al. (2020)’s published figures. Any behavioral shift observed at $\phi > 0$ can then be attributed to the added menu cost frictions.

The market consists of two symmetric and independent Q-learning algorithms. These agents operate without a shared memory, a centralized controller or a direct communication channel. The only information exchanged is the publicly observable price action. Any coordination that emerges therefore represents tacit collusion, not explicit joint profit maximization.

Demand follows the same Multinomial Logit model as in the baseline (Calvano et al., 2020). In each period t , a unit mass of consumers chooses between firm 1’s product, firm 2’s product or an outside option (firm 0). Because the mass is continuous, individual choice probabilities aggregate directly into each firm’s market share. Firm i ’s demand, given the price vector $(p_{i,t}, p_{-i,t})$, is:

$$q_i(p_{i,t}, p_{-i,t}) = \frac{e^{(a_i - p_{i,t})/\mu}}{e^{a_0/\mu} + e^{(a_i - p_{i,t})/\mu} + e^{(a_{-i} - p_{-i,t})/\mu}} \quad (1)$$

(Calvano et al., 2020, Eq. 5, p. 3273)

a_i is the product’s quality index, a_0 is the value of the outside option, and μ measures how differentiated the two products are. The outside good’s utility is normalized to $a_0 = 0$. This also bounds the model economically. If both firms price too high, consumers can exit the market, which caps the achievable monopoly price. Both firms face a constant marginal cost of $c = 1$, quality indices $a_1 = a_2 = 2$, and differentiation $\mu = 0.25$ (Calvano et al., 2020, p. 3274).⁴

3.2 State Space and Action Space Discretization

Tabular Q-learning requires a discrete environment. To achieve this continuous economic variables are mapped onto finite grids following Calvano et al. (2020)’s hyperparameters.

³The simulation environment is implemented in the `collusion_env.py` module. The full source code of the modules named in this chapter is submitted as a separate attachment to this thesis.

⁴These baseline hyperparameters are standardized in `config.py`. A comprehensive mapping of the theoretical parameters to their programmatic variables is provided in Tables A.9 and A.10 in Appendix A.1.

The Action Space ($\mathcal{A}$). In each period, the algorithms select a price from a finite grid of $m = 15$ actions. (Calvano et al., 2020, p. 3274) The grid is evenly spaced and bounded to include both the Bertrand-Nash price (p^N) and the monopoly price (p^M). Because the Logit model has no closed-form solution for either reference price, the model calculates p^N and p^M before building the grid. The grid boundaries follow from a padding parameter $\xi = 0.1$, extending the interval 10% above p^M and 10% below p^N :

$$\left[p^N - \xi(p^M - p^N), p^M + \xi(p^M - p^N) \right] \quad (2)$$

(Calvano et al., 2020, p. 3274)

This buffer allows the algorithms room to explore undercutting as well as coordinated prices.

The State Space ($\mathcal{S}$). The algorithms carry a one-period memory ($k = 1$), consistent with the baseline (Calvano et al., 2020). The state observed by firm i at time t is therefore limited to both firms' prices from the previous period: $s_t = (p_{i,t-1}, p_{-i,t-1})$. With $m = 15$ actions this results in a state space of $15 \times 15 = 225$ states.

3.3 Departure from the Baseline: The Net Profit Function and Menu Costs

π^N and π^M are the per-firm profits at these two prices:

$$\pi^N = (p^N - c) q_i(p^N, p^N) \quad \text{and} \quad \pi^M = (p^M - c) q_i(p^M, p^M). \quad (3)$$

The monopoly price p^M is found by maximizing the two firms combined profit and π^M is one firm's share of that outcome.

As mentioned above the environment matches the frictionless baseline exactly except one main difference that is the added menu cost penalty for changing prices. While the baseline pricing model allows agents to maximize gross profit by changing prices as many times as they choose to without any penalties or restrictions, this model instead requires agents to maximize a net profit function that penalizes price changes.

Firm i 's reward (net profit) at time t equals its gross profit minus the menu cost penalty:

$$\pi_{i,t}^{net} = (p_{i,t} - c) \cdot q_i(p_{i,t}, p_{-i,t}) - \left[\phi \cdot (\pi^M - \pi^N) \right] \cdot \mathbf{1}(p_{i,t} \neq p_{i,t-1}) \quad (4)$$

The indicator function $\mathbf{1}(\cdot)$ activates only when an agent changes its price from the previous period, so the penalty only triggers if a firm actively changes prices.⁵

Rather than a fixed currency amount, the menu cost ϕ is defined as a proportion of the collusive premium ($\pi^M - \pi^N$). This keeps the friction economically meaningful regardless

⁵Because the environment initializes each session at a random price, the first step is excluded; otherwise, agents would be penalized for moving from an arbitrary starting price.

of the underlying demand parameters. If μ or c changed, a fixed fee would carry a different relative weight, while ϕ would not, which keeps comparisons across possible setups consistent. ϕ denotes this menu cost parameter throughout.

Once the baseline ($\phi = 0$) was validated, ϕ was set across a range of other values to examine how friction affects market behavior. This range was not chosen to be a fixed, evenly spaced grid. A preliminary, low-density search across a broad set of candidate values identified friction levels worth closer examination. These were then simulated as full Monte Carlo runs, with additional menu cost tiers added as the analysis progressed. The result is an uneven grid: dense at low friction, where behavior shifts quickly and sparse at high friction, where the market does not change as rapidly with changing ϕ values.

The distinct ϕ tiers tested are:

[0.0, 0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 0.1, 0.11, 0.12, 0.13, 0.14, 0.15, 0.2, 0.25, 0.35, 0.5, 0.7]⁶

3.4 Agent Architecture and Q-Learning Mechanics

The pricing algorithms use independent tabular Q-learning. (Calvano et al., 2020)⁷

At initialization the agents start with no information about the market (the demand function, the market structure, or the competitor’s possible actions). They rely entirely on the rewards returned by the environment after each action to gradually map the state-action space. Each firm maximizes its cumulative discounted profit by updating the 225×15 Q-table using the standard temporal-difference Bellman equation. After taking action a_t in state s_t , observing the net reward $\pi_{i,t}^{net}$, and transitioning to state s_{t+1} , an agent updates its Q-value as follows:

$$Q_{t+1}(s_t, a_t) = (1 - \alpha)Q_t(s_t, a_t) + \alpha \left[\pi_{i,t}^{net} + \delta \max_{a'} Q_t(s_{t+1}, a') \right] \quad (5)$$

(Calvano et al., 2020, Eq. 4, p. 3271)

This update rule’s structure follows Calvano et al. (2020)’s Eq. 4 directly, but substituting the net reward $\pi_{i,t}^{net}$ for the gross reward is the addition made in this paper.

α is the learning rate, determining how much new information overrides existing knowledge δ is the discount factor, determining how much the agent values future rewards relative to immediate ones.

The discount factor $\delta = 0.95$ corresponds directly to Calvano et al. (2020)’s baseline value.

The learning rate $\alpha = 0.15$ together with the exploration-decay parameter β introduced below is taken from a single representative point from Calvano et al. (2020) that they use to illustrate deviations and punishment⁸. This thesis is designed to isolate the effect of the menu cost friction ϕ , rather than to conduct an independent sensitivity analysis over the learning parameters, so α and β were chosen as fixed points. This is applicable

⁶Every simulated tier is detailed in Appendix Table A.3.

⁷Agent learning logic is implemented in the `q_agent.py` module.

⁸This was taken from the paper’s main worked example

as Calvano et al. (2020) reported that their findings on deviations and punishment hold across their entire α/β grid, and that the resulting profit gain is not sensitive to either parameter outside extreme values. This robustness check was conducted in Calvano et al. (2020)’s frictionless setting only, so adopting their values here was an assumption rather than an independently verified result. Testing hyperparameter sensitivity alongside the friction grid would require a multidimensional search, which falls outside the scope of this thesis. This simulation therefore inherits Calvano et al. (2020)’s values and treats ϕ as the sole treatment variable.

Action selection follows an ϵ -greedy rule and during exploitation, the algorithm selects the action with the highest Q-value. A known limitation is the following: because the agents can only play pure strategies, ties are broken by selecting the lower price (a rule taken directly from Calvano et al. (2020)), which means that agents may oscillate around a long-run target price that is not itself a feasible grid point. (Calvano et al., 2020, p. 3274)

The exploration rate decays exponentially, $\epsilon_t = e^{-\beta t}$, with $\beta = 4 \times 10^{-6}$, again reflecting Calvano et al. (2020)’s representative value. (Calvano et al., 2020, Eq. 7, p. 3274) (Calvano et al., 2020, p. 3280)

Following the baseline, the Q-tables are optimistically initialized, so the agents continue exploring the state-action space rather than converging prematurely on a suboptimal policy. The starting Q-values are the discounted expected payoffs an agent would receive if its opponent played uniformly at random (Calvano et al., 2020, Eq. 8, p. 3275). This thesis applies the same substitution here as in the update rule. The menu cost is accounted for in the initialisation by using the expected net profit instead of the gross profit. With 15 actions a randomly acting agent changes its price 14 times out of 15, so the expected per-period penalty is $\frac{14}{15}\phi(\pi^M - \pi^N)$:

$$Q_0(a) = \frac{1}{m(1-\delta)} \sum_{a_{-i}} \left[\pi(a, a_{-i}) - \frac{m-1}{m} \phi(\pi^M - \pi^N) \right] \quad (6)$$

The expression does not depend on the state, so every one of the 225 states receives the same initial values. The penalty is discounted alongside the profit, so the starting values reflect an agent that expects to repeatedly pay the menu cost in future periods. At $\phi = 0$, this falls back exactly to Calvano et al. (2020)’s original initialization. For the friction runs it keeps agents from starting disproportionately optimistic simply because their environment has a higher ϕ value, which could then otherwise inflate their convergence time.

3.5 Data Collection Pipeline and the Two-Phase Evaluation Strategy

The simulation uses a data collection architecture that keeps the learning phase separate from the final market outcomes. Every Monte Carlo run seeds its random number generators from a fixed master seed prior to execution, so results remain reproducible.⁹

Each agent’s environment is nonstationary, since the other agent explores and learns simultaneously. Q-learning therefore carries no general theoretical guarantee of convergence in repeated games (Calvano et al., 2020). Convergence can only be established empirically,

⁹Simulation execution, convergence tracking, and evaluation extraction are managed through the `main.py` script.

ex post, which is precisely the purpose of the convergence-detection architecture described below.

The simulation runs in two sequential phases. Figure 1 illustrates the execution pipeline, showing how the process steps and responsibilities are distributed across the agents, the environment, and the main simulation controller.

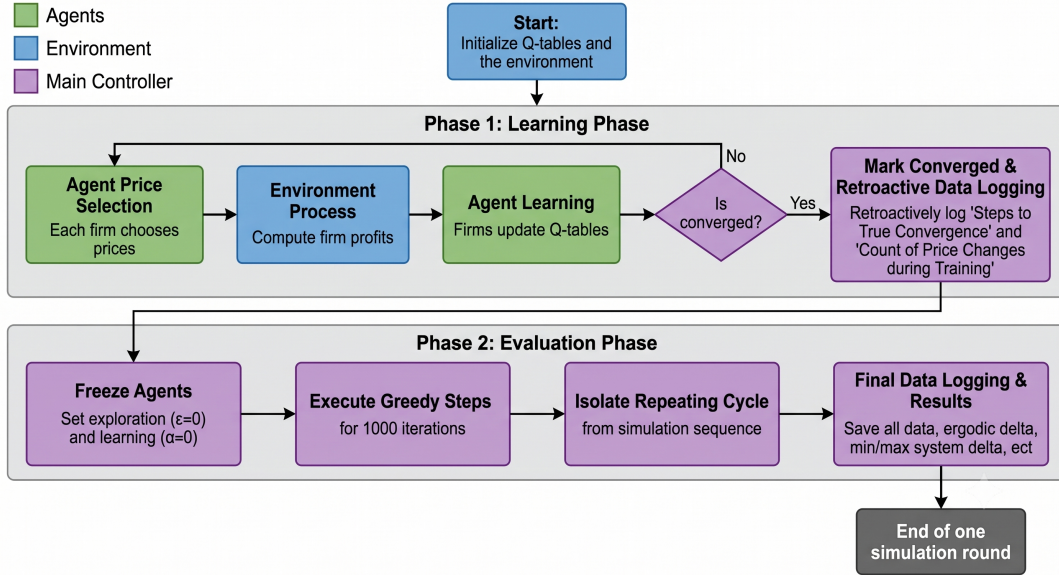


Figure 1: Pseudocode for a complete simulation run.

Phase 1: The Learning Phase. In Phase 1, both agents explore the environment and continue updating their Q-tables. Following Calvano et al. (2020), a session is considered converged once an agent’s optimal policy remains unchanged for 100,000 consecutive steps. This thesis additionally terminates any session that fails to converge within 10,000,000 steps and logs it as non-convergent. A custom safeguard was added to this thesis’s simulation to prevent the 100,000-step counter from starting until the exploration rate ϵ has decayed below 1% ($\epsilon < 0.01$), otherwise an accidental period of policy stability during the exploration phase could be mistaken for genuine convergence.

Left unadjusted, these safeguards would influence the measured learning efficiency (learning time and price changes), so the logging pipeline retroactively isolates the true learning phase. It locks in the key metrics at the exact step of the agent’s final policy change, excluding both the 100,000-step verification window and the wait for ϵ to decay.

Phase 2: Post-Convergence Data Extraction. Once both convergence conditions are met, Phase 1 ends and Phase 2 begins. Learning and exploration are frozen ($\epsilon = 0, \alpha = 0$), and the algorithms execute their greedy policies for exactly 1,000 evaluation steps.

The agents train on the net profit function (π^{net}), but the economic outcomes actually logged in Phase 2 are calculated using gross profit. Following Calvano et al. (2020)’s Eq. 9 the Profitability Index (Δ) is:

$$\Delta = \frac{\pi - \pi^N}{\pi^M - \pi^N} \quad (7)$$

(Calvano et al., 2020, Eq. 9, p. 3277)

where π is the firm’s profit, normalized against the Nash (π^N) and monopoly (π^M) benchmarks. Because π here is gross profit, this keeps the collusion measure free of the menu costs’ direct effect.

Q-learning agents can also be expected to settle into repeating price cycles (Edgeworth cycles), so cutting the evaluation off at an arbitrary point would risk capturing only parts of a cycle. In this case a plain average of the final steps would understate profitability. To avoid this, the evaluation pipeline extracts the exact repeating cycle from the end of the evaluation window and averages gross profit only over that cycle. This is the Ergodic Profitability Index (Δ). It is the same Δ from Calvano et al. (2020)’s Eq. 9, computed on the cycle-averaged profit. Isolating the cycle also allows the pipeline to extract the highest and lowest system profits reached within it, which later chapters can utilize to analyze punishment severity and price-war dynamics.

3.5.1 Master Registry and Final Data Export

After each simulation is run, specific values and files are saved for later analysis and for reproducibility. After a full Monte Carlo run is complete, the main simulation controller also logs the output to a master registry table, organized into three tiers:

- **Tier 1: Algorithmic Behavior (Phase 1).** How efficiently the agents learned: steps to convergence (`step_of_first_greedy_freeze`), price changes during training (`true_learning_price_changes_1`, `true_learning_price_changes_2`), and a frozen snapshot of policy entropy (`snapshot_entropy`).
- **Tier 2: Firm-Level Outcomes (Phase 2).** Each firm’s Ergodic Profitability Index, computed on gross profit (`ergodic_delta_1`, `ergodic_delta_2`) and the price pair the cycle settled into (`cycle_string`).
- **Tier 3: Market-Level Outcomes (Phase 2).** Both firms treated as a single industry: the combined Ergodic Profitability Index (`avg_ergodic_delta`) and the bounds of system profit (`system_min_delta`, `system_max_delta`).¹⁰

After Phase 2, the system also exports the underlying algorithmic states and full session histories: the final trained Q-tables (`.npy`), the frozen Phase 1 visitation matrices (`.npy`), and turn-by-turn price and profit trajectories for both phases (`.csv`).¹¹

¹⁰The master registry is exported as a consolidated `.csv` file via the `export_registryData` method of the `SimulationLogger` class in `data_logger.py`.

¹¹Export paths and timestamped file routing are managed by the `export_worker_trajectory`, `export_frozen_matrix`, and `export_q_tables` methods in `data_logger.py`.

3.6 Statistical Methods used for the Analysis

This section describes the statistical methods used to analyse the simulation output. It covers the structure of the dataset, the statistical tests applied and why each of them was chosen. Additionally the conventions followed in creating the figures are described for more context. This section aims to explain the underlying methods to enable a more streamlined explanation and focus on the results that are presented in Section 4.5.

All analysis was carried out in Python. Pandas and NumPy were used for data handling, SciPy and Statsmodels for the statistical procedures, and Matplotlib and Seaborn for creating the figures.¹²

3.6.1 The Analysis Panel

The analysis uses a single panel of 1,000 independent markets with their own seed. Each one is simulated at each of the 21 menu cost tiers.

The same seeds are used for every friction tier's Monte Carlo runs. This means that for a given seed the only difference between two simulation runs is the ϕ value the environment applies. This is the common random numbers (CRN) design often used in simulations to reduce the variance of the estimated difference between two runs, so the outputs can be compared with less noise. This method is widely used because it is simple to implement, although the benefits depend on the structure of the comparison and are not guaranteed in general. The method produces the strongest results where monotonicity and continuity conditions are satisfied (Glasserman & Yao, 1992).

The differences between two tiers can therefore be attributed to the menu cost friction itself rather than to random variation between unrelated runs. The pairing was verified before the analysis and all 1,000 seeds are present at every tier, so no comparison was done on partially matched samples.

Two characteristics of this design influence the methodological and statistical choices in the analysis. Each seed generated one group of outputs for each menu cost tier, so observations are not independent of each other. Additionally the profitability measure is negatively skewed rather than normally distributed, with a skewness of -0.65 at the frictionless baseline. This is why non-parametric tests were used to respect this structure.

3.6.2 Net Profitability

The ergodic profitability index (ergodic Δ) is measured on gross profit, as described above. This keeps the collusive outcomes free of the menu cost's direct cost effect. However, parts of the analysis require a net profit value. This was calculated from the logged values using the converged price cycle information per seed. The share of periods in which each firm changes its price was calculated from the repeating price cycle recorded in the evaluation phase, and then the two firms' rates were averaged.

¹²The full notebook, including the code for every test and figure, is in the `analysis.ipynb`.

The calculation is exact as the menu cost is defined as $\phi(\pi^M - \pi^N)$, and Δ is normalised by that same collusive premium.

$$\Delta^{net} = \Delta - \bar{r} \phi \quad (8)$$

where $\bar{r}$ is the average share of periods in which a firm changes its price within the converged cycle.

For markets that are not cycling and are settled on a static price, $\bar{r}$ is zero so the net and the gross measures are the same. The distinction between these two measures therefore only matters for markets that are actively cycling in the evaluation phase.

3.6.3 Paired and Independent Comparisons

Two kinds of comparison are often used in the analysis and they require different tests. Both are applied as two-sided tests, and both are often repeated across many tiers.

Where the same market is observed at two friction levels (for example $\phi = 0.0$ and $\phi = 0.08$ compared), the observations are paired, because every seed has a direct counterpart at the baseline and at each ϕ . These comparisons use the Wilcoxon signed-rank test, which works on the differences within each pair, looking at whether those differences point consistently in one direction (Wilcoxon, 1945).

Where two separate groups of markets are compared (for example comparing cycling and non-cycling markets within one friction tier), no such pairing exists and the Mann-Whitney U test is used instead (Mann & Whitney, 1947).

3.6.4 Multiple Comparisons and Inclusion Thresholds

As most comparisons in the analysis are repeated across friction tiers, each group of repeated comparisons is using the Benjamini-Hochberg correction. This controls the false discovery rate, because the expected amount of false positives would otherwise increase with the number of friction tiers analysed together (Benjamini & Hochberg, 1995).

The Benjamini-Hochberg method was also shown to be able to handle the already mentioned dependencies between groups, as long as those dependencies are positive. A seed that performs well at the baseline pushes every comparison involving it in the same direction (Benjamini & Yekutieli, 2001).

Which menu cost tiers are used differs between tests, and is determined separately for each one. For example, for tests that analyse cycling-static behaviour a tier is only included if at least ten cycling markets remain at that given friction tier. Tiers that are too small for a group comparison to be informative are excluded regardless of which test is applied.

3.6.5 Correlation and Spatial Measures

Where the analysis needs to test whether two measures move together rather than comparing groups, the Spearman rank correlation is used (Spearman, 1987). The measure captures a consistently monotonic pattern without assuming that the relationship is linear. This matters because several of the relationships tested are clearly monotonic but not straight. Working with ranked values rather than raw ones also suits the skewed distributions this simulation produced.

Spatial exploration is measured by the Shannon entropy (Shannon, 1948) of a firm's cumulative visitation distribution across the price grid, taken at the point of its last policy change. A higher Shannon entropy value reflects a policy that covers a wider range of price states (meaning that agents explored more of the price grid), and if the entropy value is lower a policy is more confined to a narrow region of the grid.

3.6.6 Visualisation Conventions

Some general goals and guidelines were used when creating the graphs and visuals. They are stated here as they are generally valid and do not need to be repeated with each visual.

Where a figure shows a subset of friction tiers rather than all available menu cost values, the tiers were chosen to help clearly present the data and the underlying trends. The selections often use menu cost values at which the general behaviour and data trend changes. These are not statistical claims on their own. The statistical tests reported alongside visuals are not restricted to the tiers visually shown, and aim to cover the full set of tiers if applicable to mathematically support the trends visible on the graphs.

4 Results

This chapter reports the simulation results in four steps. Section 4.1 checks that the frictionless environment reproduces the established benchmark before any friction is introduced. Section 4.2 then examines what menu costs do to market outcomes in terms of profit levels, and whether markets price dynamically or settle on fixed prices. Sections 4.3 and 4.4 turn from outcomes to the learning process of the agents, and Section 4.5 closes by comparing how much of the profit loss comes from the fee itself and how much comes from the learning phase impacts.

4.1 Baseline Validation

Before the effects of the added menu costs can be assessed, it needs to be validated that the simulated environment correctly replicates the baseline. The simulation needs to reproduce the results reported in the established frictionless benchmark (Calvano et al., 2020).

At $\phi = 0$ friction all 1,000 simulated seeds converge. This is consistent with the baseline

paper. Even if the convergence rate is given qualitatively rather than as an exact figure, Calvano et al. (2020) report that nearly all sessions converged.

The average firm captures 84.5% of the available monopoly rent (mean ergodic Δ is 0.845, and the median is 0.868)¹³. Calvano et al. (2020) report an average profit gain of 0.849 across all the sessions of their experiment, with Δ ranging from 70% to 90% over their full grid of learning and experimentation parameters. The output of the simulation in this thesis for $\phi = 0$ is inside that range and within half a percentage point of their reported average.

The population itself is not homogeneous. Markets use static pricing in 62.6% of the runs at the baseline $\phi = 0$. The remaining 37.4% are actively cycling in the observed evaluation phase. Calvano et al. (2020) report a closely comparable split for their experiment, with 64.3% of sessions settled at a constant price.

This split shows that a genuine share of the population ends up using a cyclical pricing pattern rather than converging on a fixed price, which is the same mechanism Calvano et al. (2020) identify as underpinning collusion in their model.

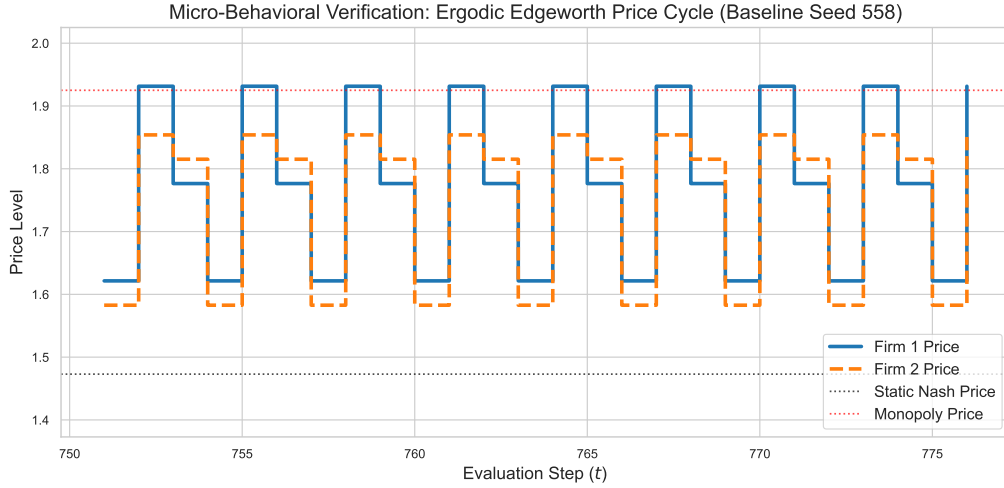


Figure 2: Frictionless baseline Edgeworth price cycle.

Figure 2 shows one such case directly. It plots a single manually selected seed's post-convergence price trajectory, with both firms stepping between several distinct price levels in a repeating pattern.¹⁴

After validating that the baseline values are correctly replicated in the current simulation, the friction tiers examined in the rest of this chapter were selected as described in Section 3.3.

¹³For more details see Appendix Table A.3

¹⁴This seed cycles over multiple price levels. Calvano et al. (2020) report that most of the cycles in their data have a cycling period of two, which is also valid for the current thesis's simulation results; the seed here was chosen for the visual clarity with which it illustrates the mechanism, not for being the most representative case.

4.2 Market Outcomes and the Phase Transition

Section 4.1 established that the frictionless environment correctly works as the baseline model predicts. This section introduces the menu cost friction and analyses what happens to market outcomes regarding profit levels and the static and cycling behaviour of the markets.

4.2.1 Continuous Profit Degradation

Table 2: Continuous Profit Degradation: Wilcoxon Signed-Rank Test vs. Baseline

Menu Cost (ϕ)	N	Med. Δ (Base)	Med. Δ (Fric.)	Med. Diff.	p (raw)	p (BH)	Sig. ^a
0.030	1000	0.868	0.860	0.008	0.463	0.514	False
0.040	1000	0.868	0.848	0.019	0.008	0.010	True
0.080	1000	0.868	0.828	0.039	5.124e-05	6.833e-05	True
0.700	1000	0.868	0.654	0.214	1.045e-128	2.089e-127	True

^a Significant at the 5% level after Benjamini-Hochberg correction.

Note. For additional data, see Table A.1.

The first question examines if profit falls as menu costs rise, and if so, from what point. At the lowest friction levels ($\phi = 0.01$ to 0.03) no decline can be distinguished from ordinary variation. Menu costs this small appear to be absorbed without disturbing the collusive outcome. From $\phi = 0.04$ onward a measurable decline appears and this trend also grows as friction rises. The median profit gap against the baseline widens from roughly 0.02 at $\phi = 0.04$ to 0.21 at the highest level of friction tested at $\phi = 0.70$ (Table 2). Menu costs therefore are shown not to act as a smooth, proportional tax on profit. They have a negligible effect below a certain threshold and get increasingly relevant after a certain turning point gets passed.

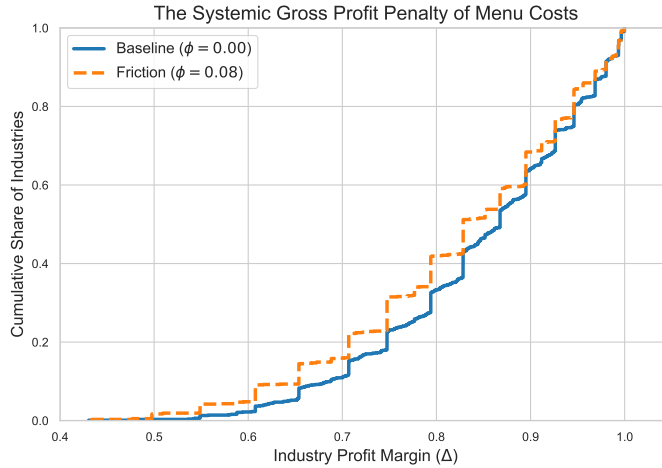


Figure 3: The systemic gross profit penalty of menu costs.

Knowing that the median falls still leaves an important question open. A single summary figure, such as the median loss of 0.039 at $\phi = 0.08$, describes the centre of the population but says nothing about its edges. The decline could be broad, or it could be driven by a subset of markets collapsing while the rest stay largely unaffected. Figure 3 compares

the full range of outcomes at baseline against $\phi = 0.08$ and shows the entire distribution visually shifting downward together, from the best-performing markets to the worst. The decline is not concentrated in a less performing minority. The paired tests confirm the fall in the median separately, and at every tier from $\phi = 0.04$ upward (Table A.1).

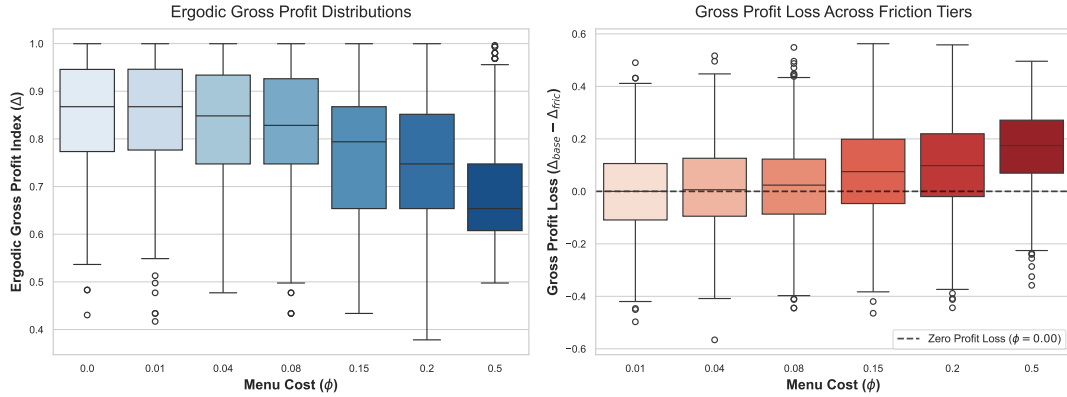


Figure 4: Ergodic gross profit distributions (left) and gross profit loss across friction tiers (right).

The two panels of Figure 4 answer different questions, which is why both are shown. The left panel plots the profit level each market reaches at a given friction tier. Its spread reflects a mix of two things: the effect of friction, and the ordinary market-to-market variation that is already present at the frictionless baseline.

The right panel removes that second component. Because the common random numbers design pairs every market with itself, each market's loss is measured against its own frictionless outcome rather than against the population. This is why the baseline tier plot was removed entirely (was kept as the zero profit loss reference line only) and with that what is depicted on the plot is the variation in how much effect the friction had at each tier.

The right panel also shows that the cost of friction is not distributed evenly. Some markets lose very little, but others lose substantially more, despite a clearly established average loss across the population. Meaning that the impact of menu costs is not a fixed quantity determined by the friction level alone, but something that plays out differently across individual markets.

4.2.2 The Extinction Threshold

The previous section established that profit falls as friction rises. In this section it is examined at what level of friction the cycling behaviour effectively disappears.

At the frictionless baseline only about 37% of simulated markets sustain cycling behaviour, so dynamic pricing is already the less common market state before any menu cost is introduced into the simulation.

The probability that a market is still cycling is estimated with a population-averaged Generalized Estimating Equations (GEE) logit model, with the menu cost used as a continuous variable and observations grouped by seed. This is because the analysis requires

a model that accounts for the same seeds appearing at every tier (Liang & Zeger, 1986). Grouping by seed produces standard errors that remain valid even when a market’s outcomes are correlated with its own outcomes at other friction levels.

As shown in Table 3 the fitted coefficient on menu cost is negative and highly significant, providing statistical evidence that the probability of a market sustaining cycling behaviour falls sharply as friction rises.

Table 3: Population-Averaged GEE Panel Logit: Cycle Survival

Coefficient	Estimate	Std. Error	z	$P > z $	[0.025	0.975]
Intercept	-0.4194	0.0375	-11.1694	5.760e-29	-0.4930	-0.3458
menu cost fraction	-28.6588	0.6999	-40.9444	< 1e-300	-30.0306	-27.2869

Because the share of cycling markets starts below 50%, there is no midpoint for a conventional breakpoint to cross. Therefore a 5% extinction threshold is used here.

Solving the fitted model for the friction level at which cycling probability drops to 5% gives $\phi_{\text{extinct}} \approx 0.088$. Past this point, dynamic pricing behaviour is practically absent from the population.

$$\phi_{\text{extinct}} = \frac{\ln\left(\frac{0.05}{0.95}\right) - b_0}{b_1} \quad (9)$$

where b_0 and b_1 are the fitted intercept and menu cost coefficient also reported in Table 3.

The consequence is that later tests restricted to cycling markets can only run over a narrow range of friction. Those tests require at least ten surviving cyclers in a tier, which limits them to $\phi \leq 0.13$.

Across the range plotted in Figure 5, the empirical cycling rate at each individual menu cost level tracks the fitted curve closely, which suggests the logistic decay describes the pattern accurately in this simulation rather than just forcing a shape on it.

The steepness of this decline is consistent with the context that prices are strategic complements in the simulated environment. Each firm’s incentive to adjust depends on the other’s behaviour, so the inaction of one firm reduces the benefits of price adjustments for the other one.

4.2.3 Profitability of Cycling versus Static Behavior

The previous section left open why cycling disappears. An explanation could be that if friction makes cycling unprofitable relative to holding a price steady, then firms would be abandoning this strategy to maximize profits. This section tests this explanation directly on the simulated outputs.

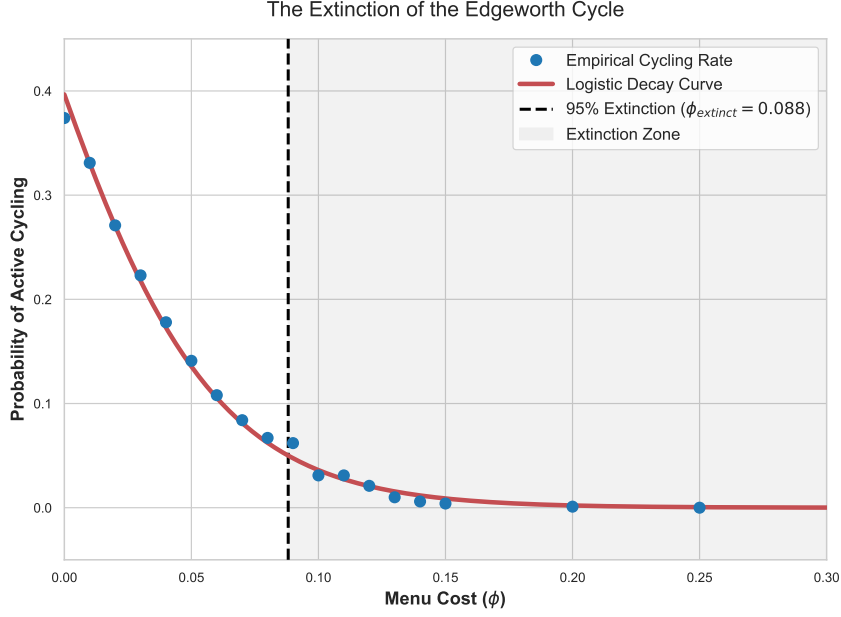


Figure 5: The extinction of the Edgeworth cycle: empirical cycling rate, fitted logistic decay curve, and the extinction threshold ϕ_{extinct} , at which cycling probability falls to 5%.

At the frictionless baseline, cyclers earn significantly less than statics (mean Δ 0.820 versus 0.860; $p < 0.001$). This is consistent with Calvano et al. (2020), who also report lower average profit for cycling sessions than for sessions settling on a constant price. The expectation in this thesis was that menu costs would widen this gap, because a strategy built on frequent price changes would have to pay the most in menu cost fees. Additionally, any surviving price wars would become rarer but more severe, since a firm cutting back on price changes would have fewer opportunities to engage in price wars so it would need to make the remaining price changes count.

Looking at the results based on gross profit, the opposite happens. As friction rises the raw Gross cycler premium changes sign, with cyclers moving from earning less than statics at baseline to earning more from $\phi \approx 0.05$ onward (Table 4). This positive gap is statistically confirmed at one tier only, $\phi = 0.10$, but it is also never once significantly negative after the sign has changed.

Table 4: Static vs. Cycler values across friction tiers

Menu Cost (ϕ)	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.10	0.11	0.12	0.13
Cycler (%)	37.4	33.1	27.1	22.3	17.8	14.1	10.8	8.4	6.7	6.2	3.1	3.1	2.1	1.0
Global Mean Δ	0.845	0.848	0.844	0.840	0.832	0.831	0.819	0.819	0.823	0.812	0.802	0.789	0.792	0.780
Static Mean Δ	0.860	0.854	0.852	0.842	0.835	0.831	0.816	0.817	0.822	0.809	0.800	0.787	0.790	0.779
Cycler Mean Δ	0.820	0.834	0.823	0.833	0.819	0.834	0.845	0.841	0.836	0.844	0.867	0.833	0.861	0.843
Cycler Mean Net Δ	0.820	0.825	0.805	0.808	0.784	0.794	0.797	0.785	0.776	0.776	0.792	0.740	0.771	0.742
Cycler Premium Mean Δ	-0.040	-0.020	-0.029	-0.009	-0.016	+0.003	+0.029	+0.024	+0.013	+0.034	+0.067	+0.046	+0.070	+0.063
Cycler Premium Mean Net Δ	-0.040	-0.029	-0.046	-0.034	-0.050	-0.037	-0.019	-0.032	-0.046	-0.034	-0.008	-0.047	-0.020	-0.038

Note. For additional data, see Table A.3.

At first glance this would suggest that menu costs make cycling comparatively more beneficial. The gross figure (from its definition) however, only measures what a cycling firm earns before paying for the price changes that cycling requires. After the net measure calculated with the menu costs subtracted, it results in the opposite answer.

The Net premium is significantly negative at every tier from the baseline through $\phi = 0.09$,

and again at $\phi = 0.11$ (Table 5). It never once reaches significance in the positive direction anywhere in the tested range. This confirms that cycling never becomes the better strategy under friction. The markets that keep cycling are not doing so because it is the better strategy, they keep cycling despite it.

Table 5: Static vs. Cycler Statistical tests

Menu Cost (ϕ)	N Cyc.	N Stat.	Gross Prem.	p_G^a	Sig. $_G^b$	Net Prem.	p_N^a	Sig. $_N^b$
0.000	374	626	-0.040	0.000	True	-0.040	0.000	True
0.050	141	859	0.003	0.650	False	-0.037	0.000	True
0.080	67	933	0.013	0.663	False	-0.046	0.003	True
0.100	31	969	0.067	0.017	True	-0.008	0.679	False

^a BH-corrected p-value.

^b Significant at the 5% level after BH correction. G = Gross, N = Net

Note. For additional data, see Table A.2.

The comparison also becomes progressively harder to make. At baseline it weighs 374 cyclers against 626 statics, but at $\phi = 0.13$ only 10 cyclers remain. The non-significant results at the upper tiers therefore likely reflect the shrinking sample size of cyclers.

Taken together, these results indicate that the apparent gross profit inversion largely reflects the cost of cycling itself rather than the economic success of cycling behaviour against the static markets.¹⁵

The rising gross figure still has some relevance for understanding market behaviour. The first is that cyclers adapt. To keep cycling at all under friction, a market would have to develop a more efficient cycling strategy than before, because it would need to extract more gross profit per price change to account for the menu costs. The second is that the population changes composition, meaning that seeds that cannot figure out a profitable enough cycling strategy ultimately abandon cycling and become static markets. The remaining cyclers with the strongest underlying pricing strategies still continue to cycle, which raises the group's average gross figure without requiring any change in how an individual strategy behaves.¹⁶

Nothing in the profit comparison can separate these two explanations so far, because both predict the same rise in gross profit. They differ in what they predict about behaviour, which the following section examines.

4.2.4 The Erosion of Credible Punishment

The two explanations from the previous section differ in what they predict about punishment behaviour. If cyclers adapted to the fee, their price wars should have changed shape and got deeper. If they were only filtered out of a shrinking population, the still remaining cyclers should engage in punishment cycles like they did at the baseline. This section aims

¹⁵The gross comparison loses significance across the middle range of ϕ (roughly 0.03 to 0.09), where the gross premium is itself close to zero and changes sign. An effect near zero is inherently the hardest to distinguish from sampling noise regardless of sample size, and this should not be read as a separate effect requiring its own explanation.

¹⁶This concerns only the population wide composition. Whether the seeds that stop cycling are the same ones from tier to tier was not tested at seed level.

to find out which of these two options is happening in the data.

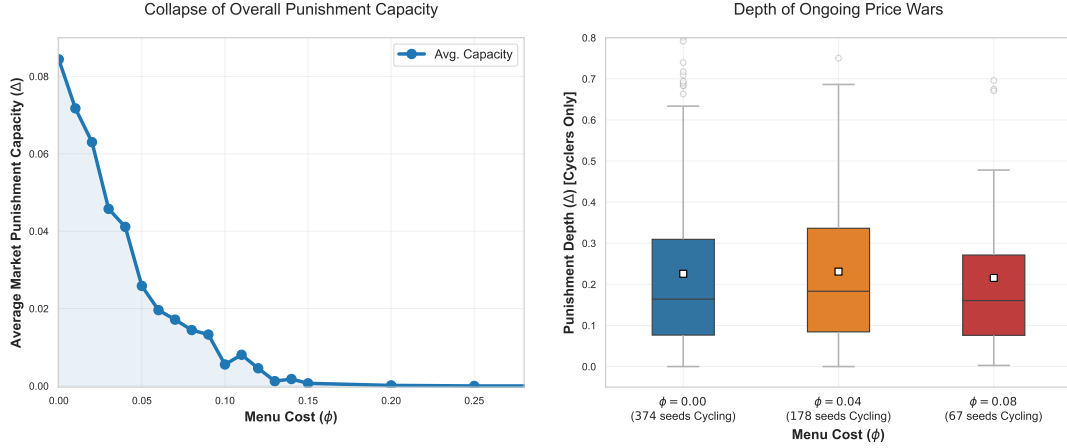


Figure 6: Collapse of overall punishment capacity and depth of ongoing price wars.

As already established the market’s overall capacity to sustain punishment declines steeply as friction rises. It approaches zero across the range shown in the left panel of Figure 6. This mirrors the extinction curve from the previous section, and not by coincidence. The measure on this graph is built as the product of how many markets are still engaging in punishment cycles, and how severe that punishment is when it happens. The results below show that the second factor barely moves with changing ϕ values. What the left panel plots is therefore close to the extinction curve shown in Figure 5, expressed in cycling punishment depth instead of cycling probability.

This leads to the distinction this section is built on. Punishment can become rarer, or it can become weaker, and the two are easily confused visually. The right boxplot panel is also a good example for this. It compares surviving price wars at the baseline $\phi = 0.0$, $\phi = 0.04$ and $\phi = 0.08$ levels. The whiskers and the number of outliers visibly shrink on the figure from left to right. That shrinkage comes from the falling number of surviving cyclers ($374 \rightarrow 178 \rightarrow 67$), not from a change in depth. A smaller sample is simply less likely to contain extreme values. In contrast the median and the interquartile range are far less sensitive to sample size, and they stay stable across all three tiers.

Table 6: Punishment Depth: Baseline Cyclers compared to Cyclers at Each Tier

Menu Cost (ϕ)	N Cyclers	Mean Depth	p (raw)	p (BH)	Significant (BH)
0.010	331	0.217	0.692	0.743	False
0.040	178	0.231	0.435	0.706	False
0.080	67	0.216	0.743	0.743	False
0.130	10	0.124	0.148	0.615	False

Note. There are not enough cyclers above $\phi = 0.13$ to test reliably. For additional data, see Table A.4.

The formal tests agree with what the boxplot shows. Surviving cyclers’ punishment depth was compared against the baseline at every valid friction tier from $\phi = 0.01$ through $\phi = 0.13$. No tier shows a statistically significant difference, even after correcting for the number of tests performed (Table 6; the full sweep is in Table A.4). No systematic change

in punishment depth shows up anywhere in the tested range.

This rejects a specific prediction from classical (S, s) pricing theory (Sheshinski & Weiss, 1977), and the expectation set out at the start of the previous section.¹⁷ Price wars surviving under higher friction do not become more severe. Punishment depth does not rise as friction increases, which also means friction is not selectively removing the weaker punishers. Cycling markets disappear regardless of how strong their punishment strategies are.

This answers the question the previous section left open. Adaptation of this kind would have shown up as a change in punishment behaviour, and this was not observed. So in conclusion the menu cost friction removes whole cycling markets rather than reshaping the price wars inside the ones that keep a cycling market behaviour.

4.3 Temporal Mechanics and Premature Calcification

The previous sections established how the introduced menu cost friction changes how firms behave in their pricing strategy. This section shifts the focus from outcomes to the process by which agents get there, examining how menu costs alter the path to convergence.

The frictionless environment already validated in Section 4.1 reproduces the baseline setting studied by Calvano et al. (2020). One feature of that setting, which the authors raise themselves, is that convergence takes an extremely large number of repetitions. On their own estimate, even if a single period lasted only a few minutes, reaching convergence would correspond to several years in real time (Calvano et al., 2020). They also argue that this matters less than it first appears, since the algorithms begin to collude well before they converge. Additionally how short a period is also depends on the markets.

Against the $\phi = 0$ baseline, the friction tiers examined converge in fewer steps, and they converge even faster as the friction rises. The added menu cost therefore reduces the above mentioned limitation of the baseline environment, without fully resolving it. How the shorter convergence time emerges is the subject of this section.

4.3.1 Premature Strategy Calcification

Exploration is not free once menu costs are introduced. Every price change made during learning carries a fee, which lowers the value the firm attaches to changing price. A policy that keeps updating itself is therefore penalised relative to one that settles for a constant value, so the greedy policy would be expected to stop changing earlier as friction rises.

The distribution of convergence times is described with a complementary empirical cumulative distribution function (ECDF). The curve plotted is the probability of a market still being in the exploration phase at a given point. It shows for each number of training steps the share of markets that have not yet converged.

¹⁷Sheshinski and Weiss (1977) show that a higher cost per price change makes adjustments less frequent and larger. In their model, firms adjust prices because inflation erodes the real value of the current price, but the effect of adjustment costs is of relevance here, not the underlying reason for adjustment.

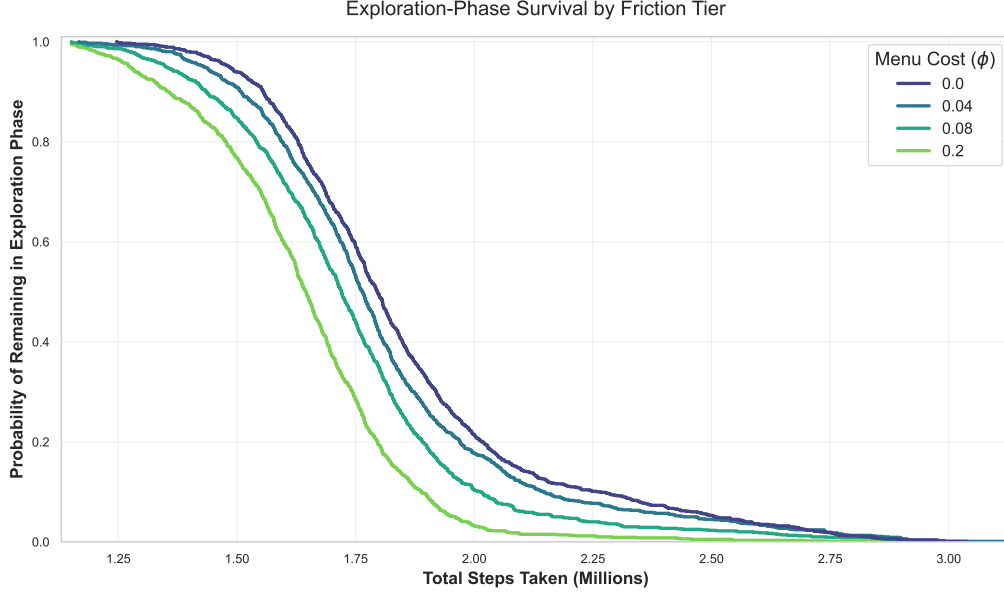


Figure 7: Exploration-phase survival by friction tier.

Figure 7 shows this pattern. At all four friction levels the curve falls as training goes on, and the whole curve shifts left as friction rises. Markets under higher menu costs reach a stable policy after fewer steps than markets at lower friction and at the frictionless baseline.

The same decline appears across the full set of friction tiers. The median step at which a market's policy last changed falls from 1,695,734 at the baseline to 1,223,686 at $\phi = 0.70$ (Table A.3).¹⁸ Not every tier falls below the one before it, and a few sit marginally higher than the preceding one. The direction is still consistent across the full range.

As every market in the simulation converged, no survival estimator is required here, because no observation is cut short or is left unresolved. The empirical distribution describes the data completely on its own. The curves are therefore used descriptively, establishing a pattern rather than a statistically tested effect. They do not show firms reaching the same endpoint faster. The policy stops changing sooner, which shortens the whole learning phase. This fits the current setting where every price change costs money. A change that would otherwise look beneficial now has lower returns once the menu cost penalty is calculated in, so the policy settles on a fixed choice earlier. Menu costs therefore shape the learning process, not only its outcome.

4.3.2 The Cumulative Friction Tax

If exploration ends earlier under friction, the next question is to review what that exploration is costing for the different training length, and whether the cost depended on which strategy (cycling or static) a market was heading toward.

¹⁸The figure plots total steps taken, which includes the fixed verification window used to confirm convergence. The appendix table reports the freeze step itself, meaning the point at which the greedy policy last changed. The two differ by the length of that verification window, and both fall as friction rises.

Table 7: Training-Phase Price-Change Rate for Cyclers and Statics with Mann-Whitney U Test

Menu Cost (ϕ)	N Statics	N Cyclers	Med. Static Rate	Med. Cycler Rate	Rate p (BH)	Rate Sig. ^a
0.00	626	374	0.5846	0.8975	5.972e-130	True
0.01	669	331	0.5768	0.8731	3.168e-125	True
0.02	729	271	0.5734	0.8715	1.018e-110	True
0.03	777	223	0.5591	0.8386	9.768e-94	True
0.04	822	178	0.5449	0.8380	2.833e-86	True
0.05	859	141	0.5400	0.8096	2.533e-66	True
0.06	892	108	0.5324	0.7680	2.747e-55	True
0.07	916	84	0.5267	0.7736	8.951e-45	True
0.08	933	67	0.5152	0.7363	2.402e-32	True
0.09	938	62	0.4994	0.7143	1.959e-34	True
0.10	969	31	0.4947	0.7096	5.502e-19	True
0.11	969	31	0.4883	0.7212	5.502e-19	True
0.12	979	21	0.4738	0.6897	3.023e-12	True
0.13	990	10	0.4707	0.6766	4.021e-07	True

^a Significant at the 5% level after Benjamini–Hochberg correction.

Note. There are not enough cyclers above $\phi = 0.13$ to test reliably.

The results show that markets that ended up cycling changed price far more often during training than markets that ended up static. At the baseline $\phi = 0$ the median rate is 0.8975 for cyclers against 0.5846 for statics, and the gap holds at every tier where enough cyclers remain to be compared (Table 7). The comparison runs to $\phi = 0.13$, the last tier with at least ten cyclers. Every one of those comparisons is significant after correcting for the number of tests. The difference is therefore not something friction produces. It is already present at the frictionless baseline, where no fee exists at all, and it persists at every tier that can be tested. Cyclers are not markets that changed their learning phase experimentation more because of the menu cost. They are a type that always experimented more. What friction changes is what being a cycling market type costs.

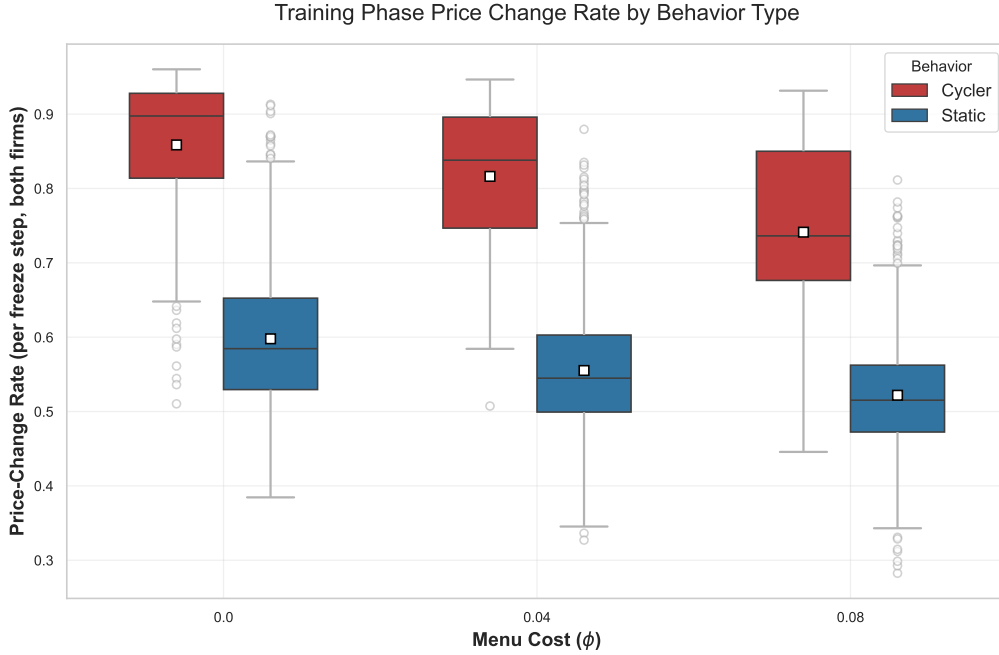


Figure 8: Training-phase price-change rate by behaviour type.

Figure 8 shows the two groups at the baseline, $\phi = 0.04$ and $\phi = 0.08$. The gap between the cycling and the static population is clearly visible at all three levels, and both groups drift downward as friction rises.

Table 8: Cumulative Training-Phase Menu Costs for Cyclers and Statics with Mann-Whitney U Test

Menu Cost (ϕ)	Med. Static Tax	Med. Cyclers Tax	Tax p (BH)	Tax Sig. ^a
0.00	—	—	—	—
0.01	9,522	14,401	1.014e-116	True
0.02	18,742	28,533	5.464e-106	True
0.03	27,284	41,919	3.332e-94	True
0.04	35,509	58,204	3.296e-90	True
0.05	43,557	67,902	1.644e-68	True
0.06	51,292	81,170	1.691e-57	True
0.07	58,528	96,162	2.001e-49	True
0.08	64,773	106,323	1.318e-39	True
0.09	71,453	116,372	6.801e-37	True
0.10	77,572	123,648	1.645e-20	True
0.11	83,361	157,223	3.174e-20	True
0.12	88,993	143,011	1.165e-14	True
0.13	94,794	158,922	3.998e-07	True

^a Significant at the 5% level after Benjamini–Hochberg correction.

Note. There are not enough cyclers above $\phi = 0.13$ to test reliably.

Cyclers also pay more in total in training. Expressed in the same normalised units used elsewhere in this thesis, the median cumulative training-phase menu cost paid by cyclers runs between 1.5 and 1.9 times that paid by statics, and this gap is significant at every comparable tier (Table 8). The rate and the total could in principle disagree. A market with a short training phase could change prices often and still have a low total cost paid during the training phase, but cyclers score high on both, so their higher rate is not an artefact of shorter training. The totals build up over more than a million training steps, so the ratio between the two groups is a clear explanation here.

This gives the profitability results of Section 4.2 a more direct cost interpretation. It would be easy to read the gap between cycler and static markets as an operating cost, where cyclers keep paying the fee each time they change prices under their converged policy and statics do not. The training-phase figures show the disadvantage starts earlier than that. Reaching a cycling strategy is the more expensive path to take, well before any of that strategy is used in the evaluation phase.¹⁹

The training-phase rate turns out to predict something else beyond costs. It also predicts how much of the price grid a market learns. Markets with higher training-phase price-change rates finish with higher policy entropy levels, meaning a more broadly explored final pricing policy. This relationship is strong at every friction level (Table A.5). It is weakest at the frictionless baseline, at $\rho = 0.786$, strengthens to around 0.930 across the middle tiers, and then weakens again at the highest friction levels, falling to $\rho = 0.500$ at $\phi = 0.50$.²⁰

¹⁹Markets are classified by the strategy they converged on, so the higher training-phase rate describes markets that ended up cycling. Whether heavier experimentation leads to cycling, or markets heading towards cycling experiment more, cannot be separated here.

²⁰These per-tier correlations are reported without a multiple-comparison correction. Given the size of

At any given friction level, the markets that changed price more often during training are the ones that end up knowing more of the price grid. Section 4.3 showed that friction shortens the training phase, and the rates reported above fall as friction rises. This section shows that markets which experiment less during training end up knowing less of the price grid. What a market ends up knowing is therefore not fixed by the environment, it depends on how much experimentation agents did in the training phase before policies stopped changing. How much the knowledge of the market differs as friction rises is discussed in the next section.

4.4 Spatial Blindness and Information Decay

Section 4.3 measured when learning stopped and what it cost along the way. This section turns to the resulting policies and examines two dimensions: how much policies are changing, and how much of the price grid they end up covering when converged.

4.4.1 Policy Volatility and Freezing

The first measure captures how stable a policy is during training. Policy volatility measures the share of training steps at which the best action for a given state changes. A higher value indicates that the policy continues to revise its choices and a lower value indicates that it has largely stabilised.²¹

Table 9: Median Policy Volatility (Greedy Inertia Rate) by Friction Tier

Menu Cost (ϕ)	Median Policy Volatility
0.00	0.2162
0.04	0.2060
0.08	0.1943
0.70	0.0447

Note. For additional data, see Table A.6.

It is compared across all 21 tiers using the Friedman test, a rank-based alternative to the analysis of variance for data classified by two criteria (Friedman, 1937). The test treats each seed as its own block and compares the tiers within it, which is the design that common random numbers produce (Cooley & Cooley, 1980). This requires every seed to be present at every tier and would mean that a seed missing at any single tier could not be used at all, because there would be no complete trajectory to compare. Therefore only seeds that converged at all 21 tiers are kept, so every comparison draws on the same fully balanced panel. In the current simulation results all seeds converged across all friction tiers so no listwise deletion was needed, meaning that the balanced panel is the full simulated population.²²

Volatility falls as friction rises. At the frictionless baseline the median firm revises its

the associated p -values, correction would probably still not change which of them are significant.

²¹The underlying variable is named `greedy_inertia_rate` in the simulation output, but it counts policy changes rather than periods of inertia. Higher values therefore indicate less inertia.

²²The retention rate is still reported in Table A.3 to rule out any survivorship concern.

greedy policy at 21.6% of training steps. At $\phi = 0.08$ this falls to 19.4%, and at the highest tier tested, $\phi = 0.70$, to 4.5% (Table 9). The test confirms at a significance level beyond ordinary reporting precision that these differences are not sampling noise. Volatility tracks friction across the whole range, not only between the extremes.

The decline is also largely consistent. Of the twenty steps between consecutive tiers, nineteen fall below the tier before them. The single exception is at $\phi = 0.02$, which sits marginally above $\phi = 0.01$ at 0.2146 against 0.2138. Several other trends reported here show more tier-to-tier noise than this one. That regularity is itself informative. Profit showed no detectable decline until $\phi = 0.04$ (Section 4.2.1), and cycling collapsed around a threshold rather than fading evenly (Section 4.2.2). Policy volatility does neither. The medians fall from the first tier onward, in small steps, with no point at which the response changes character. Whatever connects friction to policy revision appears to act smoothly rather than switching on once costs pass a threshold.

Because volatility is a rate rather than a count, this result is different from showing the shorter training phase established in Section 4.3. The number of policy revisions is divided by the number of steps taken, and both fall as friction rises. For this rate to decline, policy revisions must become less frequent more quickly than the training window shortens.²³

This adds a second dimension to the calcification result. Section 4.3 showed that training ends sooner under friction. This result shows that within that shorter phase, revisions also become less frequent. A firm facing high friction is therefore constrained in two ways. It has fewer steps in learning, and it changes its policy on fewer of the steps it does have.

4.4.2 The Decay of Spatial Awareness

The second measure captures coverage rather than stability of the agents' policies. Where policy volatility measured how often a firm adjusts its optimal policy, this section looks at how much of the price grid is explored during training. This is measured using the Shannon entropy of the greedy visitation matrix. At every training step the matrix records which price pair the two greedy policies would select, accumulating across the whole run, and it is captured as it stood at the last policy change. The Shannon entropy is high when a firm has visited many price combinations and low when it has concentrated only on a few.²⁴

²³The denominator includes the fixed verification window used to confirm convergence, during which no policy change can occur. As runs shorten, that window makes up a slightly larger share of the total, which pushes the rate down mechanically. The effect is small, moving from roughly 5.6% of the run at the baseline to 7.6% at $\phi = 0.70$, therefore it cannot account for a decline of the size reported above.

²⁴The grid holds 225 price combinations, so the measure cannot exceed $\log_2(225) \approx 7.81$ bits. Reaching that value would require a firm to have visited every combination equally often.

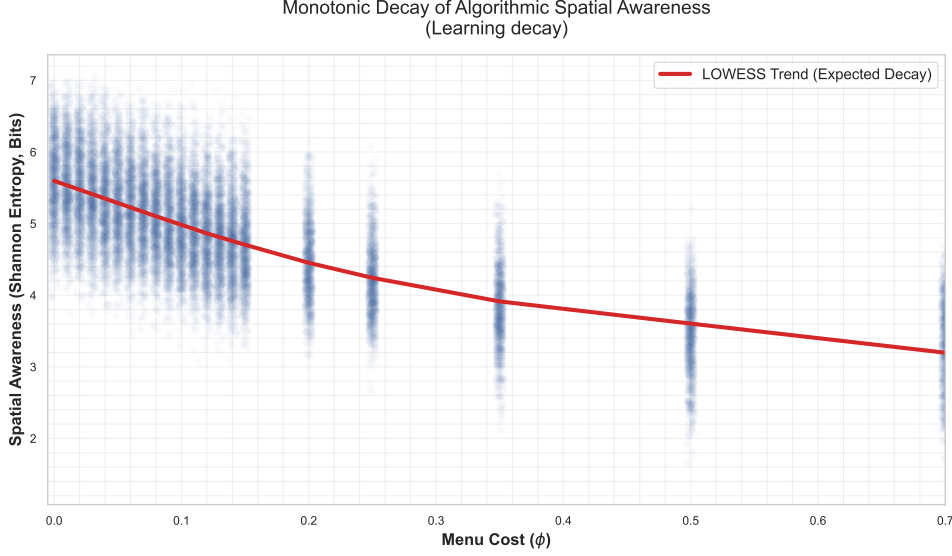


Figure 9: Decay of visitation entropy across friction tiers.

Pooling every seed’s final entropy against its friction level shows a clear downward trend (Figure 9). Firms that face higher menu costs learn measurably less of the price grid by the time they complete the training phase.

The decline is not evenly spread across the friction range. It is steepest at low friction and flattens as friction rises. In the lower menu cost scenarios the median entropy falls by 0.49 bits, from 5.59 bits at the baseline to 5.10 bits at $\phi = 0.08$. Over the tested range with higher menu costs, from $\phi = 0.50$ to $\phi = 0.70$, a span more than twice as wide, it falls by only 0.11 bits (Table A.3). Expressed as a rate per unit of ϕ , the decline at low friction is around ten times steeper than at high friction.²⁵ Entropy falls at a broadly similar rate from the baseline through to about $\phi = 0.25$, and the decline slows sharply beyond that point.

This pooled trend could allow for an alternative explanation. It could mean that every firm loses coverage steadily, or that most firms are largely unaffected while a minority declines sharply enough to pull the whole trend down. The two would look identical on Figure 9, because it mixes variation within firms together with differences in level between them.

Testing each seed’s own trajectory separates these two possible explanations. Every one of the 1,000 seeds is present at all 21 friction levels, so each firm can be followed across the whole tested range. The correlation is therefore computed within each seed separately rather than across the pooled population, giving one coefficient per seed from that seed’s own 21 observations. A single seed’s coefficient rests on only 21 observations and is too noisy to be informative on its own, but averaging across a thousand seeds removes that noise and makes the common pattern visible (Mohr & Marcon, 2005). The mean of those 1,000 coefficients is $\rho = -0.674$.

²⁵The two windows differ in width, so the rates are expressed per unit of ϕ to make them comparable. The comparison uses the most extreme tiers available in each region, since the simulated grid is not evenly spaced beyond $\phi = 0.15$.

The resulting distribution of seed-level coefficients is then tested against zero with a one-sample Wilcoxon signed-rank test (Wilcoxon, 1945), the same rank-based test used for the paired comparisons in Section 4.2, here applied to a single distribution rather than to differences between pairs. It asks specifically whether the correlations are negative.²⁶ It confirms that the distribution sits reliably below zero, at a significance level beyond ordinary reporting precision. The size of the average coefficient also indicates that the ordering is largely preserved within firms: across most of the range, tiers with higher friction tend to have lower coverage. The decline therefore occurs within individual firms' learning trajectories rather than arising from changes in the composition of firms across friction levels.

The size of the loss is substantial. Median entropy falls from 5.59 bits at the frictionless baseline to 4.45 bits at $\phi = 0.20$, by which point cycling has effectively disappeared from the population (Table A.3). That is a loss of 1.14 bits, roughly a fifth of the firm's original coverage.

4.4.3 The Contraction of the Explored State Space

The two measures reported so far show the firms learning in numbers. This section shows the learning of the agents visually on the grid of the available price combinations. It shows where the learning agents actually did spend time exploring. Each cell of the grid is one combination of the two firms' intended prices.

The seed for the market shown in Figure 10 was selected algorithmically. Each seed's entropy was compared against the population median at the four visualized reference tiers, and the seed with the smallest Euclidean distance across all four tiers was used. The market shown is therefore the one closest to a typical example across the whole range; it was not one manually selected for the shape of its result. Because in this section's visuals several matrices are shown side by side, the colour scale was set to use the same shade for the same value inside one group of visuals in every pane. This makes possible that the panels can be compared directly with each other.

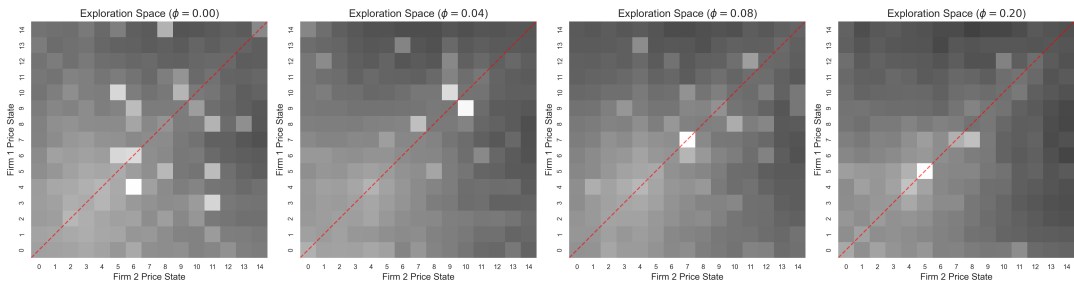


Figure 10: Exploration space of a representative seed across selected friction tiers.

The selection returned seed 281, the market lying closest to the population median across all four reference tiers. This individual market's entropy differs from the population median by at most 0.16 bits at any of the four tiers, and falls at every one of them, from 5.67 bits at the baseline $\phi = 0.0$ to 4.62 bits at $\phi = 0.20$.²⁷ Figure 10 shows what those numbers

²⁶This is the only one-sided test in the analysis, all other tests are two-sided.

²⁷These are the values for seed 281. The population medians reported in Table A.3 are 5.59 and 4.45

correspond to on the grid visually. The darker an area is, the less those price combinations were explored, so the contrast shows the differences in how much each region was visited during the learning phase.²⁸ The visitation pattern shows at the frictionless baseline that the agents have tested a wide region of the grid. As friction rises the bright region shrinks, this shows visually that the knowledge of the agents is narrowing in this area. In some part because the agents did not change prices as often in higher friction scenarios and in some part because the learning phase itself got shorter with higher menu costs.

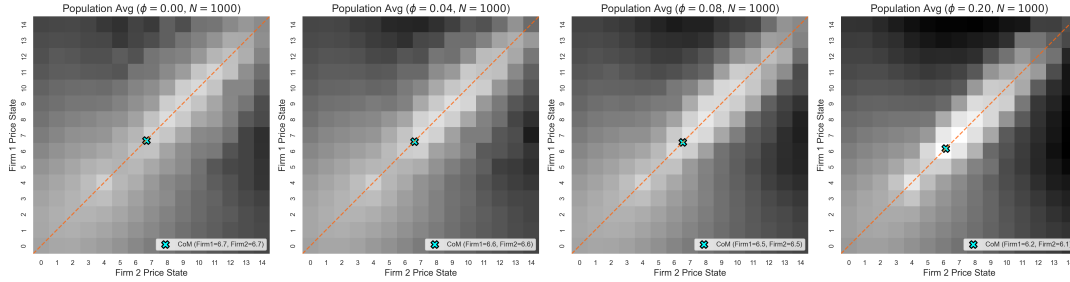


Figure 11: Population-averaged state-visitation matrices across selected friction tiers.

The second set of panels averages all 1,000 markets, which cancels out the noise of individual trajectories. Each market's matrix is normalised first²⁹, so the panels show where a typical firm explores. No market dominates the average simply by having trained longer, which results in a clearer view of which areas the firms generally learn better.

At the baseline the visits spread widely. As friction rises they pull into a narrow band along the diagonal³⁰. The centre-of-mass markers show that the band does not slide anywhere while it narrows, moving only from (6.7, 6.7) to (6.2, 6.1) between the baseline and $\phi = 0.20$. On a fifteen-point grid this is only about half a step difference.³¹

The panels also suggest where exploration is concentrated on the grid. This interpretation is descriptive rather than statistically tested. It is still discussed to explore and understand more about the learning behaviour of the agents.

Because both firms face the same problem, the pattern is close to symmetric about the diagonal, and the variation worth noting runs along it instead. Low indices are prices near the competitive level, high indices near the joint-monopoly level.³² Below the midpoint the brighter region covers a broader distance off the diagonal, meaning that firms tried more combinations in which one priced below the other. Above the midpoint the white region stays narrow around the diagonal, meaning exploration near monopoly prices concentrates

respectively.

²⁸The shading is logarithmic, so a step in shade corresponds to a change in the order of magnitude of visit counts rather than a fixed number of visits. The number of visitation counts is still relevant. As the total number of visits declines as training length gets shorter at higher friction, the matrices show darker areas, because they essentially tried less of the possible price combinations as well.

²⁹The visitation matrices that are averaged across markets were created only after each market's own matrix has been normalised, so that markets trained for longer do not get more weight in the average.

³⁰Cells on the diagonal are those where both firms charge the same price. It is marked with a red line on the visuals.

³¹The centre of mass is computed from the averaged matrices at each tier, so there is one calculated value per tier. It is reported here only as an additional information to the figures and is not tested further statistically for different ϕ tiers.

³²On the fifteen-point grid the Nash price falls at index 1.17 and the monopoly price at index 12.83, so the midpoint between them sits at index 7.

in options where both firms match the other more closely. As friction rises the upper region fades fastest, which is consistent with the centre-of-mass movement described above.

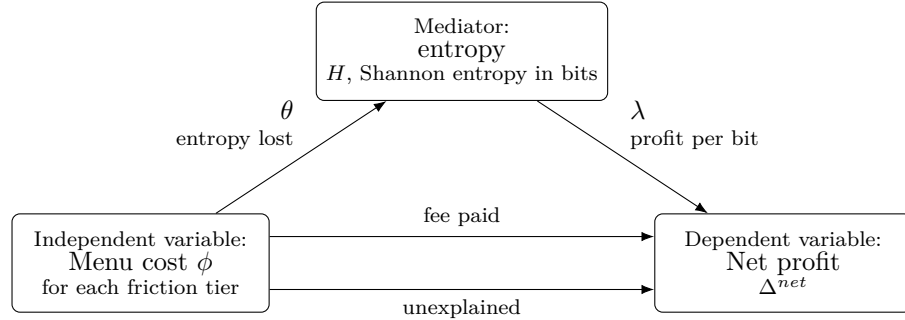
It is important to note that these matrices cover the whole training phase, not only the final converged policies of the markets. A bright region shows where a market spent its learning phase, not where it ended up pricing. Many of those steps might be taken long before the firms colluded. The converged outcomes using the final policies are reported in Section 4.2, the two concepts should not be confused.

Firms under friction are therefore not exploring a substantially different part of the price grid. They are exploring a smaller part of the same region, shifted slightly downward. The agents become progressively less aware of the pricing options away from where their final policy settles.

Taken together with Section 4.3, these results explain three dimensions of the calcification. It is shown that with increasing menu costs the training ends sooner, the policy revises itself less often, and the area of the price grid that is explored by the agents contracts. In the final section it will be explored how the policy entropy loss and the menu cost fee play together in causing the total profit loss established in Section 4.2.

4.5 Decomposing the Profit Loss

The final part of the analysis asks not only how much profit is lost as friction rises, but through which channels. The loss at each tier is split into three parts: the fee actually paid, the part caused by reduced exploration, and a remainder that neither of these parts accounts for. To get these values a mediation analysis is used.



$$\begin{aligned} \text{Indirect effect} &= \theta \times \lambda & \text{Direct effect} &= \text{fee paid} + \text{unexplained residual} \\ \text{Total effect} &= \text{Indirect effect} + \text{Direct effect} \end{aligned}$$

Figure 12: Decomposition of net profit loss (following MacKinnon et al. (2004))

The fee component requires no estimation, as it follows directly from each market's own recorded price changes. The exploration component is estimated with three regressions: the first measures the total profit loss at each tier, the second calculates the corresponding shift in entropy, and the third includes entropy alongside friction to measure the profit associated with one unit of entropy once friction is held fixed. Multiplying the second estimate by the third results in the portion of the profit loss attributed to reduced exploration. This

product is the indirect effect in the standard mediation terminology.

The standard model only splits the loss into two parts, a direct and an indirect effect (MacKinnon et al., 2004). Splitting it into three parts is possible here because the menu cost fee is known and does not have to be estimated, so the direct effect can be separated into the menu cost paid and an unexplained residual component.

Everything in this section is reported as a loss, so positive numbers mean profit lost relative to the frictionless baseline.

On Figure 13 the upper panel shows the three components against the total loss, and the lower panel plots the entropy share of the total loss on its own, including its confidence band. The result indicates that the majority of the profit loss is not due to the direct amounts paid for the menu costs. The two lowest tiers, $\phi = 0.01$ and $\phi = 0.02$, are left out of the figure, because the total loss there is not statistically distinguishable from zero. At every tier above them, the entropy decay accounts for the majority of it, running between 51% and 83% of the total net profit loss³³ (for detailed data see Table A.7 in the Appendix). Figure 13 shows the range up to $\phi = 0.25$, where most of the movements happen. The tiers above $\phi = 0.25$ continue the same pattern and are reported in Table A.7.

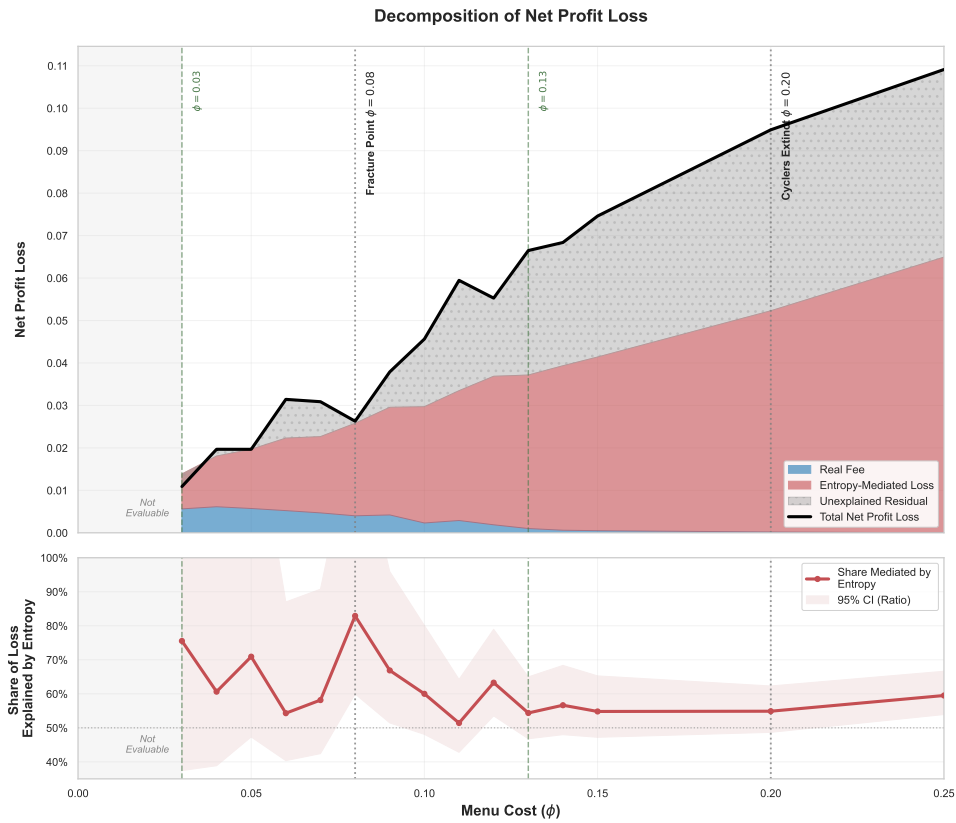


Figure 13: Decomposition of net profit loss (top), and the share of the loss mediated by entropy with the 95% bootstrap confidence band (bottom).

³³Training duration and policy volatility were also considered as mediators. Both decline with friction alongside entropy and are too collinear with it for the analysis to separate their individual contributions. The entropy channel should therefore be read as the exploration mechanism in general.

The menu cost fee behaves in a way worth noting, because it does not grow with ϕ , which is what the obvious intuition would suggest given that ϕ is the price of each change. It peaks at 0.0061 at $\phi = 0.04$, then falls steadily, rounding to zero from $\phi = 0.25$ onward. This pattern follows directly from how the fee is calculated. The fee is the number of price changes multiplied by the price of making a change, and Section 4.3 showed that firms change price far less often as friction rises. Beyond a certain point, the number of price changes falls faster than the cost per change rises. The menu cost fee therefore has a counterintuitive effect: firms avoid it by changing prices less, but avoiding price changes then is exactly what costs them more in the long run, because the cost of the lost market knowledge is not directly apparent.

On the other hand the entropy-mediated loss steadily rises at every friction tier, from 0.0018 at $\phi = 0.01$ to 0.1084 at $\phi = 0.70$. In the upper panel of Figure 13 the two are also visibly moving in opposite directions. The menu cost's direct effect band decreases to zero while the entropy band constantly covers a significant part of the total profit loss. At $\phi = 0.04$ the effect of the menu cost fee is still worth about a third of the total profit loss, but by $\phi = 0.20$ it is worth less than one percent of it.

In the equations below j is a single market and $D_{\phi j}$ is set to one if that market was observed at tier ϕ , otherwise it is set to zero. H is the entropy in bits. τ_ϕ is the profit difference at tier ϕ compared to the frictionless baseline, θ_ϕ the entropy difference at that same tier, and λ the profit associated with one bit of entropy. The terms κ_1 , κ_2 and κ_3 are the intercepts, giving the baseline values at $\phi = 0$, and ε_j , u_j and v_j are the error terms.

$$\Delta_j^{net} = \kappa_1 + \sum_{\phi > 0} \tau_\phi D_{\phi j} + \varepsilon_j \quad (10)$$

$$H_j = \kappa_2 + \sum_{\phi > 0} \theta_\phi D_{\phi j} + u_j \quad (11)$$

$$\Delta_j^{net} = \kappa_3 + \sum_{\phi > 0} \tau'_\phi D_{\phi j} + \lambda H_j + v_j \quad (12)$$

The first measures how much profit was lost at each menu cost tier against the frictionless baseline. The second calculates how much entropy was lost relative to the $\phi = 0$ tier. The third equation is less straightforward. It looks at the entropy values within the individual friction tiers for the same regression, which means λ is estimated from markets that faced identical friction but still ended up with different entropy. Estimating it from the pooled data instead would only confirm that high-friction markets have both lower entropy and lower profit, which was already shown, but would say nothing about what a bit of entropy loss is worth. The coefficients τ'_ϕ are estimated but not used further. This decomposition follows the indirect effect model of MacKinnon et al. (2004).³⁴

Because each tier is its own category, no linear relationship between friction and profit is assumed. Standard errors are clustered by seed, since observations from the same seed are correlated across tiers (Cameron & Miller, 2015). A single λ is estimated for the whole model rather than one per tier, so the tier-specific indirect effects vary only with the change in entropy at each tier.

³⁴The two path coefficients θ and λ are written as α and β from MacKinnon et al. (2004). In this thesis they were renamed because both are already used as learning parameters in Section 3.4.

The direct menu cost fees are calculated from the recorded price-change rate multiplied by ϕ , and averaging across the markets at the given tier gives that tier’s fee paid. The entropy-mediated loss and the residual are calculated with the unexplained residual defined as the part not accounted for by the other two components.

The entropy-mediated share of total profit loss peaks at 82.9% at $\phi = 0.08$, close to the friction level at which cycling effectively disappears (Section 4.2.2), and remains between 51% and 67% across the tiers above it.

The unexplained residual is the third component and it grows steadily, from effectively zero at the lowest significant tiers to 0.0755 at $\phi = 0.70$, where it is larger than the entire menu cost fee ever accounted for.

As described above, this component is visible only because the menu cost fee can be calculated exactly rather than being estimated. In a standard two-part decomposition it would be absorbed into the direct effect and never appear separately. It is the part of the loss that neither the fee paid nor the entropy channel accounts for, and the analysis could not identify what produces it. Its size and its growth are therefore a finding on their own, and identifying its source is suggested for further research.

Because the entropy-mediated loss is a product of two estimated coefficients, its distribution is not normal, so resampling gives more accurate results than a standard error would³⁵ (MacKinnon et al., 2004). Confidence intervals were therefore calculated with a cluster bootstrap method of 1,000 resamples. On each iteration whole seeds were drawn rather than individual observations, so that each resample keeps every seed’s full set of 21 tiers together to be able to carry the correlation between a seed’s own tiers into the resampled data (Cameron & Miller, 2015).

The menu cost fee is measured directly, but the entropy-mediated loss has been reported as a single best estimate. It is the product of two estimated coefficients, so it carries the uncertainty of both. Therefore it is worth asking how much of this could just be noise. The bootstrap (Table A.8 in the Appendix) answers two separate questions about this: The first part is whether the entropy channel is real. At 19 of the 20 friction tiers, the confidence interval does not include zero. The only exception is $\phi = 0.01$, where the total profit loss is also not statistically distinguishable from zero. The channel is therefore a real effect and not just produced by two estimates multiplied together.

The second part is the question of how much the entropy-mediated part of the total net profit loss could move if the sample included different markets. Here the answer is less exact. At $\phi = 0.20$, for example the best estimate is that the entropy channel accounts for 54.9% of the total loss (plotted on the visuals Figure 13), but the bootstrap places the plausible range between 48.7% and 62.3% (confidence range marked on the lower graph of Figure 13). That range shows how far the share could shift depending on which markets happen to be drawn. As the lower panel of Figure 13 shows, the range is very wide at low friction and narrows as friction rises. What the data supports is therefore not a precise percentage that can be reported directly as the entropy-mediated loss, but a clear statement about its overall impact. The share of the loss explained by exploration decay is around or above half of the total loss at every tier where it can be measured, and it is never close to

³⁵The intervals reported here are percentile intervals. MacKinnon et al. (2004) report better performance for the bias-corrected variant. The distinction has little bearing here, since the intervals are built from 1,000 resampled seeds and already exclude zero at 19 of the 20 tiers.

zero.

Menu cost friction does not eliminate collusive pricing. Even at the highest tier tested, mean Δ remains at 0.661, far above the competitive benchmark. It is confirmed that an added menu cost friction reduces market profits and removes the cycling form of collusion. However, the menu cost fee itself is a minor and eventually vanishing part in its direct effect on the profit decrease. The larger part is that the menu cost penalty mechanism indirectly influences the achieved market profits and behaviour. Section 4.2 showed that friction removes cycling markets, and Sections 4.3 and 4.4 showed that it shortens training and narrows the agents' knowledge about the market. What this adds up to is that the effect of menu costs is more significant indirectly, through altering the market behaviour and the agents' learning and knowledge.

5 Discussion and Limitations

5.1 Selection instead of adaptation

Two explanations could account for why cycling disappears as friction rises. Cyclers could either be adapting their behaviour to the new cost, or friction could be removing cycler markets without changing how the surviving ones behave. The results support the second explanation. Punishment depth among the markets that keep cycling does not substantially change across friction tiers. The price wars that survive are as deep as the ones at the frictionless baseline. And cyclers were already the more experimental type before any fee existed: at $\phi = 0$ they already changed price far more often during training than the markets that ended up static. The difference therefore predates the friction. What friction changes is what that difference costs, not how those markets behave.

This also tests a specific prediction from Sheshinski and Weiss (1977) that says a higher cost per adjustment should make price changes both rarer and larger. Rarer holds, but larger does not. The statistically unchanged punishment depth is the direct evidence against it. Part of the reason is in how the fee is built in this thesis. The penalty in Equation (4) triggers on every price change and it is also calculated as a fixed adjustment cost independent of the adjustment's size, so nothing in the reward gives a firm reason to keep its price wars shallow in order to save money.

5.2 Collusion and cycling markets

Cycling behaviour becomes extinct already at a low level of friction ($\phi_{\text{extinct}} \approx 0.088$), but the profitability index stays far above the competitive benchmark at every tier tested. Collusion persists; only its cycling form disappears.

This also rules out one explanation for the profit loss. At the frictionless baseline, cycling markets are the less profitable of the two types (mean Δ 0.820 against 0.860 for static markets). Friction removes exactly those markets, so the surviving population shifts toward the more profitable behaviour, but the total profit falls anyway. The loss therefore cannot be attributed to a change in the composition of the population, which left Section 4.5 to

identify that the agents' learning phase has a bigger effect on the final market result.

5.3 The effect of less exploration

Abada et al. (2024) and this thesis both examine what happens when agents explore less, but they do not approach it in the same way. They tested changing how much random exploration remains at the end of the decay process, holding the speed of that decay fixed. If that remaining rate is too low, the agents settle on the cooperative action before they have learned enough to know that competing could pay more, and they go on colluding. In their setting, less exploration therefore produced more collusion.

The lever in this thesis is a different one. The exploration-decay parameter β is identical at every tested friction tier, and what changes is only the cost of exploring. A menu cost makes every price change expensive, which in turn shortens training and leaves the agents covering less of the price grid. In this setting less exploration goes together with lower profitability, not higher.

Abada et al. (2024) report a cooperation rate normalised between the competitive and the cartel benchmark, which is the same construction as the profitability index used here, so the two results point in opposite directions. What differs is what the agents fail to learn. In their setting the agents never learn that undercutting would pay off for whichever firm tried it, so both keep cooperating and both end up earning more than they would by competing. Here the agents do not explore enough of the price grid to find good prices at all, so they earn less. Establishing whether a shrinking exploration rate and shrinking coverage act through the same channel would require varying β alongside ϕ , which this thesis does not do and which remains open for further research.

5.4 Shortened training time

Den Boer et al. (2026) argue that Q-learning reaches collusive outcomes only after a number of training steps far beyond anything relevant to a firm's real-world planning horizon. Calvano et al. (2020) acknowledge the same limitation of their own simulation. In this thesis it was shown that friction shortens the median training length substantially compared to the Calvano et al. (2020) baseline, but it does not completely close the gap between modelling and the real-world use case. Even under the highest friction tested, training still runs into the millions of steps, so the findings weaken this objection without fully resolving it.

5.5 Intervention placement

The decomposition in Section 4.5 shows that the direct menu cost fee a firm actually pays is a minor and eventually vanishing part of the profit loss, while the indirect effect through reduced exploration accounts for most of it. Most of what the friction does therefore happens during the learning phase, before any converged policy exists.

Section 2.3 separates two kinds of proposed responses. One places constraints directly in

an algorithm’s code or liability on its publisher (Beneke & Mackenrodt, 2021; Li & Virtosu, 2025), the other intervenes in the learning process itself. Abada and Lambin (2023) find that intervening during the learning phase produces a substantial welfare gain. The results here point the same way. Because menu cost friction is present to varying degrees in almost every real-world market and has most of its effects on the outcomes already before convergence, the learning phase appears to be where an intervention has more potential.

5.6 Limitations

As described in Section 3, this thesis used Calvano et al. (2020) as the baseline. The values taken from one representative point of this benchmark paper were not independently re-verified under friction.

The simulation is restricted to two firms. Whether the reported effects hold in markets with more competitors is untested, and the initialisation used here is only defined for a duopoly.

The only friction added to the baseline is the menu cost ϕ . It is a fixed amount per change, independent of the size of the price change as described in Section 5.1. A cost that scaled with the size of the price change, or one structured as a budget on the number of changes within a fixed period rather than a fee per change (Feldman & Segev, 2025), would likely produce different results.

The action space is bounded and discrete, a grid of m evenly spaced prices. den Boer et al. (2026) show what this means for how Δ should be read. Because the grid is built to sit between the Nash and monopoly prices, two firms choosing prices completely at random would already average above the Nash price, landing roughly halfway between the competitive and the monopoly benchmark ($\Delta \approx 0.50$). A high Δ is therefore partly a property of how the grid was set, and not only of what the agents learned. Even at the highest friction tested the mean index stays above this level, so the outcomes reported here are not accounted for by the construction of the grid alone. The comparisons across friction tiers are unaffected, since the grid is identical at every tier.

As described in Section 4.5, the third component of the profit loss is attributed to neither the menu cost fee nor the entropy channel, and at high friction it exceeds the fee. The size is reported and substantial at higher ϕ values, but its source is not identified.

Consumer surplus is not computed anywhere in the analysis, so nothing here supports a claim about who bears the cost of the friction. The results only speak to what firms earn and how they behave differently once the friction is added.

As the data is collected from simulated markets, the resulting findings and behaviours can only be handled as a starting point for expectations about real industries, but are not direct evidence about real-world behaviour.

6 Conclusion

This thesis set out to answer the question: *How do menu costs alter the stability and structural dynamics of algorithmic collusion in a Q-learning pricing duopoly?* The short answer is that the friction reshaped collusion without ever preventing it. The added menu cost mechanism acted on two things at once (described just below), but it did not affect whether collusion occurred. Collusion held at every friction level tested.

The first is stability, and it rises in two different senses. The agents' policies settle earlier and revise less often. The median market stops changing its policy after 1,223,686 training steps at the highest friction tested, against 1,695,734 at the frictionless baseline, and the share of training steps at which the policy changes during learning falls from 21.6% to 4.5%. Because every price change has a cost attached to it, an agent that keeps revising its policy is penalised. With this learning also ends before the agents have covered as much of the price grid as in lower or no-friction scenarios.

The second is the structural form that colluded markets have. At the frictionless baseline 37.4% of markets sustain edgeworth cycles. That share falls steeply as friction rises and is effectively eliminated by $\phi \approx 0.088$. Cycling includes frequent repricing, so it is the behaviour that loses more because of the menu costs. Because of this the cycling behaviour gets removed rather than reshaped³⁶. Collusion therefore does not become less common under friction, it becomes less varied.

Together these two effects lower what firms earn without changing what they are doing. Mean profitability drops from 0.845 at the baseline to 0.661 at the highest friction tested, roughly a fifth of the collusive premium, and most of that loss travels through the learning channel rather than the fee itself. What firms actually pay in menu costs is a minor and eventually vanishing part of what the total profit loss is. Meaning what costs them is what they fail to learn while avoiding the fee in training.

Ultimately, neither of these prevented collusion. A friction that eliminates cycling entirely, reduces profit by about a fifth and shortens learning by more than a quarter still did not stop a market from reaching a supracompetitive outcome. Mean profitability at every friction level tested stayed well above the competitive benchmark.

This thesis introduced an explicit per-change price-adjustment (menu cost) friction into a multi-agent reinforcement learning model of tacit collusion, and that last result is the one with the clearest practical reading. A cost that firms in almost every real market already face did not prevent collusion here at any level tested, so real-world frictions should not be expected to contain algorithmic collusion on their own. Where the friction did most of its work was before convergence, which points to the learning phase as the stage at which an intervention has the most potential to act on.

Several questions follow directly from these results, and are left open for further research.

³⁶The markets that keep cycling under higher friction are those whose pricing strategies were already efficient enough to remain viable, not ones that adapted to the fee. Their punishment depth remains unchanged across friction tiers, and, even after accounting for menu costs, they earn less than static markets at every tier for which a comparison is possible.

The clearest concerns the form of the cost. Section 2.1 established that a fixed charge per change, a cost that is proportional to the size of the change, and a cost paid once per pricing occasion each produce different adjustment behaviour. This thesis implemented the fixed per-change form of the menu cost. Running the same simulation with one of the other cost forms would test whether the effects reported here are specific to a fixed per-change fee or hold across specifications.

A related question comes from keeping ϕ constant throughout the simulation. Stamatopoulos et al. (2021) showed that the cost of repricing also depends on the technology a firm uses. Adopting electronic shelf labels made price changes more frequent and smaller, which means a firm can choose to change what it costs to adjust a price. A model in which the menu cost could change would therefore answer a different question: how firms behave when the friction is something they can influence.

The decomposition in Section 4.5 also left an open question about the part of the profit loss that could not be attributed to the fee or the exploration channel. Because that component is substantial at high friction, identifying what produces it would be worth pursuing.

Exploring the region below $\phi = 0.04$ is also interesting, since no profit loss could be distinguished from ordinary variation there, leaving open whether a very small friction has no effect or a different one. At $\phi = 0.01$ the mean profitability index is marginally above the frictionless baseline (0.848 against 0.845, Table A.3), though the difference cannot be distinguished from ordinary variation at the tier spacing used here. A small friction preventing some unprofitable price wars is one possible explanation, but testing it would require a focused analysis in this region.

Two possible extensions to the setup are also worth mentioning.

Modelling more than two firms would test whether these effects survive in a less concentrated market.

Setting different menu costs between competitors could be a closer model of real markets, where firms of different size and technology do not face the same repricing cost.

Documentation of Tools and Aids Used

The simulation was written in Python. The environment follows the structural logic of the Gymnasium framework and the numerical work uses NumPy. Data aggregation and the statistical analysis were carried out with Pandas, SciPy and Statsmodels in a Jupyter notebook, and the figures were produced with Matplotlib and Seaborn. The code was written in Visual Studio Code with GitHub Copilot providing code completion and suggestions, and is version-controlled on GitHub.

Consensus, an AI-assisted literature search tool, was used alongside conventional database searches to identify candidate papers. Its role was limited to retrieval and suggestion. Which papers to use and their interpretation were my decisions. Zotero was used to manage references and produce the bibliography.

The thesis was written and formatted in Prism LaTeX editor. Claude (AI tool from Anthropic) was used as a writing and review assistant throughout. This covered suggestions on

how to structure sections, helping me review and revise my text formulations, grammatical checks, translation of the German abstract and LaTeX formatting. It was also used as a discussion partner, to test my own ideas against counterarguments and to find weaknesses in my reasoning before committing to an approach. This was the most frequent way I used it.

These tools were used strictly as aids and not as a substitute for my own work. The research question, the design of the model, the implementation of the simulation, the analysis of the results and the conclusions drawn are my own. The listed tools only supported how the work was carried out, not the final result presented in the thesis.

Bibliography

- Abada, I., & Lambin, X. (2023). Artificial Intelligence: Can Seemingly Collusive Outcomes Be Avoided? *Management Science*, 69(9), 5042–5065. <https://doi.org/10.1287/mnsc.2022.4623>
- Abada, I., Lambin, X., & Tchakarov, N. (2024). Collusion by mistake: Does algorithmic sophistication drive supra-competitive profits? *European journal of operational research*, 318, 953. <https://doi.org/10.1016/j.ejor.2024.06.006>
- Anderson, E. T., Jaimovich, N., & Simester, D. (2015). Price Stickiness: Empirical Evidence of the Menu Cost Channel. *Review of Economics and Statistics*, 97, 813–826. https://doi.org/10.1162/rest_a_00507
- Asker, J., Fershtman, C., & Pakes, A. (2024). The impact of artificial intelligence design on pricing. *Journal of Economics & Management Strategy*, 33(2), 276–304. <https://doi.org/10.1111/jems.12516>
- Assad, S., Calvano, E., Calzolari, G., Clark, R., Denicolò, V., Ershov, D., Johnson, J., Pastorello, S., Rhodes, A., Xu, L., & Wildenbeest, M. (2021). Autonomous algorithmic collusion: Economic research and policy implications. *Oxford review of economic policy*, 37, 478. <https://doi.org/10.1093/oxrep/grab011>
- Assad, S., Clark, R., Ershov, D., & Xu, L. (2024). Algorithmic Pricing and Competition: Empirical Evidence from the German Retail Gasoline Market. *The Journal of political economy*, 132, 771. <https://doi.org/10.1086/726906>
- Athey, S., Bagwell, K., & Sanchirico, C. (2004). Collusion and Price Rigidity. *The Review of Economic Studies*, 71, 349. <https://doi.org/10.1111/0034-6527.00286>
- Beneke, F., & Mackenrodt, M.-O. (2021). Remedies for algorithmic tacit collusion. *Journal of antitrust enforcement*, 9, 176. <https://doi.org/10.1093/jaenfo/jnaa040>
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B, Methodological*, 57, 300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
- Benjamini, Y., & Yekutieli, D. (2001). The Control of the False Discovery Rate in Multiple Testing under Dependency. *The Annals of statistics*, 29, 1188. <https://doi.org/10.1214/aos/1013699998>
- Bernhardt, L., & Dewenter, R. (2020). Collusion by code or algorithmic collusion? When pricing algorithms take over. *European competition journal*, 16, 342. <https://doi.org/10.1080/17441056.2020.1733344>
- Bichler, M., Durmann, J., & Oberlechner, M. (2025). Algorithmic Pricing and Algorithmic Collusion. *Business & Information Systems Engineering*, 67(6), 971–979. <https://doi.org/10.1007/s12599-025-00965-z>

- Bils, M., & Klenow, P. J. (2004). Some Evidence on the Importance of Sticky Prices. *The Journal of political economy*, 112, 985. <https://doi.org/10.1086/422559>
- Brown, Z. Y., & MacKay, A. (2023). Competition in Pricing Algorithms. *American economic journal. Microeconomics*, 15, 156. <https://doi.org/10.1257/mic.20210158>
- Calvano, E., Calzolari, G., Denicolò, V., & Pastorello, S. (2020). Artificial Intelligence, Algorithmic Pricing, and Collusion. *American Economic Review*, 110(10), 3267–3297. <https://doi.org/10.1257/aer.20190623>
- Calvano, E., Calzolari, G., Denicolò, V., & Pastorello, S. (2023). Algorithmic collusion: Genuine or spurious? *International journal of industrial organization*, 90. <https://doi.org/10.1016/j.ijindorg.2023.102973>
- Cameron, A. C., & Miller, D. L. (2015). A Practitioner’s Guide to Cluster-Robust Inference. *The Journal of human resources*, 50, 372. <https://doi.org/10.3368/jhr.50.2.317>
- Chakrabarti, R., & Scholnick, B. (2007). The mechanics of price adjustment: New evidence on the (un)importance of menu costs. *Manage. Decis. Econ*, 28, 668. <https://doi.org/10.1002/mde.1374>
- Cooley, B. J., & Cooley, J. W. (1980). Data Analysis For Simulation Experiments: Application Of a Distribution-Free Multiple Comparisons Pprocedure. *Decision sciences*, 11, 492. <https://doi.org/10.1111/j.1540-5915.1980.tb01153.x>
- den Boer, A., Meylahn, J. M., & Schinkel, M. P. (2026). Artificial Collusion: Examining Supracompetitive Pricing by Q-Learning Algorithms. *Management science*. <https://doi.org/10.1287/mnsc.2024.08557>
- Dutta, S., Bergen, M., Levy, D., & Venable, R. (1999). Menu Costs, Posted Prices, and Multiproduct Retailers. *Journal of money, credit and banking*, 31, 703. <https://doi.org/10.2307/2601217>
- Epivent, A., & Lambin, X. (2024). On algorithmic collusion and reward–punishment schemes. *Economics Letters*, 237, 111661–111661. <https://doi.org/10.1016/j.econlet.2024.111661>
- Feldman, J., & Segev, D. (2025). Dynamic Pricing with Menu Costs: Approximation Schemes and Applications to Grocery Retail. *Manufacturing & service operations management*, 27, 1106. <https://doi.org/10.1287/msom.2024.0832>
- Fourberg, N. (2023). Let’s lock them in: Collusion under consumer switching costs. *International journal of industrial organization*, 91. <https://doi.org/10.1016/j.ijindorg.2023.103018>
- Friedman, M. (1937). The Use of Ranks to Avoid the Assumption of Normality Implicit in the Analysis of Variance. *Journal of the American Statistical Association*, 32, 701. <https://doi.org/10.1080/01621459.1937.10503522>
- Garcia, D., Tolvanen, J., & Wagner, A. K. (2023). Strategic Responses to Algorithmic Recommendations: Evidence from Hotel Pricing. (10849). <https://hdl.handle.net/10419/282537>
- Garrod, L. (2012). Collusive price rigidity under price-matching punishments. *International journal of industrial organization*, 30, 482. <https://doi.org/10.1016/j.ijindorg.2012.03.003>
- Glasserman, P., & Yao, D. D. (1992). Some Guidelines and Guarantees for Common Random Numbers. *Management science*, 38, 908. <https://doi.org/10.1287/mnsc.38.6.884>
- Green, E. J., & Porter, R. H. (1984). Noncooperative Collusion under Imperfect Price Information. *Econometrica*, 52(1), 87–100. <https://doi.org/10.2307/1911462>
- Gu, C. (2026). Can dynamic pricing algorithm facilitate tacit collusion? An experimental study using deep reinforcement learning in airline revenue management: Can dynamic pricing algorithm facilitate tacit collusion? An experimental study. *Journal*

- of revenue and pricing management*, 25, 327. <https://doi.org/10.1057/s41272-025-00562-5>
- Hall, S., Walsh, M., & Yates, A. (2000). Are UK companies' prices sticky? *Oxf. Econ. Pap*, 52, 446. <https://doi.org/10.1093/oep/52.3.425>
- Hansen, K. T., Misra, K., & Pai, M. M. (2021). Frontiers: Algorithmic Collusion: Supra-competitive Prices via Independent Algorithms. *Marketing Science*, 40(1), 1–12. <https://doi.org/10.1287/mksc.2020.1276>
- Jiménez, J. L., & Perdiguero, J. (2012). Does Rigidity of Prices Hide Collusion? *Review of industrial organization*, 41, 248. <https://doi.org/10.1007/s11151-012-9337-9>
- Kano, K. (2013). Menu costs and dynamic duopoly. *International journal of industrial organization*, 31, 118. <https://doi.org/10.1016/j.ijindorg.2012.11.002>
- Klein, T. (2021). Autonomous algorithmic collusion: Q-learning under sequential pricing. *The RAND Journal of Economics*, 52(3), 538–558. <https://doi.org/10.1111/1756-2171.12383>
- Klemperer, P. (1987). Markets with Consumer Switching Costs. *The Quarterly Journal of Economics*, 102, 394. <https://doi.org/10.2307/1885068>
- Knotek, E. S. (2024). The roles of price points and menu costs in price rigidity. *Journal of monetary economics*, 145. <https://doi.org/10.1016/j.jmoneco.2024.103559>
- Levy, D., Bergen, M., Dutta, S., & Venable, R. (1997). The Magnitude of Menu Costs: Direct Evidence from Large U.S. Supermarket Chains. *The Quarterly journal of economics*, 112, 825. <https://doi.org/10.1162/003355397555352>
- Lewis, M., & Noel, M. (2011). The Speed Of Gasoline Price Response In Markets With And Without Edgeworth Cycles. *The review of economics and statistics*, 93, 682. https://doi.org/10.1162/REST_a_00176
- Li, C., & Virtosu, I. (2025). Competition law in the age of AI: Confronting algorithmic collusion in the smart economy. *Smart Cities and Regional Development Journal*, 9, 54. <https://doi.org/10.25019/qnjgmk44>
- Liang, K., & Zeger, S. (1986). Longitudinal Data-Analysis Using Generalized Linear-Models. *Biometrika*, 73, 22. <https://doi.org/10.2307/2336267>
- Loots, T., & denBoer, A. V. (2023). Data-driven collusion and competition in a pricing duopoly with multinomial logit demand. *Production and operations management*, 32, 1186. <https://doi.org/10.1111/poms.13919>
- MacKinnon, D. P., Lockwood, C. M., & Williams, J. (2004). Confidence Limits for the Indirect Effect: Distribution of the Product and Resampling Methods. *Multivariate Behav Res*, 39, 128. <https://doi.org/10.1207/s15327906mbr3901-4>
- Mankiw, N. G. (1985). Small Menu Costs and Large Business Cycles: A Macroeconomic Model of Monopoly. *The Quarterly Journal of Economics*, 100(2), 529–538. <https://doi.org/10.2307/1885395>
- Mann, H. B., & Whitney, D. R. (1947). On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. *The Annals of mathematical statistics*, 18, 60. <https://doi.org/10.1214/aoms/1177730491>
- Maskin, E., & Tirole, J. (1988). A Theory of Dynamic Oligopoly, II: Price Competition, Kinked Demand Curves, and Edgeworth Cycles. *Econometrica*, 56(3), 571–599. <https://doi.org/10.2307/1911701>
- Meylahn, J. M., & V. den Boer, A. (2022). Learning to Collude in a Pricing Duopoly. *Manufacturing & service operations management*, 24, 2594. <https://doi.org/10.1287/msom.2021.1074>
- Miklós-Thal, J., & Tucker, C. (2019). Collusion by Algorithm: Does Better Demand Prediction Facilitate Coordination Between Sellers? *Management science*, 65, 1561. <https://doi.org/10.1287/mnsc.2019.3287>

- Mohr, D. L., & Marcon, R. A. (2005). Testing for a 'within-subjects' association in repeated measures data. *Journal of nonparametric statistics*, 17, 363. <https://doi.org/10.1080/10485250500038694>
- Nakamura, E., & Steinsson, J. (2008). Five Facts about Prices: A Reevaluation of Menu Cost Models. *The Quarterly Journal of Economics*, 123, 1464. <https://doi.org/10.1162/qjec.2008.123.4.1415>
- Noel, M. (2009). Do retail gasoline prices respond asymmetrically to cost shocks? The influence of Edgeworth Cycles. *The Rand journal of economics*, 40, 595. <https://doi.org/10.1111/j.1756-2171.2009.00078.x>
- Noel, M. D. (2007a). Edgeworth Price Cycles, Cost-based Pricing, and Sticky Pricing in Retail Gasoline Markets. *The review of economics and statistics*, 89, 334. <https://doi.org/10.1162/rest.89.2.324>
- Noel, M. D. (2007b). Edgeworth Price Cycles: Evidence from the Toronto Retail Gasoline Market. *The Journal of industrial economics*, 55, 92. <https://doi.org/10.1111/j.1467-6451.2007.00303.x>
- Rotemberg, J. J. (1982). Sticky Prices in the United States. *Journal of Political Economy*, 90(6), 1187–1211. <http://www.jstor.org/stable/1830944>
- Sanchez-Cartas, J. M., & Katsamakos, E. (2022). Artificial Intelligence, Algorithmic Competition and Market Structures. *IEEE access*, 10, 10584. <https://doi.org/10.1109/ACCESS.2022.3144390>
- Schwalbe, U. (2018). Algorithms, Machine Learning, and Collusion. *Journal of competition law & economics*, 14, 607. <https://doi.org/10.1093/joclec/nhz004>
- Seele, P., Dierksmeier, C., Hofstetter, R., & Schultz, M. D. (2021). Mapping the Ethicality of Algorithmic Pricing: A Review of Dynamic and Personalized Pricing. *Journal of business ethics*, 170, 719. <https://doi.org/10.1007/s10551-019-04371-w>
- Setianingrum, R. B. (2025). Artificial Intelligence and Antitrust Regulation: Lessons from the European Union, United States and China. *Indonesian comparative law review*, 8, 60. <https://doi.org/10.18196/iclr.v8i1.28832>
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27, 423. <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>
- Sheshinski, E., & Weiss, Y. (1977). Inflation and Costs of Price Adjustment. *The Review of Economic Studies*, 44(2), 287–303. <https://doi.org/10.2307/2297067>
- Spearman, C. (1987). The proof and measurement of association between two things. By C. Spearman, 1904. *Am J Psychol*, 100, 471. <https://doi.org/10.2307/1422689>
- Stamatopoulos, I., Bassamboo, A., & Moreno, A. (2021). The Effects of Menu Costs on Retail Performance: Evidence from Adoption of the Electronic Shelf Label Technology. *Management science*, 67, 256. <https://doi.org/10.1287/mnsc.2019.3551>
- Stella, A. (2020). Economies of scope in price setting: A moment inequalities estimation. *Journal of monetary economics*, 110, 61. <https://doi.org/10.1016/j.jmoneco.2018.12.006>
- Tesauro, G., & Kephart, J. O. (2002). Pricing in Agent Economies Using Multi-Agent Q-Learning. *Autonomous agents and multi-agent systems*, 5, 304. <https://doi.org/10.1023/A:1015504423309>
- Waltman, L., & Kaymak, U. (2008). Q-learning agents in a Cournot oligopoly model. *Journal of Economic Dynamics and Control*, 32(10), 3275–3293. <https://doi.org/10.1016/j.jedc.2008.01.003>
- Wilcoxon, F. (1945). Individual Comparisons by Ranking Methods. *Biometrics bulletin*, 1, 83. <https://doi.org/10.2307/3001968>
- Zbaracki, M. J., Ritson, M., Levy, D., Dutta, S., & Bergen, M. (2004). Managerial and Customer Costs of Price Adjustment: Direct Evidence from Industrial Mar-

A Appendix: Supplementary Tables

Table A.1: Continuous Profit Degradation: Wilcoxon Signed-Rank Test vs. Baseline

Menu Cost (ϕ)	N	Med. Δ (Base)	Med. Δ (Fric.)	Med. Diff.	p (raw)	p (BH)	Sig. ^a
0.010	1000	0.868	0.868	0.000	0.600	0.631	False
0.020	1000	0.868	0.866	0.001	0.976	0.976	False
0.030	1000	0.868	0.860	0.008	0.463	0.514	False
0.040	1000	0.868	0.848	0.019	0.008	0.010	True
0.050	1000	0.868	0.852	0.016	0.017	0.020	True
0.060	1000	0.868	0.828	0.039	1.788e-05	2.554e-05	True
0.070	1000	0.868	0.828	0.039	2.703e-06	4.159e-06	True
0.080	1000	0.868	0.828	0.039	5.124e-05	6.833e-05	True
0.090	1000	0.868	0.828	0.039	1.323e-08	2.205e-08	True
0.100	1000	0.868	0.828	0.039	2.441e-13	4.438e-13	True
0.110	1000	0.868	0.794	0.074	2.642e-21	5.871e-21	True
0.120	1000	0.868	0.794	0.074	2.673e-20	5.346e-20	True
0.130	1000	0.868	0.794	0.074	2.127e-29	5.317e-29	True
0.140	1000	0.868	0.794	0.074	4.670e-30	1.334e-29	True
0.150	1000	0.868	0.794	0.074	1.961e-33	6.537e-33	True
0.200	1000	0.868	0.747	0.120	5.655e-55	2.262e-54	True
0.250	1000	0.868	0.747	0.120	1.087e-72	5.436e-72	True
0.350	1000	0.868	0.707	0.161	8.081e-97	5.387e-96	True
0.500	1000	0.868	0.654	0.214	3.888e-123	3.888e-122	True
0.700	1000	0.868	0.654	0.214	1.045e-128	2.089e-127	True

^a Significant at the 5% level after Benjamini-Hochberg correction.

Table A.2: Static vs. Cyclical Statistical tests Full Sweep ($\phi \leq 0.13$)

Menu Cost (ϕ)	N Cyc.	N Stat.	Gross Prem.	p_G^a	Sig. _G ^b	Net Prem.	p_N^a	Sig. _N ^b
0.000	374	626	-0.040	0.000	True	-0.040	0.000	True
0.010	331	669	-0.020	0.001	True	-0.029	0.000	True
0.020	271	729	-0.029	0.000	True	-0.046	0.000	True
0.030	223	777	-0.009	0.139	False	-0.034	0.000	True
0.040	178	822	-0.016	0.056	False	-0.050	0.000	True
0.050	141	859	0.003	0.650	False	-0.037	0.000	True
0.060	108	892	0.029	0.150	False	-0.019	0.027	True
0.070	84	916	0.024	0.252	False	-0.032	0.014	True
0.080	67	933	0.013	0.663	False	-0.046	0.003	True
0.090	62	938	0.034	0.146	False	-0.034	0.033	True
0.100	31	969	0.067	0.017	True	-0.008	0.679	False
0.110	31	969	0.046	0.139	False	-0.047	0.033	True
0.120	21	979	0.070	0.056	False	-0.020	0.497	False
0.130	10	990	0.063	0.176	False	-0.038	0.380	False

^a BH-corrected p-value.

^b Significant at the 5% level after BH correction. G = Gross, N = Net

Table A.3: Macroeconomic and Computational Taxonomy of Algorithmic Pricing under Friction

Menu Cost (ϕ)	Runs (N)	Convergence (%)	Static (N)	Static (%)	Cycler (N) ^d	Cycler (%)	Global Mean Δ^a	Global Median Δ^a	Static Mean Δ^a	Static Median Δ^a	Cycler Mean Δ^a	Cycler Median Δ^a	Cycler Mean Net Δ^e	Cycler Premium Mean Δ	Cycler Premium Median Δ	Cycler Premium Mean Net Δ^f	Median Freeze Step ^b	Median Training Price Changes ^b	Median Price Changes per Freeze Step ^c	Median Eval-Phase Price Change Rate ^g	Median Entropy (bits) ^h
0.000	1,000	100%	626	62.6%	374	37.4%	0.845	0.868	0.860	0.895	0.820	0.835	0.820	-0.040	-0.060	-0.040	1,695,734	1,090,983	0.6629	1.0000	5.59
0.010	1,000	100%	669	66.9%	331	33.1%	0.848	0.868	0.854	0.868	0.834	0.845	0.825	-0.020	-0.023	-0.029	1,684,678	1,037,046	0.6314	1.0000	5.52
0.020	1,000	100%	729	72.9%	271	27.1%	0.844	0.866	0.852	0.868	0.823	0.830	0.805	-0.029	-0.037	-0.046	1,661,277	994,123	0.6135	1.0000	5.51
0.030	1,000	100%	777	77.7%	223	22.3%	0.840	0.860	0.842	0.868	0.833	0.847	0.808	-0.009	-0.021	-0.034	1,677,284	956,140	0.5874	1.0000	5.37
0.040	1,000	100%	822	82.2%	178	17.8%	0.832	0.848	0.835	0.858	0.819	0.825	0.784	-0.016	-0.033	-0.050	1,664,494	915,524	0.5669	1.0000	5.31
0.050	1,000	100%	859	85.9%	141	14.1%	0.831	0.852	0.831	0.852	0.834	0.844	0.794	+0.003	-0.007	-0.037	1,655,891	890,054	0.5541	1.0000	5.27
0.060	1,000	100%	892	89.2%	108	10.8%	0.819	0.828	0.816	0.828	0.845	0.857	0.797	+0.029	+0.029	-0.019	1,648,386	866,205	0.5439	1.0000	5.20
0.070	1,000	100%	916	91.6%	84	8.4%	0.819	0.828	0.817	0.828	0.841	0.864	0.785	+0.024	+0.036	-0.032	1,632,985	844,197	0.5356	0.9750	5.20
0.080	1,000	100%	933	93.3%	67	6.7%	0.823	0.828	0.822	0.828	0.836	0.853	0.776	+0.013	+0.025	-0.046	1,620,618	815,294	0.5187	0.8333	5.10
0.090	1,000	100%	938	93.8%	62	6.2%	0.812	0.828	0.809	0.828	0.844	0.843	0.776	+0.034	+0.014	-0.034	1,628,026	799,011	0.5064	0.8333	4.99
0.100	1,000	100%	969	96.9%	31	3.1%	0.802	0.828	0.800	0.828	0.867	0.889	0.792	+0.067	+0.061	-0.008	1,593,987	778,362	0.4972	0.8333	4.99
0.110	1,000	100%	969	96.9%	31	3.1%	0.789	0.794	0.787	0.794	0.833	0.844	0.740	+0.046	+0.050	-0.047	1,594,076	759,709	0.4911	1.0000	4.90
0.120	1,000	100%	979	97.9%	21	2.1%	0.792	0.794	0.790	0.794	0.861	0.885	0.771	+0.070	+0.091	-0.020	1,605,216	742,940	0.4757	0.7500	4.83
0.130	1,000	100%	990	99.0%	10	1.0%	0.780	0.794	0.779	0.794	0.843	0.846	0.742	+0.063	+0.052	-0.038	1,586,865	730,341	0.4719	0.8542	4.80
0.140	1,000	100%	994	99.4%	6	0.6%	0.778	0.794	0.777	0.794	0.820	0.832	0.718	+0.043	+0.038	-0.059	1,575,116	711,757	0.4608	0.6875	4.75
0.150	1,000	100%	996	99.6%	4	0.4%	0.771	0.794	0.771	0.794	0.889	0.909	0.762	+0.118	+0.115	-0.008	1,578,094	700,939	0.4534	0.9375	4.69
0.200	1,000	100%	999	99.9%	1	0.1%	0.751	0.747	0.750	0.747	0.898	0.898	0.698	+0.148	+0.151	-0.052	1,544,998	643,491	0.4220	1.0000	4.45
0.250	1,000	100%	1,000	100%	0	0.0%	0.736	0.747	0.736	0.747	—	—	—	—	—	—	1,537,832	596,876	0.3883	—	4.18
0.350	1,000	100%	1,000	100%	0	0.0%	0.707	0.707	0.707	0.707	—	—	—	—	—	—	1,433,282	514,237	0.3565	—	3.80
0.500	1,000	100%	1,000	100%	0	0.0%	0.682	0.654	0.682	0.654	—	—	—	—	—	—	1,295,680	427,842	0.3369	—	3.46
0.700	1,000	100%	1,000	100%	0	0.0%	0.661	0.654	0.661	0.654	—	—	—	—	—	—	1,223,686	387,976	0.3216	—	3.35

^a Δ is moderately skewed, so both mean and median are shown. The mean is the main figure used.

^b These two are more strongly skewed, so only the median is shown.

^c Calculation: Each market's own price changes divided by its own freeze step, then the median is calculated across markets. Simply dividing the two columns to the left would give a different and incorrect number, because it would mix the training length between markets.

^d A dash means no cycling markets remained at that tier, which is the case from $\phi = 0.25$ onward. At $\phi = 0.20$ only one remains, so that row's cycler figures rest on a single market.

^e Gross Δ minus the menu costs the market actually paid. Statics have no such column: once converged they stop changing price, so their net and gross values are identical.

^f The cycler premium recalculated on net rather than gross Δ , showing whether cyclers still come out ahead once their fees are subtracted.

^g How often price changes within a market's converged cycle, with cyclers only. Strongly skewed, so the median is shown. This describes converged behaviour.

^h How widely a market's learning spread across the 225 price combinations. The highest possible value is $\log_2(225) \approx 7.81$ bits.

Table A.4: Punishment Depth: Baseline Cyclers compared to Cyclers at Each Tier Full ($\phi \leq 0.13$)

Menu Cost (ϕ)	N Cyclers	Mean Depth	p (raw)	p (BH)	Significant (BH-corrected)
0.010	331	0.217	0.692	0.743	False
0.020	271	0.233	0.589	0.743	False
0.030	223	0.205	0.359	0.706	False
0.040	178	0.231	0.435	0.706	False
0.050	141	0.183	0.113	0.615	False
0.060	108	0.181	0.162	0.615	False
0.070	84	0.204	0.237	0.615	False
0.080	67	0.216	0.743	0.743	False
0.090	62	0.214	0.726	0.743	False
0.100	31	0.179	0.217	0.615	False
0.110	31	0.260	0.405	0.706	False
0.120	21	0.217	0.610	0.743	False
0.130	10	0.124	0.148	0.615	False

Note. There are not enough cyclers above $\phi = 0.13$ to test reliably.

Table A.5: Training Phase Price Change Rate Spearman Correlation with Final Policy Entropy by Friction Tier

Menu Cost (ϕ)	N	Spearman ρ	p
0.00	1000	0.786	2.749e-210
0.01	1000	0.814	5.350e-238
0.02	1000	0.858	2.430e-291
0.03	1000	0.868	< 1e-300
0.04	1000	0.887	< 1e-300
0.05	1000	0.907	< 1e-300
0.06	1000	0.904	< 1e-300
0.07	1000	0.919	< 1e-300
0.08	1000	0.925	< 1e-300
0.09	1000	0.934	< 1e-300
0.10	1000	0.938	< 1e-300
0.11	1000	0.920	< 1e-300
0.12	1000	0.939	< 1e-300
0.13	1000	0.927	< 1e-300
0.14	1000	0.932	< 1e-300
0.15	1000	0.934	< 1e-300
0.20	1000	0.898	< 1e-300
0.25	1000	0.873	< 1e-300
0.35	1000	0.749	2.422e-180
0.50	1000	0.500	3.088e-64
0.70	1000	0.555	5.997e-82

Table A.6: Median Policy Volatility (Greedy Inertia Rate) by Friction Tier

Menu Cost (ϕ)	Median Policy Volatility
0.00	0.2162
0.01	0.2138
0.02	0.2146
0.03	0.2084
0.04	0.2060
0.05	0.2044
0.06	0.2015
0.07	0.2003
0.08	0.1943
0.09	0.1892
0.10	0.1869
0.11	0.1825
0.12	0.1751
0.13	0.1734
0.14	0.1674
0.15	0.1646
0.20	0.1412
0.25	0.1206
0.35	0.0913
0.50	0.0628
0.70	0.0447

Table A.7: Three-Part Causal Decomposition of Net Profit Loss

Menu Cost (ϕ)	Total Profit Loss	Real Fee	Entropy-Mediated Loss	Unexplained Residual	% Mediated	Sig.
0.010000	0.000600	0.002900	0.001800	-0.004000	284.415900	False
0.020000	0.006400	0.004700	0.003700	-0.002100	57.919600	False
0.030000	0.010900	0.005600	0.008200	-0.003000	75.532400	True
0.040000	0.019700	0.006100	0.011900	0.001600	60.631500	True
0.050000	0.019700	0.005700	0.013900	-0.000000	70.918000	True
0.060000	0.031400	0.005200	0.017100	0.009100	54.302100	True
0.070000	0.030900	0.004700	0.018000	0.008200	58.156900	True
0.080000	0.026300	0.004000	0.021800	0.000500	82.903600	True
0.090000	0.037900	0.004200	0.025300	0.008300	66.874200	True
0.100000	0.045600	0.002300	0.027400	0.015900	59.997800	True
0.110000	0.059500	0.002900	0.030600	0.026000	51.412400	True
0.120000	0.055300	0.001900	0.035000	0.018400	63.276200	True
0.130000	0.066500	0.001000	0.036100	0.029300	54.367900	True
0.140000	0.068400	0.000600	0.038700	0.029000	56.643400	True
0.150000	0.074600	0.000500	0.040900	0.033200	54.809600	True
0.200000	0.094900	0.000200	0.052100	0.042600	54.881600	True
0.250000	0.109100	0.000000	0.064900	0.044200	59.499100	True
0.350000	0.138600	0.000000	0.083700	0.055000	60.348800	True
0.500000	0.163800	0.000000	0.102000	0.061800	62.253800	True
0.700000	0.183900	0.000000	0.108400	0.075500	58.930700	True

Table A.8: Bootstrap 95% Confidence Intervals: Entropy-Mediated Loss and Percent Mediated

Menu Cost (ϕ)	Loss CI Lo. ^a	Loss CI Hi. ^a	Excl. 0	N Resamp.	% Med. CI Lo.	% Med. CI Hi.
0.010000	-0.000500	0.004100	False	1000	-23.317200	918.947300
0.020000	0.001100	0.006200	True	1000	14.735700	645.371300
0.030000	0.005800	0.010800	True	1000	37.438800	320.035900
0.040000	0.009200	0.014600	True	1000	38.850300	125.627900
0.050000	0.011500	0.016500	True	1000	47.312300	149.082000
0.060000	0.014600	0.019900	True	1000	40.351800	87.000800
0.070000	0.015300	0.020700	True	1000	42.408200	90.756300
0.080000	0.019100	0.024300	True	1000	59.949300	139.372600
0.090000	0.022700	0.028000	True	1000	51.446000	95.942200
0.100000	0.024400	0.030300	True	1000	48.078400	80.084200
0.110000	0.027500	0.033500	True	1000	42.799200	64.229000
0.120000	0.031900	0.038000	True	1000	53.508800	78.969600
0.130000	0.033000	0.039300	True	1000	46.776400	64.972200
0.140000	0.035400	0.042200	True	1000	47.984400	68.349700
0.150000	0.037600	0.044300	True	1000	47.212900	65.267800
0.200000	0.048300	0.055800	True	1000	48.669200	62.300400
0.250000	0.060500	0.069200	True	1000	53.901600	66.631700
0.350000	0.078400	0.088800	True	1000	55.353300	66.474700
0.500000	0.095900	0.108300	True	1000	57.460700	67.481000
0.700000	0.101800	0.115200	True	1000	54.505100	63.789600

^a Entropy-mediated loss, 95% bootstrap CI ($N = 1000$ resamples, cluster bootstrap by seed).

A.1 Simulation Parameters and Code Mapping

The following tables map the theoretical economic and algorithmic parameters to their programmatic implementation in the simulation environment.

Table A.9: Economic environment parameters and code mapping

Parameter	Explanation	Python variable	Comments	Baseline value
m	Number of discrete prices available	<code>self.m / m_prices</code>	Defines the size of the action space and Q-table.	15
k	Memory length	Implied by the observation space	Agents observe only the previous period's price choices.	1
ξ	Grid-buffer parameter	<code>config - xi</code>	Adds padding above the monopoly price and below the Nash price, allowing exploration of severe undercutting and high-price strategies.	0.1
p^N	Bertrand–Nash equilibrium price	<code>p_nash</code>	Competitive benchmark calculated dynamically using a SciPy optimizer.	Approx. 1.47
p^M	Monopoly price	<code>p_monopoly</code>	Joint-profit-maximizing price calculated dynamically	Approx. 1.92
a_i	Product-quality index	<code>self.a / a_quality</code>	The firms are symmetric and therefore have the same baseline quality.	2.0
a_0	Quality of the outside good	<code>self.a0 / a0_outside_good</code>	Baseline utility obtained when a consumer exits the market.	0.0
μ	Index of horizontal differentiation	<code>self.mu / mu_differentiation</code>	Lower values imply closer substitutes and more price competition.	0.25
c_i	Marginal cost	<code>self.c / c_marginal_cost</code>	Constant and symmetric cost of producing one unit.	1.0

Note. Approximate equilibrium prices are generated numerically and may vary with the precise implementation and numerical tolerance.

Table A.10: Learning parameters and code mapping

Parameter	Explanation	Python variable	Comments	Baseline value
α	Learning rate	<code>config - alpha</code>	Determines the weight assigned to new information relative to existing Q-values in the Bellman update.	0.15
δ	Discount factor	<code>config - gamma</code>	Determines the weight assigned to expected future rewards. Reinforcement-learning literature conventionally uses γ Calvano et al. (2020) use δ .	0.95
β	Exploration-decay rate	<code>config - beta</code>	Exponential decay parameter in $\epsilon_t = e^{-\beta t}$, determining how quickly exploration approaches zero.	4×10^{-6}
ϕ	Menu cost penalty fraction	<code>config - menu_cost_fraction</code>	Extension introduced in this thesis. The cost of changing price is scaled proportionally to the collusive profit premium.	0.0 (baseline)