Big Data Project: Revisiting Hotelling's Spatial Competition with Reinforcement Learning

PIERRON, Marie
BACHER, Quentin
PRATA LEAL, Paul
SKOVGAARD, Anémone

January 20, 2026

1 Introduction

This paper studies spatial competition between two firms using a reinforcement learning framework. We start from the classical Hotelling (1929) model with two firms, *Ben* and *Jerry*, competing in location space under fixed prices and linear transportation costs. We then extend the model to better match modern markets by introducing (i) a discrete location grid, (ii) "beam" relocation where a firm can jump to any grid position in one step, (iii) customer loyalty modeled as customer mass that follows a firm when it moves, (iv) stochastic demand shocks, and (v) relocation frictions implemented as quadratic movement costs together with an opportunity cost of moving (no sales when relocating). We explore the resulting dynamics using two agents trained simultaneously with Q-learning.

2 Basic Hotelling Model

The project is based on the basic Hotelling model with the following assumptions:

- Consumers are uniformly distributed on the interval [0,1], and each consumer purchases exactly one unit (unit demand).
- Firms choose locations $x_i \in [0,1]$ on a continuous line.
- There are two sellers, Ben and Jerry, indexed by $i \in \{1,2\}$, choosing locations $x_i \in \mathcal{X}$.
- Prices are fixed and identical across firms: $p = 1$.
- Consumers incur a linear transportation cost $t|x - x_i|$, where x denotes the consumer's location.
- The utility of a consumer located at x when buying from firm i is:

$$U_i(x) = v - p - t|x - x_i|\,,$$

  where v is sufficiently large to ensure full market coverage.

Market Boundary and Demand

Under the assumption that $x_1 < x_2$, the indifferent consumer $\hat{x}$ is defined by:

$$1 - t|\hat{x} - x_1| = 1 - t|\hat{x} - x_2|$$

Solving yields:

$$\hat{x} = \frac{x_1 + x_2}{2}$$

Firm demands are therefore:

$$D_1 = \hat{x}, \quad D_2 = 1 - \hat{x}$$

Since prices are fixed, firm profits are proportional to demand:

$$\pi_1 = pD_1 = \frac{x_1 + x_2}{2}$$

$$\pi_2 = p(1 - D_1)$$

Solution Concept

The classical Hotelling model is solved using a Nash equilibrium in pure strategies over location space. Firms choose locations that maximize profits given the location of the rival firm. Under these assumptions, both firms have an incentive to move closer to the center of the market. The unique Nash equilibrium is:

$$x_1^* = x_2^* = \frac{1}{2}$$

This outcome is known as the *principle of minimum differentiation.*

3 Extended Model: Beam Relocation, Demand Shocks, Loyalty, and Relocation Costs

We extend the classical model in the following ways:

- **Discrete state space:** locations lie on the grid $\mathcal{X}$ (11 positions).
- **Relocation (beam):** relocation allows a firm to move directly to any grid position in a single step ("beam" relocation).
- **Customer loyalty:** a fraction of customers follow a seller after relocation.
- **Relocation costs:** moving uses the period and yields zero profit (opportunity cost) as well as entails a quadratic cost in physical distance.
- **Stochastic shocks:** Instead of uniform consumers, the market is represented by a probability vector of customer mass over grid positions.

Beam relocation

The market is discretized into a finite set of 11 locations,

$$\mathcal{X} = \left\{0,\tfrac{1}{10},\tfrac{2}{10},\ldots,1\right\}.$$

In each time step, each firm chooses a grid *index*

$$a_i(t) \in \{0,1,\ldots,10\}, \qquad x_i(t+1) = \frac{a_i(t)}{10}.$$

In the implementation, grid index k is mapped to physical distance on a beach of length 100 via

$$\phi(k) = k \cdot \frac{100}{10}.$$

This allows both small adjustments and long-distance "beam" relocations in a single step.

Loyalty as Customer Mass Transport

Customer loyalty is implemented as a spatial transport mechanism rather than as intertemporal demand smoothing. When a firm relocates from position $x_i(t)$ to $x_i(t+1)$, a fraction $\lambda_i(t) \in [0,1]$ of the customer mass located at $x_i(t)$ follows the firm to its new location. Formally, let $w_t(k)$ denote the customer density at grid position k. If firm i relocates, the density updates as:

$$w_t(x_i(t)) \leftarrow w_t(x_i(t)) - \lambda_i(t) w_t(x_i(t)),$$

$$w_t(x_i(t+1)) \leftarrow w_t(x_i(t+1)) + \lambda_i(t) w_t(x_i(t)).$$

The loyalty rate $\lambda_i(t)$ is agent-specific and depends on tenure, defined as the number of consecutive periods the firm has remained at its current location. Different loyalty strategies map tenure into retention rates, allowing firms to differ in brand strength.

Quadratic relocation costs and mutual exclusion

Relocation entails a quadratic cost in physical distance:

$$C_i(t) = c \cdot \left|\phi(x_i(t+1)) - \phi(x_i(t))\right|^2,$$

with scaling parameter $c =$ `cost_scaling`. Additionally, the environment imposes mutual exclusion: a firm that moves does not sell in that step. Therefore, if firm i moves at time t, its revenue component is zero and its reward is purely the negative relocation cost:

$$r_i(t) = -C_i(t) \quad \text{if } x_i(t+1) \neq x_i(t).$$

If the firm stays, it earns a revenue term proportional to market share, scaled by a stochastic multiplier to represent demand intensity:

$$r_i(t) = S_i(t) \cdot M_t \quad \text{if } x_i(t+1) = x_i(t),$$

where

$$M_t = \max\{1, \mathcal{N}(\mu, \sigma_M^2)\},$$

with (μ, σ_M) corresponding to `reward_mean` and `reward_std`. Thus the overall implemented per-step reward is:

$$r_i(t) = \begin{cases} -S_i\text{-independent:} \ -C_i(t), & \text{if move,} \\ S_i(t) \cdot M_t, & \text{if stay.} \end{cases}$$

Market demand as a weighted crowd with stochastic shocks

Instead of uniform consumers, the market is represented by a probability vector of customer mass over grid positions:

$$w_t = (w_t(0), \ldots, w_t(N-1)), \qquad \sum_{k=0}^{N-1} w_t(k) = 1,\ w_t(k) \geq 0.$$

At each step, the distribution is perturbed by a demand shock anchored to a baseline distribution w^0:

$$\tilde{w}_t(k) = w^0(k) \cdot \big(1 + \varepsilon_t(k)\big), \qquad \varepsilon_t(k) \sim \mathcal{N}(0, \sigma^2),$$

followed by clipping at zero and renormalization to sum to one. The parameter σ corresponds to `demand_volatility` in the code. Importantly, shocks are applied relative to a fixed baseline distribution w^0, ensuring that the market retains a stable structural shape while remaining stochastic.

Dynamic objective

Agents learn policies that maximize discounted expected rewards:

$$\mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t r_i(t)\right],$$

where $\gamma \in (0,1)$ is the discount factor used in Q-learning.

Implications for the solution concept

In the classical one-shot model, moving toward the center increases profits immediately. In the extended model, relocation implies negative reward in the current period and affects only future payoffs. A move is attractive only if the discounted future gain compensates the current loss:

$$\gamma\, \Delta\pi \geq \pi_{\text{current}},$$

where $\Delta\pi$ denotes the expected increase in future profits from relocating and π_{current} is the forgone profit from staying. Loyalty further changes incentives because a relocating firm can transport part of the customer mass at its previous position to the new position. This reduces the downside of moving and can create path dependence, since early relocations can reshape the future customer density landscape. Because the environment is dynamic and stochastic, the relevant solution concept shifts from a one-shot Nash equilibrium to equilibrium in a Markov game (e.g. Markov perfect equilibrium). In practice, we use reinforcement learning to approximate effective policies.

Why reinforcement learning

In this extended Hotelling setting, firms face a dynamic decision problem with stochastic state transitions and delayed rewards. Optimal behavior depends on expectations about future states and the opponent's evolving strategy, making analytical solutions difficult. Reinforcement learning provides a natural framework: agents learn policies through repeated interaction, updating actions based on realized rewards without requiring full knowledge of the transition kernel P or the rival's strategy. This makes RL well suited for studying spatial competition with adjustment frictions, stochastic relocation, and loyalty.

Discrete location space

For implementation with reinforcement learning, the continuous location space is discretized into a finite grid:

$$\mathcal{X} = \left\{0, \frac{1}{10}, \frac{2}{10}, \ldots, 1\right\}.$$

This discretization makes the state space finite and enables the use of standard multi-agent reinforcement learning algorithms, while preserving the spatial intuition of the Hotelling model.

Experimental design: fluid vs. rigid markets

To isolate the effect of relocation frictions, we run two scenarios that differ only in the movement cost scaling c:

- **Fluid market (low cost):** small c encourages frequent relocation.

- **Rigid market (high cost):** large c discourages relocation, generating persistence.

We additionally vary loyalty strategies across agents to compare which retention regime yields higher accumulated payoffs.

4 Implementation

To empirically test the extended Hotelling model, we developed a simulation framework using Python and the Gymnasium library. The implementation is divided into three core modules: the environment dynamics (`hotelling_env.py`), the reinforcement learning agents (`hotelling_agent.py`), and a visualization engine (`hotelling_viz.py`) to analyze the spatial evolution of the market.

Environment Architecture (ExtendedHotellingEnv)

The environment is modeled as a discrete state space where the "beach" is divided into $N = 11$ distinct positions, indexed by $\{0, \ldots, 10\}$.

State Space. The state at time t, denoted S_t, is observed as a tuple

$$S_t = (x_{\text{Ben}}, x_{\text{Jerry}}),$$

representing the current discrete locations of both firms.

Action Space. The action space is discrete and defined as

$$A = \{0, 1, \ldots, 10\}.$$

Unlike the classical Hotelling model where movement is continuous or incremental, agents in our framework may "beam" (teleport) to any location on the grid in a single step, subject to relocation costs.

Q-Learning Agents

We employ independent Q-learning agents for Ben and Jerry. The agents do not share information and learn exclusively through interaction with the environment.

Q-Table. Each agent maintains a Q-table of dimensions $(11 \times 11 \times 11)$, corresponding to

$$(\text{Own}_{\text{pos}}, \text{Opp}_{\text{pos}}, \text{Action}).$$

Learning Rule. Expected action values are updated using the Bellman equation:

$$Q(s,a) \leftarrow Q(s,a) + \alpha \left[r + \gamma \max_{a'} Q(s', a') - Q(s,a) \right].$$

The learning rate is set to $\alpha = 0.2$, the discount factor to $\gamma = 0.95$, and the exploration rate ϵ decays linearly from 1.0 to 0.05.

Loyalty Strategies. Agents are polymorphic and support multiple loyalty profiles (e.g., `HighBrandLoyalty`, `ModerateBrandLoyalty`, `WeakBrandLoyalty`). These profiles determine the transport rate ℓ, which depends on the agent's tenure at a given location.

Training and Visualization

The training loop implemented in `hotelling_train.py` consists of two distinct phases:

- **Exploration (Training):** Agents follow an ϵ-greedy policy for 5,000 episodes to populate their Q-tables.
- **Exploitation (Validation):** The exploration rate is set to $\epsilon = 0$, and agents play 30 games to demonstrate converged behavior.

A specialized `HotellingVisualizer` generates heatmaps of crowd density evolution and trajectory plots of agent locations. These visual diagnostics allow us to assess whether the Minimum Differentiation Principle holds or fails under the extended model's constraints.

Experimental Scenarios

To isolate the effect of friction on spatial equilibrium, we define two experimental scenarios within the training script (`hotelling_train.py`). In both cases, loyalty is held constant using a "Fair" configuration in which both agents employ `HighBrandLoyalty`. Consequently, observed differences in outcomes are driven solely by relocation costs.

Scenario 1: The Fluid Market (Low Friction).

- **Parameter:** `cost_scaling` = 0.001.
- **Economic Interpretation:** This scenario represents a market with negligible barriers to movement, such as pop-up retail or digital services.
- **Hypothesis:** We expect high volatility and frequent relocation as agents pursue marginal gains in market share, potentially approximating the classical Hotelling prediction of centripetal convergence.

Scenario 2: The Rigid Market (High Friction).

- **Parameter:** `cost_scaling` = 0.15.
- **Economic Interpretation:** This scenario models markets with substantial infrastructure costs, such as manufacturing plants or brick-and-mortar retail.
- **Hypothesis:** We expect strong risk aversion. Agents should settle into stable local monopolies earlier, as the cost of relocation outweighs the potential gains from capturing the rival's market share, generating a pronounced lock-in effect.

5 Analysis

This section analyzes the empirical findings obtained from the reinforcement learning simulations. We contrast the outcomes of the extended model with the theoretical baseline in order to assess how customer loyalty and relocation costs reshape spatial competition.

Baseline Validation

Before introducing demand shocks, loyalty transport, and movement costs, we validated the Q-learning setup using a simplified Hotelling environment with uniform demand and no relocation penalty (see `Hotelling_model_base`). Consistent with the theoretical *Principle of Minimum Differentiation*, both agents converged to the unique Nash equilibrium located at the center of the market ($x = 0.5$).

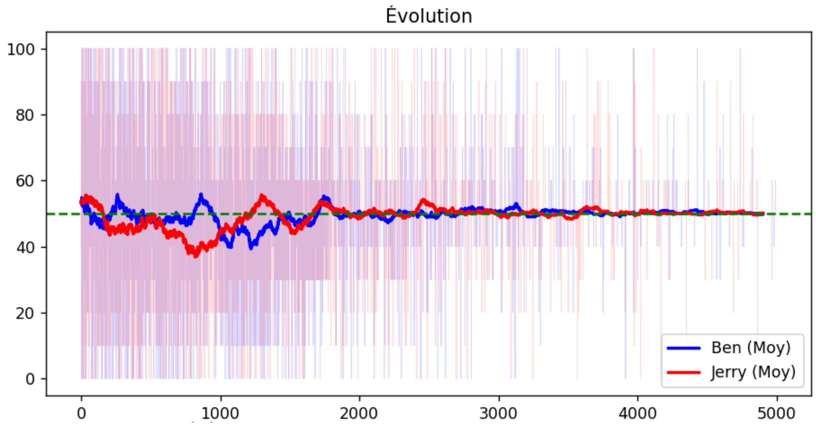

Figure 1: Evolution of Basic Hotelling's model

This result confirms that the agents are capable of learning optimal strategies in a well-understood setting and provides a reliable baseline against which the extended model can be evaluated.

Scenario 1: The Fluid Market (Low Friction)

In this scenario, customer loyalty and stochastic demand shocks are introduced while relocation costs remain negligible (`cost_scaling` $= 0.001$). This configuration represents a highly mobile market in which firms can reposition freely in response to changing demand conditions.

Performance Overview

Across a validation set of 30 games, competition remains tightly contested, with no agent achieving sustained dominance:

- **Ben Wins:** 17 games
- **Jerry Wins:** 9 games
- **Draws:** 4 games

Cumulative rewards further highlight this competitive parity:

$$\text{Total Rewards: Ben } = 2707.45 \quad \text{vs.} \quad \text{Jerry } = 2700.06.$$

The near-identical reward totals indicate that, despite stochastic demand fluctuations, both agents effectively split the market value in the long run.

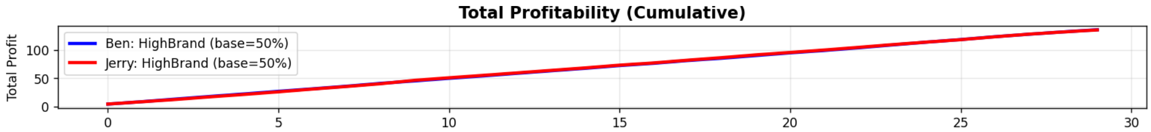

Figure 2: Cumulative rewards evolution

Divergence from the Classical Equilibrium

Unlike the static central agglomeration observed in the baseline model, the Fluid Market exhibits persistent dynamic instability.

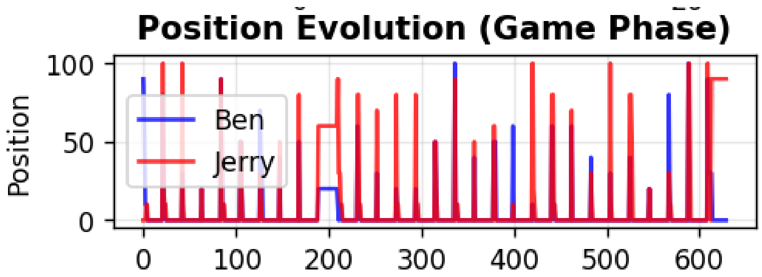

Figure 3: Evolution of the agent's position

Oscillation. As illustrated by the position convergence plots, agents do not settle permanently at the market center. Instead, they oscillate around the median location. This behavior is driven by the negligible cost of movement: in the absence of meaningful relocation penalties, agents are incentivized to continuously micro-adjust their positions to capture transient demand shocks or to appropriate the opponent's loyal customer base.

Cyclical Competition. The high frequency of relocation events confirms that remaining stationary is rarely optimal in a frictionless market with loyalty. Agents engage in a perpetual cat-and-mouse dynamic, reacting strategically to the opponent's previous move rather than converging to a fixed point.

The Role of Loyalty

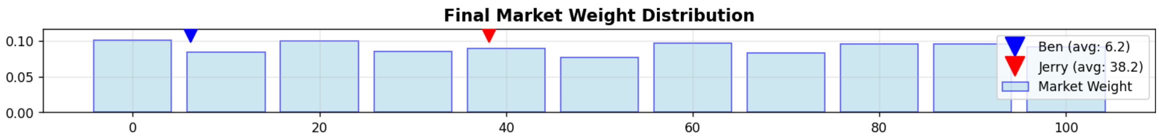

Figure 4: Final market weight distribution

Heatmaps of the market weight distribution reveal that customer density is no longer uniform over space. As agents relocate, they transport a fraction of their loyal customers with them, creating mobile clusters of demand.

This mechanism changes strategic incentives: strict central positioning becomes suboptimal if an opponent has successfully pulled a substantial mass of loyal customers toward an off-center location. These results validate our hypothesis that loyalty generates localized monopolies, thereby disrupting the pure centripetal force predicted by the classical Hotelling model.

Scenario 2: The Rigid Market (High Friction)

In this second experiment, we substantially increase the cost of relocation (`cost_scaling` = 0.15) while keeping the loyalty parameters identical to those used in Scenario 1. This configuration is intended to simulate markets with heavy infrastructure requirements, such as brick-and-mortar retail or manufacturing, where relocation constitutes a large and largely irreversible capital expenditure.

Performance Overview

Results from the validation phase (30 games) reveal a striking asymmetry that was absent in the fluid market:

- **Ben Wins:** 8 games
- **Jerry Wins:** 21 games

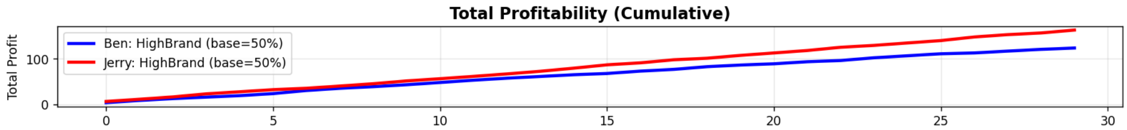

Figure 5: Rewards evolution

Cumulative rewards further emphasize this divergence:

$$\text{Total Rewards: Ben } = 2462.54 \quad \text{vs.} \quad \text{Jerry } = 3247.60.$$

Unlike the balanced competition observed in Scenario 1, the rigid market produces a clear winner. In high-friction environments, early strategic advantages or favorable initial positioning become decisive, as the prohibitive cost of movement prevents the disadvantaged agent from easily correcting its location.

Strategic "Lock-In" and Stability

The most salient result of this scenario is the emergence of pronounced spatial rigidity.

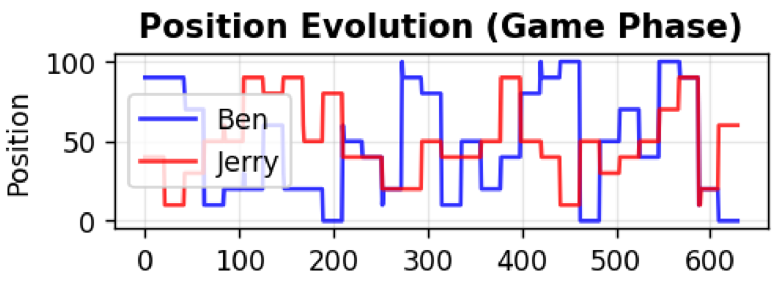

Figure 6: Evolution of the agent's position

Oscillation.

Step-Function Behavior. The position evolution plots exhibit extended plateaus in which agents remain at fixed locations for dozens of consecutive periods, producing a step-like trajectory. This behavior stands in sharp contrast to the persistent oscillations observed in the fluid market.

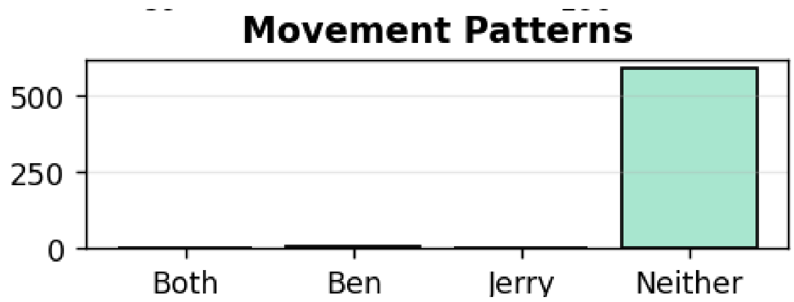

Figure 7: Movement patterns

Relocation Aversion. Histograms of movement patterns confirm that in the vast majority of time steps, neither agent relocates. Movement thus becomes a rare, high-stakes strategic decision rather than a routine tactical adjustment.

Failure of Minimum Differentiation

Crucially, this scenario illustrates a breakdown of Hotelling's *Principle of Minimum Differentiation* in the presence of high friction.

Spatial Separation. Rather than converging to the market center ($x = 0.5$), agents settle at distinct and persistent locations. Ben's average position is 22.6, while Jerry stabilizes around 47.6.

Local Monopolies. The agents effectively partition the market space. Ben captures the left segment of the beach (approximately $[0, 35]$), while Jerry dominates the center-right region. High relocation costs create a strategic "moat": although relocating toward the center could theoretically increase market share, the upfront cost of movement

$$\text{Cost} \propto \text{distance}^2$$

exceeds the discounted value of future gains.

Market Dynamics

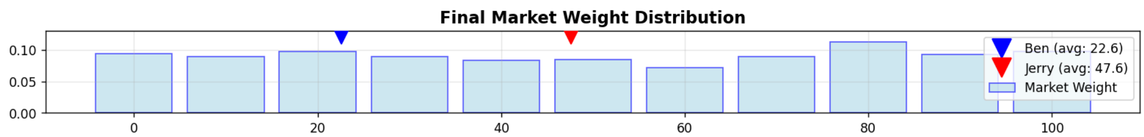

Figure 8: Final market weight distribution

Heatmaps for Scenario 2 display persistent vertical bands, indicating that customer density remains stable over time. Because agents rarely relocate, they are able to accumulate and retain high levels of local loyalty without this demand being dissipated by frequent spatial adjustments. These dynamics confirm that high friction facilitates the emergence of stable, non-centrist local monopolies.

Comparative Conclusion

The contrast between the two scenarios highlights the central role of adjustment costs in shaping spatial competition:

- **Low Friction (Fluid Market):** Competition is characterized by instability and continuous oscillation around the market center. While the classical Nash equilibrium ($x = 0.5$) remains an attractor, customer loyalty prevents agents from remaining there permanently.
- **High Friction (Rigid Market):** Competition exhibits stable differentiation. Agents become locked into suboptimal locations because the cost of reaching the Nash equilibrium is prohibitive.

These findings suggest that in real-world markets with substantial entry or relocation costs, persistent product differentiation and local monopolies are more likely outcomes than the central clustering predicted by classical Hotelling theory.

6 Limitations and Conclusion

This study revisited Hotelling's model of spatial competition through a reinforcement learning framework, incorporating modern market frictions such as customer loyalty and relocation costs. While the results clearly demonstrate the fragility of the *Principle of Minimum Differentiation*, several modeling choices impose limitations that also suggest promising directions for future research.

Limitations

Fixed Prices and Inelastic Demand

The model assumes fixed prices ($p = 1$) and full market coverage with unit demand. In practice, firms compete jointly on price and location. Allowing endogenous pricing could change strategic outcomes: a spatially disadvantaged firm—such as one located at the periphery in the rigid market scenario—might partially offset its positional disadvantage by lowering prices. This price flexibility could weaken or even eliminate the lock-in effects observed under high relocation costs.

Discretization and "Beam" Movement

To enable tabular Q-learning, the continuous market line was discretized into 11 locations, and agents were allowed to relocate instantaneously to any position.

Granularity. A coarse grid amplifies the impact of a single relocation. In a continuous space, firms might engage in gradual micro-adjustments rather than the binary "stay or relocate" decisions observed in the rigid market regime.

Spatial Physics. The teleportation mechanism abstracts away from the time and cost of traversing intermediate space. A "walking" model, in which agents must pass sequentially through positions $x_{t+1}, x_{t+2}, \ldots$, could generate richer strategic behavior, including spatial deterrence and gradual encroachment strategies.

Short-Term Market Memory

As discussed in the results section, the environment introduces stochastic demand shocks that partially reset the weight distribution at each time step. While this prevents the market from becoming degenerate, it limits how persistent loyalty can become over long horizons. Consequently, the model likely underestimates first-mover advantages and the persistence of brand dominance observed in real-world markets, where early incumbents can build durable customer bases over decades.

Agent Complexity

The analysis relies on standard tabular Q-learning agents. While this approach is sufficient for small state spaces (on the order of 11^3 states), it does not scale to richer environments. Extending the framework to continuous spaces, endogenous pricing, or more than two competing firms would require function approximation methods such as Deep Q-Networks (DQN) or policy-gradient approaches (e.g., PPO).

Conclusion

The classical Hotelling model predicts a singular and compelling outcome: in equilibrium, firms converge to the center of the market in order to minimize differentiation and maximize demand coverage.

Our reinforcement learning simulations demonstrate that this prediction is highly contingent on the absence of frictions.

By introducing relocation costs and customer loyalty, we uncover two qualitatively distinct market regimes. In low-friction environments, competition becomes dynamically unstable: although the market center remains an attractor, firms are unable to maintain a central position due to incentives to poach loyal customers, resulting in perpetual oscillation. In contrast, high-friction environments exhibit persistent spatial differentiation. Here, relocation costs act as a barrier to equilibrium, locking firms into locally optimal but globally suboptimal positions and giving rise to stable local monopolies.

Taken together, these findings suggest that the *Principle of Minimum Differentiation* should be interpreted not as a universal law, but as a result that emerges only in frictionless settings. In realistic markets—characterized by switching costs, brand loyalty, and sunk investments—spatial competition is path-dependent, history matters, and equilibrium outcomes may be both inefficient and highly asymmetric. Reinforcement learning thus provides a powerful tool for exploring how bounded rationality and market frictions reshape classical economic predictions.

References

[1] Hotelling, H. (1929). Stability in Competition. *The Economic Journal*, 39(153), 41–57. Oxford University Press on behalf of the Royal Economic Society.

[2] Balvers, R., & Szerb, L. (1996). Location in the Hotelling duopoly model with demand uncertainty. *European Economic Review*, 40(7), 1453–1461.

[3] Biscaia, R., & Mota, I. (2012). Models of spatial competition: A critical review. *International Journal of Industrial Organization*, 30(6), 573–589.