The Limits of AI Trading:

Market Sophistication and Algorithmic Herding

Winston Wei Dou Itay Goldstein Zigang Li Liyan Yang *

June 14, 2026

Abstract

We develop a theoretical framework of trading competition between AI-powered investors and rational investors with heterogeneous sophistication. Rational investors have superior private information but differ in their ability to infer information from prices and anticipate others' trading behavior, modeled through level-(k) reasoning. AI investors do not observe private signals or reason through belief hierarchies; they learn from realized trading profits through reinforcement learning. Despite heterogeneous algorithms and independent exploration, AI investors endogenously converge to common trading rules, generating algorithmic herding. In the limit, their learned demand coincides with the rational-expectations demand of uninformed investors who correctly extract fundamental information from prices. This convergence does not imply algorithmic dominance. The most sophisticated rational investors can outperform AI investors because AI learns from market data generated by average, not frontier, investor sophistication. AI profitability is limited by rational investors' private-information advantage, the price-stabilizing trades of the most sophisticated rational investors, and the rising price impact of AI trading as algorithmic herding increases AI market share. These forces identify the limits of algorithmic superiority in financial markets.

Keywords: Heterogeneous-Agent Asset Pricing, Training Data Limitations, Bounded Rationality, Level-K Reasoning, AI Dominance, Reinforcement Learning, Private Soft Information.

JEL Classification: C72, D80, G10, L19.

*Dou (wdou@wharton.upenn.edu) and Goldstein (itayg@wharton.upenn.edu): University of Pennsylvania and NBER; Li (zigang.li@rotman.utoronto.ca) and Yang (Liyan.Yang@rotman.utoronto.ca): University of Toronto. We thank seminar and confidence participants at AFA Panel, Conference on the Frontiers of AI in Economics, and Wolfe AI and Financial Market Workshop.

1 Introduction

Financial markets are undergoing a profound transformation driven by AI-powered trading. Modern AI trading systems combine algorithmic execution with reinforcement learning, allowing them to interact repeatedly with markets, learn from realized trading outcomes, and autonomously adjust their strategies over time. Their rapid expansion raises a central question for market competition: can AI-powered investors systematically outperform other investors and erode their trading profits.

We develop a theoretical framework of trading competition between AI-powered investors and rational investors with heterogeneous sophistication. Existing research has primarily studied competition, strategic interaction, and collusive behavior among oligopolistic AI traders. By contrast, we focus on direct competition between AI investors and rational investors who possess superior private information but differ in their ability to reason strategically about price formation. The model is a discrete-time, infinite-horizon trading environment with a continuum of rational investors, a continuum of AI investors, and noise traders. All AI and rational investors submit demand schedules and seek to maximize expected trading profits net of trading costs.

The key distinction between the two investor groups is the nature of their advantage. Rational investors have an informational advantage. They can collect proprietary data, acquire soft information, conduct fundamental research, and interact directly with firms and other market participants. We model this advantage as private signals about future asset payoffs that are not directly observed by AI algorithms. Rational investors, however, differ in sophistication. Some investors are limited in their ability to infer information from prices and to anticipate the trading behavior of others. AI investors, in contrast, do not observe private signals and do not begin with knowledge of the payoff distribution or the demand schedules of other investors. They learn from realized trading profits through reinforcement learning and gradually optimize their price-contingent demand rules.

Both types of investors trade based on beliefs about price formation. Rational investors combine private signals with information inferred from prices, but their inference depends on their perceived price-formation rule. Less sophisticated investors may make larger errors when assessing the informational content of prices, because they are more likely to misperceive the trading behavior of other market participants. AI investors' perceived price-formation rule is not represented by an explicit belief hierarchy. Instead, it is embedded in the demand policies they learn from repeated interaction with market-clearing prices. Market equilibrium is therefore jointly determined by private information, strategic reasoning, reinforcement learning, and the market-clearing condition.

We model heterogeneity in rational investors' sophistication through a level-k reasoning structure. Level-0 investors are cursed investors: they use their own private signals but do not correctly account for the information about other investors' signals embedded in prices. A level-k strategic investor, with $k \geq 1$, best responds to a perceived environment in which other strategic investors use level-$(k-1)$ demand schedules, cursed investors continue to use their level-0 demand schedules, and AI investors follow the policies learned in the corresponding level-$(k-1)$ environment. This recursive structure differs from the standard level-k hierarchy because the market also contains a positive measure of autonomous learning algorithms. AI investors do not reason through the hierarchy directly; they learn demand rules from the price and profit data generated by each perceived environment.

We first characterize AI learning within any fixed level-k environment. Using tabular Q-learning with Boltzmann exploration as the leading example, we show that heterogeneous AI investors who learn independently from their own realized profits converge to a common demand rule. This limiting rule coincides with the rational-expectations demand of uninformed investors who correctly extract fundamental information from equilibrium prices. Thus, reinforcement learning generates algorithmic herding: despite heterogeneous algorithms, independent exploration, and no communication, AI investors endogenously herd on a common price-contingent trading rule. We prove this result analytically and provide simulation evidence showing convergence of a large finite population of AI investors to the rational-expectations demand rule.

The central question is whether this learning advantage implies algorithmic dominance. Our answer is no. AI investors can learn rational-expectations price inference, but their performance remains constrained by the sophistication of the market environment from which they learn. They do not observe private signals directly. Instead, they infer fundamentals from prices, and the information in prices is generated by the trading behavior of rational investors. As a result, AI learns from market data generated by average, not frontier, investor sophistication. The most sophisticated rational investors can therefore outperform AI investors even when AI investors learn the correct price-inference rule.

We identify three forces that limit AI profitability. The first is the private-information advantage of rational investors. When private signals are precise, rational investors trade on information that is more accurate than the information AI investors can recover from prices. AI investors may infer fundamentals from equilibrium prices, but prices are noisy aggregates of private information and noise-trader demand. The second force is the stabilizing role of the most sophisticated rational investors. As the sophistication of rational investors increases, their demand schedules become more disciplined and their trading stabilizes prices. This reduces the price variation and residual informational rents from which AI investors learn

and profit. The third force is AI investors' own price impact. Algorithmic herding makes AI demand more correlated, thereby increasing the price impact of AI trading and reducing its profitability, especially when the AI sector becomes larger.

These mechanisms imply that AI superiority in financial markets is inherently limited. AI investors can learn from prices and converge to uninformed rational-expectations demand, but their information is indirect, their learning target is shaped by the sophistication of other investors, and their own market share reduces the profitability of their trades. The model therefore separates algorithmic learning from algorithmic dominance. Reinforcement learning can make AI investors behave like uninformed rational-expectations traders, but it does not allow them to dominate informed and highly sophisticated rational investors.

Contributions and related literature. First, this paper contributes to the growing body of work on how AI market participants compete and influence the market environment. A closely related stream of this research investigates the impact of AI on financial markets, with a specific focus on the interactions among algorithms.[1] For instance, Dou, Goldstein, and Ji (2025) study informed speculators who use Q-learning algorithms. They find that these agents can autonomously learn to sustain collusive, supra-competitive profits without explicit communication, which harms competition and market efficiency. Colliard, Foucault, and Lovo (2022) focus on algorithmic market makers using Q-learning algorithms to set prices. They also show that these algorithms fail to learn competitive pricing strategies, a failure they attribute to limited experimentation and noisy feedback. Routledge (1999) and Routledge (2001) explore whether adaptive algorithms can converge to a rational expectations equilibrium within a repeated Grossman and Stiglitz (1980) model. They prove convergence for adaptive learning and provide examples for genetic algorithms, showing that both can converge to the rational expectations equilibrium. In these cases, their algorithms learn to make correct inferences about a signal from the market-clearing price. Other notable works studying the impact of AI algorithms on financial markets include Marimon, McGrattan, and Sargent (1990), Cartea et al. (2022), and Cartea, Chang, and Penalva (2022).

Our work differs from the existing literature in several important ways. First, while much of the literature focuses on the interaction among multiple reinforcement learning agents or the convergence of algorithms to rational expectations, we center our analysis on the competition between humans and AI. Our paper provides a framework for understanding the distinct comparative advantages of humans and AI and for analyzing how the market environment affects this human-AI competition. Second, whereas most existing works rely

[1] Another stream of literature focuses on algorithmic collusion in retail markets, e.g., Calvano et al. (2020, 2021); Johnson, Rhodes, and Wildenbeest (2023); Waltman and Kaymak (2008); Hansen, Misra, and Pai (2021); Abada and Lambin (2023); Banchio and Mantegazza (2024).

on simulation-based analysis, we provide an analytical characterization of the equilibrium and how changes in key factors affect competition and market outcomes. We also establish analytical conditions for AI algorithms to learn and converge to rational expectations in our setting. Third, unlike research that focuses on a single algorithm (e.g., Q-learning), our analysis is not restricted to one type but applies to a broad class of algorithms. We thereby offer a more general theoretical framework.

A related contemporaneous work by Banerjee and Szydlowski (2025) studies a market with rational investors and a single Q-learning trader. They find that the Q-learner's feedback-driven trading generates stochastic volatility and predictable returns, and can sometimes improve overall investor utility despite increasing price volatility. While their focus on human-AI interaction is similar to ours, our paper differs and contributes in three significant ways. First, we incorporate information asymmetry, allowing humans and AI to possess distinct informational advantages. This asymmetry captures a key difference between humans and AI in financial markets and is crucial for understanding the competition between them. Second, our analysis applies to a wide range of reinforcement learning algorithms post-convergence to a stable policy. We abstract from the proprietary details of both the specific type and the training process of the algorithm. It is unrealistic for human investors to know the details of the algorithm being used by their AI competitors, such as the exact type of algorithm, specific hyperparameters, or its stage of training, all of which greatly affect how the algorithm evolves and converges. Our focus on the post-convergence phase therefore reflects a more realistic scenario where investors compete with stable, deployed trading strategies used by other market participants. Third, we introduce the concept of bounded strategic thinking for human investors, allowing them to have misperceptions about others' behavior when facing a complex market environment that includes AI traders. By using the CH framework, we provide a comprehensive framework to model human perception of AI. This approach more closely captures the diverging sophistication levels of real-world investors.

Furthermore, this paper contributes to the theory literature on imperfect competition and bounded strategic reasoning in financial markets. Our work introduces different levels of bounded strategic thinking into an imperfectly competitive financial market in the spirit of Kyle (1989).[2,3] We model human investors' strategic thinking using the Cognitive Hier-

[2]See Crawford, Costa-Gomes, and Iriberri (2013) for a review on recent theory and evidence on strategic thinking and the applications of level-k models. Many other experimental papers show direct evidence of level-k thinking (Stahl and Wilson, 1994, 1995; Nagel, 1995; Costa-Gomes, Crawford, and Broseta, 2001; Costa-Gomes and Crawford, 2006).

[3]A recent literature studies the impact of bounded strategic thinking in macroeconomics (García-Schmidt and Woodford, 2019; Farhi and Werning, 2019; Angeletos and Lian, 2023).

archy framework of Camerer, Ho, and Chong (2004), which posits that agents have iterated levels of reasoning about others. This approach is related to, but distinct from, the level-k thinking models used in papers like Zhou (2022). Unlike standard level-k models where a player believes all others are level-(k-1), the CH framework assumes a player best responds to a distribution of lower-level types. Zhou (2022) uses level-k thinking to model human speculators and focuses on how bounded strategic reasoning can generate momentum and contrarian trading strategies. In contrast, our application of the CH model provides a game-theoretic foundation for understanding how the distribution of human sophistication levels affects the competition between humans and AI. Other works study the effect of higher-order beliefs and the perception of information in asset prices on financial markets (Allen, Morris, and Shin, 2006; Han and Kyle, 2018; Eyster, Rabin, and Vayanos, 2019). Eyster, Rabin, and Vayanos (2019) studies how the presence of investors who do not fully invert prices to uncover others' information (cursed) affects the financial markets and considers the degree of cursedness as the measure of sophistication. This paper considers sophistication as the level of iterated reasoning, which incorporates human investors' reasoning about AI investors' behavior.

2 Model

2.1 Model Setup

Assets. Time is discrete and infinite, indexed by $t = 1, 2, \ldots$. A single risky asset is traded in each period. The asset has a per capita supply of S. Its payoff, v_t, is realized at the end of the period and follows a normal distribution: $v_t \sim \mathcal{N}(\bar{v}, \sigma_v^2)$.

Rational investors with heterogeneous sophistication. There is a continuum of rational investors with heterogeneous sophistication, with total measure ψ.[4] Rational investors are Bayesian expected-utility maximizers given their subjective beliefs. These beliefs need not coincide with the true objective distribution unless we impose rational expectations. Thus, rationality requires optimality conditional on beliefs, while rational expectations additionally requires belief correctness in equilibrium.

Each rational investor, indexed by $i \in \mathcal{M}$, maximizes the expected present value of net

[4]Rational investors can have subjective beliefs that differ from the true objective distribution; in economic theory, rationality usually requires internal consistency and optimal decision-making given beliefs, not necessarily that beliefs equal the objective data-generating process.

trading profits:

$$\mathbb{E}\left[\sum_{t=0}^{\infty}\rho^{t}\pi_{i,t}\right], \tag{1}$$

where $0<\rho<1$ is the discount rate. The trading profit $\pi_{i,t}$ is determined by

$$\pi_{i,t}=(v_t-p_t)\,x_{i,t}-\frac{1}{2}\gamma x_{i,t}^{2},$$

where p_t is the equilibrium asset price, and $x_{i,t}$ is the number of shares of the risky asset that investor i purchases at the beginning of period t. The term $\frac{1}{2}\gamma x_{i,t}^{2}$ represents transaction costs, and γ controls the level of these costs.[5]

Rational investors with heterogeneous sophistication may possess informational advantages. They observe private signals about future fundamentals, such as soft corporate information from management interactions, customer and supplier conditions, channel checks, expert-network insights, and proprietary forecasts. These signals are neither publicly disclosed nor directly accessible to AI systems through internet-based data, allowing investors to form superior beliefs about future asset payoffs.

Specifically, at the beginning of each period t, each rational investor i observes a private signal about the end-of-period payoff of the risky asset. The private signal is given by

$$\eta_{i,t}=v_t+e_{i,t},$$

where $e_{i,t}$ are i.i.d. and $e_{i,t}\sim\mathcal{N}(0,\sigma_M^2)$.

Despite their informational advantages, rational investors may differ in strategic reasoning capacity. All rational investors choose optimally given their subjective beliefs. However, investors with limited strategic reasoning may hold incorrect beliefs about the sophistication and behavior of other market participants. As a result, they may misperceive how equilibrium prices aggregate private information and therefore form imperfect posterior beliefs from prices.

We model this heterogeneity using two groups of rational investors. The first group consists of cursed investors, with measure ψ_c, indexed by $i\in\mathcal{M}_c$. These investors do not fully account for the information about fundamentals that is revealed through equilibrium prices. The second group consists of relatively sophisticated level-k thinking investors, also referred to as strategic investors, with measure ψ_s, indexed by $i\in\mathcal{M}_s$. These investors

[5]The transaction cost term is similar to that in Gârleanu and Pedersen (2013). This setup simplifies the comparison of profitability between rational and AI investors. Assuming investors have CARA or mean-variance utility over end-of-period wealth with a given level of precision of rational investors' private signals would yield similar results and not change our analysis in the following sections.

perform a finite number of iterative reasoning steps when inferring information from prices and anticipating the behavior of other investors. Thus, $\psi_c + \psi_s = \psi$, $\mathcal{M}_c \cup \mathcal{M}_s = \mathcal{M}$, and $\mathcal{M}_c \cap \mathcal{M}_s = \emptyset$.

Because investors submit limit orders, a cursed investor i submits a demand schedule that depends on the trading price and the investor's private signal:

$$x_{i,t}^{M,c}(p_t; \eta_{i,t}) \equiv \mu^{M,c} + \beta_\eta^{M,c}\eta_{i,t} - \beta_p^{M,c}p_t, \tag{2}$$

where the coefficients $\mu^{M,c}$, $\beta_\eta^{M,c}$, and $\beta_p^{M,c}$ are chosen optimally given the cursed investor's subjective beliefs. Cursed investors use their private signals to form beliefs about fundamentals, but they do not use prices to infer the private information or trading behavior of other market participants.

A strategic investor i, who is a level-k thinking investor, submits a linear limit-order schedule that depends on both the trading price and the investor's private signal:

$$x_{i,t}^{M,k}(p_t; \eta_{i,t}) \equiv \mu^{M,k} + \beta_\eta^{M,k}\eta_{i,t} - \beta_p^{M,k}p_t. \tag{3}$$

The coefficients $\mu^{M,k}$, $\beta_\eta^{M,k}$, and $\beta_p^{M,k}$ are chosen optimally under the investor's level-k subjective price-formation rule. Thus, a level-k investor uses the price both as the trading price at which demand is submitted and as an informational signal about fundamentals and other investors' trading behavior. Let $\mathcal{P}_{M,k}(\cdot)$ denote the price-formation rule that is perceived by level-k thinking investors.

AI investors. AI investors execute algorithmic trading based on the autonomous, self-learned trading strategies by reinforcement learning (RL), rather than rigid human-defined trading strategies. There is a continuum of atomistic AI investors in the market, with total mass of ϕ. By design of the algorithms, each AI investor, indexed by $j \in \mathcal{A}$, aims to maximize its total expected discounted trading profits:

$$\mathbb{E}\left[\sum_{t=0}^{\infty} \rho^t \pi_{j,t}\right]. \tag{4}$$

The net trading profit, $\pi_{j,t}$, is determined by

$$\pi_{j,t} = (v_t - p_t)\, x_{j,t} - \frac{1}{2}\gamma x_{j,t}^2, \tag{5}$$

where γ controls the level of an AI investor's transaction cost, and $x_{j,t}$ represents the shares of the risky asset that it chooses to purchase at the beginning of period t.

Each AI investor begins with no direct knowledge of the market environment. It does not know the distribution of asset payoffs, the price-formation rule, or the strategies of other market participants. Instead, it learns from repeated market interactions through reinforcement learning. In each period, the algorithm chooses a demand schedule, observes the realized trading profit as its reward, and updates its action values and trading policy to maximize expected discounted profits.

Let $\mathcal{P}_{A,k}(\cdot)$ denote the price-formation rule that is implicit in the AI investor's learned value function and trading policy in a trading environment populated with level-k thinking investors. Unlike rational investors with heterogeneous sophistication, who form explicit subjective models of price formation, an AI investor does not reason through an explicit perceived pricing rule. Instead, the relevant price-formation mapping is embedded indirectly in the algorithm's learned action values: demand schedules that systematically generate higher profits under the realized market-clearing prices receive higher values and are chosen more frequently over time.

This paper studies relatively less sophisticated AI-powered trading algorithms adopted by a broad set of atomistic investors, such as retail investors trading through platforms like Robinhood. We therefore model AI investors as using stateless Q-learning algorithms, in the spirit of Colliard, Foucault, and Lovo (2022) and Banerjee and Szydlowski (2025). This focus contrasts with Dou, Goldstein, and Ji (2025, 2026), who study intertemporally sophisticated AI trading by oligopolistic, advanced informed speculators and its implications for market efficiency and stability. Specifically, following the definition in Watkins (1989), stateless Q-learning is detailed as follows. Let $\mathcal{X}$ be the action space. For AI investor j, the problem is to choose a sequence of demand schedule $x_{j,t}(\cdot) \in \mathcal{X}$, where $\mathcal{X}$ is a linear functional space to maximize the expected total discounted rewards, using the following algorithm for Q-value:

$$Q_{j,t+1}\left(x_{j,t}(\cdot)\right) = (1-\alpha_{j,t})Q_{j,t}\left(x_{j,t}(\cdot)\right) + \alpha_{j,t}\left[\pi_{j,t} + \rho \max_{\tilde{x}_j(\cdot)\in\mathcal{X}} Q_{j,t}\left(\tilde{x}_j(\cdot)\right)\right], \tag{6}$$

where $\pi_{j,t}$ is defined in (5), and $\alpha_{j,t} \in [0,1]$ controls the learning rate of the Q-function with $\alpha_{j,t} \to 0$ as $t \to \infty$. A larger learning rate means that the algorithm puts more weight on the new observation and updates the Q-function more quickly. This temporal-difference update method bootstraps from current estimates and does not require a model of the environment's dynamics.

Practitioners often do not set learning rates, or equivalently step sizes, to converge literally to zero, especially in modern deep reinforcement learning. Sutton and Barto (1998) emphasize that shrinking learning-rate sequences are primarily used in theoretical work and are seldom used in applications and empirical research. They further note that constant learn-

ing rates are natural in potentially nonstationary environments because they place sufficient weight on recent observations, a consideration particularly relevant for financial markets. Consistent with this practice, constant-learning-rate reinforcement learning is widely used in applications and studied as a practically relevant alternative to diminishing-learning-rate algorithms, because it is simple to implement, can improve convergence speed over finite training horizons, and remains more responsive to new data. Our objective here, however, is theoretical: to characterize asymptotic algorithmic herding and the limits of atomistic RL trading performance. We therefore impose the standard Robbins–Monro restrictions on the learning-rate sequence, which provide the appropriate benchmark for convergence analysis.

Each AI investor a explores using a Boltzmann (softmax) policy over its learned action values. Let $Q_{j,t}(x)$ denote AI investor a's estimated Q-value of demand schedule $x(\cdot)$ at time t, and let $\kappa_j > 0$ be its exploration temperature. The decision policy assigns to a demand schedule, $x(\cdot)$, with the probability

$$q_j(x(\cdot)) \equiv \frac{\exp\big(Q_{j,t}(x(\cdot))/\kappa_j\big)}{\sum_{\tilde{x}(\cdot)\in\mathcal{X}} \exp\big(Q_{j,t}(\tilde{x}(\cdot))/\kappa_j\big)}. \tag{7}$$

A higher temperature κ_j flattens the distribution toward uniform exploration, while a lower temperature concentrates choice on the highest-valued actions. Boltzmann exploration keeps every action's probability strictly positive, so each demand schedule continues to be sampled and its value estimate keeps improving. We allow the temperatures κ_j to differ across AI investors, capturing heterogeneity in how aggressively each algorithm explores.[6]

The linear demand-schedule space $\mathcal{X}$ is two-dimensional. In an environment populated with level-k thinking investors, each AI investor j's demand schedule is given by

$$x_{j,t}^{A,k}(p_t) \equiv \mu_{j,t}^{A,k} - \beta_{j,p,t}^{A,k} p_t. \tag{8}$$

Hence, an action in the Q-learning problem can be represented by the coefficient pair

$$\left(\mu_{j,t}^{A,k}, \beta_{j,p,t}^{A,k}\right).$$

Noise traders. A unit measure of noise traders trade for non-informational reasons, such as hedging needs, estimation errors, or sentiment. Their aggregate demand is z_t shares of the risky asset, where $z_t \sim \mathcal{N}(0, \sigma_z^2)$. The demand z_t is independent of other shocks and across time.

[6]The smoothness of these choice probabilities in the value estimates, rather than the hard switching of an ε-greedy rule, is what makes the population learning dynamics a contraction and delivers the convergence results below (see Appendix A).

Equilibrium. An equilibrium in a trading environment populated by cursed investors, level-k strategic investors, and AI investors consists of a sequence of cursed-investor demand-schedule collections,

$$\left\{\left(x_{i,t}^{M,c}(\cdot)\right)_{i\in\mathcal{M}_c}\right\}_{t\geq 0},$$

a sequence of level-k strategic-investor demand-schedule collections,

$$\left\{\left(x_{i,t}^{M,k}(\cdot)\right)_{i\in\mathcal{M}_s}\right\}_{t\geq 0},$$

a sequence of AI-investor demand-schedule collections,

$$\left\{\left(x_{j,t}^{A,k}(\cdot)\right)_{j\in\mathcal{A}}\right\}_{t\geq 0},$$

and a sequence of market-clearing prices,

$$\{p_t\}_{t\geq 0},$$

such that the following conditions hold in each period t:

1. Rational investors update beliefs according to Bayes' rule. Each cursed investor i combines its prior and its private signal $\eta_{i,t}$. Thus, its information set is

   $$\mathcal{I}_{i,t}^{M,c} = \{\eta_{i,t}\}.$$

   Each rational investor i combines its prior, its private signal $\eta_{i,t}$, and the information inferred from the equilibrium price under its perceived price-formation rule. Thus, its information set is

   $$\mathcal{I}_{m,t}^{M,k} = \left\{\eta_{i,t}, \mathcal{P}_{M,k}^{-1}(p_t)\right\}.$$

   AI investors update their implicit beliefs through the Q-value updating rule in (6), which revises the valuation of trading actions based on realized rewards. Thus, its effective, implicit information set is

   $$\mathcal{I}_{j,t}^{A,k} = \left\{\mathcal{P}_{A,k}^{-1}(p_t)\right\}.$$

2. At the beginning of period t, each cursed investor $i \in \mathcal{M}_c$ chooses a demand schedule $x_{i,t}^{M,c}(\cdot)$ to maximize expected trading profits, as in (1), taking prices as trading terms but not as signals of other investors' private information. Each level-k strategic investor

$i \in \mathcal{M}_s$ chooses a demand schedule $x_{i,t}^{M,k}(\cdot)$ to maximize expected trading profits, as in (1), under its level-k subjective price-formation rule.

3. At the beginning of period t, each AI investor $j \in \mathcal{A}$ chooses a demand schedule $x_{j,t}^{A,k}(\cdot)$ according to the Q-learning decision rule in (7).

4. The market clears at the end of period t. Total demand from cursed investors, level-k strategic investors, AI investors, and noise traders equals the asset supply S:

$$\int_{\mathcal{M}_c} x_{i,t}^{M,c}(p_t; \eta_{i,t})\, di + \int_{\mathcal{M}_s} x_{i,t}^{M,k}(p_t; \eta_{i,t})\, di + \int_{\mathcal{A}} x_{a,t}^{A,k}(p_t)\, da + z_t = S. \tag{9}$$

Equilibrium price formation. We focus on an equilibrium in which the Q-learning algorithms of AI investors converge to their steady-state action values in the environment populated by cursed investors and level-k strategic investors. Let $Q_j^{(k)}$ denote the limiting Q-value function of AI investor j in this level-k environment.

Given the equilibrium demand schedules of the different investor types, the market-clearing condition implies a noisy linear relationship between the equilibrium price and the fundamental payoff. In particular, suppose cursed investors use the demand schedule coefficients

$$\left(\mu^{M,c}, \beta_\eta^{M,c}, \beta_p^{M,c}\right),$$

level-k strategic investors use

$$\left(\mu^{M,k}, \beta_\eta^{M,k}, \beta_p^{M,k}\right),$$

and AI investors use

$$\left(\mu^{A,k}, \beta_p^{A,k}\right).$$

Then the equilibrium price aggregates information from the private signals of cursed and level-k strategic investors, the learned demand of AI investors, and noise-trader supply. As a result, an uninformed rational-expectations investor can infer payoff information from the equilibrium price by correctly inverting the true price-formation rule. Level-k strategic investors also extract information from prices, but they do so under their subjective price-formation rule. In particular, a level-k investor believes that other strategic investors use level-$(k-1)$ demand schedules and that AI investors follow the policies learned in the corresponding level-$(k-1)$ environment. Hence, relative to a fully rational-expectations investor, the level-k investor may misperceive the informativeness of prices because it underestimates the sophistication of other strategic investors' demand schedules and the sophistication embedded in the AI policies learned from interacting with those investors.

The price is determined by the market-clearing condition in (9). By the exact law of large numbers for a continuum of i.i.d. private-signal errors in the cross section,

$$\int_{\mathcal{M}_c} \eta_{i,t}\, di = \psi_c v_t, \quad \text{and} \quad \int_{\mathcal{M}_s} \eta_{i,t}\, di = \psi_s v_t. \tag{10}$$

Using the demand schedules of cursed investors, level-k strategic investors, and AI investors, the market-clearing condition can therefore be written as

$$\xi^{(k)} p_t = \zeta^{(k)} v_t + z_t + \mu^{(k)}, \tag{11}$$

where the superscript (k) means the equilibrium of a trading environment populated with level-k thinking investors, and the ceofficients are

$$\zeta^{(k)} \equiv \psi_c \beta_\eta^{M,c} + \psi_s \beta_\eta^{M,k}, \tag{12}$$

$$\xi^{(k)} \equiv \psi_c \beta_p^{M,c} + \psi_s \beta_p^{M,k} + \phi \beta_p^{A,k}, \tag{13}$$

$$\mu^{(k)} \equiv \psi_c \mu^{M,c} + \psi_s \mu^{M,k} + \phi \mu^{A,k} - S. \tag{14}$$

Equivalently,

$$p_t = \frac{\mu^{(k)}}{\xi^{(k)}} + \frac{\zeta^{(k)}}{\xi^{(k)}} v_t + \frac{1}{\xi^{(k)}} z_t. \tag{15}$$

Thus, the equilibrium price is a noisy linear signal of the fundamental payoff. The ratio $\zeta^{(k)}/\xi^{(k)}$ is the loading of the equilibrium price on fundamentals, and $1/\xi^{(k)}$ is the loading of the equilibrium price on noise-trader demand.

The noisy relationship between equilibrium price and the fundamental, characterized by (15), implies that a signal about v_t as follows:

$$s_{p,t}^{(k)} \equiv \frac{\xi^{(k)}}{\zeta^{(k)}} \left(p_t - \frac{\mu^{(k)}}{\xi^{(k)}} \right) = v_t + \frac{1}{\zeta^{(k)}} z_t. \tag{16}$$

The noise standard deviation of this price-implied signal is

$$\sigma_p^{(k)} = \frac{\sigma_z}{|\zeta^{(k)}|}, \tag{17}$$

and its precision is

$$\tau_p^{(k)} = \frac{1}{\left(\sigma_p^{(k)}\right)^2} = \left(\zeta^{(k)}\right)^2 \tau_z. \tag{18}$$

Price informativeness about v_t is therefore governed by the magnitude of the aggregate signal loading $\zeta^{(k)}$: stronger informed non-AI demand makes the price-implied signal less noisy and

increases the precision with which prices reveal fundamentals.

2.2 Equilibrium Demand Schedules

Rational investors differ in sophistication, which we define as their capacity to reason strategically about price formation. This heterogeneity affects both how investors extract information about fundamentals from prices and how they form beliefs about the trading behavior of other market participants, including cursed investors, more sophisticated rational investors, and AI investors. In this section, we characterize the equilibrium demand schedules of rational investors with different degrees of sophistication. We begin with level-0 investors, whom we specify as cursed investors. These investors use their own private signals when forming beliefs about fundamentals, but do not correctly account for the information about other investors' signals that is embedded in equilibrium prices. We then construct level-k strategic investors recursively. A level-1 investor best responds to a perceived environment in which other strategic investors behave as level-0 investors and AI investors follow the policies learned in that level-0 environment. More generally, a level-k investor best responds to a perceived environment in which other strategic investors behave according to level-$(k-1)$ reasoning, AI investors follow the policies learned in the corresponding level-$(k-1)$ environment, and cursed investors continue to trade according to their level-0 demand schedules. This recursive structure differs from the standard level-k hierarchy because the market also contains a positive measure of autonomous Q-learning investors. AI investors do not reason through an explicit finite hierarchy of beliefs. Instead, for each perceived level-k environment, they learn trading policies from repeated market interactions generated by the demand schedules of rational investors and by market-clearing prices. The level-k iteration is illustrated in Figure 1.

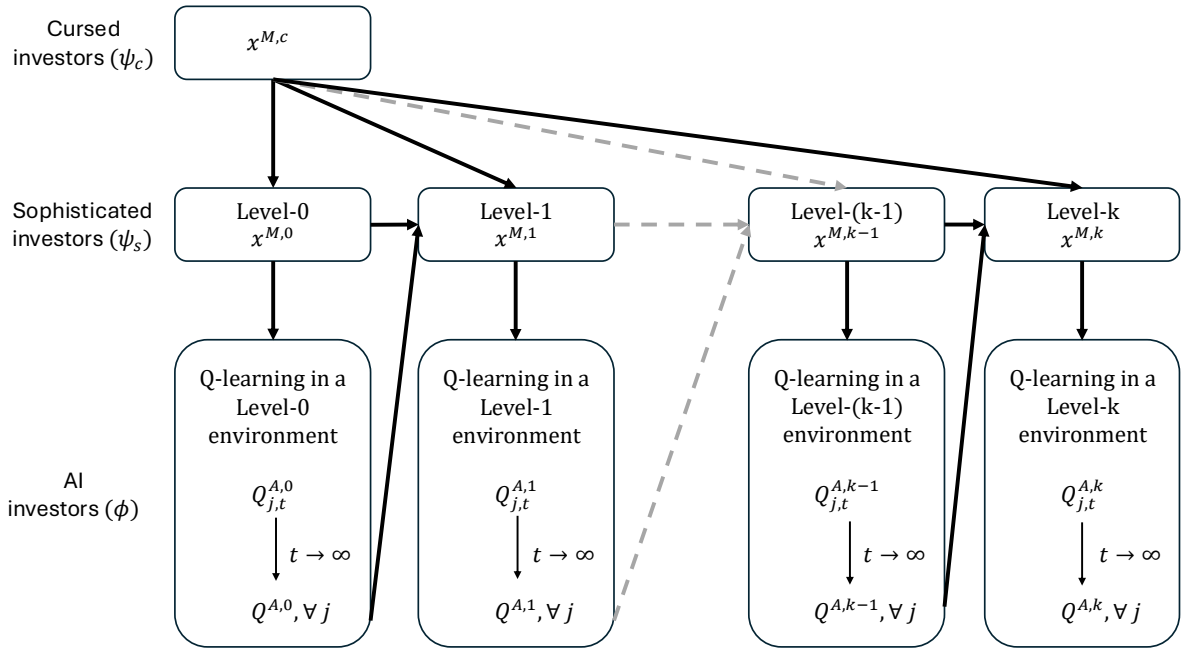
Cursed investors. They do not understand the information contained in prices at all. They only use their private signal $\eta_{i,t}$ to form their posterior belief about v_t and make trading decisions. Specifically, the conditional expectation of v_t given the private signal $\eta_{i,t}$ is

$$\mathbb{E}\left[v_t \mid \eta_{i,t}\right] = h_v^c \bar{v} + h_\eta^c \, \eta_{i,t}, \tag{19}$$

where the coefficients h_η^c and h_v^c are the Bayesian update weights

$$h_\eta^c \equiv \frac{\tau_M}{\tau_v + \tau_M} \quad \text{and} \quad h_v^c \equiv \frac{\tau_v}{\tau_v + \tau_M}, \tag{20}$$

Figure 1: The Level-k Iteration



with precision parameters defined as

$$\tau_v \equiv \sigma_v^{-2} \quad \text{and} \quad \tau_M \equiv \sigma_M^{-2}. \tag{21}$$

The optimal demand schedule of the cursed investor i is:

$$x^{M,c}\left(p_t; \eta_{i,t}\right) = \frac{\mathbb{E}\left[v_t \mid \eta_{i,t}\right] - p_t}{\gamma} \tag{22}$$

$$= \mu^{M,c} + \beta_\eta^{M,c} \eta_{i,t} - \beta_p^{M,c} p_t, \tag{23}$$

where

$$\beta^{M,c} = \gamma^{-1} h_v^c \bar{v}, \quad \beta_\eta^{M,c} = \gamma^{-1} h_\eta^c, \quad \text{and} \quad \beta_p^{M,c} = \gamma^{-1}. \tag{24}$$

Cursed investors are not noise traders. They use their own private signals correctly when forming beliefs about fundamentals, but they do not fully internalize the price-formation process. In particular, they do not correctly account for the existence and behavior of AI investors, nor do they account for the information about other investors' private signals that is embedded in equilibrium prices. As a result, they under-infer the information about fundamentals contained in prices. Their optimal demand schedule is characterized by (23).

AI investors: Algorithmic herding in a level-k environment. Section 3 shows that heterogeneous AI investors do not remain dispersed across idiosyncratic trading rules. Each AI investor learns only from its own realized trading profits and explores independently. Nevertheless, market-clearing price feedback makes the profitability of any demand schedule depend on the aggregate behavior of AI investors. This feedback aligns the learning incentives of different AI investors and drives their population demand toward a common rule. The limiting common rule has a sharp interpretation. In an environment populated by cursed investors, level-k strategic investors, AI investors, and noise traders, all AI investors' algorithms converge to the same one, and the aggregate AI demand schedule converges to the demand schedule of an uninformed rational-expectations investor who observes the equilibrium price, understands the true price-formation rule, and uses the price to infer payoff information. Thus,

$$x^{A,k}(p) = \frac{\mathbb{E}^{(k)}[v \mid p] - p}{\gamma} = \beta_0^{A,k} - \beta_p^{A,k} p.$$

Here, $\mathbb{E}^{(k)}[v \mid p]$ denotes the posterior expectation of the payoff v formed by an uninformed rational-expectations investor after observing price p in the level-k environment, which can be expressed as $\mathbb{E}[v \mid s_p^{(k)}]$. The superscript (k) emphasizes that the price-formation rule, price informativeness, and the learned AI demand coefficients all depend on the sophistication of the strategic investors in that environment. This result links algorithmic herding to rational-expectations price inference. AI investors do not directly observe the private signals of rational investors, nor do they reason through an explicit hierarchy of beliefs. Instead, they learn from realized profits which price-contingent demand schedules perform well. In the limit, this learning process selects the same demand schedule that an uninformed rational-expectations investor would choose when extracting information from prices, as in the benchmark of Kyle (1989). This characterization is also the key input into the level-k recursion. When a level-k strategic investor forms beliefs about price formation, she assumes that AI investors follow the demand rule learned in the perceived level-$(k-1)$ environment:

$$x^{A,k-1}(p) = \mu^{A,k-1} - \beta_p^{A,k-1} p.$$

Thus, the recursive step is precise: level-k strategic investors best respond to a perceived market in which other strategic investors use level-$(k-1)$ demand schedules, cursed investors continue to use their level-0 demand schedules, and AI investors use the rational-expectations demand rule learned from interacting with the level-$(k-1)$ environment.

We now focus on characterizing $\mu^{A,k}$ and $\beta_p^{A,k}$ for every $k \geq 1$. In the limit, all AI investors become homogeneous and the representative AI investor in a trading environment

populated with level-k investors behave as if it correctly understands the ψ_c cursed investors would behave according to (23) and perceives the remaining ψ_s level-k investors would use level-k strategies. From the implied price signal $s_{p,t}^{(k)}$ in (16), the AI investor behaves as if it forms the belief using the posterior

$$\mathbb{E}[v_t \mid s_{p,t}^{(k)}] = h_p^{A,k}\, s_{p,t} + (1 - h_p^{A,k})\bar{v}.$$

where the AI's Bayesian weight on the price signal,

$$h_p^{A,k} \equiv \frac{\tau_p^{(k)}}{\tau_v + \tau_p^{(k)}}, \tag{25}$$

with $\tau_p^{(k)}$ is the precision coefficient in (18).

The first-order condition gives a linear demand,

$$x^{A,k}(p) \equiv \frac{\mathbb{E}[v \mid s_p^{(k)}] - p}{\gamma}. \tag{26}$$

Therefore, according to the expression of $s_p^{(k)}$ in terms of p, it holds that

$$\beta_p^{A,k} = \frac{1 - h_p^{A,k} \frac{\psi_c \beta_p^{M,c} + \psi_s \beta_p^{M,k}}{\zeta^{(k)}}}{\gamma + h_p^{A,k} \frac{\phi}{\zeta^{k-1}}}, \tag{27}$$

$$\mu^{A,k} = \frac{(1 - h_p^{A,k})\bar{v} - h_p^{A,k} \frac{\psi_c \mu^{M,c} + \psi_s \mu^{M,k} - S}{\zeta^{(k)}}}{\gamma + h_p^{A,k} \frac{\phi}{\zeta^{(k)}}}. \tag{28}$$

These expressions show how AI demand depends on the information embedded in prices. The numerator of $\beta_p^{A,k}$ reflects the direct response of demand to price, adjusted for the fact that price also carries information about fundamentals. The denominator captures the equilibrium feedback from the positive mass ϕ of AI investors: because AI demand affects the market-clearing price, the representative learned AI rule internalizes the price signal generated by aggregate AI trading.

Thus, despite starting from heterogeneous learning algorithms and exploring independently, AI investors endogenously converge to a common demand rule. This common rule coincides with the optimal trading rule of uninformed rational-expectations investors in a rational-expectations equilibrium (REE) of the trading environment populated by level-k strategic investors. Algorithmic herding therefore carries the AI population to the rational-expectations demand rule associated with the level-k environment.

Strategic investors: Level-k thinking investors with $k \geq 1$. Level-k thinking investors, with $k \geq 1$, know the population shares of cursed investors and AI investors in the market. Their reasoning is anchored at level 0, which we specify as cursed-investor behavior. For $k \geq 1$, a level-k investor forms beliefs about price formation by assuming that other sophisticated investors reason at level $k-1$, while AI investors follow the trading policies learned in the corresponding perceived environment. In particular, the level-k investor believes that AI investors have been trained in an environment with measure ψ_s of level-$(k-1)$ investors. Given this perceived price-formation rule, the level-k investor chooses a demand schedule that maximizes expected trading profits. Higher k therefore corresponds to a deeper ability to infer information from prices and to anticipate the trading behavior of other market participants.

Starting from level-0 investors, we recursively construct level-k investor strategies and AI policies learned in the corresponding level-k environment, for $k = 0, 1, 2, \ldots$. This recursion differs from the standard level-k reasoning hierarchy because the market contains a positive measure ϕ of autonomous Q-learning investors. These AI investors do not reason through a finite hierarchy of beliefs. Instead, for each perceived level-k environment, they learn trading policies from repeated market interactions generated by the strategies of level-k investors.

For each $k \geq 1$, consider a level-k investor's perceived market environment. The investor assumes that non-AI investors other than itself use the demand schedules associated with level-$(k-1)$ reasoning. She also takes as given the AI trading policy that would be learned when AI investors are trained in an environment populated by a measure ψ_s of level-$(k-1)$ investors. Thus, in the level-k investor's subjective model, prices are formed from the aggregate order flow generated by level-$(k-1)$ non-AI investors, AI investors following the policy learned in the corresponding level-$(k-1)$ environment, and any exogenous noise supply.

Given this perceived price-formation rule, the level-k investor chooses her own demand schedule to maximize expected trading profits. The expectation is taken under her perceived joint distribution of fundamentals, private signals, aggregate order flow, AI trading, and prices. Its optimization therefore treats the strategies of lower-level non-AI investors and the learned AI policy as fixed components of the pricing rule, while choosing her own demand schedule as the best response to that perceived rule.

The Bayesian posterior is

$$\mathbb{E}[v_t \mid \eta_{i,t}, p_t] = h_\eta^{M,k}\, \eta_{i,t} + h_p^{M,k}\, s_{p,t}^{(k-1)} + h_v^{M,k}\, \bar{v}, \tag{29}$$

where the posterior weights are

$$h_\eta^{M,k} \equiv \frac{\tau_M}{\tau_v + \tau_M + \tau_p^{(k-1)}}, \qquad h_p^{M,k} \equiv \frac{\tau_p^{(k-1)}}{\tau_v + \tau_M + \tau_p^{(k-1)}}, \qquad h_v^{M,k} \equiv \frac{\tau_v}{\tau_v + \tau_M + \tau_p^{(k-1)}}.$$

The first-order condition gives $x^{M,k}(p_t) \equiv (\mathbb{E}[v \mid \eta_{i,t}, p_t] - p_t)/\gamma$. Therefore, according to the expression of $s_p^{(k-1)}$ in terms of p, it holds that

$$\beta_\eta^{M,k} = \frac{h_\eta^{M,k}}{\gamma} = \frac{\tau_M}{\gamma(\tau_v + \tau_M + \tau_p^{(k-1)})}, \tag{30}$$

$$\beta_p^{M,k} = \frac{1 - h_p^{M,k} \frac{\xi^{(k-1)}}{\zeta^{(k-1)}}}{\gamma}, \tag{31}$$

$$\mu^{M,k} = \frac{h_v^{M,k}\overline{v} - h_p^{M,k} \frac{\mu^{(k-1)}}{\zeta^{(k-1)}}}{\gamma}. \tag{32}$$

3 Algorithmic Herding

We now show that, in any fixed level-k market environment, AI investors who learn through reinforcement learning endogenously converge to a common demand rule. This common rule coincides with the optimal trading rule of uninformed rational-expectations investors in a rational-expectations equilibrium (REE) of that environment. Throughout this section, we hold fixed the distribution of rational investors with heterogeneous sophistication.

We say that AI investors' learned strategies are consistent with rational expectations if two conditions hold. First, each AI investor's limiting demand rule is optimal given the fundamental information it can infer from prices. Second, the price-formation rule embedded in the AI investors' learned policies, denoted by $\mathcal{P}_{A,k}$, coincides with the actual market-clearing price rule generated by aggregate demand. Thus, when AI investors use $\mathcal{P}_{A,k}^{-1}(p_t)$ to extract payoff information from the equilibrium price p_t, this inference is correct under the true price-formation rule.

We approximate the original trading environment by a sequence of finite, discretized economies that are tailored to the tabular Q-learning problem. In each approximating economy, the admissible AI demand coefficients are restricted to a finite grid $\mathcal{X}_\Delta$, and the exogenous shocks are truncated at finite bounds. The market-clearing equation and the demand schedules of rational investors with heterogeneous sophistication are otherwise kept fixed at their values in the original model.[7] As the grid becomes finer and the truncation bounds

[7]This approximation mirrors the algorithmic environment faced by tabular Q-learning. The algorithm does not choose from a continuous action space; it selects from a finite grid of actions and updates action

grow, these finite approximating economies converge to the original trading environment.

For each fixed approximating economy, the limiting object of AI learning is a mean-field fixed point. This object imposes two consistency requirements. First, taking the population-average AI demand rule as given, each atomistic AI investor learns action values from the trading profits generated by the corresponding market-clearing prices and chooses demand coefficients according to its limiting learning policy. Second, the population-average demand rule induced by these individual limiting policies must coincide with the population-average AI demand rule that was used to generate prices. Thus, the fixed point is not a new economic equilibrium concept separate from market clearing. It is the consistency condition that closes the learning problem in a finite economy with a continuum of atomistic AI investors.

This formulation follows the standard mean-field logic: each individual AI investor is negligible and takes the aggregate state as given, while the aggregate state must be reproduced by the population of individual decisions. In our setting, the relevant aggregate state is the population-average AI demand schedule, because it enters the market-clearing price and therefore determines the realized rewards from which AI investors learn. The rational-expectations demand rule of the original model is then obtained as the limit of these learned mean-field fixed points as the auxiliary economies are refined.

Definition 1 (Discretized bounded environment)**.** *Fix a level-k trading environment. A discretized bounded environment is a finite approximating economy for the AI learning problem in that environment. In this approximating economy, AI investors choose demand coefficients*

$$\left(\mu_{\Delta}^{A,k}, \beta_{p,\Delta}^{A,k}\right) \in \mathcal{X}_{\Delta} \subset [-B_{\mu}, B_{\mu}] \times [-B_{\beta}, B_{\beta}],$$

where $(\mu_{\Delta}^{A,k}, \beta_{p,\Delta}^{A,k})$ *represents the linear demand schedule*

$$x_{\Delta}^{A,k}(p) \equiv \mu_{\Delta}^{A,k} - \beta_{p,\Delta}^{A,k} p.$$

The payoff shock and the noise-trader shock are truncated at finite bounds (B_v, B_z)*. The demand schedules of cursed investors and level-k strategic investors, together with the market-clearing equation, are kept the same as in the original level-k trading environment. Prices and realized trading profits are evaluated using this market-clearing equation under the discretized AI action space and the truncated shock distributions.*

Definition 2 (Mean-field fixed point in a discretized bounded environment)**.** *Fix a discretized bounded environment associated with the level-k trading environment. A mean-field*

values using realized trading profits.

fixed point is a population-average AI demand-coefficient pair

$$\left(\bar{\mu}_{\Delta}^{A,k}, \bar{\beta}_{p,\Delta}^{A,k}\right) \in \mathrm{co}(\mathcal{X}_{\Delta}),$$

where $\mathrm{co}(\mathcal{X}_{\Delta})$ *denotes the convex hull of the finite action grid. Given* $(\bar{\mu}_{\Delta}^{A,k}, \bar{\beta}_{p,\Delta}^{A,k})$*, market-clearing prices are determined by the fixed demand schedules of cursed investors and level-k strategic investors, the truncated shocks, and the population-average AI demand schedule*

$$\bar{x}_{\Delta}^{A,k}(p; \bar{x}_{\Delta}^{A,k}) \equiv \bar{\mu}_{\Delta}^{A,k} - \bar{\beta}_{p,\Delta}^{A,k} p.$$

The pair $(\bar{\mu}_{\Delta}^{A,k}, \bar{\beta}_{p,\Delta}^{A,k})$*, or equivalently the associated average demand schedule* $\bar{x}_{\Delta}^{A,k}(p)$*, is a mean-field fixed point if the following two conditions hold. First, taking as given the price process generated by the market-clearing equation under* $\bar{x}_{\Delta}^{A,k}(p)$*, each atomistic AI investor's limiting Q-learning policy is optimal with respect to the expected trading profits induced by that price process. That is, the limiting Q-values coincide with the expected continuation values of the corresponding demand coefficient pairs in* $\mathcal{X}_{\Delta}$*, evaluated at the market-clearing prices generated by* $\bar{x}_{\Delta}^{A,k}(p)$*. Second, the population average of the demand coefficient pairs induced by these limiting individual policies is exactly* $(\bar{\mu}_{\Delta}^{A,k}, \bar{\beta}_{p,\Delta}^{A,k})$*. Thus, the aggregate AI demand schedule that enters the market-clearing equation is reproduced by the individual learning policies that are optimal under the price process generated by that same aggregate demand schedule.*

3.1 Theoretical Results

The next two results establish the main technical claims on convergence and algorithmic herding for heterogeneous, atomistic Q-learning investors. Proposition 1 studies learning in a fixed discretized bounded environment with level-k strategic investors and shows that heterogeneous AI investors converge to a unique mean-field fixed point. Proposition 2 then studies the approximation step and shows that, as the discretized bounded environments converge to the original level-k trading environment, the corresponding mean-field fixed points converge to the rational-expectations demand rule.

Proposition 1 (Convergence of fixed-environment learning)**.** *Fix a finite population and a discretized bounded environment populated with level-k strategic investors. Suppose the Boltzmann exploration temperatures lie in a compact subset of* $(0, \infty)$*, the learning step sizes satisfy the stochastic-approximation conditions in Appendix A, and the initial state is locally denominator-safe, in the sense that the relevant deterministic and stochastic learning paths remain in a denominator-safe state set. Let* $\left(\bar{\mu}_{\Delta,t}^{A,k}, \bar{\beta}_{\Delta,t}^{A,k}\right)$ *be the learned population-average*

AI demand coefficient pair after period t. Then

$$\lim_{t\to\infty}\left(\bar{\mu}^{A,k}_{\Delta,t},\bar{\beta}^{A,k}_{\Delta,t}\right)=\left(\bar{\mu}^{A,k}_{\Delta},\bar{\beta}^{A,k}_{\Delta}\right),\qquad \textit{almost surely,}$$

where $m^{*}_{\Delta,B}$ is the environment's unique learned mean-field equilibrium.

This proposition is pointwise in a fixed discretized bounded environment. Heterogeneity in exploration temperatures affects the limiting mixed policies over grid actions, while heterogeneity in learning rates affects the speed of adjustment. Under the stated conditions, however, neither source of heterogeneity changes the population-average limit.

Boltzmann exploration makes the population learning operator smooth and contractive on the relevant denominator-safe state set, namely the set of learning states in which aggregate price sensitivity is bounded away from zero. On this set, market clearing determines a finite and stable price process. The contraction implies uniqueness of the mean-field fixed point and almost-sure convergence of the stochastic Q-learning dynamics to that fixed point.

The economic force behind the contraction is market-clearing price feedback. When AI investors collectively tilt their demand toward a particular set of coefficients, the market-clearing price adjusts against that aggregate tilt, reducing the incremental profitability of the deviation. This feedback makes the best-response learning map stable and pulls heterogeneous learners toward the same population-average demand rule. Thus, within any fixed discretized bounded environment, heterogeneous AI investors herd onto a single mean-field fixed point rather than drifting toward persistently different trading rules. The proof is in Appendix A.

Proposition 2 (Mean-field fixed points converge to rational-expectations demand)**.** *Suppose the original level-k trading environment satisfies the regularity condition stated in Appendix A. Fix a compact interval of exploration temperatures and a compact neighborhood $\mathcal{K}$ of the rational-expectations AI demand coefficients*

$$\left(\bar{\mu}^{A,k},\bar{\beta}^{A,k}_{p}\right).$$

Then, for every $\varepsilon>0$, there exist sufficiently large action and shock bounds and a sufficiently fine action grid, chosen independently of the population size and of the particular exploration temperatures, such that, for every finite population of AI investors, the corresponding discretized bounded environment admits a unique mean-field fixed point

$$\left(\bar{\mu}^{A,k}_{\Delta},\bar{\beta}^{A,k}_{p,\Delta}\right)\in\mathcal{K},$$

and this fixed point satisfies

$$\left\| \left(\bar{\mu}_{\Delta}^{A,k}, \bar{\beta}_{p,\Delta}^{A,k} \right) - \left(\bar{\mu}^{A,k}, \bar{\beta}_{p}^{A,k} \right) \right\|_{\infty} < \varepsilon.$$

Equivalently, the mean-field fixed points of sufficiently accurate discretized bounded environments converge locally to the rational-expectations AI demand coefficients of the original level-k trading environment as the action grid becomes finer and the truncation bounds grow.

In the original level-k trading environment, expected trading profit is a strictly concave quadratic function of the AI demand coefficients. Therefore, for any price process induced by aggregate demand, AI investors have a unique optimal demand-coefficient pair. The rational-expectations AI demand coefficients

$$\left(\bar{\mu}^{A,k}, \bar{\beta}_{p}^{A,k} \right)$$

are the fixed point of this optimal-demand map: they are optimal when prices are generated by the aggregate AI demand rule associated with those same coefficients.

The discretized bounded environments approximate this continuous optimization problem. In each discretized environment, Boltzmann exploration smooths the choice over grid actions. The exploration temperature affects how dispersed individual choices are across nearby actions, but it does not alter the limiting target as the grid becomes fine and the shock bounds grow. Appendix A shows that the discretized Boltzmann mean maps, together with their derivatives, converge uniformly on $\mathcal{K}$ to the continuous optimal-demand map. This uniform convergence yields local existence, local uniqueness, and convergence of the mean-field fixed point to the rational-expectations AI demand coefficients.

Economically, strict concavity gives AI investors a well-defined optimal demand rule against the market-clearing price function, while the discretization and truncation errors vanish as the approximating environments are refined. Hence, the demand rule learned by AI investors in the discretized bounded environments converges to the rational-expectations demand rule in the original level-k trading environment.

Theorem 3 (Algorithmic herding toward rational-expectations demand)**.** *Fix a level-k trading environment. Suppose the assumptions of Proposition 2 hold. Then, for every $\varepsilon > 0$, there exist sufficiently large action and shock bounds and a sufficiently fine action grid such that the following statement holds. Consider any discretized bounded environment constructed with these bounds and this grid, indexed by Δ that implicitly includes both the grid fineness and the truncation bounds. Suppose the fixed-environment learning assumptions of Proposition 1 hold in this environment. Let $\left(\bar{\mu}_{\Delta,t}^{A,k}, \bar{\beta}_{p,\Delta,t}^{A,k} \right)$ denote the population-average AI demand*

coefficients induced at time t by the Q-learning policies of the AI investors. Then

$$\lim_{t\to\infty} \left\| \left(\bar{\mu}^{A,k}_{\Delta,t}, \bar{\beta}^{A,k}_{p,\Delta,t}\right) - \left(\bar{\mu}^{A,k}, \bar{\beta}^{A,k}_{p}\right) \right\|_{\infty} < \varepsilon, \qquad \textit{almost surely.}$$

The theorem combines two different convergence statements. The first is a learning result in a fixed discretized bounded environment. By Proposition 1, the stochastic Q-learning dynamics of heterogeneous atomistic AI investors converge almost surely to the unique mean-field fixed point of that environment. This step holds with the action grid and the shock bounds fixed. The second is an approximation result across discretized bounded environments. By Proposition 2, as the action grid becomes sufficiently fine and the shock bounds become sufficiently large, the mean-field fixed point of the discretized bounded environment lies arbitrarily close to the rational-expectations AI demand coefficients of the original level-k trading environment. Hence, after first taking the learning limit within a fixed approximating environment, the learned population-average AI demand can be made arbitrarily close to the rational-expectations AI demand rule by refining the approximating environment. This is an iterated-limit statement. The limit $t \to \infty$ is taken first, holding fixed the discretized bounded environment. The approximation parameters are then chosen so that the limiting mean-field fixed point of that fixed environment lies within the prescribed ε-neighborhood of the rational-expectations AI demand coefficients. The theorem does not assert almost-sure convergence along an arbitrary sequence of simultaneously expanding action grids and growing shock bounds. The result formalizes algorithmic herding. Although AI investors may differ in exploration temperatures, learning rates, and initial Q-values, their population-average demand converges to a single mean-field fixed point in any fixed sufficiently accurate discretized bounded environment. As the approximating environment converges to the original level-k trading environment, this common learned demand rule converges to the rational-expectations AI demand rule. Thus, algorithmic herding and rational-expectations price inference arise as the two components of the same limiting argument: learning selects a common rule, and approximation identifies that rule with rational-expectations demand.

3.2 Limit of Level-k Strategic Reasoning

We next consider the limit of the level-k reasoning hierarchy as $k \to \infty$. This limit corresponds to the case in which strategic investors perform arbitrarily many rounds of reasoning about price formation. The key question is whether the recursive construction of perceived price-formation rules, strategic demand schedules, and the associated learned AI demand rules converges to a well-defined limiting level-∞ trading environment.

Proposition 4 (Convergence of the level-k recursion)**.** *The level-k recursion converges to a unique fixed point for all* $\sigma_M > 0$ *if*

$$\psi_s \leq \frac{27\rho_M^2\sigma_z^2}{8\psi\sigma_v^2}. \tag{33}$$

Proposition 4 provides a sufficient condition under which the hierarchy of level-k strategic reasoning is well defined in the limit. The condition requires the measure of strategic investors, ψ_s, not to be too large relative to noise-trader risk and the strength of fundamental uncertainty. Under this condition, the recursive mapping from level-$(k-1)$ beliefs and demand schedules to level-k demand schedules is stable and admits a unique limiting fixed point. Hence, as $k \to \infty$, the economy converges to a well-defined level-∞ trading environment.

3.3 Simulation Evidence

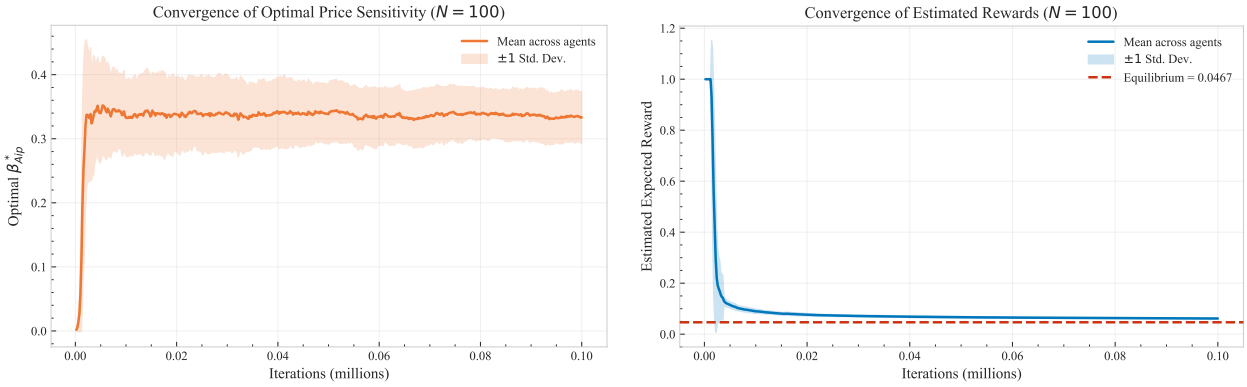
Theorem 3 establishes the convergence and herding result analytically. We now provide complementary numerical evidence by simulating the tabular Q-learning algorithm described in (6). The simulation asks whether a large finite population of independently learning AI investors converges to the rational-expectations AI demand rule predicted by the theory.

We simulate N atomistic AI investors trading in the level-k market environment. The continuum of AI investors in the model is approximated by a large finite population. Each AI investor learns independently from its own realized trading profits. To match the tabular implementation, we restrict the AI action space to a finite grid of price-sensitivity coefficients $\beta_{p,t}^{A,k}$, while fixing the intercept at $\mu_t^{A,k} \equiv 0$. In each period t, an AI investor selects a grid point according to the Boltzmann rule and updates the estimated value of the chosen coefficient toward the realized trading profit generated by that choice.

Market clearing links the learning problems of the AI investors. In each period, the price clears the market given the realized population-average AI demand, the demand schedules of rational investors with heterogeneous sophistication, and noise-trader demand. Thus, although AI investors update their Q-values independently, their realized rewards are jointly determined through the market-clearing price. The simulation therefore directly implements the mechanism behind the theory: individual reinforcement learning interacts through prices, and price feedback can cause heterogeneous AI investors to herd toward the rational-expectations demand rule.

Simulation setup. Each AI investor chooses a price-sensitivity coefficient $\beta_{p,t}^{A,k}$ from a finite grid on $[0, 0.6]$ with mesh size 0.002. To focus on learning over price sensitivity, we

Figure 2: Learning Convergence to the Rational-Expectations Benchmark



(a) Learned price sensitivity $\beta_p^{A,k}$

(b) Estimated expected reward

hold the intercept fixed at $\mu_t^{A,k} \equiv 0$. AI investors use a common Boltzmann exploration temperature κ. For each grid point, an investor updates its estimated value as the running average of realized trading profits, corresponding to the reward-averaging version of tabular Q-learning in this stateless bandit environment.

We simulate $N = 100$ AI investors. The remaining market parameters follow the benchmark calibration: $\sigma_v = 2$, $\sigma_z = 0.2$, $\gamma = 0.2$, asset supply $S = 0$, total rational-investor measure $\psi = 1.5$, and aggregate rational-investor coefficients $\bar{\beta}_p^k = 4.0$ and $\bar{\beta}_\eta^k = 4.5$,.

Results. The learning dynamics converge to the rational-expectations demand rule. Figure 2 shows that the cross-sectional average of the learned price-sensitivity coefficients converges to the rational-expectations coefficient $\beta_p^{A,k}$ in Panel (a), and that the average estimated expected reward converges to its equilibrium level in Panel (b).

This convergence occurs despite heterogeneity in initial value estimates and independent exploration across AI investors. Over the learning horizon, the cross-sectional dispersion of learned coefficients declines and the population concentrates on a common trading rule close to the rational-expectations demand. The simulation therefore provides direct numerical evidence of algorithmic herding: independently learning AI investors, linked only through market-clearing prices, endogenously herd on the same near-rational demand rule. Appendix A.9 reports the full cross-sectional distribution of learned coefficients and documents its concentration over time.

4 Market Sophistication and AI Performance

This section studies AI performance in markets that differ in investor sophistication. First, we compute the expected trading profits of rational investors and AI investors in any given trading environment. Second, we study limits on AI performance in an environment with cursed investors, level-∞ sophisticated investors, and AI investors, comparing AI investors separately with level-∞ investors and cursed investors.

4.1 Expected Trading Profits

Fix a market. Given the investor composition (ψ_c, ψ_s, ϕ) and the sophistication level k of strategic investors, the equilibrium price rule is summarized by ξ, ζ, and μ: aggregate price sensitivity, aggregate signal loading, and the aggregate intercept.[8] The formulas below evaluate a single investor's net expected profit holding the environment fixed. They apply to any admissible rational-investor coefficients and any AI coefficients evaluated against the same price rule.

Proposition 5 (Rational-investor expected profit)**.** *Consider a rational investor with demand coefficients $(\beta_\eta^M, \beta_p^M, \mu^M)$. The investor's single-period net expected profit is*

$$\begin{aligned}\Pi_i^M = \frac{1}{\xi^2}\Big\{&[(\xi-\zeta)\bar{v}-\mu]\left[(\beta_\eta^M\xi-\beta_p^M\zeta)\bar{v}+\xi\mu^M-\beta_p^M\mu\right]\\ &+(\xi-\zeta)(\beta_\eta^M\xi-\beta_p^M\zeta)\sigma_v^2+\beta_p^M\sigma_z^2\Big\}\\ -\frac{\gamma_M}{2\xi^2}\Big\{&\left[(\beta_\eta^M\xi-\beta_p^M\zeta)\bar{v}+\xi\mu^M-\beta_p^M\mu\right]^2\\ &+(\beta_\eta^M\xi-\beta_p^M\zeta)^2\sigma_v^2+\xi^2(\beta_\eta^M)^2\sigma_M^2+(\beta_p^M)^2\sigma_z^2\Big\}. \end{aligned} \tag{34}$$

The formula applies to cursed investors and to strategic level-k investors after substituting their demand coefficients.

Proposition 6 (AI expected profit)**.** *Consider an AI investor with demand coefficients (β_p^A, μ^A). The AI investor's single-period net expected profit is*

$$\begin{aligned}\Pi_j^A = \frac{1}{\xi^2}\Big\{&[(\xi-\zeta)\bar{v}-\mu]\left[\xi\mu^A-\beta_p^A(\zeta\bar{v}+\mu)\right]\\ &-\beta_p^A\zeta(\xi-\zeta)\sigma_v^2+\beta_p^A\sigma_z^2\Big\}\\ -\frac{\gamma_A}{2\xi^2}\Big\{&\left[\xi\mu^A-\beta_p^A(\zeta\bar{v}+\mu)\right]^2+(\beta_p^A)^2(\zeta^2\sigma_v^2+\sigma_z^2)\Big\}. \end{aligned} \tag{35}$$

[8] For ease of notation, we drop superscripts on (ξ, ζ, μ) in this section.

Proofs, including the gross-profit and trading-cost decompositions, are in Appendix C. The first brace in each formula is gross expected trading revenue. Its first product is the unconditional mean component: unconditional average excess payoff times unconditional average demand. The remaining gross-profit terms measure how the investor's position covaries with payoff-relevant shocks. A rational investor earns covariance profits from private-signal exposure β_η^M and from price sensitivity β_p^M. An AI investor has only the price-sensitivity channel, so its covariance term depends on how price movements load on fundamentals and noise trader demand.

The second brace in each formula is the expected quadratic trading cost. For rational investors, costs rise with the mean position, with exposure to fundamental variation, with idiosyncratic private-signal noise, and with positions induced by noise trading. For AI investors, costs rise with the mean position and with the price variation generated by fundamentals and noise trading. These decompositions identify the forces behind the limits below: prices can be informative, but exploiting price information requires taking positions that also absorb noise-trader demand and incur trading costs.

4.2 Limits of AI Performance

We now focus on an environment with cursed investors, level-∞ sophisticated investors, and AI investors. Level-∞ investors provide the rational-expectations benchmark from the previous section: they understand the population composition, the equilibrium price function, and the price-inference problem. Cursed investors use their private signals but fail to extract the information contained in prices. AI investors condition on prices and therefore depend on the information that equilibrium prices aggregate.

Throughout this subsection, impose equal trading costs, $\gamma_M = \gamma_A \equiv \gamma$. This benchmark isolates information-channel limits. Profit rankings then reflect differences in what each investor learns from private signals and prices, rather than differences in the cost of taking a given position.

4.2.1 AI and Level-∞ Investors

Level-∞ investors are the strict benchmark for AI because they have rational expectations and a finer information set. They condition on the same equilibrium price as AI investors and also observe private signals. This comparison therefore asks whether price-based learning can match an investor who understands the price system and observes the signal that helps generate it.

Proposition 7 (Level-∞ investors always outperform AI)**.** *In any nondegenerate convergent environment with cursed investors, level-∞ investors, and AI investors,*

$$\Pi^{M,\infty} - \Pi^{A} = \frac{\tau_M}{2\gamma(\tau_v + \tau_p)(\tau_v + \tau_M + \tau_p)} > 0 \tag{36}$$

for all $\tau_M > 0$, where $\tau_p = \zeta^2 \tau_z$ is equilibrium price precision.

The proof is in Appendix C. Level-∞ investors strictly dominate AI because they observe a richer information set. Both types observe the price and face the same trading cost, but level-∞ investors also observe private signals. Conditional on the price, those signals still lower posterior uncertainty about v, and the lower conditional variance raises the expected square of the optimal trading return. The profit gap in (36) is exactly the value of that extra signal after conditioning on the price signal.

The result also shows why AI performance is limited by the market environment that trains it. AI investors learn from prices, but prices are equilibrium objects produced by other investors' trades. When sophisticated investors infer from prices correctly, their trades move prices closer to fundamentals. This stabilizes prices and removes some of the mispricing variation that an AI investor could otherwise exploit. More sophistication among rational investors therefore has two effects that both work against AI: it creates competitors with better information, and it makes the price signal less profitable as a trading object.

Figure 3 visualizes the uniform limit in (36). The plotted object is $\Pi^{A} - \Pi^{M,\infty}$, so negative values correspond to level-∞ dominance. The heatmap remains negative throughout the parameter region. Changes in private-signal noise and noise-trading volatility change the magnitude of the gap, but not its sign. When private signals are precise, the extra signal is valuable. When prices are precise, AI learns more from prices, but level-∞ investors condition on that same price and retain the additional private-signal channel.

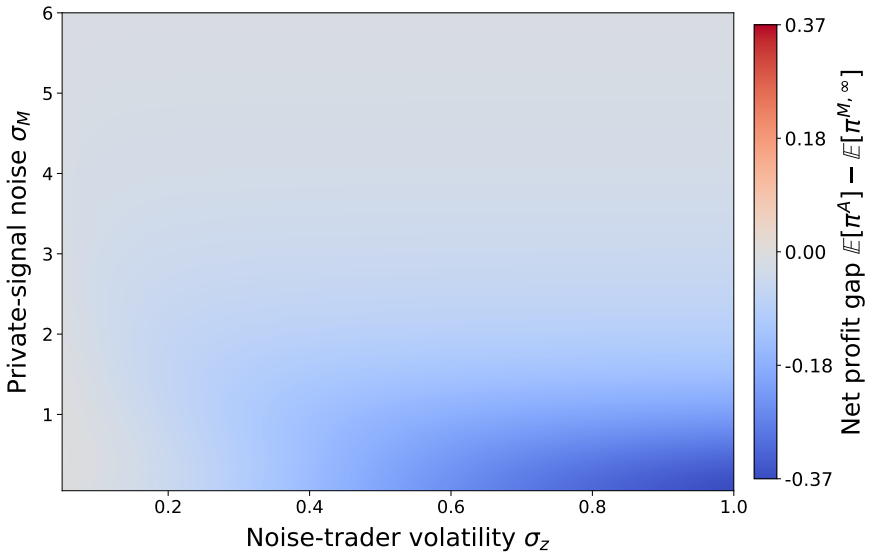
4.2.2 AI and Cursed Investors

The comparison with cursed investors isolates a different AI limit. Cursed investors have biased price inference but receive private signals. AI investors use the price signal unbiasedly but do not receive private signals. The next result reduces the ranking to a comparison between the precision of the private signal and the precision of the price signal.

Lemma 8 (Cursed versus AI: profit comparison)**.** *In any informative nondegenerate environment, under the maintained equal-cost assumption,*

$$\Pi^{M,c} - \Pi^{A} = \frac{h - h_p}{2\gamma\tau_v}, \qquad h \equiv \frac{\tau_M}{\tau_v + \tau_M}, \qquad h_p \equiv \frac{\tau_p}{\tau_v + \tau_p}. \tag{37}$$

Figure 3: A Uniform Limit: Level-∞ Investors Outperform AI



Notes: The figure plots the difference between AI net expected profit and level-∞ investor net expected profit, $\Pi^A - \Pi^{M,\infty}$. The horizontal axis is noise-trading volatility σ_z, and the vertical axis is private-signal noise σ_M. Negative values mean level-∞ investors earn higher net expected profits than AI investors. Parameters are $\bar{v} = 0$, $S = 0$, $\sigma_v = 1$, $\gamma_M = \gamma_A = 1$, $\psi_c = 0.30$, $\psi_s = 0.30$, and $\phi = 0.40$. The heatmap ranges are $\sigma_M \in [0.05, 6]$ and $\sigma_z \in [0.05, 1]$.

Therefore,

$$\Pi^{M,c} > \Pi^A \quad \Longleftrightarrow \quad \tau_M > \tau_p.$$

The proof is in Appendix C. The result compares two usable signal channels. Cursed investors do not learn from prices: a high price lowers the trading surplus they perceive, but it does not raise their estimate of the asset payoff. Their information advantage comes only from the private signal. The weight h is the value of that private-signal channel. It is close to zero when the private signal is mostly noise, and it approaches one when the private signal is very precise. Cursed investors outperform AI only when this private-signal channel is more valuable than the price information they fail to use.

AI investors use the opposite channel. They update from the equilibrium price signal, and h_p measures the value of that price-signal channel. This value is high when informed trading makes prices precise and noise trader demand is small; it is low when prices mostly reflect noise trading. Because both investor types trade the same asset at the same price and face the same trading cost, the ranking does not come from unconditional average mispricing or from the price paid per share. It comes from which signal improves payoff forecasts more. The profit ranking therefore reduces to whether the private signal is more precise than the

equilibrium price signal.

Proposition 9 (Large noise trading: cursed investors dominate)**.** *If*

$$\sigma_z > \frac{\psi \sigma_v}{2\gamma},$$

then, for every $\sigma_M > 0$*,*

$$\Pi^{M,c} > \Pi^{A}.$$

The proof is in Appendix C. Large noise trading makes the price channel too noisy for AI. AI investors benefit from private information only when other investors' trades make prices sufficiently informative. When noise trader demand is volatile, price movements mostly reflect nonfundamental demand rather than payoff information. The equilibrium price then remains less precise than the private signal for every level of private-signal quality.

Cursed investors still misunderstand prices, but the result says that their remaining information source is always better than AI's only information source in this region. If private signals are weak, prices contain little payoff information for AI to learn from. If private signals are strong, noise trader demand is still large enough to keep prices too noisy. Across both cases, the AI price channel is dominated by the private-signal channel.

Proposition 10 (Small noise trading: AI dominance window)**.** *Suppose the objective environment is nondegenerate, the fixed-point price precision* $\tau_p^*(\tau_M)$ *is well defined for each* $\tau_M > 0$*, and the intercept channel converges. If*

$$\sigma_z < \frac{\psi_c \sigma_v}{2\gamma},$$

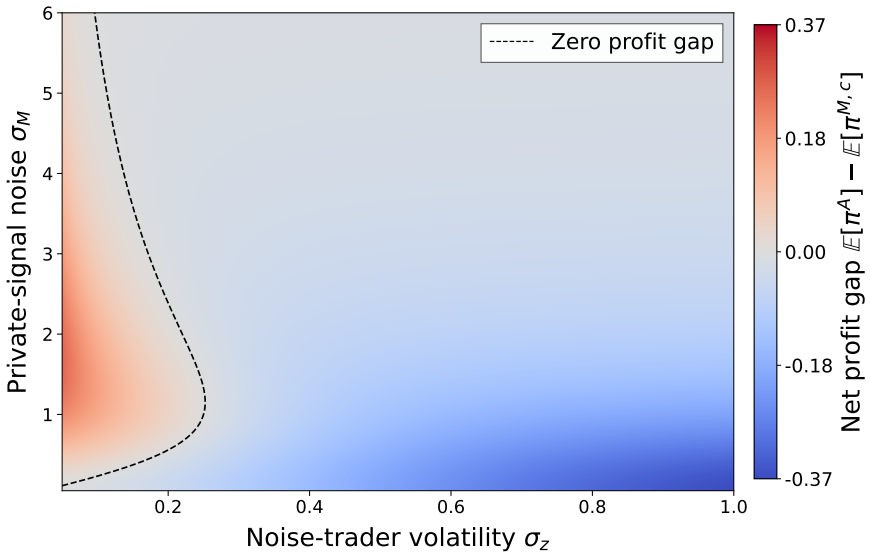
then there exist thresholds $\sigma_M^L < \sigma_M^H$ *such that*

$$\Pi^{M,c} < \Pi^{A} \quad \Longleftrightarrow \quad \sigma_M \in (\sigma_M^L, \sigma_M^H).$$

The proof is in Appendix C. AI dominance over cursed investors is a window, not a universal result. When private signals are very precise, cursed investors can have accurate forecasts of the payoff. They fail to extract information from prices, but this mistake does not offset the value of a highly accurate private signal. AI observes only the price, which aggregates information indirectly through equilibrium trading, so the direct private-signal advantage dominates.

When private signals are very noisy, cursed investors also dominate. Their own signal is weak, but that signal is also the input that makes prices informative. Cursed and level-∞ investors then trade only weakly on payoff information, so prices contain little useful

Figure 4: Cursed Investors and the AI-Dominance Window



Notes: The figure plots the difference between AI net expected profit and cursed-investor net expected profit, $\Pi^A - \Pi^{M,c}$. The horizontal axis is noise-trading volatility σ_z, and the vertical axis is private-signal noise σ_M. Positive values indicate AI dominance over cursed investors. Moving right raises noise-trading volatility and illustrates the large-noise region in which cursed investors dominate. At low σ_z, the positive region appears for intermediate σ_M, illustrating the AI-dominance window. Parameters are $\bar{v} = 0$, $S = 0$, $\sigma_v = 1$, $\gamma_M = \gamma_A = 1$, $\psi_c = 0.30$, $\psi_s = 0.30$, and $\phi = 0.40$. The heatmap ranges are $\sigma_M \in [0.05, 6]$ and $\sigma_z \in [0.05, 1]$.

variation for AI to exploit. AI gains little from observing prices when the trades that form prices carry little information about fundamentals.

AI can outperform cursed investors only in the middle. Private signals must be informative enough to enter prices through informed trading, but not so precise that the direct private-signal channel dominates the price-information channel. Small noise trading makes this middle region possible: prices can aggregate private information without being swamped by noise trader demand. The AI advantage is therefore bounded on both sides by the quality of the information that prices aggregate.

Figure 4 shows the two limits in the same comparison. On the right side of the heatmap, high noise-trading volatility keeps prices too noisy, and cursed investors earn higher profits. On the left side, where noise trading is small, AI dominance appears only for intermediate σ_M. The bottom of the figure corresponds to very precise private signals: cursed investors' direct information is too strong for AI to match. The top corresponds to very noisy private signals: prices inherit too little payoff information from informed trading. The positive region appears between these two forces, where prices aggregate enough private information

for AI but private signals are not so precise that cursed investors dominate directly.

References

Abada, I., and X. Lambin. 2023. Artificial intelligence: Can seemingly collusive outcomes be avoided? *Management Science* 69:5042–65.

Allen, F., S. Morris, and H. S. Shin. 2006. Beauty contests and iterated expectations in asset markets. *The Review of Financial Studies* 19:719–52.

Angeletos, G.-M., and C. Lian. 2023. Dampening general equilibrium: incomplete information and bounded rationality. In *Handbook of Economic Expectations*, 613–45. Elsevier.

Banchio, M., and G. Mantegazza. 2024. Artificial intelligence and spontaneous collusion. *arXiv preprint arXiv:2202.05946* .

Banerjee, S., and M. Szydlowski. 2025. Trading against algorithms: Price dynamics and risk-sharing in a market with q-learners. *Available at SSRN 5380152* .

Benaïm, M. 1999. Dynamics of stochastic approximation algorithms. *Seminaire de Probabilites XXXIII* 28:1–.

Calvano, E., G. Calzolari, V. Denicolo, and S. Pastorello. 2020. Artificial intelligence, algorithmic pricing, and collusion. *American Economic Review* 110:3267–97.

Calvano, E., G. Calzolari, V. Denicoló, and S. Pastorello. 2021. Algorithmic collusion with imperfect monitoring. *International journal of industrial organization* 79:102712–.

Camerer, C. F., T.-H. Ho, and J.-K. Chong. 2004. A cognitive hierarchy model of games. *The Quarterly Journal of Economics* 119:861–98.

Cartea, Á., P. Chang, M. Mroczka, and R. Oomen. 2022. Ai-driven liquidity provision in otc financial markets. *Quantitative Finance* 22:2171–204.

Cartea, Á., P. Chang, and J. Penalva. 2022. Algorithmic collusion in electronic markets: The impact of tick size. *Available at SSRN 4105954* .

Colliard, J.-E., T. Foucault, and S. Lovo. 2022. Algorithmic pricing and liquidity in securities markets. *HEC Paris Research Paper No. FIN-2022-1459* .

Costa-Gomes, M., V. P. Crawford, and B. Broseta. 2001. Cognition and behavior in normal-form games: An experimental study. *Econometrica* 69:1193–235.

Costa-Gomes, M. A., and V. P. Crawford. 2006. Cognition and behavior in two-person guessing games: An experimental study. *American economic review* 96:1737–68.

Crawford, V. P., M. A. Costa-Gomes, and N. Iriberri. 2013. Structural models of nonequilibrium strategic thinking: Theory, evidence, and applications. *Journal of Economic Literature* 51:5–62.

Dou, W. W., I. Goldstein, and Y. Ji. 2025. Ai-powered trading, algorithmic collusion, and price efficiency. *Jacobs Levy Equity Management Center for Quantitative Financial Research Paper, The Wharton School Research Paper* .

———. 2026. Financial market fragility in the era of ai planning. *Available at SSRN 5763222* .

Eyster, E., M. Rabin, and D. Vayanos. 2019. Financial markets where traders neglect the informational content of prices. *The Journal of Finance* 74:371–99.

Farhi, E., and I. Werning. 2019. Monetary policy, bounded rationality, and incomplete markets. *American Economic Review* 109:3887–928.

García-Schmidt, M., and M. Woodford. 2019. Are low interest rates deflationary? a paradox of perfect-foresight analysis. *American Economic Review* 109:86–120.

Gârleanu, N., and L. H. Pedersen. 2013. Dynamic trading with predictable returns and transaction costs. *The Journal of Finance* 68:2309–40.

Grossman, S. J., and J. E. Stiglitz. 1980. On the impossibility of informationally efficient markets. *American Economic Review* 70:393–408.

Han, J., and A. S. Kyle. 2018. Speculative equilibrium with differences in higher-order beliefs. *Management Science* 64:4317–32.

Hansen, K. T., K. Misra, and M. M. Pai. 2021. Frontiers: Algorithmic collusion: Supra-competitive prices via independent algorithms. *Marketing Science* 40:1–12.

Johnson, J. P., A. Rhodes, and M. Wildenbeest. 2023. Platform design when sellers use pricing algorithms. *Econometrica* 91:1841–79.

Kyle, A. S. 1989. Informed speculation with imperfect competition. *The Review of Economic Studies* 56:317–55.

Marimon, R., E. McGrattan, and T. J. Sargent. 1990. Money as a medium of exchange in an economy with artificially intelligent agents. *Journal of Economic dynamics and control* 14:329–73.

Nagel, R. 1995. Unraveling in guessing games: An experimental study. *The American economic review* 85:1313–26.

Routledge, B. R. 1999. Adaptive learning in financial markets. *The Review of Financial Studies* 12:1165–202.

———. 2001. Genetic algorithm learning to choose and use information. *Macroeconomic dynamics* 5:303–25.

Stahl, D. O., and P. W. Wilson. 1994. Experimental evidence on players' models of other players. *Journal of Economic Behavior & Organization* 25:309–27.

———. 1995. On players' models of other players: Theory and experimental evidence. *Games and Economic Behavior* 10:218–54.

Sutton, R. S., and A. G. Barto. 1998. *Reinforcement learning: An introduction*, vol. 1. MIT press Cambridge.

Waltman, L., and U. Kaymak. 2008. Q-learning agents in a cournot oligopoly model. *Journal of Economic Dynamics and Control* 32:3275–93.

Wang, X., and R. Jia. 2021. Mean field equilibrium in multi-armed bandit game with continuous reward. In Z.-H. Zhou, ed., *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21*, 3118–24. International Joint Conferences on Artificial Intelligence Organization. doi:10.24963/ijcai.2021/429. Main Track.

Watkins, C. J. C. H. 1989. Learning from delayed rewards. *PhD thesis, Cambridge University* .

Zhou, H. 2022. Informed speculation with k-level reasoning. *Journal of Economic Theory* 200:105384–.

Appendix

A Proofs of Algorithmic Convergence and Herding

This appendix proves Proposition 1, Proposition 2, and Theorem 3. These results connect the learning of a continuum of heterogeneous AI investors to the rational expectations equilibrium (REE) of the benchmark environment. Only AI investors converge to rational expectations. Human demand is held at the fixed (possibly boundedly rational) coefficients $\bar{\beta}_\eta^M, \bar{\beta}_p^M, \bar{\mu}^M$, while AI investors optimize against the actual market-clearing price function. The benchmark environment is unbounded, and AI demand coefficients can be chosen from the continuous space $\mathbb{R}^2$. The proof has two distinct parts. The first part fixes one discretized bounded environment and shows that heterogeneous AI learning converges to that environment's mean-field equilibrium. The second part refines the auxiliary environments and shows that their local deterministic mean-field equilibria converge to the benchmark coefficient vector.

The argument proceeds in four steps. First, we place the financial market in a discretized bounded environment with a finite grid of AI investor demand choices and truncated shocks. In this environment, the market-clearing feedback from actual AI demand to prices is Lipschitz, and Boltzmann exploration turns the learning operator into a contraction under an explicit condition. Second, the heterogeneous Q-learning recursion is a stochastic approximation to stable learning dynamics, so learned aggregate AI demand converges almost surely to the mean-field equilibrium in the fixed discretized bounded environment. Third, in the benchmark environment, expected trading profit is quadratic, the Boltzmann mean equals the unique best response, and the best-response fixed point gives the benchmark demand coefficient vector. Fourth, deterministic mean-field equilibria in sufficiently accurate discretized bounded environments converge uniformly to this benchmark map as the grid is refined and the bounds grow.

The stochastic learning theorem is pointwise in each fixed discretized bounded environment, while the approximation theorem is deterministic and varies the auxiliary environment. Thus, for any tolerance ε, a sufficiently accurate discretized bounded environment has a local mean-field equilibrium within ε of the REE demand coefficient vector. If that same fixed environment also satisfies the fixed-environment learning conditions stated below, including stochastic localization, heterogeneous AI investors who learn within it converge to aggregate demand coefficients within ε of the REE. [9]

[9]The notation separates three coefficient objects. For any learning state $\mathbf{s}$, $m_{\Delta,B}(\mathbf{s})$ is the demand coefficient pair induced by the policy-implied expected profile in one discretized bounded environment. The

A.1 Discretized bounded environment

The first step is to define a class of discretized bounded environments that approximate the benchmark environment faced by AI investors. The discretized bounded environment keeps the economic structure of the main model—human demand, AI demand, noise-trader demand, and market clearing—but restricts the coefficient choices in the AI demand function to a finite grid and truncates the shocks. These restrictions make prices, demand coefficients, and rewards uniformly bounded, which is what the stochastic approximation argument requires. Later sections refine these restrictions at the deterministic approximation stage.

Brief summary of market structure. Time is discrete, $t = 1, 2, \ldots$. A single risky asset with per capita supply S is traded each period. The asset payoff is $v_t \sim \mathcal{N}(\bar{v}, \sigma_v^2)$, i.i.d. across t. Noise traders submit $z_t \sim \mathcal{N}(0, \sigma_z^2)$, independent of (v_t, e_{it}). Write $\tau_v = \sigma_v^{-2}$, $\tau_M = \sigma_M^{-2}$, $\tau_z = \sigma_z^{-2}$, and $\widehat{\tau}^c = \tau_v + \tau_M$. The maintained primitives satisfy $\gamma_A > 0$, $\gamma_M > 0$, $\phi \geq 0$, and $\sigma_v^2, \sigma_M^2, \sigma_z^2 > 0$.

Human investors. A measure ψ^c of cursed humans and a measure ψ^l of level-k humans trade with linear demands. Each human observes $\eta_i = v + e_i$, where e_i is idiosyncratic with mean zero. In the continuum, the idiosyncratic errors average out, so aggregate human demand loads on v through the average coefficient on η_i. The cursed demand is

$$x_i^{M,c}(\eta_i, p) = \beta_\eta^{M,c}\eta_i - \beta_p^{M,c}p + \mu^{M,c}, \tag{A1}$$

with

$$\beta_\eta^{M,c} = \frac{\tau_M}{\gamma_M \widehat{\tau}^c}, \qquad \beta_p^{M,c} = \frac{1}{\gamma_M}, \qquad \mu^{M,c} = \frac{\tau_v \bar{v}}{\gamma_M \widehat{\tau}^c}. \tag{A2}$$

The level-k demand is

$$x_i^{M,k}(\eta_i, p) = \beta_\eta^{M,k}\eta_i - \beta_p^{M,k}p + \mu^{M,k},$$

where the coefficients come from the level-k recursion in the main text. Those coefficients are finite and fixed.

Aggregate human demand. Define the fixed aggregate human coefficients[10]

$$\bar{\beta}_\eta^M := \psi^c\beta_\eta^{M,c} + \psi^l\beta_\eta^{M,k}, \qquad \bar{\beta}_p^M := \psi^c\beta_p^{M,c} + \psi^l\beta_p^{M,k}, \qquad \bar{\mu}^M := \psi^c\mu^{M,c} + \psi^l\mu^{M,k}. \tag{A3}$$

learned fixed point in that environment is denoted $m^*_{\Delta,B}$. The local deterministic fixed point of the raw-profit Boltzmann coefficient map is denoted $m_{\Delta,B}$, without a star. The benchmark REE demand coefficient vector is m^{REE}.

[10] In the notation of the main text, $\bar{\beta}_\eta^M = \zeta^{env}$.

AI investors and discretized bounded demand choices. A continuum of AI investors of measure ϕ observes only prices and uses linear demand[11]

$$x^A(p) = -\beta_p^A p + \mu^A.$$

For the discretized bounded environment, fix bounds $B_\beta, B_\mu > 0$ and mesh sizes $\Delta_\beta, \Delta_\mu > 0$. Assume the bounds are mesh-compatible, so $2B_\beta/\Delta_\beta$ and $2B_\mu/\Delta_\mu$ are nonnegative integers, and define

$$\mathcal{B}_\Delta = \{-B_\beta, -B_\beta + \Delta_\beta, \ldots, B_\beta\}, \qquad \mathcal{U}_\Delta = \{-B_\mu, -B_\mu + \Delta_\mu, \ldots, B_\mu\}.$$

The arm set is

$$\mathcal{M}_\Delta := \mathcal{B}_\Delta \times \mathcal{U}_\Delta, \qquad M := |\mathcal{M}_\Delta|.$$

Finite number of AI investors. The learning argument in discretized bounded environments represents the AI sector by a finite number of AI investors, $i = 1, \ldots, N$, each carrying an equal share $1/N$ of the measure-ϕ AI continuum. Each investor i may differ in its degree of exploration and learning speed. The number of AI investors N is finite from this section through Section A.7; the limit as $N \to \infty$ is discussed at the close of the appendix in Section A.8. Equivalently, the analysis is uniform in N: the approximation thresholds derived below do not depend on N, so the $N \to \infty$ statement reduces to a term-by-term application of the finite-N result.

Population profiles on the demand grid. A population profile $f \in \Delta^{M-1}$ is a probability distribution over the AI investor demand choices in $\mathcal{M}_\Delta$. It indicates the proportion of AI investors who choose each arm. Each arm $a \in \{1, \ldots, M\}$ corresponds to a pair $(\beta_A(a), \mu_A(a)) \in \mathcal{M}_\Delta$, so

$$|\beta_A(a)| \le B_\beta, \qquad |\mu_A(a)| \le B_\mu.$$

Given a profile f, the mean AI demand coefficients are

$$\bar{\beta}_A(f) := \sum_{a=1}^{M} f(a)\beta_A(a), \qquad \bar{\mu}_A(f) := \sum_{a=1}^{M} f(a)\mu_A(a). \tag{A4}$$

[11] In the appendix, β_A denotes this AI price-sensitivity coefficient; it corresponds to β_p^A in the main-text notation.

Write the corresponding aggregate price sensitivity as

$$\xi(f) := \bar{\beta}_p^M + \phi \bar{\beta}_A(f). \tag{A5}$$

For a coefficient pair $m = (\bar{\beta}_A, \bar{\mu}_A)$, use the same notation

$$\xi(m) := \bar{\beta}_p^M + \phi \bar{\beta}_A.$$

Shock truncation. For the bounded-model learning theorem we also truncate the exogenous shocks:

$$v_t \sim \mathcal{N}(\bar{v}, \sigma_v^2) \text{ truncated to } |v_t - \bar{v}| \le B_v, \qquad z_t \sim \mathcal{N}(0, \sigma_z^2) \text{ truncated to } |z_t| \le B_z.$$

Throughout, truncated means the conditional Gaussian law given the displayed truncation event, with v_t and z_t truncated independently. Hence

$$|v_t| \le V_{\max} := |\bar{v}| + B_v, \qquad |z_t| \le B_z$$

almost surely.

Price and profit induced by a profile. Market clearing gives

$$p(f; v, z) = \frac{\bar{\beta}_\eta^M v + z + \bar{\mu}^M + \phi \bar{\mu}_A(f) - S}{\xi(f)}. \tag{A6}$$

If arm a is chosen, demand and profit are

$$x(a; f; v, z) = -\beta_A(a) p(f; v, z) + \mu_A(a), \tag{A7}$$

$$\pi(a; f; v, z) = (v - p(f; v, z)) x(a; f; v, z) - \frac{\gamma_A}{2} x(a; f; v, z)^2. \tag{A8}$$

Assumption 1 (Local denominator safety in the fixed learning environment)**.** *For the fixed discretized bounded environment used in the learning theorem, there exist $\underline{\xi}_B > 0$ and a nonempty closed set of learning states $\mathcal{S}_B \subset [0,1]^{N \times M}$ such that:*

$$\xi(f(\mathbf{s})) \ge \underline{\xi}_B \quad \textit{for all } \mathbf{s} \in \mathcal{S}_B.$$

The assumption is pointwise for one fixed learning environment.[12]

[12] At this stage it only identifies the local region on which prices and rewards are evaluated. After the ODE is introduced, the same condition can be read as a restriction on initial states whose learning paths

Under Assumption 1, define

$$A_{\max} := |\bar{\beta}_\eta^M| V_{\max} + B_z + |\bar{\mu}^M - S| + \phi B_\mu,$$

$$P_{\max} := \frac{A_{\max}}{\underline{\xi}_B}, \qquad X_{\max} := B_\beta P_{\max} + B_\mu, \qquad \bar{\Pi} := (V_{\max} + P_{\max}) X_{\max} + \frac{\gamma_A}{2} X_{\max}^2. \tag{A9}$$

Then $|p| \le P_{\max}$, $|x| \le X_{\max}$, and $|\pi| \le \bar{\Pi}$ uniformly along $\mathcal{S}_B$. For profiles induced by states in $\mathcal{S}_B$, write the raw expected profit as

$$R_B(f, a) := \mathbb{E}_{v,z}\big[\pi(a; f; v, z)\big], \tag{A10}$$

and the normalized expected reward as

$$r_B(f, a) := \mathbb{E}_{v,z}\left[\frac{\pi(a; f; v, z) + \bar{\Pi}}{2\bar{\Pi}}\right] = \frac{R_B(f, a) + \bar{\Pi}}{2\bar{\Pi}}. \tag{A11}$$

For every $f = f(\mathbf{s})$ with $\mathbf{s} \in \mathcal{S}_B$, this normalized reward belongs to $[0, 1]$.

Definition 3 (Discretized bounded environment)**.** *A discretized bounded environment is the finite-grid, shock-truncated environment described in this section. AI investors choose coefficient pairs from the grid $\mathcal{M}_\Delta$; arm a corresponds to $(\beta_A(a), \mu_A(a))$. Exogenous shocks are truncated at (B_v, B_z), the price clears according to (A6), and payoffs are the normalized trading profits (A11). Local denominator safety in Assumption 1 makes the environment denominator-admissible on $\mathcal{S}_B$. The learning theorem additionally imposes the self-map condition (Assumption 2), the contraction condition, and ODE- and stochastic-localization conditions stated below; when these hold, the environment is learning-admissible on $\mathcal{S}_B$.*

A.2 Policy-implied profile map

The next step links individual learning states to the population demand profile that clears the market. Each AI algorithm stores learned rewards for the grid arms, and Boltzmann choice maps those rewards into arm probabilities. Aggregating these probabilities gives the policy-implied expected population profile. Working with this conditional mean field isolates the price-feedback channel; realized arm draws and realized trading profits are handled later by stochastic approximation.

Recall the N AI investors of equal mass $1/N$ introduced in Section A.1, with exploration temperatures κ_i and learning-speed multipliers c_i. The profile $f(\mathbf{s})$ below is the policy-implied expected profile used by the conditional mean-field learning problem.

remain in this denominator-safe region.

States and policies. AI investor i's reward-estimate vector is $s^i = [s^i(1), \ldots, s^i(M)] \in [0,1]^M$. The full state profile is

$$\mathbf{s} = [s^1, \ldots, s^N] \in [0,1]^{N \times M}.$$

AI investor i uses Boltzmann exploration

$$\sigma_i(s^i, a) = \frac{\exp(s^i(a)/\kappa_i)}{\sum_{j=1}^{M} \exp(s^i(j)/\kappa_i)}. \tag{A12}$$

Expected population profile. The policy-implied expected profile induced by the current state profile is

$$f(\mathbf{s})(a) := \frac{1}{N} \sum_{i=1}^{N} \sigma_i(s^i, a), \qquad a = 1, \ldots, M. \tag{A13}$$

This is the conditional mean-field object used below.

Discount-factor irrelevance. In the bandit environment, heterogeneous discount factors only add an arm-independent continuation constant to fixed-point Q-values.

Lemma 11 (Discount-factor irrelevance)**.** *At any fixed point of Q-learning in the bandit environment with $\rho_i \in [0,1)$, the Boltzmann policy of AI investor i and the mean-field fixed point is independent of $\{\rho_i\}$.*

Proof. The bandit structure makes discounting irrelevant for the policy. At a fixed point, the continuation value is the same no matter which arm is chosen, so discounting adds a common constant to all arm values. Boltzmann probabilities depend only on value differences, and this common term cancels.

The fixed-point Q-values satisfy

$$Q_i^*(a) = r_B(f^*, a) + \rho_i \sum_{a'} \sigma_i(Q_i^*, a') Q_i^*(a').$$

Let

$$V_i^* := \sum_{a'} \sigma_i(Q_i^*, a') Q_i^*(a').$$

Then $Q_i^*(a) = r_B(f^*, a) + \rho_i V_i^*$ for all a, so

$$V_i^* = \sum_{a'} \sigma_i(Q_i^*, a') r_B(f^*, a') + \rho_i V_i^* = \bar{r}_i + \rho_i V_i^*,$$

where $\bar{r}_i := \sum_{a'} \sigma_i(Q_i^*, a') r_B(f^*, a')$. Thus $V_i^* = \bar{r}_i/(1-\rho_i)$ and

$$Q_i^*(a) = r_B(f^*, a) + \frac{\rho_i \bar{r}_i}{1-\rho_i}.$$

The second term is constant across arms, hence cancels in the softmax. □

The preceding argument requires $\rho_i \in [0, 1)$. By Lemma 11, the remainder of Sections A.2–A.5 sets $\rho_i = 0$. Define the harmonic-mean normalized temperature

$$\bar{\kappa}_H := \left(\frac{1}{N} \sum_{i=1}^{N} \frac{1}{\kappa_i} \right)^{-1}, \tag{A14}$$

so $\underline{\kappa} \leq \bar{\kappa}_H \leq \bar{\kappa}$.

Finally define the conditional expected normalized-reward operator

$$G_B(\mathbf{s})^i(a) := r_B(f(\mathbf{s}), a). \tag{A15}$$

The dependence on i enters only through the profile map $f(\mathbf{s})$.

Definition 4 (Mean-field equilibrium in a discretized bounded environment)**.** *Using the usual fixed-point definition of mean-field equilibrium, a state profile* $\mathbf{s}$ *is a mean-field equilibrium of the fixed discretized bounded environment if*

$$\mathbf{s} = G_B(\mathbf{s}).$$

The induced demand-profile equilibrium is $f(\mathbf{s})$*, and the induced aggregate AI demand coefficient pair is*

$$m_{\Delta,B}(\mathbf{s}) := \big(\bar{\beta}_A(f(\mathbf{s})), \bar{\mu}_A(f(\mathbf{s}))\big).$$

A.3 Lipschitz continuity of the normalized expected reward

The contraction proof needs one bound: a small change in the population profile cannot have an arbitrarily large effect on expected profits. The reason is market clearing. A change in the profile changes aggregate AI demand, which changes the price and, in turn, affects each arm's trading profit. In the discretized bounded environment the denominator is bounded away from zero on the relevant learning-state set and all positions are bounded, so this feedback is Lipschitz. The Lipschitz constant derived here is the primitive input into the fixed-point argument.

Lemma 12 (Lipschitz normalized reward)**.** *Under Assumption 1, for each arm a,*

$$|r_B(f,a) - r_B(f',a)| \le \theta_B \|f - f'\|_1 \qquad \forall f = f(\mathbf{s}),\ f' = f(\mathbf{s}'),\ \mathbf{s},\mathbf{s}' \in \mathcal{S}_B,$$

where

$$\theta_B := \frac{L_{\pi,B}}{2\bar{\Pi}}, \qquad L_{\pi,B} := \big((V_{\max} + P_{\max} + \gamma_A X_{\max})B_\beta + X_{\max}\big)L_{p,B}, \tag{A16}$$

and

$$L_{p,B} := \phi\left(\frac{A_{\max}B_\beta}{\underline{\xi}_B^2} + \frac{B_\mu}{\underline{\xi}_B}\right). \tag{A17}$$

Proof. We first bound how much a change in the population profile can move the market-clearing price. We then pass that price movement through a fixed arm's demand and profit. The final step divides by the normalization that maps raw profits into $[0,1]$.

Fix (v,z) and write $p(f) = A(f)/\xi(f)$ with

$$A(f) = \bar{\beta}_\eta^M v + z + \bar{\mu}^M + \phi\bar{\mu}_A(f) - S.$$

For any f, f',

$$\frac{A(f)}{\xi(f)} - \frac{A(f')}{\xi(f')} = \frac{A(f) - A(f')}{\xi(f)} + A(f')\frac{\xi(f') - \xi(f)}{\xi(f)\xi(f')}.$$

Hence

$$|p(f) - p(f')| \le \frac{|A(f) - A(f')|}{\underline{\xi}_B} + \frac{A_{\max}}{\underline{\xi}_B^2}|\xi(f) - \xi(f')|.$$

Since

$$|\bar{\beta}_A(f) - \bar{\beta}_A(f')| \le B_\beta\|f - f'\|_1, \qquad |\bar{\mu}_A(f) - \bar{\mu}_A(f')| \le B_\mu\|f - f'\|_1,$$

we obtain

$$|p(f) - p(f')| \le L_{p,B}\|f - f'\|_1. \tag{A18}$$

Now

$$|x(a;f;v,z) - x(a;f';v,z)| \le B_\beta|p(f) - p(f')|.$$

Using (A8), the bounds $|v| \le V_{\max}$, $|p| \le P_{\max}$, and $|x| \le X_{\max}$ yield

$$\begin{aligned}|\pi(a;f;v,z) - \pi(a;f';v,z)| &\le \big((V_{\max} + P_{\max} + \gamma_A X_{\max})B_\beta + X_{\max}\big)|p(f) - p(f')| \\ &\le L_{\pi,B}\|f - f'\|_1.\end{aligned}$$

Normalization and expectation preserve the bound:

$$|r_B(f,a) - r_B(f',a)| \le \frac{L_{\pi,B}}{2\bar{\Pi}}\|f - f'\|_1 = \theta_B\|f - f'\|_1.$$

□

A.4 Contraction and fixed point

With the reward feedback bounded, the discretized bounded environment has a unique deterministic learning fixed point when exploration is sufficiently smooth. Boltzmann choice maps learned rewards into expected arm shares. Lower temperatures make those shares more sensitive to small value changes; higher temperatures flatten the response. With heterogeneous temperatures, aggregate sensitivity is governed by the average inverse temperature, equivalently the harmonic-mean temperature.

Lemma 13 (Heterogeneous-temperature contraction)**.** *Suppose $r_B(f, a)$ is θ_B-Lipschitz in f with respect to $\|\cdot\|_1$. Then, for every arm a and every pair of state profiles $\mathbf{s}, \tilde{\mathbf{s}} \in \mathcal{S}_B$,*

$$|r_B(f(\mathbf{s}), a) - r_B(f(\tilde{\mathbf{s}}), a)| \leq \frac{2\theta_B}{\bar{\kappa}_H} \|\mathbf{s} - \tilde{\mathbf{s}}\|_\infty.$$

Define the contraction condition as

$$\frac{2\theta_B}{\bar{\kappa}_H} < 1. \tag{A19}$$

Under (A19), G_B has Lipschitz modulus $2\theta_B/\bar{\kappa}_H < 1$ on $\mathcal{S}_B$.

Proof. The contraction comes from the Lipschitz sensitivity of the expected population profile to learned values. For each investor, the reward effect equals the profile Lipschitz constant times the softmax sensitivity of that investor's policy. Summing over investors produces the average inverse temperature, which is the reciprocal of the harmonic-mean temperature.

Fix an arm a. Because $f(\cdot)$ is the expected profile map,

$$\begin{aligned} \|f(\mathbf{s}) - f(\tilde{\mathbf{s}})\|_1 &= \sum_{j=1}^{M} \left| \frac{1}{N} \sum_{i=1}^{N} \big[\sigma_i(s^i, j) - \sigma_i(\tilde{s}^i, j)\big] \right| \\ &\leq \frac{1}{N} \sum_{i=1}^{N} \|\sigma_i(s^i, \cdot) - \sigma_i(\tilde{s}^i, \cdot)\|_1. \end{aligned}$$

The Jacobian of the softmax is

$$\frac{\partial \sigma_i(s, j)}{\partial s(\ell)} = \frac{1}{\kappa_i} \sigma_i(s, j) \big[\delta_{j\ell} - \sigma_i(s, \ell)\big].$$

Hence

$$\sum_{j,\ell} \left| \frac{\partial \sigma_i(s, j)}{\partial s(\ell)} \right| = \frac{2}{\kappa_i} \left(1 - \sum_{\ell=1}^{M} \sigma_i(s, \ell)^2 \right) \leq \frac{2}{\kappa_i}.$$

The mean-value theorem therefore gives

$$\|\sigma_i(s^i,\cdot)-\sigma_i(\tilde{s}^i,\cdot)\|_1 \le \frac{2}{\kappa_i}\|s^i-\tilde{s}^i\|_\infty \le \frac{2}{\kappa_i}\|\mathbf{s}-\tilde{\mathbf{s}}\|_\infty.$$

Therefore

$$\|f(\mathbf{s})-f(\tilde{\mathbf{s}})\|_1 \le \frac{2}{\bar{\kappa}_H}\|\mathbf{s}-\tilde{\mathbf{s}}\|_\infty.$$

Applying Lemma 12 to the two endpoint state profiles, both of which lie in $\mathcal{S}_B$, gives

$$|r_B(f(\mathbf{s}),a)-r_B(f(\tilde{\mathbf{s}}),a)| \le \frac{2\theta_B}{\bar{\kappa}_H}\|\mathbf{s}-\tilde{\mathbf{s}}\|_\infty.$$

□

The Lipschitz bound in Lemma 13 does not give a contraction mapping unless G_B also sends $\mathcal{S}_B$ into itself. Although $G_B(\mathbf{s}) \in [0,1]^{N\times M}$ holds automatically because $r_B \in [0,1]$, Assumption 1 only restricts the denominator at points $\mathbf{s} \in \mathcal{S}_B$, not at the images $G_B(\mathbf{s})$. The reward $r_B(f(\mathbf{s}),\cdot)$ ranks arms by current-profile profitability, and the softmax over those rewards can shift the aggregate $\bar{\beta}_A$ outside the denominator-safe range, so $\xi(f(G_B(\mathbf{s}))) \ge \underline{\xi}_B$ need not follow from Assumption 1 alone. Forward invariance is therefore a separate hypothesis.

Assumption 2 (Forward invariance of $\mathcal{S}_B$ under G_B)**.** *The learning-state set $\mathcal{S}_B$ from Assumption 1 satisfies*

$$G_B(\mathcal{S}_B) \subseteq \mathcal{S}_B.$$

Assumption 2 holds under two leading parametric cases.

(i) Global denominator safety. If $\bar{\beta}_p^M > \phi B_\beta$, then $\xi(f) \ge \bar{\beta}_p^M - \phi B_\beta > 0$ for every $f \in \Delta^{M-1}$. Setting $\mathcal{S}_B = [0,1]^{N\times M}$ and $\underline{\xi}_B = \bar{\beta}_p^M - \phi B_\beta$ satisfies Assumptions 1 and 2 jointly, since $G_B(\mathbf{s}) \in [0,1]^{N\times M}$ for every $\mathbf{s}$.

(ii) Near-equilibrium ball displacement. Suppose $\mathbf{s}_0 \in \mathcal{S}_B$ and $\delta > 0$ satisfy $\{\mathbf{s} : \|\mathbf{s}-\mathbf{s}_0\|_\infty \le \delta\} \subseteq \mathcal{S}_B$ and

$$\|G_B(\mathbf{s}_0)-\mathbf{s}_0\|_\infty \le \left(1-\frac{2\theta_B}{\bar{\kappa}_H}\right)\delta.$$

For any $\mathbf{s}$ with $\|\mathbf{s}-\mathbf{s}_0\|_\infty \le \delta$, the triangle inequality and Lemma 13 give

$$\|G_B(\mathbf{s})-\mathbf{s}_0\|_\infty \le \|G_B(\mathbf{s})-G_B(\mathbf{s}_0)\|_\infty + \|G_B(\mathbf{s}_0)-\mathbf{s}_0\|_\infty \le \frac{2\theta_B}{\bar{\kappa}_H}\delta + \left(1-\frac{2\theta_B}{\bar{\kappa}_H}\right)\delta = \delta.$$

Hence G_B maps the closed ball $\{\mathbf{s} : \|\mathbf{s}-\mathbf{s}_0\|_\infty \le \delta\}$ into itself, and $\mathcal{S}_B$ may be taken to be this ball.

Combining Lemma 13 with Assumption 2, G_B is a contraction mapping on $\mathcal{S}_B$ with modulus $2\theta_B/\bar{\kappa}_H < 1$ under (A19).

Theorem 14 (Unique local deterministic fixed point in the discretized bounded environment)**.** *Under Assumption 2 and (A19), G_B has a unique fixed point $\mathbf{s}^* \in \mathcal{S}_B$. Moreover, every fixed point in $\mathcal{S}_B$ is common across AI investors:*

$$s^{1,*} = \cdots = s^{N,*} \equiv s^* \in [0,1]^M, \qquad s^*(a) = r_B(f^*, a), \tag{A20}$$

where

$$f^*(a) = \frac{1}{N}\sum_{i=1}^{N} \sigma_i(s^*, a). \tag{A21}$$

Proof. Lemma 13 and Assumption 2 together turn the normalized-reward operator into a contraction on a complete local state space, so Banach's theorem gives existence and uniqueness on $\mathcal{S}_B$. The common-value conclusion shows that, once the aggregate profile is fixed, all AI investors face the same expected reward vector. Heterogeneous temperatures affect how they randomize over arms, not the reward vector they learn.

The set $\mathcal{S}_B$ is closed in the complete space $[0,1]^{N\times M}$, so it is complete under $\|\cdot\|_\infty$. Lemma 13 and Assumption 2 make G_B a contraction on $\mathcal{S}_B$, so Banach's fixed-point theorem gives a unique fixed point $\mathbf{s}^*$. Because $G_B(\mathbf{s})^i(a)$ does not depend on i, any fixed point satisfies

$$s^{i,*}(a) = r_B(f(\mathbf{s}^*), a) \qquad \forall i, a,$$

hence all AI investors share the same fixed-point reward vector. □

For later reference, when the hypotheses of Theorem 14 hold, write the coefficient pair induced by its unique deterministic learning fixed point as

$$m^*_{\Delta,B} := \big(\bar{\beta}_A(f(\mathbf{s}^*)), \bar{\mu}_A(f(\mathbf{s}^*))\big).$$

A.5 Mean-field Q-learning with heterogeneous learning speeds

We next show that the fixed point is the limit of the heterogeneous Q-learning through stochastic approximation. Each AI investor updates only the arm it plays, so realized trading profits add martingale noise around the conditional expected drift. Bounded rewards keep the noise controlled, and Boltzmann choice assigns positive probability to every arm. The Q-learning recursion therefore tracks an ODE whose unique global attractor is the contraction fixed point.

Consider the stochastic mean-field Q-learning recursion for the discretized bounded environment.

Assumption 3 (Common-base stepsize schedule)**.** *There exist constants* $0 < c_{\min} \le c_i \le c_{\max} < \infty$ *and a scalar sequence* $a_t > 0$ *such that*

$$0 < \alpha_{ti} = c_i a_t \le 1, \qquad i = 1, \dots, N,\ t \ge 0,$$

and

$$\sum_{t=0}^{\infty} a_t = \infty, \qquad \sum_{t=0}^{\infty} a_t^2 < \infty.$$

Stochastic recursion. At time t, AI investor i draws an arm a_t^i according to $\sigma_i(s_t^i, \cdot)$. Conditional on $\mathbf{s}_t$, the mean-field environment uses the policy-implied expected profile $f(\mathbf{s}_t)$. The realized normalized reward sample for arm a is

$$\tilde{r}_t^i(a) := \frac{\pi(a; f(\mathbf{s}_t); v_t, z_t) + \bar{\Pi}}{2\bar{\Pi}},$$

where (v_t, z_t) are independent truncated exogenous draws. Thus the stochastic recursion below learns from rewards evaluated at the conditional expected mean field $f(\mathbf{s}_t)$. The update rule is

$$s_{t+1}^i(a) = s_t^i(a) + \alpha_{ti} \mathbf{1}\{a_t^i = a\}\big(\tilde{r}_t^i(a) - s_t^i(a)\big). \tag{A22}$$

On any path that remains in $\mathcal{S}_B$, both $s_t^i(a)$ and $\tilde{r}_t^i(a)$ lie in $[0,1]$; with $0 < \alpha_{ti} \le 1$, the update is a convex combination and remains in $[0,1]^{N \times M}$. The theorem below imposes the additional local requirement that the path relevant for the fixed learning problem remains in $\mathcal{S}_B$.

Associated ODE. Define the drift

$$F_i(\mathbf{s})(a) := c_i \sigma_i(s^i, a)\big[r_B(f(\mathbf{s}), a) - s^i(a)\big]. \tag{A23}$$

The limiting ODE is

$$\dot{s}^i(a) = F_i(\mathbf{s})(a), \qquad i = 1, \dots, N,\ a = 1, \dots, M. \tag{A24}$$

Assumption 4 (ODE localization)**.** *The local set* $\mathcal{S}_B$ *is positively invariant for the ODE* (A24)*: every ODE solution with initial condition in* $\mathcal{S}_B$ *exists for all* $t \ge 0$ *and remains in* $\mathcal{S}_B$*.*

Assumption 5 (Stochastic localization)**.** *For the local set $\mathcal{S}_B$ and initial state $\mathbf{s}_0$ used in the fixed-environment learning theorem, define the no-exit event*

$$\mathcal{E}_B(\mathbf{s}_0) := \{\mathbf{s}_t \in \mathcal{S}_B \text{ for all } t \geq 0\}.$$

The stochastic recursion is localized on $\mathcal{S}_B$ from $\mathbf{s}_0$ if

$$\Pr\big(\mathcal{E}_B(\mathbf{s}_0)\big) = 1.$$

[13]

Theorem 15 (Almost-sure convergence in the fixed conditional-mean-field environment)**.** *Suppose Assumptions 1, 2, 3, and 4 hold, and (A19) holds. Start from $\mathbf{s}_0 \in \mathcal{S}_B$. Suppose stochastic localization holds from $\mathbf{s}_0$ as in Assumption 5. Then:*

1. *the ODE (A24) has the unique equilibrium $\mathbf{s}^*$ on $\mathcal{S}_B$ from Theorem 14;*
2. *$\mathbf{s}^*$ is globally exponentially stable for ODE trajectories in $\mathcal{S}_B$;*
3. *the stochastic recursion (A22) converges almost surely:*

$$\mathbf{s}_t \to \mathbf{s}^*;$$

4. *the policy-implied expected profile converges:*

$$f(\mathbf{s}_t) \to f^*.$$

Proof. The learning recursion is a stochastic approximation to the conditional mean-field dynamics. We decompose each update into its conditional mean and a bounded martingale difference. The conditional mean defines an ODE whose fixed points coincide with the contraction fixed points. A standard sup-norm Lyapunov comparison gives global exponential stability. Benaïm's asymptotic-pseudotrajectory theorem then transfers this ODE convergence to the stochastic recursion. This parallels Wang and Jia (2021)'s ODE approximation, while verifying the bounded martingale-difference conditions for the present price-clearing reward.

Step 1: Drift-plus-noise decomposition. Let $\mathcal{F}_t$ be the filtration generated by the history

[13] This is a path condition, not a consequence of the initial state alone. It can be verified by an absorbing state space, projection, or a model-specific no-exit argument; no such device is imposed here.

up to time t. From (A22) and Assumption 3,

$$s^i_{t+1}(a) - s^i_t(a) = a_t\Big(F_i(\mathbf{s}_t)(a) + M^i_{t+1}(a)\Big),$$

where

$$M^i_{t+1}(a) := c_i \mathbf{1}\{a^i_t = a\}\big(\tilde{r}^i_t(a) - s^i_t(a)\big) - c_i\sigma_i(s^i_t, a)\big[r_B(f(\mathbf{s}_t), a) - s^i_t(a)\big].$$

Conditional on $\mathcal{F}_t$, the action draw and the exogenous shocks are independent, and the profile entering the reward is the conditional expected quantity $f(\mathbf{s}_t)$. Hence

$$\mathbb{E}[M^i_{t+1}(a) \mid \mathcal{F}_t] = 0.$$

Also $|\tilde{r}^i_t(a) - s^i_t(a)| \le 1$ and $|r_B(f(\mathbf{s}_t), a) - s^i_t(a)| \le 1$, so

$$|M^i_{t+1}(a)| \le 2c_{\max}.$$

Thus the recursion is a bounded stochastic approximation with drift F and martingale-difference noise.

Step 2: Fixed points of the ODE. If $\mathbf{s}$ is an equilibrium of (A24), then

$$0 = c_i\sigma_i(s^i, a)\big[r_B(f(\mathbf{s}), a) - s^i(a)\big].$$

Since $c_i > 0$ and $\sigma_i(s^i, a) > 0$ for every arm,

$$s^i(a) = r_B(f(\mathbf{s}), a) \qquad \forall i, a.$$

Therefore equilibria of the ODE coincide with fixed points of the operator G_B. By Theorem 14, the ODE has the unique equilibrium $\mathbf{s}^*$ on $\mathcal{S}_B$.

Step 3: Global exponential stability of the ODE on $\mathcal{S}_B$. Because $\mathcal{S}_B$ is compact and $\xi(f(\mathbf{s})) \ge \underline{\xi}_B$ on $\mathcal{S}_B$, continuity gives an open neighborhood of $\mathcal{S}_B$ on which the denominator is still bounded away from zero. On that neighborhood, the softmax is smooth and $r_B(f(\mathbf{s}), a)$ is locally Lipschitz in $\mathbf{s}$; restricting back to the compact local state space gives a Lipschitz drift F. Hence every ODE solution starting in $\mathcal{S}_B$ is unique and absolutely continuous up to any possible exit time, and Assumption 4 keeps the solution in $\mathcal{S}_B$ for all $t \ge 0$. Define

$$V(\mathbf{s}) := \|\mathbf{s} - \mathbf{s}^*\|_\infty.$$

For an absolutely continuous trajectory $t \mapsto \mathbf{s}(t)$, the upper right Dini derivative[14] of $V(\mathbf{s}(t))$ satisfies

$$D^+V(\mathbf{s}(t)) \le \max_{(i,a)\in I(t)} \operatorname{sgn}\big(s^i(t,a) - s^{i,*}(a)\big)\, \dot{s}^i(t,a),$$

where $I(t)$ is the set of active coordinates attaining the sup norm. Fix t and choose an active coordinate $(i^\star, a^\star) \in I(t)$ that attains this maximum. The case $V(\mathbf{s}(t)) = 0$ is immediate, so suppose $V(\mathbf{s}(t)) > 0$. Let

$$\chi^\star := \operatorname{sgn}\big(s^{i^\star}(t,a^\star) - s^{i^\star,*}(a^\star)\big).$$

Then $\chi^\star\big(s^{i^\star}(t,a^\star) - s^{i^\star,*}(a^\star)\big) = V(\mathbf{s}(t))$. Along the ODE,

$$\chi^\star \dot{s}^{i^\star}(t,a^\star) = c_{i^\star}\sigma_{i^\star}(s^{i^\star}(t),a^\star)\chi^\star\Big(r_B(f(\mathbf{s}(t)),a^\star) - s^{i^\star}(t,a^\star)\Big).$$

Using $s^{i^\star,*}(a^\star) = r_B(f^*,a^\star)$,

$$D^+V(\mathbf{s}(t)) \le c_{i^\star}\sigma_{i^\star}(s^{i^\star}(t),a^\star)\Big(\chi^\star\big[r_B(f(\mathbf{s}(t)),a^\star) - r_B(f^*,a^\star)\big] - V(\mathbf{s}(t))\Big).$$

By Lemma 13,

$$|r_B(f(\mathbf{s}(t)),a^\star) - r_B(f^*,a^\star)| \le \frac{2\theta_B}{\bar{\kappa}_H} V(\mathbf{s}(t)).$$

Hence

$$D^+V(\mathbf{s}(t)) \le c_{i^\star}\sigma_{i^\star}(s^{i^\star}(t),a^\star)\left(\frac{2\theta_B}{\bar{\kappa}_H} - 1\right) V(\mathbf{s}(t)).$$

Because $s^i(a) \in [0,1]$ and $\kappa_i \ge \underline{\kappa}$,

$$\sigma_i(s^i,a) \ge \frac{1}{Me^{1/\underline{\kappa}}} =: \sigma_{\min} > 0.$$

Therefore

$$D^+V(\mathbf{s}(t)) \le -\nu V(\mathbf{s}(t)), \qquad \nu := c_{\min}\sigma_{\min}\left(1 - \frac{2\theta_B}{\bar{\kappa}_H}\right) > 0.$$

The scalar comparison inequality yields

$$V(\mathbf{s}(t)) \le e^{-\nu t} V(\mathbf{s}(0)).$$

So the ODE is globally exponentially stable on $\mathcal{S}_B$.

Step 4: Almost-sure convergence of the recursion. By stochastic localization, the realized

[14] The sup norm is not differentiable at ties across active coordinates, so we use its upper right Dini derivative. This is only a convenient way to write the usual sup-norm Lyapunov comparison. Because the sup norm is the maximum of finitely many absolute coordinate values, Danskin's theorem gives the displayed inequality on each D^+V.

recursion remains in $\mathcal{S}_B$ almost surely. The set $\mathcal{S}_B$ is compact and denominator-safe. On this set the drift F is Lipschitz, and the noise sequence is a bounded martingale difference; in particular, it has uniformly bounded second moments. Together with Assumption 3, these are the standard Robbins–Monro/Benaïm hypotheses: the stochastic path is confined to a compact no-exit set, the drift is Lipschitz there, martingale increments have uniformly bounded second moments, $\sum_t a_t = \infty$, and $\sum_t a_t^2 < \infty$. ODE localization ensures that the limiting flow is defined on the compact set visited by the stochastic path. Hence the linear interpolation of $\mathbf{s}_t$ is an asymptotic pseudo-trajectory of (A24); see Benaïm (1999). Because the ODE has the unique globally attracting equilibrium $\mathbf{s}^*$ on $\mathcal{S}_B$, global exponential stability implies that the only internally chain-transitive subset of $\mathcal{S}_B$ is $\{\mathbf{s}^*\}$ Benaïm (1999). Hence the asymptotic-pseudotrajectory limit set equals $\{\mathbf{s}^*\}$, and the recursion converges almost surely to $\mathbf{s}^*$. If localization held only on an event with probability less than one, the same argument would give convergence on that no-exit event. Finally, the map $\mathbf{s} \mapsto f(\mathbf{s})$ is continuous, so $f(\mathbf{s}_t) \to f^*$. □

Remark 1 (Stepsize perturbations for fixed N)**.** *The exact proportionality $\alpha_{ti} = c_i a_t$ keeps notation clean. For any fixed N, the same argument allows $0 < \alpha_{ti} \leq 1$ and $\max_{1 \leq i \leq N} |\alpha_{ti}/a_t - c_i| \to 0$. Writing $\alpha_{ti} = a_t(c_i + \varepsilon_{ti})$, the recursion becomes*

$$s_{t+1}^i(a) - s_t^i(a) = a_t\big(F_i(\mathbf{s}_t)(a) + M_{t+1}^i(a) + \zeta_t^i(a)\big),$$

where

$$\zeta_t^i(a) := \varepsilon_{ti}\mathbf{1}\{a_t^i = a\}\big(\tilde{r}_t^i(a) - s_t^i(a)\big).$$

Since $|\tilde{r}_t^i(a) - s_t^i(a)| \leq 1$, $\sup_{i,a} |\zeta_t^i(a)| \leq \max_i |\varepsilon_{ti}| \to 0$. Thus the weaker condition adds only a uniformly vanishing adapted perturbation, which leaves the asymptotic-pseudotrajectory property intact Benaïm (1999). A joint limit in which N changes would require a separate triangular-array condition that is uniform in N.

Corollary 1 (Initial-state buffer for deterministic local denominator safety)**.** *Fix one discretized bounded environment with deterministic fixed point $\mathbf{s}^*$ and induced coefficient pair*

$$m^* := \big(\bar{\beta}_A(f^*), \bar{\mu}_A(f^*)\big).$$

Define the state-to-denominator Lipschitz constant

$$L_{\xi,s} := \frac{2\phi B_\beta}{\bar{\kappa}_H}. \tag{A25}$$

If, for some $\underline{\xi}_B > 0$,

$$\xi(m^*) - L_{\xi,s}\|\mathbf{s}_0 - \mathbf{s}^*\|_\infty \geq \underline{\xi}_B, \tag{A26}$$

then the closed ball

$$\mathcal{S}_B(\mathbf{s}_0) := \{\mathbf{s} : \|\mathbf{s} - \mathbf{s}^*\|_\infty \leq \|\mathbf{s}_0 - \mathbf{s}^*\|_\infty\}$$

is denominator-safe:

$$\xi(f(\mathbf{s})) \geq \underline{\xi}_B \qquad \forall\, \mathbf{s} \in \mathcal{S}_B(\mathbf{s}_0).$$

If, in addition, (A19) *holds on this ball, then every ODE path starting in* $\mathcal{S}_B(\mathbf{s}_0)$ *remains in* $\mathcal{S}_B(\mathbf{s}_0)$ *and*

$$\xi(f(\mathbf{s}(t;\tilde{\mathbf{s}}_0))) \geq \underline{\xi}_B \qquad \text{for all } t \geq 0.$$

The buffer guarantees only deterministic ODE invariance; realized action and reward shocks can move the stochastic recursion (A22) *outside any ball that the mean ODE never leaves, so Assumption 5 must be imposed separately when applying Theorem 15.*

Proof. For any two state profiles $\mathbf{s}, \mathbf{s}'$, the profile-sensitivity bound proved in Lemma 13 gives

$$\|f(\mathbf{s}) - f(\mathbf{s}')\|_1 \leq \frac{2}{\bar{\kappa}_H}\|\mathbf{s} - \mathbf{s}'\|_\infty.$$

Since $|\beta_A(a)| \leq B_\beta$ for every arm,

$$|\xi(f(\mathbf{s})) - \xi(f(\mathbf{s}'))| \leq \phi B_\beta \|f(\mathbf{s}) - f(\mathbf{s}')\|_1 \leq L_{\xi,s}\|\mathbf{s} - \mathbf{s}'\|_\infty.$$

For any $\mathbf{s} \in \mathcal{S}_B(\mathbf{s}_0)$, this Lipschitz bound gives

$$\xi(f(\mathbf{s})) \geq \xi(f^*) - L_{\xi,s}\|\mathbf{s} - \mathbf{s}^*\|_\infty \geq \xi(m^*) - L_{\xi,s}\|\mathbf{s}_0 - \mathbf{s}^*\|_\infty \geq \underline{\xi}_B.$$

Thus $\mathcal{S}_B(\mathbf{s}_0)$ is denominator-safe. If the contraction condition holds on this ball, the comparison argument in the proof of Theorem 15 gives, for any initial condition $\tilde{\mathbf{s}}_0 \in \mathcal{S}_B(\mathbf{s}_0)$ and for some $\nu > 0$,

$$\|\mathbf{s}(t;\tilde{\mathbf{s}}_0) - \mathbf{s}^*\|_\infty \leq e^{-\nu t}\|\tilde{\mathbf{s}}_0 - \mathbf{s}^*\|_\infty \leq \|\mathbf{s}_0 - \mathbf{s}^*\|_\infty.$$

An exit-time argument justifies applying the comparison up to the first possible boundary exit and then rules that exit out. Hence every ODE path starting in the ball remains in $\mathcal{S}_B(\mathbf{s}_0)$, and the denominator bound holds along each such path. □

The corollary is stated for one fixed learning environment. The REE-neighborhood version of the buffer is stated after the compact rectangle $\mathcal{K}$ has been introduced in Section A.7; see Corollary 3.

Corollary 2 (Convergence of learned coefficients in the discretized bounded environment)**.** *Under the assumptions of Theorem 15, define the learned mean coefficient pair*

$$m_t := \big(\bar{\beta}_A(f(\mathbf{s}_t)), \bar{\mu}_A(f(\mathbf{s}_t))\big), \qquad m^* := \big(\bar{\beta}_A(f^*), \bar{\mu}_A(f^*)\big).$$

Then $m_t \to m^*$ almost surely. Moreover, if raw temperatures are defined by $\vartheta_i = 2\bar{\Pi}\kappa_i$, then m^* is a fixed point of the raw Boltzmann coefficient map for the discretized bounded environment, defined on coefficient pairs with positive denominator by

$$\mathcal{T}_B^{raw}(m) := \frac{1}{N}\sum_{i=1}^{N}\sum_{a=1}^{M}\frac{\exp(R_B(m,a)/\vartheta_i)}{\sum_{k=1}^{M}\exp(R_B(m,k)/\vartheta_i)}(\beta_A(a), \mu_A(a)),$$

where $R_B(m,a)$ denotes the truncated raw expected profit obtained from (A10) after replacing $(\bar{\beta}_A(f), \bar{\mu}_A(f))$ by the candidate pair m in the price formula.

Proof. The corollary translates state profile convergence into the demand coefficient convergence. Since average demand coefficients are continuous functions of the arm probabilities, convergence of the learned profile implies convergence of aggregate coefficients. The raw-profit fixed-point statement then follows from the affine relation between normalized rewards and raw profits: softmax choice is invariant to common payoff shifts, with the multiplicative rescaling absorbed into temperature.

The convergence $m_t \to m^*$ follows from Theorem 15 and continuity of the finite-grid averaging maps in (A4). At the fixed point, $s^*(a) = r_B(f^*, a)$ for every arm and $R_B(f^*, a) = R_B(m^*, a)$. Because $r_B(f^*, a) = R_B(m^*, a)/(2\bar{\Pi}) + 1/2$, with the additive $1/2$ and the multiplicative $1/(2\bar{\Pi})$ both arm-independent, the softmax at normalized temperature κ_i coincides with the raw-profit softmax at temperature $\vartheta_i = 2\bar{\Pi}\kappa_i$ (Lemma 17). Thus f^* is the raw Boltzmann profile generated by m^*, and averaging the arm coefficients under that profile gives $m^* = \mathcal{T}_B^{raw}(m^*)$. □

Scope of the fixed-environment learning theorem. Theorem 15 and Corollary 2 are stated for one fixed finite action grid, one fixed shock truncation, and initial states in a set $\mathcal{S}_B$ satisfying Assumptions 1 and 2, the contraction condition (A19), ODE localization, and stochastic localization. In such a fixed discretized bounded environment,

$$m_t^{\Delta,B} \to m_{\Delta,B}^* \qquad \text{almost surely.}$$

Thus the learning theorem gives pointwise dynamic convergence for each learning-admissible discretized bounded environment.

A.6 Continuous-action benchmark and equilibrium identification

The discretized bounded environment has a well-defined learning limit. In the benchmark environment, expected profit from an AI demand rule is a strictly concave quadratic in the demand coefficients. Raw-profit Boltzmann choice is therefore Gaussian around the unique best response. Temperatures scale dispersion around the optimum, but they do not change the mean.[15] The aggregate Boltzmann mean is consequently the best-response map, and its fixed point gives the benchmark coefficient vector.

Admissible aggregate pairs. For $m = (\bar{\beta}_A, \bar{\mu}_A) \in \mathbb{R}^2$, write

$$\zeta := \bar{\beta}_\eta^M, \qquad \xi(m) := \bar{\beta}_p^M + \phi\bar{\beta}_A, \qquad \mu_p(m) := \bar{\mu}^M + \phi\bar{\mu}_A - S.$$

An aggregate pair m is *admissible* if $\xi(m) > 0$. All formulas below are understood on the admissible set.

For explicit equilibrium identification, impose the benchmark regularity condition

$$\gamma_A > 0, \qquad \sigma_v^2 > 0, \qquad \sigma_z^2 > 0, \qquad \phi \geq 0, \qquad \bar{\beta}_\eta^M > 0, \qquad \gamma_A\bar{\beta}_p^M + \phi > 0. \tag{A27}$$

This regularity condition focuses attention on the financially admissible equilibrium: Theorem 16 shows that it implies $\xi(m^{REE}) > 0$.

Continuous-action expected profit. Let $(\beta, \mu) \in \mathbb{R}^2$ be a generic AI action and let $m = (\bar{\beta}_A, \bar{\mu}_A)$ be admissible. The price has the same form as in the main model:

$$p_m = \frac{\zeta}{\xi(m)}\, v + \frac{1}{\xi(m)}\, z + \frac{\mu_p(m)}{\xi(m)}. \tag{A28}$$

Write

$$m_p(m) := \mathbb{E}[p_m], \qquad q(m) := \mathbb{E}[p_m^2], \qquad d(m) := \bar{v} - m_p(m),$$

and

$$E_{vp}(m) := \mathbb{E}[(v - p_m)p_m]. \tag{A29}$$

The raw expected profit of action (β, μ) against aggregate pair m is

$$R(\beta, \mu; m) = \mu d(m) - \beta E_{vp}(m) - \frac{\gamma_A}{2}\big(\mu^2 - 2\mu\beta m_p(m) + \beta^2 q(m)\big). \tag{A30}$$

[15] The relevant temperature parameter in this section is the raw-profit temperature.

Raw-temperature Boltzmann densities. Fix raw-profit temperatures $\vartheta_i \in [\underline{\vartheta}, \bar{\vartheta}] \subset (0, \infty)$, $i = 1, \ldots, N$. For an admissible aggregate pair m, define AI investor i's continuous-action Boltzmann density with respect to Lebesgue measure in the chosen (β, μ) coordinates[16]:

$$\varphi_i(\beta, \mu \mid m) \propto \exp\left(\frac{R(\beta, \mu; m)}{\vartheta_i}\right). \tag{A31}$$

Theorem 16 (Continuous-action benchmark equilibrium)**.** *Suppose the benchmark regularity condition* (A27) *holds, and fix raw temperatures $\vartheta_i \in (0, \infty)$. Then:*

1. *for every admissible aggregate pair m, the function $(\beta, \mu) \mapsto R(\beta, \mu; m)$ is a strictly concave quadratic and has a unique maximizer, denoted $T(m) = (\beta^*(m), \mu^*(m))$;*

2. *for every AI investor i, $\varphi_i(\cdot \mid m)$ is Gaussian. Its mean is $T(m)$ and its covariance is $\vartheta_i H(m)^{-1}$, where*

$$H(m) := \gamma_A \begin{pmatrix} q(m) & -m_p(m) \\ -m_p(m) & 1 \end{pmatrix};$$

3. *the aggregate Boltzmann mean map*

$$\mathcal{T}(m) := \frac{1}{N} \sum_{i=1}^{N} \int_{\mathbb{R}^2} (\beta, \mu)\, \varphi_i(\beta, \mu \mid m)\, d\beta\, d\mu$$

equals $T(m)$ for every admissible m, so it is independent of the temperature vector $\{\vartheta_i\}$;[17]

4. *the self-consistency equation*

$$m = \mathcal{T}(m)$$

has the unique admissible solution $m^{REE} = (\bar{\beta}_A^*, \bar{\mu}_A^*)$, where

$$\bar{\beta}_A^* = \frac{(\bar{\beta}_\eta^M)^2 \sigma_v^2 + \sigma_z^2 - \bar{\beta}_p^M \bar{\beta}_\eta^M \sigma_v^2}{\gamma_A\big((\bar{\beta}_\eta^M)^2 \sigma_v^2 + \sigma_z^2\big) + \phi \bar{\beta}_\eta^M \sigma_v^2}, \tag{A32}$$

$$\bar{\mu}_A^* = \frac{\bar{\beta}_\eta^M \sigma_v^2 (S - \bar{\mu}^M) + \sigma_z^2 \bar{v}}{\gamma_A\big((\bar{\beta}_\eta^M)^2 \sigma_v^2 + \sigma_z^2\big) + \phi \bar{\beta}_\eta^M \sigma_v^2}. \tag{A33}$$

[16] The coordinate/base-measure convention is part of the approximation: the finite grids below use equal rectangular cell weights in these same coordinates. A reparameterization of (β, μ) would generate a different base measure and Boltzmann mean; the present convention matches the discretization in Section A.7, which assigns equal weight to each rectangular cell in (β, μ) coordinates.

[17] Heterogeneous raw temperatures change the covariance matrices of the Gaussian Boltzmann policies, but not their means. Accordingly, they affect cross-sectional dispersion around the optimum, not the population mean.

Its implied denominator is

$$\xi(m^{REE}) = \frac{(\gamma_A \bar{\beta}_p^M + \phi)\big((\bar{\beta}_\eta^M)^2\sigma_v^2 + \sigma_z^2\big)}{\gamma_A\big((\bar{\beta}_\eta^M)^2\sigma_v^2 + \sigma_z^2\big) + \phi\bar{\beta}_\eta^M\sigma_v^2}. \tag{A34}$$

Under (A27), this quantity is strictly positive, so m^{REE} is admissible and therefore is the unique admissible solution.

Proof. Gaussian shocks and the affine price rule make expected profit quadratic in (β, μ). Positive price variance makes the Hessian negative definite and gives a unique best response. Completing the square shows that each Boltzmann density is Gaussian around this best response; temperature changes only the covariance matrix. The aggregate Boltzmann mean therefore equals the best-response map, and the fixed point reduces to a two-equation linear system that matches the benchmark coefficients.

Fix an admissible aggregate pair m and suppress its dependence in the notation. From (A28),

$$\mathrm{Var}(p_m) = \left(\frac{\zeta}{\xi(m)}\right)^2 \sigma_v^2 + \frac{\sigma_z^2}{\xi(m)^2} > 0, \qquad q(m) = m_p(m)^2 + \mathrm{Var}(p_m).$$

The Hessian of R with respect to (β, μ) is $-H(m)$, where

$$H(m) = \gamma_A \begin{pmatrix} q(m) & -m_p(m) \\ -m_p(m) & 1 \end{pmatrix}.$$

Because

$$\det(H(m)) = \gamma_A^2\big(q(m) - m_p(m)^2\big) = \gamma_A^2\mathrm{Var}(p_m) > 0, \qquad \mathrm{tr}(H(m)) = \gamma_A\big(q(m) + 1\big) > 0,$$

$H(m)$ is positive definite (for a symmetric 2×2 matrix, positive determinant and positive trace are jointly equivalent to positive definiteness). Hence $R(\cdot, \cdot; m)$ is a strictly concave quadratic and has a unique maximizer $T(m) = (\beta^*(m), \mu^*(m))$. This establishes part (1).

The first-order conditions are

$$\frac{\partial R}{\partial \mu} = d(m) - \gamma_A\big(\mu - \beta m_p(m)\big) = 0, \tag{A35}$$

$$\frac{\partial R}{\partial \beta} = -E_{vp}(m) + \gamma_A\big(\mu m_p(m) - \beta q(m)\big) = 0. \tag{A36}$$

Because R is quadratic and $H(m)$ is positive definite,

$$R(\beta,\mu;m) = R(\beta^*(m),\mu^*(m);m) - \frac{1}{2}\begin{pmatrix}\beta-\beta^*(m)\\ \mu-\mu^*(m)\end{pmatrix}^\top H(m)\begin{pmatrix}\beta-\beta^*(m)\\ \mu-\mu^*(m)\end{pmatrix}.$$

Exponentiating shows that $\varphi_i(\cdot \mid m)$ is Gaussian with mean $T(m)$ and covariance $\vartheta_i H(m)^{-1}$. This establishes part (2).

Part (3) follows directly from part (2): each Gaussian $\varphi_i(\cdot \mid m)$ has the common mean $T(m)$ and an investor-specific covariance $\vartheta_i H(m)^{-1}$. Averaging Gaussians with a common mean preserves that mean, so $\mathcal{T}(m) = T(m)$ for every admissible m.

For part (4), impose self-consistency:

$$\bar\beta_A = \beta^*(m), \qquad \bar\mu_A = \mu^*(m).$$

Solving (A35) for $\mu = \beta m_p(m) + d(m)/\gamma_A$ and substituting into (A36) gives $\beta^*(m)\gamma_A \mathrm{Var}(p_m) = \mathrm{Var}(p_m) - \mathrm{Cov}(v,p_m)$, using $q(m) - m_p(m)^2 = \mathrm{Var}(p_m)$ and $d(m)m_p(m) - E_{vp}(m) = \mathrm{Var}(p_m) - \mathrm{Cov}(v,p_m)$. Hence the first-order conditions imply

$$\beta^*(m) = \frac{\mathrm{Var}(p_m) - \mathrm{Cov}(v,p_m)}{\gamma_A \mathrm{Var}(p_m)}, \qquad \mu^*(m) = \beta^*(m)m_p(m) + \frac{d(m)}{\gamma_A}.$$

Substituting the affine price moments into these expressions gives

$$\beta^*(m) = \frac{(\bar\beta_\eta^M)^2\sigma_v^2 + \sigma_z^2 - \bar\beta_\eta^M\bar\beta_p^M\sigma_v^2 - \phi\bar\beta_\eta^M\sigma_v^2\bar\beta_A}{\gamma_A\big((\bar\beta_\eta^M)^2\sigma_v^2 + \sigma_z^2\big)},$$

$$\mu^*(m) = \frac{\bar\beta_\eta^M\sigma_v^2(S-\bar\mu^M) - \phi\bar\beta_\eta^M\sigma_v^2\bar\mu_A + \sigma_z^2\bar v}{\gamma_A\big((\bar\beta_\eta^M)^2\sigma_v^2 + \sigma_z^2\big)}.$$

Thus the fixed-point conditions form a linear system in $(\bar\beta_A, \bar\mu_A)$. Solving that system yields the unique algebraic solution (A32)–(A33). Substituting (A32) into $\xi(m) = \bar\beta_p^M + \phi\bar\beta_A$ gives (A34). The denominator in (A34) is strictly positive because $\gamma_A > 0$, $(\bar\beta_\eta^M)^2\sigma_v^2 + \sigma_z^2 > 0$, $\phi \geq 0$, and $\bar\beta_\eta^M > 0$. Under (A27), the numerator is also strictly positive, so $\xi(m^{REE}) > 0$. Hence the algebraic fixed point is admissible, and uniqueness among admissible fixed points follows from uniqueness of the algebraic solution. □

Equivalence to the main-text benchmark. The closed-form solution (A32)–(A33) coincides with the benchmark coefficients from the main text,

$$\beta_p^{A,env} = \frac{1 - h_p^{A,env} \frac{\bar{\beta}_p^M}{\bar{\beta}_\eta^M}}{\gamma_A + h_p^{A,env} \frac{\phi}{\bar{\beta}_\eta^M}}, \tag{A37}$$

$$\mu^{A,env} = \frac{(1 - h_p^{A,env})\bar{v} - h_p^{A,env} \frac{\bar{\mu}^M - S}{\bar{\beta}_\eta^M}}{\gamma_A + h_p^{A,env} \frac{\phi}{\bar{\beta}_\eta^M}}, \tag{A38}$$

where

$$h_p^{A,env} = \frac{(\bar{\beta}_\eta^M)^2 \sigma_v^2}{(\bar{\beta}_\eta^M)^2 \sigma_v^2 + \sigma_z^2}$$

and $\bar{\beta}_\eta^M$, $\bar{\beta}_p^M$, $\bar{\mu}^M$ are the aggregate human-demand coefficients defined in (A3). Substituting the definition of $h_p^{A,env}$ into (A37), multiplying numerator and denominator by $\bar{\beta}_\eta^M\big((\bar{\beta}_\eta^M)^2\sigma_v^2 + \sigma_z^2\big)$, and using $h_p^{A,env}\big((\bar{\beta}_\eta^M)^2\sigma_v^2 + \sigma_z^2\big) = (\bar{\beta}_\eta^M)^2\sigma_v^2$ recovers (A32); the analogous calculation for $\mu^{A,env}$ recovers (A33).

A.7 Deterministic approximation to the REE

The final approximation step connects fixed-environment learning to the benchmark environment. The argument has three links. First, the learning theorem uses normalized rewards to keep Q-values in a compact state space, while the benchmark model uses raw profits to exploit the quadratic structure; in any fixed discretized bounded environment, the two Boltzmann rules coincide after a deterministic temperature rescaling. Second, in a compact admissible neighborhood of the benchmark fixed point, the raw-profit Boltzmann mean maps generated by sufficiently accurate discretized bounded environments converge uniformly to the continuous Gaussian map. Third, a local fixed-point argument identifies the unique nearby deterministic mean-field equilibrium and shows that it lies close to the benchmark coefficient vector.

Lemma 17 (Affine rescaling in a discretized bounded environment)**.** *Fix one discretized bounded environment and let* $\bar{\Pi}$ *be the bound from* (A9)*. For any raw temperature* $\vartheta_i > 0$*, set*

$$\kappa_i = \frac{\vartheta_i}{2\bar{\Pi}}.$$

Then, for every arm profile f *and every arm* a*,*

$$\frac{\exp(r_B(f,a)/\kappa_i)}{\sum_k \exp(r_B(f,k)/\kappa_i)} = \frac{\exp(R_B(f,a)/\vartheta_i)}{\sum_k \exp(R_B(f,k)/\vartheta_i)}.$$

Hence normalized-reward Boltzmann policies with temperature κ_i coincide with raw-profit Boltzmann policies with temperature ϑ_i.

Proof. By (A11) and $\kappa_i = \vartheta_i/(2\bar{\Pi})$,

$$\frac{r_B(f,a)}{\kappa_i} = \frac{R_B(f,a)}{\vartheta_i} + \frac{\bar{\Pi}}{\vartheta_i}.$$

The second term is common across arms. Thus the normalized-reward softmax is the raw-profit softmax multiplied in numerator and denominator by the same factor $\exp(\bar{\Pi}/\vartheta_i)$, which cancels. □

Definition 5 (Deterministic Boltzmann map in the discretized bounded environment)**.** *Fix raw temperatures $\vartheta_i \in [\underline{\vartheta}, \bar{\vartheta}]$. For a discretized bounded environment with action grid $\mathcal{M}_\Delta$ and shock truncation (B_v, B_z), write $m = (\bar{\beta}_A, \bar{\mu}_A)$ for a candidate aggregate pair and define the truncated raw expected profit*

$$R_{\Delta,B}(\beta, \mu; m) := \mathbb{E}^{tr}_{v,z}\Big[(v - p^{tr}_m)\big(-\beta p^{tr}_m + \mu\big) - \frac{\gamma_A}{2}\big(-\beta p^{tr}_m + \mu\big)^2\Big], \tag{A39}$$

where $\mathbb{E}^{tr}_{v,z}$ is expectation under the independent conditionally truncated Gaussian shocks described in Section A.1, and

$$p^{tr}_m := \frac{\bar{\beta}^M_\eta v + z + \bar{\mu}^M + \phi\bar{\mu}_A - S}{\bar{\beta}^M_p + \phi\bar{\beta}_A}.$$

For AI investor i, define the raw Boltzmann probabilities in this discretized bounded environment

$$\omega^{\Delta,B}_i(a \mid m) := \frac{\exp(R_{\Delta,B}(\beta_A(a), \mu_A(a); m)/\vartheta_i)}{\sum_{k=1}^M \exp(R_{\Delta,B}(\beta_A(k), \mu_A(k); m)/\vartheta_i)}. \tag{A40}$$

The associated mean-coefficient map is

$$\mathcal{T}^{N,\vartheta}_{\Delta,B}(m) := \frac{1}{N}\sum_{i=1}^N \sum_{a=1}^M \omega^{\Delta,B}_i(a \mid m)\,(\beta_A(a), \mu_A(a)). \tag{A41}$$

Given a compact admissible neighborhood $\mathcal{K}$ of m^{REE}, a local deterministic mean-field equilibrium in the discretized bounded environment is a pair $m_{\Delta,B} \in \mathcal{K}$ satisfying $m_{\Delta,B} = \mathcal{T}^{N,\vartheta}_{\Delta,B}(m_{\Delta,B})$. For a fixed bounded environment, comparison with a normalized-reward learning fixed point uses the temperature match $\vartheta_i = 2\bar{\Pi}\kappa_i$.

Throughout the deterministic approximation step, raw temperatures are kept in a common compact interval $[\underline{\vartheta}, \bar{\vartheta}] \subset (0, \infty)$, independently of the discretized environment and the

finite population size. Therefore, when a particular discretized bounded environment is interpreted as a normalized-reward learning problem, the matched normalized temperatures are environment-specific: $\kappa_i = \vartheta_i/(2\bar{\Pi})$.

For every compact set $\mathcal{K}_0 \subset \{m \in \mathbb{R}^2 : \xi(m) > 0\}$, the map $\mathcal{T}^{N,\vartheta}_{\Delta,B}$ is continuous on $\mathcal{K}_0$. Once the denominator is bounded away from zero on $\mathcal{K}_0$, the truncated expected profit in (A39) is continuous in m, and the finite-set softmax in (A40) is continuous in its payoff vector.

Remark 2 (Raw learned map equals the deterministic grid map)**.** *Fix one discretized bounded environment, raw temperatures $\vartheta_i = 2\bar{\Pi}\kappa_i$, and an aggregate coefficient pair with $\xi(m) > 0$. Because $R_B(m, a)$ in Corollary 2 and $R_{\Delta,B}(\beta_A(a), \mu_A(a); m)$ in (A39) evaluate the same trading-profit expression at the same coefficient pair and grid arm under the same independently truncated shock law, they agree arm by arm. Hence $\mathcal{T}^{raw}_B(m) = \mathcal{T}^{N,\vartheta}_{\Delta,B}(m)$ on the admissible set.*

Compact neighborhood of the benchmark fixed point. Let m^{REE} be the unique benchmark fixed point from Theorem 16. Under (A27), $\xi(m^{REE}) > 0$ by (A34), and the continuous map T from Theorem 16 satisfies $T(m^{REE}) = m^{REE}$. Since ξ is continuous and $\xi(m^{REE}) > 0$, choose a compact rectangle

$$\mathcal{K} = [\bar{\beta}^*_A - h_\beta, \bar{\beta}^*_A + h_\beta] \times [\bar{\mu}^*_A - h_\mu, \bar{\mu}^*_A + h_\mu] \tag{A42}$$

with $m^{REE} \in \operatorname{int} \mathcal{K}$ such that

$$\inf_{m \in \mathcal{K}} \xi(m) \geq \underline{\xi}_{\mathcal{K}} > 0. \tag{A43}$$

No self-map condition is imposed on T over $\mathcal{K}$. The first-order-condition simplification in Theorem 16 gives the linear identity

$$T(m) - m^{REE} = -\lambda_T\big(m - m^{REE}\big), \qquad \lambda_T := \frac{\phi\bar{\beta}^M_\eta \sigma^2_v}{\gamma_A\big((\bar{\beta}^M_\eta)^2\sigma^2_v + \sigma^2_z\big)} \geq 0, \tag{A44}$$

coordinate by coordinate. This identity makes m^{REE} an isolated, nondegenerate zero of $m - T(m)$; it does not require $\lambda_T < 1$. Fix one such admissible rectangle for the remainder of the section.

Corollary 1 gives a fixed-environment initial-state buffer in terms of the denominator at the fixed point. On the REE rectangle just defined, that denominator has a primitive lower bound.

Corollary 3 (REE-neighborhood buffer for local denominator safety)**.** *Fix one discretized bounded environment with deterministic fixed point* $\mathbf{s}^*$*, induced profile* f^**, and coefficient pair*

$$m^* := \big(\bar{\beta}_A(f^*), \bar{\mu}_A(f^*)\big).$$

Suppose $m^* \in \mathcal{K}$*, and let* $L_{\xi,s}$ *be defined by* (A25)*. If* $L_{\xi,s} > 0$ *and, for some* $\underline{\xi}_B > 0$*,*

$$\|\mathbf{s}_0 - \mathbf{s}^*\|_\infty \le \frac{\xi(m^{REE}) - \phi h_\beta - \underline{\xi}_B}{L_{\xi,s}}, \tag{A45}$$

with positive numerator, then condition (A26) *holds. If* (A19) *also holds on* $\mathcal{S}_B(\mathbf{s}_0)$*, then Corollary 1 gives deterministic denominator safety and ODE localization. To apply Theorem 15, additionally assume stochastic localization on this REE-neighborhood safe set.*

Using (A34)*, condition* (A45) *is*

$$\|\mathbf{s}_0 - \mathbf{s}^*\|_\infty \le \frac{\bar{\kappa}_H}{2\phi B_\beta}\left[\frac{(\gamma_A\bar{\beta}_p^M + \phi)\big((\bar{\beta}_\eta^M)^2\sigma_v^2 + \sigma_z^2\big)}{\gamma_A\big((\bar{\beta}_\eta^M)^2\sigma_v^2 + \sigma_z^2\big) + \phi\bar{\beta}_\eta^M\sigma_v^2} - \phi h_\beta - \underline{\xi}_B\right], \qquad \phi B_\beta > 0.$$

If $\phi B_\beta = 0$*, the denominator* ξ *does not vary with the AI demand profile, so the condition reduces to* $\bar{\beta}_p^M \ge \underline{\xi}_B > 0$*.*

Proof. If $m^* \in \mathcal{K}$, then the linear form $\xi(m) = \bar{\beta}_p^M + \phi\bar{\beta}_A$ and $\phi \ge 0$ imply

$$\xi(m^*) \ge \inf_{m\in\mathcal{K}} \xi(m) = \xi(m^{REE}) - \phi h_\beta.$$

Substituting this lower bound into the fixed-environment buffer (A26) gives (A45). The deterministic denominator-safety and ODE conclusions then follow from Corollary 1. Stochastic localization is the additional no-exit condition required by Theorem 15. Substituting (A34) gives the displayed primitive expression. The case $\phi B_\beta = 0$ is immediate because ξ is then independent of the AI demand profile. □

Proposition 18 (Uniform quadratic tail bound on $\mathcal{K}$)**.** *There exist constants* $c_\mathcal{K}, C_\mathcal{K} > 0$ *such that*

$$R(\beta, \mu; m) \le C_\mathcal{K} - c_\mathcal{K}(\beta^2 + \mu^2) \qquad \forall(\beta,\mu) \in \mathbb{R}^2,\ \forall m \in \mathcal{K}. \tag{A46}$$

Moreover, for all sufficiently large shock truncation bounds (B_v, B_z)*, the same bound holds for the truncated raw expected profits:*

$$R_{\Delta,B}(\beta, \mu; m) \le 2C_\mathcal{K} - \frac{c_\mathcal{K}}{2}(\beta^2 + \mu^2) \qquad \forall(\beta,\mu) \in \mathbb{R}^2,\ \forall m \in \mathcal{K}. \tag{A47}$$

Proof. The bound controls the tails of Boltzmann weights uniformly over the local neighborhood $\mathcal{K}$. On $\mathcal{K}$, the price denominator stays away from zero and price variance stays positive, so the negative quadratic term in expected profit is uniformly strong. The remaining terms are at most linear in (β, μ) and can be absorbed into this quadratic loss. The same argument applies to truncated shocks once the truncated moments are uniformly close to the Gaussian moments.

For $m \in \mathcal{K}$, the matrix

$$\frac{1}{\gamma_A} H(m) = \begin{pmatrix} q(m) & -m_p(m) \\ -m_p(m) & 1 \end{pmatrix}$$

is continuous in m and positive definite. Because $\mathcal{K}$ is compact and $\xi(m) \geq \underline{\xi}_{\mathcal{K}} > 0$ on $\mathcal{K}$, the smallest eigenvalue of $H(m)/\gamma_A$ has a strictly positive minimum over $\mathcal{K}$. Hence there exists $\lambda_{\mathcal{K}} > 0$ such that

$$\mu^2 - 2\mu\beta m_p(m) + \beta^2 q(m) \geq \lambda_{\mathcal{K}}(\beta^2 + \mu^2) \qquad \forall (\beta, \mu),\ \forall m \in \mathcal{K}.$$

The linear coefficients $d(m)$ and $E_{vp}(m)$ are continuous in m, hence bounded on $\mathcal{K}$. Therefore, for a constant $L_{\mathcal{K}} < \infty$,

$$R(\beta, \mu; m) \leq L_{\mathcal{K}} \|(\beta, \mu)\|_2 - \frac{\gamma_A \lambda_{\mathcal{K}}}{2} \|(\beta, \mu)\|_2^2 \qquad \forall (\beta, \mu),\ \forall m \in \mathcal{K}.$$

Choosing $\epsilon = \gamma_A \lambda_{\mathcal{K}}/2$ in $L_{\mathcal{K}} \|(\beta, \mu)\|_2 \leq L_{\mathcal{K}}^2/(2\epsilon) + (\epsilon/2)\|(\beta, \mu)\|_2^2$ yields the explicit constants $c_{\mathcal{K}} := \gamma_A \lambda_{\mathcal{K}}/4 > 0$ and $C_{\mathcal{K}} := L_{\mathcal{K}}^2/(\gamma_A \lambda_{\mathcal{K}}) < \infty$ satisfying (A46).

As the truncation bounds grow, the truncated moments converge uniformly to the untruncated moments on $\mathcal{K}$. This follows because $\xi(m)$ is bounded away from zero on $\mathcal{K}$, so p_m^{tr} is affine in (v, z) with coefficients continuous and uniformly bounded in m, and truncated Gaussian moments converge to the corresponding full Gaussian moments. Hence the coefficients of the truncated quadratic form converge uniformly to those of the untruncated one on $\mathcal{K}$.

To make the last step explicit, write

$$m_p^{tr}(m) := \mathbb{E}^{tr}[p_m^{tr}], \qquad q^{tr}(m) := \mathbb{E}^{tr}[(p_m^{tr})^2].$$

Because the truncations are symmetric around the Gaussian means,

$$\mathrm{Var}(p_m^{tr}) = \frac{(\bar{\beta}_\eta^M)^2 \mathrm{Var}^{tr}(v) + \mathrm{Var}^{tr}(z)}{\xi(m)^2}.$$

The truncated variances converge to σ_v^2 and σ_z^2, while $\xi(m)$ is bounded above and below on $\mathcal{K}$. Hence

$$\inf_{m\in\mathcal{K}} \mathrm{Var}(p_m^{tr}) > 0$$

for all sufficiently large (B_v, B_z). The smallest eigenvalue of

$$\begin{pmatrix} q^{tr}(m) & -m_p^{tr}(m) \\ -m_p^{tr}(m) & 1 \end{pmatrix}$$

therefore converges uniformly to the smallest eigenvalue of $H(m)/\gamma_A$ and is at least $\lambda_{\mathcal{K}}/2$ once (B_v, B_z) exceed the thresholds at which the truncated variances are within $\lambda_{\mathcal{K}}/4$ of σ_v^2, σ_z^2 on $\mathcal{K}$. The truncated linear coefficients $\bar{v} - m_p^{tr}(m)$ and $\mathbb{E}^{tr}[(v - p_m^{tr})p_m^{tr}]$ are uniformly bounded on $\mathcal{K}$. The factors $2C_{\mathcal{K}}$ and $c_{\mathcal{K}}/2$ in (A47) are nominal buffers that absorb this eigenvalue gap; the same completion-of-squares argument with the explicit constants of (A46) delivers them. □

Proposition 19 (Uniform C^1 convergence of discretized mean maps on $\mathcal{K}$)**.** *For a finite population size N and raw-temperature vector $\vartheta = (\vartheta_1, \ldots, \vartheta_N)$, write $\mathcal{T}_{\Delta,B}^{N,\vartheta}$ for the map in (A41). Fix a compact raw-temperature interval $[\underline{\vartheta}, \bar{\vartheta}] \subset (0, \infty)$. Along any sequence of mesh-compatible rectangular action grids and shock truncations satisfying*

$$\Delta_\beta, \Delta_\mu \to 0, \qquad B_\beta, B_\mu \to \infty, \qquad B_v, B_z \to \infty,$$

the raw Boltzmann mean maps for discretized bounded environments converge to the benchmark map in C^1 on $\mathcal{K}$, uniformly over all finite population sizes and all temperature vectors in this interval:

$$\sup_{N\geq 1} \sup_{\vartheta\in[\underline{\vartheta},\bar{\vartheta}]^N} \sup_{m\in\mathcal{K}} \left\| \mathcal{T}_{\Delta,B}^{N,\vartheta}(m) - T(m) \right\|_\infty \to 0. \tag{A48}$$

Moreover,

$$\sup_{N\geq 1} \sup_{\vartheta\in[\underline{\vartheta},\bar{\vartheta}]^N} \sup_{m\in\mathcal{K}} \left\| D_m \mathcal{T}_{\Delta,B}^{N,\vartheta}(m) - D_m T(m) \right\|_{\mathrm{op}} \to 0. \tag{A49}$$

Proof. The proof records four claims. First, an envelope controls continuous and grid tails uniformly. Second, expanding rectangular Riemann sums converge uniformly on each compact action box. Third, the quotient defining the mean is stable because partition functions stay bounded away from zero. Fourth, the same argument applies to the derivative quotients after differentiating under the integral and through the finite grid sums.

For each raw temperature $\vartheta \in [\underline{\vartheta}, \bar{\vartheta}]$ and each $m \in \mathcal{K}$, define the continuous partition

function and first moment

$$Z_\vartheta(m) := \int_{\mathbb{R}^2} \exp\left(\frac{R(\beta,\mu;m)}{\vartheta}\right) d\beta\, d\mu, \qquad Y_\vartheta(m) := \int_{\mathbb{R}^2} (\beta,\mu) \exp\left(\frac{R(\beta,\mu;m)}{\vartheta}\right) d\beta\, d\mu.$$

Theorem 16 implies that

$$\frac{Y_\vartheta(m)}{Z_\vartheta(m)} = T(m) \qquad \forall\, m \in \mathcal{K}, \forall\, \vartheta \in [\underline{\vartheta}, \bar{\vartheta}].$$

Let $h_\Delta := \Delta_\beta \Delta_\mu$ and define the weighted grid sums

$$Z_\vartheta^{\Delta,B}(m) := h_\Delta \sum_{a=1}^{M} \exp\left(\frac{R_{\Delta,B}(\beta_A(a),\mu_A(a);m)}{\vartheta}\right),$$

$$Y_\vartheta^{\Delta,B}(m) := h_\Delta \sum_{a=1}^{M} (\beta_A(a),\mu_A(a)) \exp\left(\frac{R_{\Delta,B}(\beta_A(a),\mu_A(a);m)}{\vartheta}\right).$$

Multiplying numerator and denominator by the common cell area h_Δ does not change the discrete Boltzmann mean, so for any finite N and $\vartheta = (\vartheta_1, \ldots, \vartheta_N)$,

$$\mathcal{T}_{\Delta,B}^{N,\vartheta}(m) = \frac{1}{N} \sum_{i=1}^{N} \frac{Y_{\vartheta_i}^{\Delta,B}(m)}{Z_{\vartheta_i}^{\Delta,B}(m)}.$$

Tail control. By Proposition 18 and the temperature bounds $\underline{\vartheta} \le \vartheta \le \bar{\vartheta}$, there exist constants $C_0, c_0 > 0$ such that, for all sufficiently large shock truncation bounds, all $m \in \mathcal{K}$, all $\vartheta \in [\underline{\vartheta}, \bar{\vartheta}]$, and all $(\beta,\mu) \in \mathbb{R}^2$,

$$\exp\left(\frac{R(\beta,\mu;m)}{\vartheta}\right), \quad \exp\left(\frac{R_{\Delta,B}(\beta,\mu;m)}{\vartheta}\right) \le C_0 e^{-c_0(\beta^2+\mu^2)},$$

and

$$\|(\beta,\mu)\|_\infty \exp\left(\frac{R(\beta,\mu;m)}{\vartheta}\right), \quad \|(\beta,\mu)\|_\infty \exp\left(\frac{R_{\Delta,B}(\beta,\mu;m)}{\vartheta}\right) \le C_0 \|(\beta,\mu)\|_\infty e^{-c_0(\beta^2+\mu^2)}.$$

The right-hand sides are integrable on $\mathbb{R}^2$. This integrable envelope also controls the grid tails. Let

$$G(\beta,\mu) := (1 + \|(\beta,\mu)\|_\infty) C_0 e^{-c_0(\beta^2+\mu^2)}.$$

For $L > 0$, write $\mathcal{A}_L = [-L, L] \times [-L, L]$. Then G is integrable. Because G is continuous, positive, and integrable on $\mathbb{R}^2$, its improper rectangular Riemann sums converge to its

integral, so the usual upper-sum comparison for rectangular grids gives

$$\lim_{L\to\infty} \limsup_{\Delta_\beta,\Delta_\mu\to 0,\ B_\beta,B_\mu\to\infty} h_\Delta \sum_{a:\,(\beta_A(a),\mu_A(a))\notin\mathcal{A}_L} G(\beta_A(a),\mu_A(a)) = \int_{\mathbb{R}^2\setminus\mathcal{A}_L} G \to 0$$

as $L \to \infty$ by dominated convergence. Thus both the continuous integral tails and the weighted grid-sum tails of the partition functions and first moments can be made uniformly small over $m \in \mathcal{K}$ and $\vartheta \in [\underline{\vartheta}, \bar{\vartheta}]$ by first choosing L large and then taking the grid fine with action bounds larger than L.

Compact-box convergence. On every fixed action box $\mathcal{A}_L$, the coefficients of the quadratic form $R_{\Delta,B}(\beta,\mu;m)$ converge uniformly in $m \in \mathcal{K}$ to those of $R(\beta,\mu;m)$ as $B_v, B_z \to \infty$. Hence, for each $g \in \{1,\beta,\mu\}$,

$$g(\beta,\mu)\exp\left(\frac{R_{\Delta,B}(\beta,\mu;m)}{\vartheta}\right) \to g(\beta,\mu)\exp\left(\frac{R(\beta,\mu;m)}{\vartheta}\right)$$

uniformly on $\mathcal{A}_L \times \mathcal{K} \times [\underline{\vartheta}, \bar{\vartheta}]$. Since the integrands are continuous on this compact set, the corresponding rectangular Riemann sums on $\mathcal{A}_L$ converge uniformly in (m,ϑ) to the integrals over $\mathcal{A}_L$. Combining this compact-box convergence with the preceding uniform tail bound implies

$$\sup_{\vartheta\in[\underline{\vartheta},\bar{\vartheta}]}\sup_{m\in\mathcal{K}} |Z_\vartheta^{\Delta,B}(m) - Z_\vartheta(m)| \to 0, \qquad \sup_{\vartheta\in[\underline{\vartheta},\bar{\vartheta}]}\sup_{m\in\mathcal{K}} \|Y_\vartheta^{\Delta,B}(m) - Y_\vartheta(m)\|_\infty \to 0$$

as

$$\Delta_\beta, \Delta_\mu \to 0, \qquad B_\beta, B_\mu \to \infty, \qquad B_v, B_z \to \infty.$$

Quotient convergence. Because $Z_\vartheta(m)$ is continuous and strictly positive on the compact set $[\underline{\vartheta}, \bar{\vartheta}] \times \mathcal{K}$,

$$z_* := \inf_{\vartheta\in[\underline{\vartheta},\bar{\vartheta}]}\inf_{m\in\mathcal{K}} Z_\vartheta(m) > 0.$$

Hence $Z_\vartheta^{\Delta,B}(m) \geq z_*/2$ for all sufficiently fine grids, sufficiently large bounds, all $\vartheta \in [\underline{\vartheta}, \bar{\vartheta}]$, and all $m \in \mathcal{K}$. The integrable envelope gives a common bound for the continuous and discrete numerator vectors, so the ratio map $(y,z) \mapsto y/z$ is Lipschitz on the relevant bounded set with $z \geq z_*/2$. Therefore

$$\sup_{\vartheta\in[\underline{\vartheta},\bar{\vartheta}]}\sup_{m\in\mathcal{K}} \left\| \frac{Y_\vartheta^{\Delta,B}(m)}{Z_\vartheta^{\Delta,B}(m)} - \frac{Y_\vartheta(m)}{Z_\vartheta(m)} \right\|_\infty \to 0.$$

Derivative convergence. It remains to record the same convergence for derivatives. On $\mathcal{K}$,

$\xi(m)$ is bounded away from zero. The coefficients of the quadratic polynomials $R(\beta, \mu; m)$ and $R_{\Delta,B}(\beta, \mu; m)$, and their first derivatives with respect to $m = (\bar{\beta}_A, \bar{\mu}_A)$, are therefore uniformly bounded on $\mathcal{K}$ for all sufficiently large shock truncation bounds. Hence, for each coordinate $\ell \in \{1, 2\}$, there is a constant $C_1 > 0$ such that

$$\left|\frac{\partial R(\beta, \mu; m)}{\partial m_\ell}\right|, \quad \left|\frac{\partial R_{\Delta,B}(\beta, \mu; m)}{\partial m_\ell}\right| \leq C_1(1 + \beta^2 + \mu^2)$$

uniformly over $m \in \mathcal{K}$ and over all sufficiently large truncation bounds. Moreover, on each compact action box $\mathcal{A}_L$, the same moment convergence gives uniform convergence of the derivatives:

$$\sup_{(\beta,\mu)\in\mathcal{A}_L} \sup_{m\in\mathcal{K}} \|D_m R_{\Delta,B}(\beta, \mu; m) - D_m R(\beta, \mu; m)\|_\infty \to 0.$$

Indeed, both derivatives are polynomials in (β, μ) whose coefficients are C^1 functions of m and of the first two moments of the shock law. The denominator satisfies $\inf_{m\in\mathcal{K}} \xi(m) > 0$, so differentiating the affine price coefficients with respect to $m = (\bar{\beta}_A, \bar{\mu}_A)$ is uniform on $\mathcal{K}$; the truncated first and second moments converge to their Gaussian counterparts. Combining this uniform derivative convergence with the polynomial bound and the Gaussian envelope above gives integrable envelopes for the derivative integrands:

$$(1 + \beta^2 + \mu^2)e^{-c_0(\beta^2+\mu^2)}, \qquad \|(\beta, \mu)\|_\infty(1 + \beta^2 + \mu^2)e^{-c_0(\beta^2+\mu^2)}.$$

The same tail and compact-box Riemann-sum argument applied to

$$\frac{1}{\vartheta}\frac{\partial R_{\Delta,B}}{\partial m_\ell}\exp\left(\frac{R_{\Delta,B}}{\vartheta}\right), \qquad \frac{(\beta, \mu)}{\vartheta}\frac{\partial R_{\Delta,B}}{\partial m_\ell}\exp\left(\frac{R_{\Delta,B}}{\vartheta}\right)$$

shows that

$$\sup_{\vartheta\in[\underline{\vartheta},\bar{\vartheta}]} \sup_{m\in\mathcal{K}} \|D_m Z_\vartheta^{\Delta,B}(m) - D_m Z_\vartheta(m)\|_\infty \to 0,$$

and

$$\sup_{\vartheta\in[\underline{\vartheta},\bar{\vartheta}]} \sup_{m\in\mathcal{K}} \|D_m Y_\vartheta^{\Delta,B}(m) - D_m Y_\vartheta(m)\|_{\text{op}} \to 0.$$

Here the derivative of the discrete sum is obtained by differentiating the finite sum term by term, while the derivative of the continuous integral is obtained by dominated convergence using the derivative envelopes above. Because the partition functions are uniformly bounded

below by $z_*/2$ for sufficiently accurate environments, the quotient derivative formula gives

$$D_m\left(\frac{Y_\vartheta^{\Delta,B}}{Z_\vartheta^{\Delta,B}}\right) \to D_m\left(\frac{Y_\vartheta}{Z_\vartheta}\right) = D_m T$$

uniformly over $\mathcal{K} \times [\underline{\vartheta}, \bar{\vartheta}]$.

Finally, since $Y_\vartheta(m)/Z_\vartheta(m) = T(m)$, averaging the level quotient convergence over any finite population gives

$$\sup_{m\in\mathcal{K}} \left\| \mathcal{T}_{\Delta,B}^{N,\vartheta}(m) - T(m) \right\|_\infty \le \sup_{\theta\in[\underline{\vartheta},\bar{\vartheta}]} \sup_{m\in\mathcal{K}} \left\| \frac{Y_\theta^{\Delta,B}(m)}{Z_\theta^{\Delta,B}(m)} - T(m) \right\|_\infty .$$

The right-hand side does not depend on N or on the particular temperature vector ϑ, and it converges to zero. The derivative convergence is uniform in the same way:

$$\sup_{m\in\mathcal{K}} \left\| D_m \mathcal{T}_{\Delta,B}^{N,\vartheta}(m) - D_m T(m) \right\|_{\mathrm{op}} \le \sup_{\theta\in[\underline{\vartheta},\bar{\vartheta}]} \sup_{m\in\mathcal{K}} \left\| D_m\left(\frac{Y_\theta^{\Delta,B}(m)}{Z_\theta^{\Delta,B}(m)}\right) - D_m\left(\frac{Y_\theta(m)}{Z_\theta(m)}\right) \right\|_{\mathrm{op}},$$

and the right-hand side converges to zero. This proves (A48) and (A49). □

The learning theorem below is pointwise in a fixed admissible environment.[18] The approximation theorem is a deterministic statement about local fixed points in $\mathcal{K}$, where $\xi(m)$ is bounded away from zero. Existence and uniqueness of a finite-environment learned mean-field equilibrium come from Theorem 14 under the fixed-environment conditions above; the approximation theorem below supplies a unique local deterministic fixed point in $\mathcal{K}$ and shows that this point is close to the benchmark fixed point once the discretized bounded environment is sufficiently accurate.

Theorem 20 (Deterministic approximation to the REE)**.** *Fix a compact raw-temperature interval $[\underline{\vartheta}, \bar{\vartheta}] \subset (0, \infty)$ and the compact admissible rectangle $\mathcal{K}$ from (A42)–(A43). For every $\varepsilon > 0$, there exist thresholds*

$$\bar{B}_\beta, \bar{B}_\mu, \bar{B}_v, \bar{B}_z < \infty, \qquad \bar{\Delta}_\beta, \bar{\Delta}_\mu > 0$$

such that whenever

$$B_\beta \ge \bar{B}_\beta, \quad B_\mu \ge \bar{B}_\mu, \quad B_v \ge \bar{B}_v, \quad B_z \ge \bar{B}_z, \quad \Delta_\beta \le \bar{\Delta}_\beta, \quad \Delta_\mu \le \bar{\Delta}_\mu,$$

[18] Assumption 1 is a local condition on one fixed finite learning environment.

then, for every finite N and every raw-temperature vector $\vartheta \in [\underline{\vartheta}, \bar{\vartheta}]^N$, the raw Boltzmann map $\mathcal{T}^{N,\vartheta}_{\Delta,B}$ has a unique local deterministic fixed point $m_{\Delta,B} \in \mathcal{K}$. This fixed point satisfies

$$\left\| m_{\Delta,B} - m^{REE} \right\|_\infty \leq \varepsilon.$$

Equivalently, for any sequence of finite population sizes N_j, temperature vectors $\vartheta^j \in [\underline{\vartheta}, \bar{\vartheta}]^{N_j}$, and unique local deterministic fixed points in $\mathcal{K}$ along an approximating sequence,

$$m_{\Delta,B} \to m^{REE} = (\bar{\beta}^*_A, \bar{\mu}^*_A) = (\beta_p^{A,env}, \mu^{A,env})$$

as

$$\Delta_\beta, \Delta_\mu \to 0, \qquad B_\beta, B_\mu \to \infty, \qquad B_v, B_z \to \infty.$$

Proof. The proof has three parts. First, a local index argument shows that the discretized map has a fixed point in $\mathcal{K}$ once it is uniformly close to the continuous map. Second, uniform derivative convergence makes this fixed point unique in $\mathcal{K}$. Third, a separation argument places the unique fixed point near the benchmark fixed point.

Fix an arbitrary finite N and $\vartheta \in [\underline{\vartheta}, \bar{\vartheta}]^N$. The thresholds below are chosen from Proposition 19 and therefore do not depend on this particular N or temperature vector. The sequential convergence in Proposition 19 implies such finite thresholds: if no threshold existed for one of the bounds used below, a violating sequence of grids, bounds, population sizes, and temperature vectors would contradict (A48) or (A49). Define

$$F(m) := m - T(m), \qquad F^{N,\vartheta}_{\Delta,B}(m) := m - \mathcal{T}^{N,\vartheta}_{\Delta,B}(m).$$

By (A44),

$$F(m) = (1 + \lambda_T)(m - m^{REE}).$$

Since $m^{REE} \in \operatorname{int} \mathcal{K}$, both h_β and h_μ are strictly positive. Let

$$\eta_{\mathcal{K}} := \frac{1 + \lambda_T}{2} \min\{h_\beta, h_\mu\} > 0.$$

Proposition 19 implies that, for all sufficiently fine grids and sufficiently large action and shock bounds,

$$\sup_{m \in \mathcal{K}} \| F^{N,\vartheta}_{\Delta,B}(m) - F(m) \|_\infty = \sup_{m \in \mathcal{K}} \| \mathcal{T}^{N,\vartheta}_{\Delta,B}(m) - T(m) \|_\infty < \eta_{\mathcal{K}}.$$

Hence $F^{N,\vartheta}_{\Delta,B}$ has the same outward sign as F on the opposite faces of the rectangle, where

subscripts denote the β and μ coordinates:

$$F^{N,\vartheta}_{\Delta,B,\beta}(\bar\beta^*_A - h_\beta, \bar\mu_A) < 0 < F^{N,\vartheta}_{\Delta,B,\beta}(\bar\beta^*_A + h_\beta, \bar\mu_A) \qquad \forall\, \bar\mu_A \in [\bar\mu^*_A - h_\mu, \bar\mu^*_A + h_\mu],$$

and

$$F^{N,\vartheta}_{\Delta,B,\mu}(\bar\beta_A, \bar\mu^*_A - h_\mu) < 0 < F^{N,\vartheta}_{\Delta,B,\mu}(\bar\beta_A, \bar\mu^*_A + h_\mu) \qquad \forall\, \bar\beta_A \in [\bar\beta^*_A - h_\beta, \bar\beta^*_A + h_\beta].$$

The map $F^{N,\vartheta}_{\Delta,B}$ is continuous on $\mathcal{K}$, so the Poincaré–Miranda theorem gives a zero $m_{\Delta,B} \in \mathcal{K}$. This zero is exactly a fixed point of $\mathcal{T}^{N,\vartheta}_{\Delta,B}$. The face inequalities are strict, so no zero can lie on the boundary of $\mathcal{K}$.

Next, use the C^1 part of Proposition 19. Since $D_m T(m) = -\lambda_T I_2$, we have

$$D_m F(m) = I_2 - D_m T(m) = (1 + \lambda_T) I_2 \qquad \text{for all } m \in \mathcal{K}.$$

Choose the grids fine and the bounds large enough that

$$\sup_{m \in \mathcal{K}} \big\| D_m F^{N,\vartheta}_{\Delta,B}(m) - (1 + \lambda_T) I_2 \big\|_{\mathrm{op}} < \frac{1 + \lambda_T}{2}.$$

This bound is uniform in N and in the raw-temperature vector. If $m, m' \in \mathcal{K}$, then because $\mathcal{K}$ is a rectangle the segment $\{m' + t(m - m') : t \in [0, 1]\}$ remains in $\mathcal{K}$; convexity of $\mathcal{K}$ and the fundamental theorem of calculus then give

$$F^{N,\vartheta}_{\Delta,B}(m) - F^{N,\vartheta}_{\Delta,B}(m') = \int_0^1 D_m F^{N,\vartheta}_{\Delta,B}\big(m' + t(m - m')\big)(m - m')\, dt.$$

Writing $D_m F^{N,\vartheta}_{\Delta,B} = (1 + \lambda_T) I_2 + E$ along this segment, with $\|E\|_{\mathrm{op}} < (1 + \lambda_T)/2$, the reverse triangle inequality gives, in Euclidean norm,

$$\big\| F^{N,\vartheta}_{\Delta,B}(m) - F^{N,\vartheta}_{\Delta,B}(m') \big\|_2 \geq \frac{1 + \lambda_T}{2} \|m - m'\|_2.$$

Thus $F^{N,\vartheta}_{\Delta,B}$ is injective on $\mathcal{K}$, and the zero found above is the unique zero in $\mathcal{K}$. Equivalently, $\mathcal{T}^{N,\vartheta}_{\Delta,B}$ has a unique local deterministic fixed point in $\mathcal{K}$.

Fix $\varepsilon > 0$. If

$$\{m \in \mathcal{K} : \|m - m^{REE}\|_\infty \geq \varepsilon\}$$

is empty, every point in $\mathcal{K}$ is within ε of m^{REE}, so the distance claim follows from the fixed point constructed above. It remains to consider the nonempty case. By (A44), every $m \in \mathcal{K}$

with $\|m - m^{REE}\|_\infty \geq \varepsilon$ satisfies

$$\|m - T(m)\|_\infty = (1 + \lambda_T)\|m - m^{REE}\|_\infty \geq (1 + \lambda_T)\varepsilon.$$

Proposition 19 implies that, for all sufficiently fine grids and sufficiently large action and shock bounds,

$$\sup_{m \in \mathcal{K}} \|\mathcal{T}^{N,\vartheta}_{\Delta,B}(m) - T(m)\|_\infty < (1 + \lambda_T)\varepsilon.$$

Choose the thresholds so that this separation bound, the preceding sign bound, and the derivative bound used for injectivity all hold. If the unique fixed point $m_{\Delta,B} \in \mathcal{K}$ satisfies $\|m_{\Delta,B} - m^{REE}\|_\infty \geq \varepsilon$, then

$$\|m_{\Delta,B} - T(m_{\Delta,B})\|_\infty = \|\mathcal{T}^{N,\vartheta}_{\Delta,B}(m_{\Delta,B}) - T(m_{\Delta,B})\|_\infty < (1 + \lambda_T)\varepsilon,$$

contradicting the preceding lower bound. Hence the unique local deterministic fixed point in $\mathcal{K}$ lies within ε of m^{REE}.

The equivalent sequential statement follows by applying the preceding ε argument to each fixed tolerance. □

A.8 Connection to the main-text propositions.

Theorem 20 is the technical approximation result behind Proposition 2. We prove the three main-text claims by reducing each to the technical results above.

Proof of Proposition 1. The hypotheses are exactly the fixed-environment conditions of Theorem 15, which gives almost-sure convergence of the learning state to the unique deterministic fixed point in $\mathcal{S}_B$. Corollary 2 converts state convergence into convergence of the population-average coefficient pair to $m^*_{\Delta,B}$. □

Proof of Proposition 2. This is Theorem 20. Uniform convergence of the discretized Boltzmann mean maps and their derivatives to the benchmark best-response map on $\mathcal{K}$ yields a local sign argument for existence, a derivative bound for local uniqueness, and a separation argument that places the unique local fixed point near m^{REE}. □

Proof of Theorem 3. Proposition 1 gives $m_t^{\Delta,B} \to m^*_{\Delta,B}$ almost surely. The unique learning fixed point $\mathbf{s}^*$ lies in $\mathcal{S}_B$, so the hypothesis on $\mathcal{S}_B$ gives $m^*_{\Delta,B} = m_{\Delta,B}(\mathbf{s}^*) \in \mathcal{K}$. Under $\vartheta_i = 2\bar{\Pi}\kappa_i$, Corollary 2 shows that $m^*_{\Delta,B}$ is a fixed point of the raw Boltzmann coefficient map, which coincides with $\mathcal{T}^{N,\vartheta}_{\Delta,B}$ on $\mathcal{K}$ by Remark 2. Proposition 2 gives a unique local

deterministic fixed point in $\mathcal{K}$, so $m^*_{\Delta,B}$ coincides with it and $\|m^*_{\Delta,B} - m^{REE}\|_\infty \leq \varepsilon$. The almost-sure distance bound follows. □

Iterated learning and approximation. The preceding corollary is an iterated limit. If, along a refining sequence of discretized bounded environments, the fixed-environment learning assumptions hold on denominator-safe sets satisfying $m_{\Delta,B}(\mathbf{s}) \in \mathcal{K}$ throughout, then each learned mean-field equilibrium lies in $\mathcal{K}$, local uniqueness identifies it with the unique local deterministic fixed point in $\mathcal{K}$, and

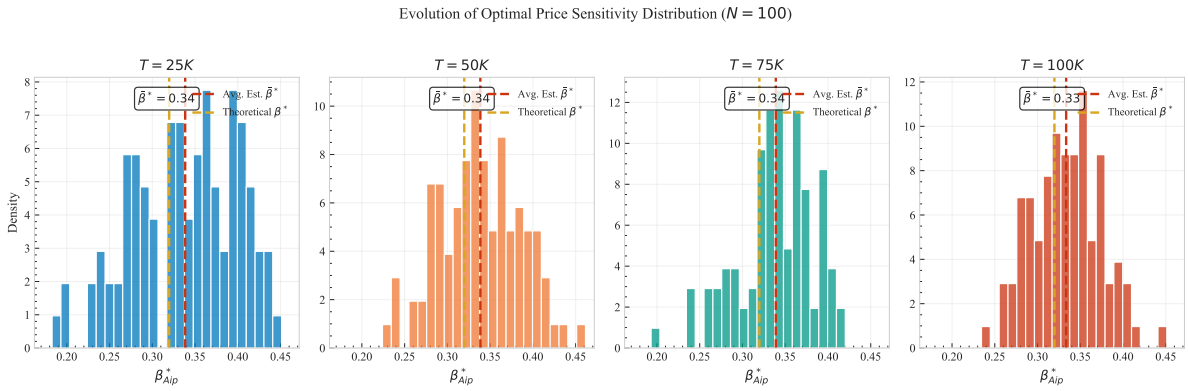
$$\lim_{\Delta_\beta,\Delta_\mu\to 0,\ B_\beta,B_\mu,B_v,B_z\to\infty} \left(\lim_{t\to\infty} m_t^{\Delta,B}\right) = m^{REE}.$$

Uniformity in the finite population size. The approximation statement of Proposition 2 is uniform over finite populations in the following sense. Given any finite N, however large, and any raw temperatures $\vartheta \in [\underline{\vartheta}, \bar{\vartheta}]^N$ in the common compact interval, a sufficiently accurate discretized bounded environment has a unique local deterministic mean-field equilibrium in $\mathcal{K}$ within any prescribed distance of the benchmark coefficient pair m^{REE}. The grid and shock-bound thresholds that deliver this ε-bound are chosen independently of N. Along any sequence $N_j \to \infty$, the same conclusion holds term by term for each finite N_j, provided the environment at step j is chosen sufficiently accurate. Equivalently, the heterogeneous mean-field analysis extends to arbitrarily large finite populations without losing the ε-approximation to the benchmark equilibrium.

A.9 Simulation evidence: algorithmic herding

Figure A1 complements the main-text convergence evidence by showing the cross-sectional distribution of the learned price-sensitivity coefficient β_{Ap} across the population of AI investors and how this distribution concentrates over the learning horizon. Although AI investors begin from heterogeneous, optimistic initial estimates and explore independently, their learned coefficients collapse onto a tight neighborhood of the benchmark rational-expectations value. This is the algorithmic-herding counterpart of Theorem 3: heterogeneous learners converge to a common, near-rational trading rule.

Figure A1: Algorithmic Herding: Cross-Sectional Distribution of Learned β_{Ap}



Notes: The panels show the cross-sectional distribution of the learned price-sensitivity coefficient β_{Ap} across $N = 100$ AI investors at increasing learning horizons T. The red dashed line marks the cross-sectional average and the golden dashed line the benchmark rational-expectations value β^*_{Ap}. As learning proceeds, the distribution concentrates around the benchmark value, illustrating algorithmic herding. Simulation parameters: $\sigma_v = 2$, $\sigma_z = 0.2$, $\gamma_A = 0.2$, $\psi = 1.5$, $\bar{\beta}_{Mp} = 4.0$, $\bar{\beta}_{M\eta} = 4.5$, $S = 0$, Boltzmann temperature $\kappa = 20$, and a β-grid on $[0, 0.6]$ with mesh 0.002 (μ_A fixed at 0).

B Proof of Proposition 4

B.1 Triangular structure and convergence channels

The update equations admit a triangular decomposition. Price informativeness τ_p^k evolves autonomously through a one-dimensional map G_c. Once this scalar path is known, the normalized price loading and intercept follow driven affine recursions. The first channel asks whether higher-order reasoning stabilizes the information content of prices, while the driven channels ask whether the portfolio coefficients remain bounded along that information path.

The informativeness map. From (30), $\beta_\eta^{M,k} = \tau_M/(\rho_M(\widehat{\tau}^c + \tau_p^{k-1}))$, so

$$\zeta^k = \psi^c \beta_\eta^{M,c} + \frac{\psi^l \tau_M}{\rho_M(\widehat{\tau}^c + \tau_p^{k-1})}.$$

Define the cursed-human signal loading $\zeta^c \equiv \psi^c \beta_\eta^{M,c} \geq 0$ and the learnable-human signal-capacity term $\chi^l \equiv \psi^l \tau_M/\rho_M > 0$. Note $\zeta^c + \chi^l/\widehat{\tau}^c = \psi \beta_\eta^{M,c}$. Since $\tau_p^k = (\zeta^k)^2 \tau_z$, the scalar map governing informativeness is

$$\tau_p^k = G_c(\tau_p^{k-1}), \qquad G_c(\tau_p) \equiv \tau_z \left(\zeta^c + \frac{\chi^l}{\widehat{\tau}^c + \tau_p}\right)^2. \tag{A50}$$

Properties of G_c.

1. G_c is continuous and strictly decreasing on $[0, \infty)$, since $\chi^l > 0$.
2. $G_c(0) = \tau_z(\zeta^c + \chi^l/\widehat{\tau}^c)^2 = (\psi\beta_\eta^{M,c})^2\tau_z$.
3. $\lim_{\tau_p \to \infty} G_c(\tau_p) = (\zeta^c)^2\tau_z = (\psi^c\beta_\eta^{M,c})^2\tau_z > 0$ when $\psi^c > 0$ (informativeness floor).
4. G_c maps the interval $\mathcal{I} \equiv [(\zeta^c)^2\tau_z,\, G_c(0)]$ into itself.

Property 3 is a key structural feature: when $\psi^c > 0$, the cursed humans' fixed signal loading $\beta_\eta^{M,c}$ contributes a positive floor to price informativeness regardless of how the level-k thinkers adjust their behavior.

[Informativeness channel] The sequence $\{\tau_p^k\}$ is bounded in $\mathcal{I}$. Let τ_p^* denote the unique fixed point of G_c on $(0, \infty)$. Define

$$\kappa_c \equiv |G_c'(\tau_p^*)| = \frac{2\chi^l\sqrt{\tau_p^*\tau_z}}{(\widehat{\tau}^c + \tau_p^*)^2}.$$

1. If $\kappa_c \leq 1$, then $\tau_p^k \to \tau_p^*$.
2. If $\kappa_c > 1$, then τ_p^k converges to an oscillation cycle (τ_p^-, τ_p^+) with $G_c(\tau_p^-) = \tau_p^+$ and $G_c(\tau_p^+) = \tau_p^-$, and $\tau_p^-, \tau_p^+ \geq (\zeta^c)^2\tau_z$.

Proof. Fixed point and bounded orbit. The fixed point solves $\tau_p = G_c(\tau_p)$. The left-hand side is strictly increasing, while G_c is strictly decreasing and satisfies $G_c(0) > 0$. Hence the two graphs cross once on $(0, \infty)$; call the crossing τ_p^*. The initial condition is $\tau_p^0 = (\psi\beta_\eta^{M,c})^2\tau_z = G_c(0)$. Because G_c maps $\mathcal{I} = [(\zeta^c)^2\tau_z, G_c(0)]$ into itself, every subsequent τ_p^k lies in $\mathcal{I}$.

Parity subsequences. Let $G_c^{(2)} \equiv G_c \circ G_c$. Since G_c is decreasing, $G_c^{(2)}$ is increasing. Moreover $\tau_p^0 > \tau_p^*$ and $\tau_p^1 = G_c(\tau_p^0) < \tau_p^*$. Iterating this argument gives $\tau_p^{2n} > \tau_p^*$ and $\tau_p^{2n+1} < \tau_p^*$ for all n. Since $G_c^{(2)}(\tau_p^0) \leq \tau_p^0$ and $G_c^{(2)}$ is increasing, the even subsequence is decreasing. Applying the decreasing map G_c to the even subsequence shows that the odd subsequence is increasing. Both are bounded, so they converge. Let their limits be τ_p^{even} and τ_p^{odd}. Continuity implies $G_c(\tau_p^{\text{even}}) = \tau_p^{\text{odd}}$ and $G_c(\tau_p^{\text{odd}}) = \tau_p^{\text{even}}$. If the two limits coincide, the common value must be τ_p^*; otherwise the two limits form a period-two cycle.

Local derivative and global restriction. Differentiating gives

$$G_c'(\tau_p) = -\frac{2\tau_z\chi^l}{(\widehat{\tau}^c + \tau_p)^2}\left(\zeta^c + \frac{\chi^l}{\widehat{\tau}^c + \tau_p}\right).$$

At the fixed point, $\zeta^c + \chi^l/(\widehat{\tau}^c + \tau_p^*) = \sqrt{\tau_p^*/\tau_z}$, so $G_c'(\tau_p^*) = -2\chi^l\sqrt{\tau_p^*\tau_z}/(\widehat{\tau}^c + \tau_p^*)^2 = -\kappa_c$. The stability of τ_p^* under $G_c^{(2)}$ is governed by $(G_c^{(2)})'(\tau_p^*) = \kappa_c^2$.

The Schwarzian derivative of G_c is

$$S(G_c)(\tau_p) = \frac{G_c'''}{G_c'} - \frac{3}{2}\left(\frac{G_c''}{G_c'}\right)^2 = -\frac{3(\chi^l)^2}{2(\widehat{\tau}^c+\tau_p)^2\left(\zeta^c(\widehat{\tau}^c+\tau_p)+\chi^l\right)^2} < 0.$$

Because $G_c'(\tau_p) \neq 0$, the chain rule for Schwarzian derivatives gives $S(G_c^{(2)}) < 0$. For an increasing map with negative Schwarzian derivative, the function $((G_c^{(2)})')^{-1/2}$ is strictly convex. Therefore $(G_c^{(2)})'(\tau_p) = 1$ has at most two solutions on $\mathcal{I}$, and Rolle's theorem implies that $\delta_2(\tau_p) \equiv G_c^{(2)}(\tau_p) - \tau_p$ has at most three zeros on $\mathcal{I}$.

Convergence when $\kappa_c \leq 1$. Suppose, toward a contradiction, that the parity limits differ. Then $G_c^{(2)}$ has three fixed points:

$$\tau_p^{\text{odd}} < \tau_p^* < \tau_p^{\text{even}}.$$

By Rolle's theorem, there are $u \in (\tau_p^{\text{odd}}, \tau_p^*)$ and $v \in (\tau_p^*, \tau_p^{\text{even}})$ with $(G_c^{(2)})'(u) = (G_c^{(2)})'(v) = 1$. Strict convexity of $((G_c^{(2)})')^{-1/2}$ then implies

$$\left((G_c^{(2)})'(\tau_p^*)\right)^{-1/2} < 1,$$

which means $(G_c^{(2)})'(\tau_p^*) > 1$. This contradicts $(G_c^{(2)})'(\tau_p^*) = \kappa_c^2 \leq 1$. Hence $\tau_p^{\text{even}} = \tau_p^{\text{odd}} = \tau_p^*$.

Oscillation cycle when $\kappa_c > 1$. When $\kappa_c > 1$, $\delta_2'(\tau_p^*) = \kappa_c^2 - 1 > 0$, so δ_2 crosses zero from negative to positive at τ_p^*. The boundary values $\delta_2((\zeta^c)^2\tau_z) \geq 0$ and $\delta_2(G_c(0)) \leq 0$ imply, by the intermediate value theorem, one zero below and one zero above τ_p^*. Since δ_2 has at most three zeros, these are the only fixed points of $G_c^{(2)}$ besides τ_p^*. They form the unique period-two cycle (τ_p^-, τ_p^+) of G_c in $\mathcal{I}$, and the monotone parity subsequences converge to the two cycle points. □

Informativeness floor. When $\psi^c > 0$, for all $k \geq 0$,

$$\tau_p^k \geq (\zeta^c)^2\tau_z = (\psi^c\beta_\eta^{M,c})^2\tau_z > 0. \tag{A51}$$

The equilibrium price is always informative about v.

The driven channels. Let $\nu^k \equiv \beta_p^{M,k}/\beta_\eta^{M,k}$ denote the normalized price loading. Since $\tau_p^k = (\zeta^k)^2\tau_z$ and $\zeta^k = \zeta^c + \chi^l/(\widehat{\tau}^c + \tau_p^{k-1})$, the scalar τ_p^k uniquely determines $\beta_\eta^{M,k}$. Writing $\zeta = \sqrt{\tau_p/\tau_z}$ for the aggregate signal loading at informativeness τ_p, the level-$(k-1)$ human signal coefficient is $\beta_\eta^{M,k-1} = (\zeta - \zeta^c)/\psi^l$, and the human aggregate price sensitivity in the

perceived environment is

$$\psi^c \beta_p^{M,c} + \psi^l \beta_p^{M,k-1} = \frac{\psi^c}{\rho_M} + \nu^{k-1}(\zeta - \zeta^c).$$

Here $h_p^A(\tau_p) \equiv \tau_p/(\tau_v + \tau_p)$ denotes the AI posterior weight on the price signal.

Substituting the AI fixed-point coefficient (27) into the definition of ξ^{k-1} and simplifying yields ξ^{k-1}/ζ^{k-1} as an affine function of ν^{k-1}. Substituting this into (31) gives the affine recursion

$$\nu^k = a_\nu(\tau_p^{k-1}) + b_\nu(\tau_p^{k-1})\, \nu^{k-1},$$

where

$$a_\nu(\tau_p) \equiv \frac{\widehat{\tau}^c + \tau_p}{\tau_M} - \frac{\tau_p}{\tau_M} \frac{\rho_A \frac{\psi^c}{\rho_M} + \phi}{\zeta\big(\rho_A + h_p^A(\tau_p)\phi/\zeta\big)}, \tag{A52}$$

$$b_\nu(\tau_p) \equiv -\frac{\tau_p}{\tau_M} \frac{\rho_A(\zeta - \zeta^c)}{\zeta\big(\rho_A + h_p^A(\tau_p)\phi/\zeta\big)} < 0. \tag{A53}$$

The slope satisfies $b_\nu < 0$: a higher normalized price loading at level $k-1$ makes prices more informative, so the level-k human shifts weight from her private signal toward the price.

Similarly, the intercept channel obeys $\mu^{M,k} = a_\mu(\tau_p^{k-1}) + b_\mu(\tau_p^{k-1})\mu^{M,k-1}$ with

$$b_\mu(\tau_p) = -\frac{h_p^M(\tau_p)}{\rho_M} \frac{\psi^l \rho_A}{\zeta\big(\rho_A + h_p^A(\tau_p)\phi/\zeta\big)} < 0,$$

where $h_p^M(\tau_p) \equiv \tau_p/(\widehat{\tau}^c + \tau_p)$.

Common stability. The normalized price-loading and intercept channels share the same asymptotic stability: at the fixed point τ_p^*, $|b_\nu(\tau_p^*)| = |b_\mu(\tau_p^*)|$. This follows by direct computation: the ratio $|b_\nu|/|b_\mu|$ simplifies to $(\widehat{\tau}^c + \tau_p^*)\rho_M(\zeta^* - \zeta^c)/(\tau_M \psi^l)$, and substituting $\zeta^* - \zeta^c = \psi^l \tau_M/(\rho_M(\widehat{\tau}^c + \tau_p^*))$ gives $|b_\nu|/|b_\mu| = 1$.

[Exceptional initial values in driven affine channels] Consider a scalar driven recursion

$$y^k = a_k + b_k y^{k-1}, \qquad b_k \neq 0,$$

and define $B_k \equiv \prod_{m=1}^k b_m$. Then

$$y^k = B_k \left(y^0 + \sum_{j=1}^k \frac{a_j}{B_j} \right).$$

If $|B_k| \to \infty$ and $\sum_{j=1}^{\infty} a_j/B_j$ converges, there is at most one initial value that can keep $\{y^k\}$ bounded:

$$y^0_{\text{exc}} = -\sum_{j=1}^{\infty} \frac{a_j}{B_j}.$$

Every $y^0 \neq y^0_{\text{exc}}$ generates $|y^k| \to \infty$. In the unstable fixed-point and period-two regimes below, the coefficients are bounded and $|B_k|$ grows geometrically, so the displayed value is the unique exceptional initial condition for the driven channel.

Proof. Iterating the affine recursion and dividing by B_k gives

$$\frac{y^k}{B_k} = y^0 + \sum_{j=1}^{k} \frac{a_j}{B_j}.$$

If $y^0 + \sum_{j=1}^{\infty} a_j/B_j \neq 0$, the right-hand side converges to a nonzero number, while $|B_k| \to \infty$; hence $|y^k| \to \infty$. Thus boundedness can hold only at the single initial value y^0_{exc}. In the fixed-point unstable case, $b_k \to b$ with $|b| > 1$; in the period-two unstable case, b_k alternates between limits whose product exceeds one. In both cases $|B_k|$ grows geometrically and the series converges because $\{a_k\}$ is bounded. At $y^0 = y^0_{\text{exc}}$,

$$y^k = -B_k \sum_{j=k+1}^{\infty} \frac{a_j}{B_j},$$

and the geometric tail is bounded in the unstable regimes just described. Hence this value is the unique bounded-path initialization. □

Assumption 6 (Regular initialization of unstable driven channels)**.** *Whenever a driven channel is in an unstable regime, the model's initial condition is not the exceptional value in Lemma B.1. In particular, for the normalized price-loading channel,*

$$\nu^0 = \frac{\beta_p^{M,c}}{\beta_\eta^{M,c}} = \frac{\widehat{\tau}^c}{\tau_M} \neq \nu^0_{\text{exc}}.$$

If $\bar{v}$ or S is nonzero, the same restriction applies to the intercept channel. In the centered case $\bar{v} = S = 0$, $\mu^{M,k} = 0$ for all k, so the intercept channel requires no regularity restriction.

Classification of asymptotic behavior. Under Assumption 6, and away from neutral boundaries, the full system has three regimes:

1. *Fixed-point convergence:* $\kappa_c \leq 1$ and $|b_\nu(\tau_p^*)| < 1$. Then $\tau_p^k \to \tau_p^*$ and both driven channels converge. The full coefficient vector approaches a single fixed point with

alternating convergence (since $b_\nu < 0$).

2. *Bounded oscillation:* $\kappa_c > 1$ and $b_\nu(\tau_p^+)b_\nu(\tau_p^-) < 1$. Then the coefficient vector oscillates between two bounded parity-specific limits.

3. *Divergence:* $\kappa_c \leq 1$ with $|b_\nu(\tau_p^*)| > 1$, or $\kappa_c > 1$ with $b_\nu(\tau_p^+)b_\nu(\tau_p^-) > 1$. The normalized price-loading channel diverges, so $|\beta_p^{M,k}| \to \infty$. The intercept channel shares the same stability away from the centered case; when $\bar{v} = S = 0$, it remains identically zero.

The remaining equalities,

$$|b_\nu(\tau_p^*)| = 1 \quad \text{or} \quad b_\nu(\tau_p^+)b_\nu(\tau_p^-) = 1,$$

are neutral boundaries where the driven channel is no longer contractive. Under Assumption 6, and away from these neutral boundaries, the fixed-point convergence criterion is exact: the iteration converges to a unique fixed point if and only if $\kappa_c \leq 1$ and $|b_\nu(\tau_p^*)| < 1$.

B.2 Convergence conditions

Under the maintained assumptions, the level-k iteration converges to a unique fixed point if and only if both $\kappa_c \leq 1$ and $|b_\nu(\tau_p^*)| < 1$. This section derives closed-form sufficient conditions on the share of level-k thinkers, ψ^l, that guarantee fixed-point convergence uniformly over all private-signal precisions $\tau_M > 0$.

B.2.1 Informativeness channel

The first stability restriction controls the feedback from price informativeness to future signal use. If current prices are too informative, the next level of thinkers will rely more on prices and less on private signals, which reduces price informativeness. If this feedback is too strong, the system can fail to converge to a fixed point and instead oscillate between two levels of informativeness.

Proposition 21 (Informativeness convergence)**.** *For all* $\sigma_M > 0$*,*

$$\kappa_c \leq \frac{8\psi\psi^l\sigma_v^2}{27\rho_M^2\sigma_z^2} = \frac{8\psi\psi^l\tau_z}{27\rho_M^2\tau_v}. \tag{A54}$$

Hence $\kappa_c \leq 1$ *for all* $\sigma_M > 0$ *whenever*

$$\psi^l \leq \frac{27\rho_M^2\sigma_z^2}{8\psi\sigma_v^2}. \tag{A55}$$

Proof. The strategy is to bound κ_c by replacing the endogenous fixed-point value τ_p^* with exogenous bounds, then optimize over τ_M.

Substituting the fixed-point identity $\zeta^c+\chi^l/(\widehat{\tau}^c+\tau_p^*) = \sqrt{\tau_p^*/\tau_z}$ into the derivative formula for G_c' and using $\zeta^c + \chi^l/(\widehat{\tau}^c + \tau_p^*) \le \zeta^c + \chi^l/\widehat{\tau}^c = \psi\tau_M/(\rho_M\widehat{\tau}^c)$ together with $\widehat{\tau}^c + \tau_p^* \ge \widehat{\tau}^c$ gives

$$\kappa_c = \frac{2\tau_z\chi^l\left[\zeta^c + \chi^l/(\widehat{\tau}^c + \tau_p^*)\right]}{(\widehat{\tau}^c + \tau_p^*)^2} \le \frac{2\tau_z\chi^l \cdot \psi\tau_M/(\rho_M\widehat{\tau}^c)}{(\widehat{\tau}^c)^2} = \frac{2\psi\psi^l\tau_M^2\tau_z}{\rho_M^2(\widehat{\tau}^c)^3}.$$

The right-hand side depends on τ_M only through $\tau_M^2/(\tau_v + \tau_M)^3$, which is maximized at $\tau_M = 2\tau_v$ with value $4/(27\tau_v)$. Substituting yields (A54). □

B.2.2 Driven channel

The second restriction controls the coefficients that are driven by the informativeness path. Even when τ_p^k converges, the price-sensitivity coefficient can diverge if each level overreacts to the previous level's price loading.

Proposition 22 (Driven-channel convergence)**.** *For all $\sigma_M > 0$,*

$$|b_\nu(\tau_p^*)| < \frac{\psi\psi^l\tau_z}{4\rho_M^2\tau_v} = \frac{\psi\psi^l\sigma_v^2}{4\rho_M^2\sigma_z^2}. \tag{A56}$$

Hence $|b_\nu(\tau_p^*)| < 1$ for all $\sigma_M > 0$ whenever

$$\psi^l \le \frac{4\rho_M^2\tau_v}{\psi\tau_z} = \frac{4\rho_M^2\sigma_z^2}{\psi\sigma_v^2}. \tag{A57}$$

Proof. From (A53),

$$|b_\nu(\tau_p)| = \frac{\tau_p}{\tau_M}\,\frac{\rho_A(\zeta - \zeta^c)}{\zeta\left(\rho_A + h_p^A(\tau_p)\phi/\zeta\right)} = \frac{\tau_p}{\tau_M}\,\frac{\rho_A(\zeta - \zeta^c)}{\rho_A\zeta + h_p^A(\tau_p)\phi}.$$

Dropping the non-negative term $h_p^A(\tau_p)\phi$ from the denominator and canceling ρ_A gives

$$|b_\nu(\tau_p)| \le \frac{\tau_p(\zeta - \zeta^c)}{\tau_M\zeta} = \frac{\zeta(\zeta - \zeta^c)\tau_z}{\tau_M},$$

where the last step uses $\tau_p = \zeta^2\tau_z$. At the fixed point, $\zeta^* - \zeta^c = \chi^l/(\widehat{\tau}^c + \tau_p^*) = \psi^l\tau_M/[\rho_M(\widehat{\tau}^c + \tau_p^*)]$, so

$$|b_\nu| \le \frac{\tau_z}{\tau_M}\,\zeta^*(\zeta^* - \zeta^c) \le \frac{\tau_z}{\tau_M} \cdot \frac{\psi\tau_M}{\rho_M\widehat{\tau}^c} \cdot \frac{\psi^l\tau_M}{\rho_M(\widehat{\tau}^c + \tau_p^*)} < \frac{\psi\psi^l\tau_M\tau_z}{\rho_M^2(\widehat{\tau}^c)^2},$$

where we used $\zeta^* \le \zeta^c + \chi^l/\widehat{\tau}^c = \psi\tau_M/(\rho_M\widehat{\tau}^c)$ and $\tau_p^* > 0$. The right-hand side depends

on τ_M only through $\tau_M/(\tau_v + \tau_M)^2$, which is maximized at $\tau_M = \tau_v$ with value $1/(4\tau_v)$. Substituting yields (A56). □

Remark. The sufficient bounds in Propositions 21 and 22 do not involve the AI parameters ρ_A or ϕ. In the driven-channel bound, ρ_A cancels algebraically and the term $h_p^A\phi \geq 0$ is dropped from the denominator. This does not mean that the AI sector is irrelevant for the exact coefficient dynamics: the exact slope $b_\nu(\tau_p)$ still contains ρ_A and ϕ through the term $\rho_A + h_p^A(\tau_p)\phi/\zeta$. The point is that the stated sufficient region is uniform over AI-sector parameters.

B.2.3 Combined condition

The level-k iteration converges to a unique fixed point for all $\sigma_M > 0$ if

$$\psi^l \leq \frac{27\rho_M^2\sigma_z^2}{8\psi\sigma_v^2}. \tag{A58}$$

This bound is sufficient because the driven-channel requirement is weaker: (A57) allows $\psi^l \leq 4\rho_M^2\sigma_z^2/(\psi\sigma_v^2)$, and $27/8 < 4$. Since the bound implies $|b_\nu(\tau_p^*)| < 1$, the regular-initialization restriction is not binding in this sufficient region.

The sufficient convergence region shrinks as σ_z decreases (i.e., as noise trading becomes less volatile).

Remark. Condition (A58) is a uniform sufficient condition, not a necessary characterization of the full parameter space. Its admissible upper bound is proportional to σ_z^2 (equivalently inversely proportional to τ_z). Thus this sufficient region shrinks as noise-trading volatility falls, reflecting the economic force behind instability: when prices are very precise, small changes in perceived signal extraction can generate large changes in higher-level price responses.

C AI Performance Proofs

C.1 Moments and Expected-Profit Formulas

Under the price rule

$$p = \frac{\zeta}{\xi}v + \frac{1}{\xi}z + \frac{\mu}{\xi},$$

with v and z independent,

$$\mathbb{E}[p] = \frac{\zeta}{\xi}\bar{v} + \frac{\mu}{\xi}, \qquad \mathrm{Var}(p) = \left(\frac{\zeta}{\xi}\right)^2 \sigma_v^2 + \frac{\sigma_z^2}{\xi^2}, \qquad \mathrm{Cov}(v,p) = \frac{\zeta}{\xi}\sigma_v^2. \tag{A59}$$

For a rational investor's private signal $\eta_i = v + e_i$,

$$\mathbb{E}[\eta_i] = \bar{v}, \qquad \mathrm{Var}(\eta_i) = \sigma_v^2 + \sigma_M^2, \qquad \mathrm{Cov}(\eta_i, p) = \frac{\zeta}{\xi}\sigma_v^2, \qquad \mathrm{Cov}(\eta_i, v) = \sigma_v^2.$$

The mean excess payoff and mean demands are

$$\bar{v} - \mathbb{E}[p] = \frac{(\xi - \zeta)\bar{v} - \mu}{\xi}, \tag{A60}$$

$$\mathbb{E}[x_i^M] = \frac{(\beta_\eta^M \xi - \beta_p^M \zeta)\bar{v} + \xi\mu^M - \beta_p^M \mu}{\xi}, \tag{A61}$$

$$\mathbb{E}[x_j^A] = \frac{\xi\mu^A - \beta_p^A(\zeta\bar{v} + \mu)}{\xi}. \tag{A62}$$

Proof of Proposition 5. The rational investor uses

$$x_i^M = \beta_\eta^M \eta_i - \beta_p^M p + \mu^M.$$

Gross expected profit decomposes as

$$\mathbb{E}[(v-p)x_i^M] = (\bar{v} - \mathbb{E}[p])\mathbb{E}[x_i^M] + \mathrm{Cov}(v - p, x_i^M).$$

Using $v - p = (1 - \zeta/\xi)v - z/\xi - \mu/\xi$ and

$$x_i^M = \beta_\eta^M(v + e_i) - \beta_p^M\left(\frac{\zeta}{\xi}v + \frac{1}{\xi}z + \frac{\mu}{\xi}\right) + \mu^M,$$

independence of v, e_i, and z gives

$$\mathrm{Cov}(v - p, x_i^M) = \left(1 - \frac{\zeta}{\xi}\right)\left(\beta_\eta^M - \beta_p^M\frac{\zeta}{\xi}\right)\sigma_v^2 + \beta_p^M\frac{\sigma_z^2}{\xi^2}. \tag{A63}$$

Substituting (A60), (A61), and (A63) gives gross profit in primitives:

$$\begin{aligned}\mathbb{E}[(v-p)x_i^M] = \frac{1}{\xi^2}\Big\{&[(\xi-\zeta)\bar{v} - \mu]\left[(\beta_\eta^M\xi - \beta_p^M\zeta)\bar{v} + \xi\mu^M - \beta_p^M\mu\right] \\ &+(\xi-\zeta)(\beta_\eta^M\xi - \beta_p^M\zeta)\sigma_v^2 + \beta_p^M\sigma_z^2\Big\}.\end{aligned} \tag{A64}$$

The variance of rational-investor demand is

$$\mathrm{Var}(x_i^M) = \left(\beta_\eta^M - \beta_p^M \frac{\zeta}{\xi}\right)^2 \sigma_v^2 + (\beta_\eta^M)^2 \sigma_M^2 + (\beta_p^M)^2 \frac{\sigma_z^2}{\xi^2}. \tag{A65}$$

By $\mathbb{E}[(x_i^M)^2] = \mathrm{Var}(x_i^M) + (\mathbb{E}[x_i^M])^2$, expected trading costs are

$$\begin{aligned}\frac{\gamma_M}{2}\mathbb{E}[(x_i^M)^2] = \frac{\gamma_M}{2\xi^2}\Big\{&\left[(\beta_\eta^M \xi - \beta_p^M \zeta)\bar{v} + \xi\mu^M - \beta_p^M \mu\right]^2 \\ &+(\beta_\eta^M \xi - \beta_p^M \zeta)^2 \sigma_v^2 + \xi^2 (\beta_\eta^M)^2 \sigma_M^2 + (\beta_p^M)^2 \sigma_z^2\Big\}.\end{aligned} \tag{A66}$$

Combining (A64) and (A66) gives (34). □

Proof of Proposition 6. The AI investor uses

$$x_j^A = -\beta_p^A p + \mu^A.$$

Gross expected profit decomposes as

$$\mathbb{E}[(v-p)x_j^A] = (\bar{v} - \mathbb{E}[p])\mathbb{E}[x_j^A] + \mathrm{Cov}(v-p, x_j^A).$$

Since x_j^A has no private-signal component,

$$\begin{aligned}\mathrm{Cov}(v-p, x_j^A) &= -\beta_p^A \mathrm{Cov}(v-p, p) \\ &= -\beta_p^A \frac{\zeta}{\xi}\left(1 - \frac{\zeta}{\xi}\right)\sigma_v^2 + \beta_p^A \frac{\sigma_z^2}{\xi^2}.\end{aligned} \tag{A67}$$

Substituting (A60), (A62), and (A67) gives

$$\begin{aligned}\mathbb{E}[(v-p)x_j^A] = \frac{1}{\xi^2}\Big\{&\left[(\xi-\zeta)\bar{v} - \mu\right]\left[\xi\mu^A - \beta_p^A\left(\zeta\bar{v} + \mu\right)\right] \\ &-\beta_p^A \zeta(\xi-\zeta)\sigma_v^2 + \beta_p^A \sigma_z^2\Big\}.\end{aligned} \tag{A68}$$

The variance of AI demand is

$$\mathrm{Var}(x_j^A) = (\beta_p^A)^2\left[\left(\frac{\zeta}{\xi}\right)^2 \sigma_v^2 + \frac{\sigma_z^2}{\xi^2}\right]. \tag{A69}$$

By $\mathbb{E}[(x_j^A)^2] = \mathrm{Var}(x_j^A) + (\mathbb{E}[x_j^A])^2$, expected trading costs are

$$\frac{\gamma_A}{2}\mathbb{E}[(x_j^A)^2] = \frac{\gamma_A}{2\xi^2}\left\{\left[\xi\mu^A - \beta_p^A\left(\zeta\bar{v} + \mu\right)\right]^2 + (\beta_p^A)^2(\zeta^2\sigma_v^2 + \sigma_z^2)\right\}. \tag{A70}$$

Combining (A68) and (A70) gives (35). □

C.2 Profit Comparisons

Throughout this subsection, impose equal trading costs $\gamma_M = \gamma_A \equiv \gamma$. In any informative nondegenerate price environment, write

$$\alpha \equiv \frac{\zeta}{\xi}, \qquad \omega \equiv \frac{\sigma_z^2}{\xi^2}, \qquad \tau_p \equiv \zeta^2 \tau_z, \qquad h \equiv \frac{\tau_M}{\tau_v + \tau_M}, \qquad h_p \equiv \frac{\tau_p}{\tau_v + \tau_p}.$$

The noncentered price is $p = \alpha v + z/\xi + \mu/\xi$. Define

$$\delta \equiv \bar{v} - \mathbb{E}[p]. \tag{A71}$$

Let $\psi \equiv \psi_c + \psi_s$.

Proof of Proposition 7. Both the level-∞ investor and the AI investor are Bayesian best-responders. The level-∞ investor observes (η_i, p), while the AI investor observes p. Thus

$$\Pi^{M,\infty} = \frac{1}{2\gamma}\mathbb{E}\left[(\mathbb{E}[v - p \mid \eta_i, p])^2\right], \qquad \Pi^{A} = \frac{1}{2\gamma}\mathbb{E}\left[(\mathbb{E}[v - p \mid p])^2\right].$$

Let $Y \equiv v - p$, $\mathcal{F} \equiv \sigma(\eta_i, p)$, and $\mathcal{G} \equiv \sigma(p)$. Since $\mathcal{G} \subset \mathcal{F}$,

$$\mathbb{E}[(\mathbb{E}[Y \mid \mathcal{F}])^2] = \delta^2 + \mathrm{Var}(\mathbb{E}[Y \mid \mathcal{F}]), \qquad \mathbb{E}[(\mathbb{E}[Y \mid \mathcal{G}])^2] = \delta^2 + \mathrm{Var}(\mathbb{E}[Y \mid \mathcal{G}]).$$

The unconditional mean term $\delta^2/(2\gamma)$ is therefore common to the two investors and cancels from the difference. By the law of total variance,

$$\mathrm{Var}(\mathbb{E}[Y \mid \mathcal{H}]) = \mathrm{Var}(Y) - \mathbb{E}[\mathrm{Var}(Y \mid \mathcal{H})]$$

for any conditioning set $\mathcal{H}$. Applying this identity to $\mathcal{F}$ and $\mathcal{G}$ gives

$$\mathrm{Var}(\mathbb{E}[Y \mid \mathcal{F}]) - \mathrm{Var}(\mathbb{E}[Y \mid \mathcal{G}]) = \mathbb{E}[\mathrm{Var}(Y \mid \mathcal{G})] - \mathbb{E}[\mathrm{Var}(Y \mid \mathcal{F})].$$

Normality makes the conditional variances constant:

$$\mathrm{Var}(v \mid \eta_i, p) = \frac{1}{\tau_v + \tau_M + \tau_p}, \qquad \mathrm{Var}(v \mid p) = \frac{1}{\tau_v + \tau_p}.$$

Since p belongs to both conditioning sets, $\mathrm{Var}(v - p \mid \cdot) = \mathrm{Var}(v \mid \cdot)$. Hence

$$\begin{aligned}\Pi^{M,\infty} - \Pi^{A} &= \frac{1}{2\gamma}\left[\frac{1}{\tau_v + \tau_p} - \frac{1}{\tau_v + \tau_M + \tau_p}\right] \\ &= \frac{\tau_M}{2\gamma(\tau_v + \tau_p)(\tau_v + \tau_M + \tau_p)}.\end{aligned}$$

The expression is strictly positive whenever $\tau_M > 0$. □

Proof of Lemma 8. Let $\widetilde{v} \equiv v - \bar{v}$ and $\widetilde{p} \equiv p - \mathbb{E}[p]$. Since

$$\widetilde{p} = \alpha\widetilde{v} + \frac{z}{\xi}, \qquad v - p = \delta + (1-\alpha)\widetilde{v} - \frac{z}{\xi},$$

the cursed investor's demand can be written as

$$x_i^{M,c} = \frac{h\eta_i + (1-h)\bar{v} - p}{\gamma} = \frac{\delta - \widetilde{p}}{\gamma} + \frac{h(\eta_i - \bar{v})}{\gamma}.$$

Thus the common prior-minus-price component is $(\delta - \widetilde{p})/\gamma$, while the investor-specific information correction is $h(\eta_i - \bar{v})/\gamma$. Equivalently,

$$x_i^{M,c} = \frac{\delta + (h-\alpha)\widetilde{v} + he_i - z/\xi}{\gamma}.$$

The gross expected return is therefore

$$\begin{aligned}\mathbb{E}[(v-p)x_i^{M,c}] &= \frac{1}{\gamma}\mathbb{E}\left[\left(\delta + (1-\alpha)\widetilde{v} - \frac{z}{\xi}\right)\left(\delta + (h-\alpha)\widetilde{v} + he_i - \frac{z}{\xi}\right)\right] \\ &= \frac{1}{\gamma}\left[\delta^2 + (1-\alpha)(h-\alpha)\sigma_v^2 + \omega\right],\end{aligned}$$

where independence eliminates all cross terms involving e_i, z, and $\widetilde{v}$. The expected trading cost is

$$\frac{\gamma}{2}\mathbb{E}[(x_i^{M,c})^2] = \frac{1}{2\gamma}\left[\delta^2 + (h-\alpha)^2\sigma_v^2 + h^2\sigma_M^2 + \omega\right].$$

The identity

$$h^2\sigma_M^2 = h(1-h)\sigma_v^2$$

follows from $h = \tau_M/(\tau_v + \tau_M)$, $\sigma_M^2 = 1/\tau_M$, and $\sigma_v^2 = 1/\tau_v$. Substituting this identity and collecting terms gives

$$\Pi^{M,c} = \frac{\delta^2}{2\gamma} + \frac{(h - 2\alpha + \alpha^2)\sigma_v^2 + \omega}{2\gamma}, \tag{A72}$$

or, equivalently,

$$\Pi^{M,c} = \frac{\delta^2}{2\gamma} + \frac{\sigma_v^2}{2\gamma}\left[h - 2\alpha + \alpha^2 + \omega\tau_v\right]. \tag{A73}$$

The AI investor is a Bayesian best-responder to p, so

$$\Pi^A = \frac{1}{2\gamma}\mathbb{E}\left[(\mathbb{E}[v - p \mid p])^2\right].$$

The equivalent price signal is

$$s_p = \frac{\xi}{\zeta}\left(p - \frac{\mu}{\xi}\right) = v + \frac{z}{\zeta},$$

so s_p is a signal about v with noise variance $1/\tau_p$. Thus

$$\mathbb{E}[v \mid p] = h_p s_p + (1 - h_p)\bar{v}, \qquad h_p = \frac{\tau_p}{\tau_v + \tau_p}.$$

Because $\widetilde{p} = \alpha(s_p - \bar{v})$,

$$\mathbb{E}[v - p \mid p] = \delta + \left(\frac{h_p}{\alpha} - 1\right)\widetilde{p}.$$

The AI demand can therefore be decomposed as

$$x_j^A = \frac{\delta - \widetilde{p}}{\gamma} + \frac{h_p(s_p - \bar{v})}{\gamma}.$$

The first term is the same prior-minus-price component that appears in cursed investor demand. The second term is the AI's price-signal correction. Since $s_p - \bar{v} = \widetilde{p}/\alpha$, the AI's total price loading differs from the cursed investor's total price loading, but the difference is exactly the price-information channel summarized by h_p.

Using $\mathrm{Var}(\widetilde{p}) = \alpha^2\sigma_v^2 + \omega$, the optimal-demand representation gives

$$\begin{aligned}\Pi^A &= \frac{1}{2\gamma}\mathbb{E}\left[\left(\delta + \left(\frac{h_p}{\alpha} - 1\right)\widetilde{p}\right)^2\right] \\ &= \frac{\delta^2}{2\gamma} + \frac{(h_p - \alpha)^2}{2\gamma\alpha^2}\left(\alpha^2\sigma_v^2 + \omega\right).\end{aligned}$$

Since $\omega/\alpha^2 = \sigma_z^2/\zeta^2 = 1/\tau_p$, and

$$\sigma_v^2 + \frac{\omega}{\alpha^2} = \frac{1}{\tau_v} + \frac{1}{\tau_p} = \frac{\sigma_v^2}{h_p},$$

the AI profit is

$$\Pi^A = \frac{\delta^2}{2\gamma} + \frac{(h_p - \alpha)^2 \sigma_v^2}{2\gamma h_p}. \tag{A74}$$

Equivalently,

$$\frac{(h_p - \alpha)^2}{h_p} = h_p - 2\alpha + \frac{\alpha^2}{h_p} = h_p - 2\alpha + \alpha^2 + \frac{\alpha^2 \tau_v}{\tau_p}.$$

Since $\omega = \alpha^2/\tau_p$, this becomes

$$\Pi^A = \frac{\delta^2}{2\gamma} + \frac{\sigma_v^2}{2\gamma}\left[h_p - 2\alpha + \alpha^2 + \omega\tau_v\right]. \tag{A75}$$

Subtracting (A75) from (A73) gives the result directly:

$$\Pi^{M,c} - \Pi^A = \frac{\sigma_v^2(h - h_p)}{2\gamma} = \frac{h - h_p}{2\gamma\tau_v}.$$

Since h and h_p are increasing transformations of τ_M and τ_p, respectively, the sign is positive if and only if $\tau_M > \tau_p$. □

Proof of Proposition 9. By Lemma 8, it suffices to show $\tau_M > \tau_p$ for all $\tau_M > 0$. The aggregate signal loading satisfies

$$\zeta \leq \psi\beta_\eta^{M,c} = \frac{\psi\tau_M}{\gamma(\tau_v + \tau_M)}.$$

Thus

$$\tau_p = \zeta^2\tau_z \leq \frac{\psi^2\tau_M^2\tau_z}{\gamma^2(\tau_v + \tau_M)^2}.$$

The inequality $\tau_M > \tau_p$ is implied by

$$\frac{(\tau_v + \tau_M)^2}{\tau_M} > \frac{\psi^2\tau_z}{\gamma^2}.$$

Let

$$f(\tau_M) \equiv \frac{(\tau_v + \tau_M)^2}{\tau_M} = \frac{\tau_v^2}{\tau_M} + 2\tau_v + \tau_M.$$

Then

$$f'(\tau_M) = 1 - \frac{\tau_v^2}{\tau_M^2},$$

so the unique minimizer is $\tau_M = \tau_v$, with minimum $4\tau_v$. Hence $\tau_M > \tau_p$ for all $\tau_M > 0$ whenever

$$4\tau_v > \frac{\psi^2\tau_z}{\gamma^2},$$

which is equivalent to $\sigma_z > \psi\sigma_v/(2\gamma)$. □

Convergence compatibility. The threshold $\sigma_z > \psi\sigma_v/(2\gamma)$, where $\psi = \psi_c + \psi_s$, is compatible with the sufficient fixed-point convergence condition

$$\psi_s \leq \frac{27\gamma^2\sigma_z^2}{8\psi\sigma_v^2}.$$

At the boundary $\sigma_z = \psi\sigma_v/(2\gamma)$, this condition becomes $\psi_s \leq 27\psi/32$.

For the AI-dominance window established below, Proposition 10 requires $\sigma_z < \psi_c\sigma_v/(2\gamma)$. At the boundary $\sigma_z = \psi_c\sigma_v/(2\gamma)$, the same sufficient convergence condition becomes $\psi_s \leq 27\psi_c^2/(32\psi)$. Proposition 10 assumes nondegenerate fixed-point convergence for each τ_M under consideration; these inequalities give sufficient conditions under which that convergence assumption holds at the respective boundaries.

Proof of Proposition 10. By Lemma 8, AI investors outperform cursed investors if and only if

$$\tau_p^*(\tau_M) > \tau_M.$$

The fixed-point map for price informativeness depends only on signal loadings:

$$G_c(\tau_p) = \tau_z\left(\zeta^c + \frac{\chi^s}{\widehat{\tau}^c + \tau_p}\right)^2, \qquad \widehat{\tau}^c \equiv \tau_v + \tau_M, \qquad \zeta^c \equiv \psi_c\frac{\tau_M}{\gamma(\tau_v + \tau_M)}, \qquad \chi^s \equiv \frac{\psi_s\tau_M}{\gamma}. \tag{A76}$$

Nonzero $\bar{v}$ and S affect the intercept channel, but the assumed convergence of that channel ensures that the noncentered objective environment is well defined. Define $g(\tau_p;\tau_M) \equiv G_c(\tau_p) - \tau_p$. Since G_c is strictly decreasing and the identity map is strictly increasing, g is strictly decreasing in τ_p. The fixed point satisfies $g(\tau_p^*(\tau_M);\tau_M) = 0$, so $\tau_p^*(\tau_M) > \tau_M$ if and only if $g(\tau_M;\tau_M) > 0$. This is equivalent to

$$\varphi(\tau_M) \equiv \frac{G_c(\tau_M)}{\tau_M} = \frac{\tau_z\tau_M}{\gamma^2}\left(\frac{\psi_c}{\tau_v + \tau_M} + \frac{\psi_s}{\tau_v + 2\tau_M}\right)^2 > 1. \tag{A77}$$

Step 1: the relative informativeness ratio vanishes at both extremes. As $\tau_M \to 0$, the cursed investor's signal coefficient satisfies

$$\beta_\eta^{M,c} = \frac{\tau_M}{\gamma(\tau_v + \tau_M)} \to 0.$$

Strategic investors' signal loadings are also bounded by the cursed loading, so the aggregate

signal loading is bounded above by $\psi\beta_\eta^{M,c}$. Hence $\zeta = O(\tau_M)$, $\tau_p^*(\tau_M) = \zeta^2\tau_z = O(\tau_M^2)$, and

$$\frac{\tau_p^*(\tau_M)}{\tau_M} \to 0.$$

As $\tau_M \to \infty$, $\beta_\eta^{M,c} \to 1/\gamma$, so $\zeta \leq \psi/\gamma$ and $\tau_p^*(\tau_M) \leq \psi^2\tau_z/\gamma^2$. This bound is independent of τ_M, so again

$$\frac{\tau_p^*(\tau_M)}{\tau_M} \to 0.$$

Step 2: the dominance region is non-empty when noise trading is small. The fixed point $\tau_p^*(\tau_M)$ is continuous in τ_M because $G_c(\tau_p) - \tau_p$ is continuous in (τ_p, τ_M) and strictly decreasing in τ_p. To find an interior point at which AI dominates, evaluate the relative price precision at $\tau_M = \tau_v$. The cursed-investor signal-loading floor gives

$$\zeta \geq \psi_c\beta_\eta^{M,c}.$$

At $\tau_M = \tau_v$, this gives

$$\frac{\tau_p^*(\tau_v)}{\tau_v} \geq \frac{(\psi_c/(2\gamma))^2\tau_z}{\tau_v} = \frac{\psi_c^2\sigma_v^2}{4\gamma^2\sigma_z^2}.$$

This exceeds one when $\sigma_z < \psi_c\sigma_v/(2\gamma)$, so the AI-dominance region is non-empty.

Step 3: the dominance set has a two-threshold structure. The function φ in (A77) is a positive constant times the square of

$$q(\tau_M) \equiv \sqrt{\tau_M}\left(\frac{\psi_c}{\tau_v + \tau_M} + \frac{\psi_s}{\tau_v + 2\tau_M}\right).$$

Because $q(\tau_M) \geq 0$, it suffices to show that q is unimodal. Differentiating q, multiplying by $2\sqrt{\tau_M}(\tau_v + \tau_M)^2(\tau_v + 2\tau_M)^2 > 0$, and setting the derivative equal to zero yields

$$\frac{\psi_c(\tau_v - \tau_M)}{(\tau_v + \tau_M)^2} = \frac{\psi_s(2\tau_M - \tau_v)}{(\tau_v + 2\tau_M)^2}.$$

The left-hand side is positive only when $\tau_M < \tau_v$, and the right-hand side is positive only when $\tau_M > \tau_v/2$. Any solution therefore lies in $(\tau_v/2, \tau_v)$. On this interval, the left-hand side is strictly decreasing from $2\psi_c/(9\tau_v)$ to 0, while the right-hand side is strictly increasing from 0 to $\psi_s/(9\tau_v)$. Hence there is exactly one critical point. The function q extends continuously to zero at $\tau_M = 0$, is positive on $(0, \infty)$, and converges to zero as $\tau_M \to \infty$. Its unique critical point is therefore a maximum, so φ is unimodal.

Because φ is unimodal, the super-level set $\{\tau_M : \varphi(\tau_M) > 1\}$ is either empty or a single interval. Step 2 shows that it is non-empty. Write it as (τ_M^a, τ_M^b). Since $\sigma_M = 1/\sqrt{\tau_M}$ is

strictly decreasing in τ_M, define

$$\sigma_M^L \equiv \frac{1}{\sqrt{\tau_M^b}}, \qquad \sigma_M^H \equiv \frac{1}{\sqrt{\tau_M^a}}.$$

Then $\Pi^{M,c} < \Pi^A$ if and only if $\sigma_M \in (\sigma_M^L, \sigma_M^H)$. □