The Limits of AI Trading

Winston Wei Dou Itay Goldstein Zigang Li Liyan Yang *

August 14, 2026

Abstract

We show that powerful autonomous learning does not ensure trading dominance in competitive markets, even when humans hold systematically biased beliefs. We characterize how information structure and market sophistication bound reinforcement-learning traders' performance relative to humans. Humans receive private signals, potentially including soft information unavailable to algorithms, and differ in their ability to extract information from prices and anticipate others' trading, as captured by level-k reasoning. AI traders learn demand schedules from prices and realized profits and may also receive private signals. Despite heterogeneous initial conditions and persistent differences in learning parameters, all AI traders converge to a common probabilistic trading policy, generating algorithmic herding, which amplifies the equilibrium-price response to aggregate AI demand and compresses profits. Without an informational advantage, AI is outperformed by the most sophisticated humans because it learns from market data generated by average, rather than frontier, human sophistication. Even the least sophisticated humans can outperform AI when their private information is either very precise or very imprecise. AI sophistication rises with that of the humans who compete with and train it.

Keywords: AI Dominance, Heterogeneous-Agent Asset Pricing, Training Data Limitations, Bounded Rationality, Level-K Reasoning, Reinforcement Learning.
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 Jean-Edouard Colliard, Snehal Banerjee, Philip Bond, Peter DeMarzo, Andrea Eisfeldt, Zhiguo He, Maureen O'Hara, Thomas Sargent, Yajun Wang, and Mao Ye, as well as seminar and conference participants at the AFA Panel, the Bristol-Essex-UCL Workshop on Decentralised Financial Markets, the Conference on the Frontiers of AI in Economics, the LHA Annual Conference, the NBER Summer Institute, the WFA, and the Wolfe AI and Financial Markets Workshop.

1 Introduction

Artificial intelligence is rapidly transforming how investors acquire information, develop trading strategies, and execute trades. Reinforcement-learning (RL) systems are especially consequential for financial trading because they learn directly from realized trading outcomes and adapt autonomously without requiring an explicit, fully specified model of the market. Their performance, however, cannot be evaluated in isolation because prices and trading profits are endogenously determined by the market's information structure and the strategic sophistication of human investors. This paper characterizes the performance of RL traders relative to human investors across market environments, identifies the conditions under which RL traders outperform humans, and explains how information structures and human strategic sophistication determine the boundaries between AI and human advantage.

Characterizing these boundaries requires separating the contribution of algorithmic learning from other potential sources of trading advantage, including superior information and market power, commonly bundled with AI. It also requires an equilibrium framework in which human investors respond strategically to AI rather than serve as a fixed background. We develop such a theoretical framework, rather than adopt the simulation-based experimental approaches in existing work (e.g., Dou, Goldstein, and Ji, 2023, 2026; Colliard, Foucault, and Lovo, 2026). By modeling all human and AI investors as atomistic price takers, we abstract from individual market power and AI-driven price manipulation. We allow either group to possess superior private information and allow human investors to differ in strategic sophistication. This structure isolates how relative performance is shaped by private-information quality, inference from prices, and human strategic sophistication, which also shapes the market environment in which AI learns.

The competitive market setting we study is also increasingly relevant in practice. Autonomous AI trading is no longer confined to large financial institutions. Platforms such as Robinhood now allow retail investors to connect AI agents directly to brokerage accounts, where the agents can make investment decisions and execute trades on their behalf. Our model captures the competitive limit of this development, in which a positive mass of atomistic investors delegates trading to autonomous RL algorithms while each investor remains a price taker. The framework therefore studies the market consequences of broad AI adoption under competition, rather than strategic interactions among a small number of AI traders with market power (e.g., Dou, Goldstein, and Ji, 2023, 2026).

The economy is a repeated competitive market for a risky asset. Continua of human and AI investors submit demand schedules and trade against noise traders. Both groups may receive conditionally independent private signals about the asset payoff. Human investors

know the relevant payoff distribution but differ in how deeply they reason about price formation. Cursed investors use their private signals correctly but neglect the information about other investors' signals embedded in prices, following the central friction in Eyster, Rabin, and Vayanos (2019). More sophisticated humans perform progressively deeper rounds of reasoning. At each higher level, strategic human investors best respond to a perceived market in which other strategic humans reason one level less deeply and AI investors follow policies learned in that lower-level market. This recursive structure builds on the level-k approach to bounded strategic reasoning[1] and extends it to human-AI competition by embedding endogenous AI learning within the level-k reasoning hierarchy.

AI investors begin without knowledge of the payoff distribution, the price-formation rule, or other investors' demand schedules. Instead, they learn complete signal- and price-contingent demand schedules from their own realized trading profits. Our leading implementation uses stateless tabular Q-learning with Boltzmann exploration. AI investors may also receive private signals whose precision can be higher or lower than that of human signals, allowing the information advantage to lie with either group. The boundary case in which AI receives no private signal provides an especially sharp benchmark. Price-only AI has no direct fundamental information and must learn entirely from equilibrium prices, thereby isolating the value and limits of algorithmic price inference.

Human reasoning and algorithmic learning are linked through market clearing. At each level of the hierarchy, human demand, learned AI demand, and noise trading jointly determine a distinct price process. AI investors learn from the resulting prices and profits, while human investors at the next level anticipate the AI policy learned in the preceding environment when forming beliefs about price formation. Humans therefore hold explicit beliefs about the price process, whereas AI investors encode an implicit model of that process in their learned action values and trading policies. The resulting level-k hierarchy embeds a positive mass of autonomous RL traders within an equilibrium model of information aggregation, making market prices endogenous training data rather than externally supplied inputs.

Our first set of results characterizes AI learning within any fixed human-sophistication environment. Each AI investor learns a complete demand schedule from realized trading profits, choosing how its position responds to its private signal and the market price. Under regularity conditions, despite heterogeneous initial Q-values and persistent differences across algorithms, the policy-implied population-average demand converges almost surely

[1] See, e.g., Stahl and Wilson (1994, 1995), Nagel (1995), Costa-Gomes and Crawford (2006), and Crawford, Costa-Gomes, and Iriberri (2013). For an application of level-k reasoning to financial markets, see Zhou (2022).

to a unique learning fixed point. In the atomistic limit, this fixed point coincides with the unique rational-expectations demand rule available to AI given its information. The rule optimally combines the AI investor's private signal, when available, with the information extracted from the equilibrium price.

These results formalize algorithmic herding at the level of trading policies, not merely at the aggregate level. All AI investors converge to the same probabilistic policy, although their realized trades need not coincide because each investor independently draws a demand schedule from that policy and may observe a different private signal. The mechanism operates through market-clearing feedback. Each algorithm learns independently from its own realized profits, but those profits depend on a market price jointly determined by aggregate human and AI demand. Prices therefore provide a common endogenous feedback channel that aligns the learned policies without communication or coordinated exploration. Because each AI investor remains an atomistic price taker with vanishing individual price impact, this convergence does not arise from market power or the internalization of price impact. Algorithmic herding therefore means convergence to a common probabilistic trading policy, not synchronization of realized trades.

To establish the fully sophisticated human benchmark, we also characterize the limit of increasingly deep human reasoning. We derive primitive conditions under which human beliefs, human demand, learned AI policies, and equilibrium prices converge. Whenever this level-k hierarchy converges as k approaches infinity, its limit is a rational-expectations equilibrium in which fully sophisticated humans and AI traders correctly understand the same price process and best respond to one another, while cursed investors continue to neglect the information embedded in prices.

Our main results map the boundaries of AI performance across information environments. We impose equal trading costs so that profit differences reflect information and the ability to extract it from prices. Because AI and fully sophisticated humans interpret the equilibrium price correctly, they share the same price information, and their ranking depends entirely on their private signals. AI outperforms fully sophisticated humans if and only if its private signal is more precise. A more informative price narrows the profit gap by improving the common information available to both groups, but cannot reverse their ranking. Algorithmic learning alone therefore cannot overcome inferior private information.

The comparison with cursed investors is different because they do not extract information from prices. AI outperforms them if and only if the combined informativeness of its private signal and the equilibrium price exceeds the informativeness of their private signal. Correct price inference can therefore compensate for inferior private information, but only when informed trading makes prices sufficiently revealing.

This distinction produces especially sharp results when AI traders have no private signal. Fully sophisticated humans always outperform price-only AI traders because they observe the same price, add a private signal, and use both sources of information correctly. Relative to cursed investors, however, AI performance is non-monotonic in human signal quality. Highly precise human signals give cursed investors a strong direct information advantage, whereas highly imprecise signals generate too little informed trading for prices to reveal much. Price-only AI can outperform cursed investors only between these extremes, when human trading makes prices informative but the direct human signal is not overwhelmingly valuable.

Noise trading further narrows the conditions under which AI has an advantage. Above an explicit threshold of noise-trader volatility, cursed investors outperform price-only AI for every level of human signal precision because noise obscures the private information embedded in prices. When noise trading is sufficiently limited, price-only AI can outperform cursed investors, but only over an intermediate range of human signal precision. Complementary numerical results show that these limits extend beyond cursed investors to humans with finite but nontrivial reasoning depth. Price-only AI outperforms these more sophisticated humans only within a narrow range of market conditions, and even informed AI can underperform them when their private signals are more precise.

These results identify three forces that limit AI's relative performance. First, AI cannot directly use information possessed by human investors but unavailable to the algorithm. When AI has inferior private information, learning from prices can compensate only if prices reveal enough of the information held by humans. Second, the information content of prices is endogenous and multifaceted. Human signals must be sufficiently precise for human trading to convey meaningful fundamental information to prices, while noise trading must be sufficiently limited for AI to recover that information. Human sophistication also matters. More sophisticated humans better understand how prices aggregate information and shade their demand accordingly, making prices less informative and weakening AI's ability to infer human private information from them. Human sophistication therefore shapes both the information available in prices and the market environment from which AI learns, ultimately influencing the sophistication of the policies learned by AI. Third, although each AI trader is an atomistic price taker with no individual price impact, endogenous algorithmic herding leads all AI traders to adopt a common probabilistic trading policy. The aggregate demand generated by this common policy affects equilibrium prices, compressing the trading opportunities available to the AI sector and limiting the profitability of all AI traders.

Together, these mechanisms explain why powerful autonomous learning does not ensure trading dominance in a competitive market, even when human investors form systematically

flawed beliefs. AI can learn the rational-expectations demand rule that makes optimal use of its available information, but learning cannot create missing private information, remove noise from prices, or fully prepare AI for sophisticated human behavior absent from its training environment. Algorithmic herding imposes a further limit. As AI traders converge to a common probabilistic policy, their policy-implied demands respond to market conditions in the same way. This alignment strengthens the response of equilibrium prices to the AI sector's aggregate demand, thereby compressing its trading profits. The effect arises at the population level even though each AI trader remains an atomistic price taker with vanishing individual price impact. Convergence to a sophisticated trading policy therefore need not translate into superior trading performance.

Our final results isolate how the sophistication of the training environment shapes AI performance at deployment. We focus on price-only AI traders, which receive no private information. We train them in markets populated by humans with different depths of reasoning, freeze the resulting policies, and deploy each pretrained policy in a common test environment. In this environment, fully sophisticated humans understand the deployed AI policy and respond optimally to the price process it generates. Market primitives and population shares are held fixed across test markets. Because price-only AI contributes no independent fundamental information to aggregate demand, differences in its deployed policies rescale and shift equilibrium prices without changing their signal-to-noise ratio. Fully sophisticated humans correctly account for these transformations when extracting information from prices. Equilibrium price informativeness is therefore identical across test markets, so differences in AI performance isolate the sophistication embedded in the policies learned during training rather than differences in the information available at deployment.

AI's relative profit in the test market separates into a fixed information component and deployment regret. The information component reflects the relative value of the information contained in prices and human private signals. Deployment regret is the profit lost because AI uses a policy learned in the training market with lower sophistication level, rather than the policy that would be optimal in the test market with fully rational-expectations human investors as sophisticated ones. The same deployment-regret term enters AI's profit gaps relative to cursed and fully sophisticated humans. Consequently, both gaps change by exactly the same amount across training environments. Under regularity conditions, we show that training against more sophisticated humans produces a policy better adapted to the test market, reduces deployment regret, and improves AI's relative performance against both human benchmarks. Strict stability guarantees that this improvement occurs at sufficiently high levels of training sophistication, even if it does not occur at every earlier level. More sophisticated training can enable price-only AI to overtake cursed investors in a favorable

low-noise environment. It cannot overcome the information advantage of a fully sophisticated human who observes the same price, adds a private signal, and reoptimizes in response to the deployed AI policy.

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.[2] 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 (2026) 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 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 mar-

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

ket 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 misperception about others' behavior when facing a complex market environment that includes AI traders. By using the Cognitive Hierarchy 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).[3][4] We model human investors' strategic thinking using the Cognitive Hierarchy 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-

[3]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).

[4]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).

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 The Baseline 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: $v_t \sim \mathcal{N}(\bar{v}, \sigma_v^2)$. Let $\tau_v \equiv \sigma_v^{-2}$ denote the payoff precision. The asset payoffs are independently and identically distributed (i.i.d.) over time.

Human investors. There is a continuum of human investors with total measure ψ who differ in sophistication. Sophistication determines how accurately an investor understands price formation and interprets the information contained in prices. Each investor forms subjective beliefs about asset payoffs and equilibrium prices and chooses a demand schedule that maximizes expected utility under those beliefs. Investors are therefore rational in the sense that their trading decisions are optimal given their beliefs, even when those beliefs are incorrect. Rational expectations imposes the additional requirement that subjective beliefs coincide with the objective equilibrium distribution.

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

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

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 $\frac{1}{2}\gamma_M x_{i,t}^2$ represents transaction costs, and γ_M controls the level of these costs.[5]

Human investors 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 human 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., $e_{i,t} \sim \mathcal{N}(0, \sigma_M^2)$, and $\tau_M \equiv \sigma_M^{-2}$ denotes the human-signal precision. The errors are independent of the fundamental and all other shocks.

Human investors also differ in strategic sophistication. All investors choose optimally given their subjective beliefs, but less sophisticated investors may misperceive the sophistication and trading behavior of other market participants. They therefore form incorrect beliefs about how equilibrium prices aggregate private information and, as a result, draw imperfect inferences about asset payoffs from prices.

We model this heterogeneity using two groups of human 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 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

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

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

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

where γ_A 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, the strategies of other market participants, or how to use available information optimally. In the informed case, the AI nevertheless observes a period-specific private signal

$$\eta_{j,t} = v_t + e_{j,t}, \qquad e_{j,t} \sim \mathcal{N}(0, \sigma_A^2),$$

where the errors are i.i.d. across AI investors and time and are independent of human signal errors and all other shocks. For finite σ_A^2, let $\tau_A \equiv \sigma_A^{-2}$ denote the AI-signal precision.

Such signals can be produced by data acquisition and information technologies paired with the trading algorithm. The human signal is more precise than the AI signal when $\sigma_M < \sigma_A$ (equivalently when $\tau_M > \tau_A$); the AI signal is more precise when $\sigma_M > \sigma_A$ (equivalently when $\tau_M < \tau_A$). Thus, the model allows the information advantage to run in either direction. We also allow the boundary $\tau_A = 0$, interpreted as $\sigma_A^2 \to \infty$, which nests the benchmark in which AI investors receive no private information and learn only from prices.

The AI learns how to combine its private signal with prices through repeated market interactions. In each period, the algorithm chooses a signal- and price-contingent 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 implicitly encoded in an AI investor's learned value function and trading policy when the training environment is populated by level-k human investors. Unlike human investors, who form explicit subjective models of price formation, AI investors do not reason through an explicit pricing rule. Instead, the relevant mapping is learned indirectly through action values. Demand schedules that generate higher realized profits under market-clearing prices receive higher values and are selected more frequently over time. The learned value function and policy therefore embed an implicit model of how prices respond to trading behavior in the level-k environment.

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 (2026) 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 learning 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 j explores using a Boltzmann (softmax) policy over its learned action values. Let $Q_{j,t}(x)$ denote AI investor j'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 three-dimensional. In an environment populated with level-k thinking investors, each AI investor j's demand schedule is given by

$$x^{A,k}_{j,t}(p_t;\eta_{j,t}) \equiv \mu^{A,k}_{j,t} + \beta^{A,k}_{\eta,j,t}\eta_{j,t} - \beta^{A,k}_{p,j,t}p_t. \tag{8}$$

Hence, each arm in the Q-learning problem is a complete signal- and price-contingent schedule, represented by the coefficient triple

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

[6]The smoothness of these choice probabilities, rather than the hard switching of an ε-greedy rule, permits the contraction and local-stability arguments under the conditions stated in Appendix A.

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 $z_t \sim \mathcal{N}(0, \sigma_z^2)$, and $\tau_z \equiv \sigma_z^{-2}$ denotes its precision. The demand z_t is independent of other shocks and across time.

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. Human 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 human 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

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

   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. Each AI conditions its chosen schedule on its private signal and on the price information encoded

by its learned price-formation rule. Thus, its effective information set is

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

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:

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

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

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 the AI population has policy-implied average coefficients

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

Then the equilibrium price aggregates information from the private signals of cursed, level-k, and AI investors, together with noise-trader supply. A rational-expectations investor can infer additional payoff information 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 price informativeness 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,

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

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 coefficients are

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

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

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

We restrict attention to price-admissible environments, in which

$$\xi^{(k)} \neq 0, \tag{15}$$

so market clearing determines a unique finite linear price. Equivalently,

$$p_t = \frac{\mu^{(k)}}{\xi^{(k)}} + \frac{\zeta^{(k)}}{\xi^{(k)}} v_t + \frac{1}{\xi^{(k)}} z_t. \tag{16}$$

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 on noise-trader demand. In particular, parameter $\xi^{(k)}$ is a measure of market liquidity – i.e., the inverse of "Kyle's lambda" (Kyle, 1985) – which captures the market depth: a higher

$\xi^{(k)}$ means that noisy trading has a smaller price impact, and so the market is deeper.[7]

The noisy relationship between equilibrium price and the fundamental then implies the signal

$$s_{p,t}^{(k)} \equiv \frac{\xi^{(k)}}{\zeta^{(k)}}\left(p_t - \frac{\mu^{(k)}}{\xi^{(k)}}\right) = v_t + \frac{1}{\zeta^{(k)}} z_t. \tag{17}$$

The standard deviation of the noise term in this price-implied signal is

$$\sigma_p^{(k)} = \frac{\sigma_z}{|\zeta^{(k)}|}, \tag{18}$$

and its precision, a measure of price informativeness, is

$$\tau_p^{(k)} = \frac{1}{\left(\sigma_p^{(k)}\right)^2} = \left(\zeta^{(k)}\right)^2 \tau_z. \tag{19}$$

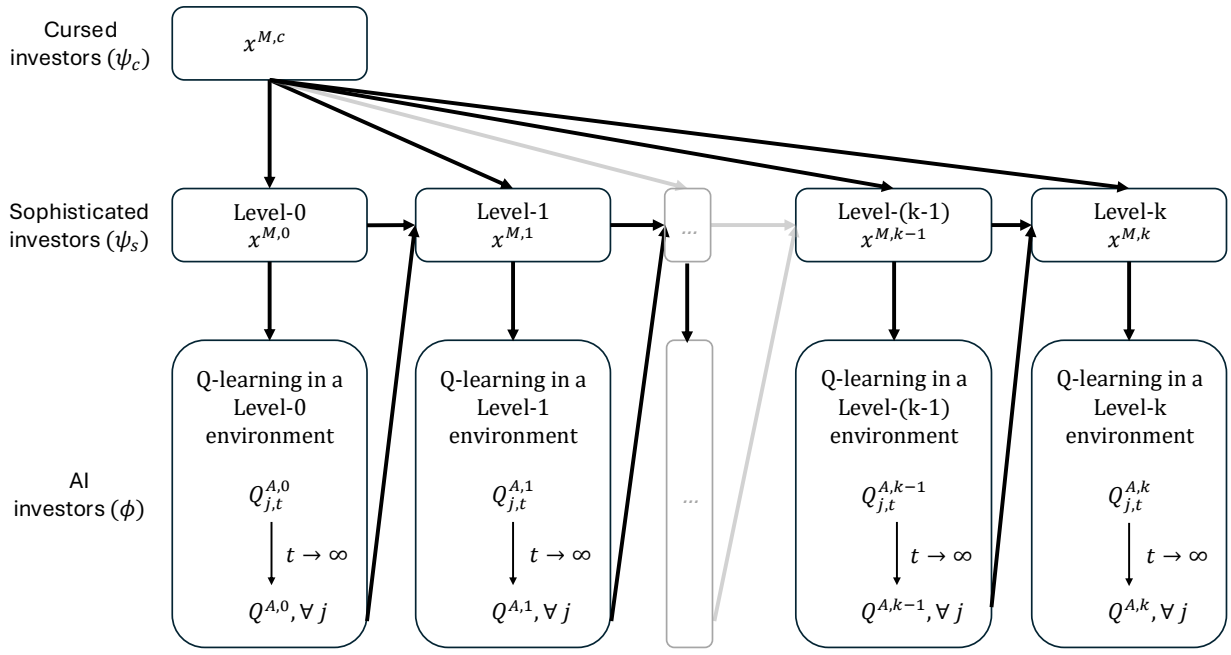
Price informativeness about v_t is therefore governed by the magnitude of the aggregate signal loading $\zeta^{(k)}$: stronger signal-based demand from either human or AI investors makes the price-implied signal less noisy and increases the precision with which prices reveal fundamentals.

2.2 Equilibrium Demand Schedules

Human 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 human investors, and AI investors. In this section, we characterize the equilibrium demand schedules of human 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

[7] Price admissibility does not itself require $\zeta^{(k)} \neq 0$. If $\zeta^{(k)} = 0$, the price is independent of v_t and we set $\tau_p^{(k)} = 0$. Along the level-k hierarchy considered below, $\zeta^{(k)} > 0$.

Figure 1: The Level-k Iteration. The AI node tracks the learned aggregate AI demand rule in each training environment, rather than any individual algorithm's realized action.



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 human 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

$$\mathbb{E}\left[v_t \mid \eta_{i,t}\right] = h_v^c \bar{v} + h_\eta^c \, \eta_{i,t}, \tag{20}$$

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

$$h_\eta^c \equiv \frac{\tau_M}{\tau_v + \tau_M} \quad \text{and} \quad h_v^c \equiv \frac{\tau_v}{\tau_v + \tau_M}. \tag{21}$$

The optimal demand schedule of the cursed investor i is:

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

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

where

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

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 how independent learning can align AI trading through market-clearing price feedback. 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, i.e., the policy-implied population average of the coefficient triple $(\mu^{A,k}, \beta_\eta^{A,k}, \beta_p^{A,k})$, toward a common rule. The limiting common rule has a sharp interpretation. Under the convergence and ordered-refinement conditions stated in Section 3, the population-average AI demand converges to the demand of an informed atomistic rational-expectations investor who observes the private signal η_j and the equilibrium price, understands the true price-formation rule, and uses the price to infer payoff information. Convergence of these aggregate coefficients does not require individual algorithms to adopt identical policies or take identical realized actions: each AI conditions its demand on its own signal.

We now focus on characterizing $(\mu^{A,k}, \beta_\eta^{A,k}, \beta_p^{A,k})$ for each level-k environment. In the limit, all AI investors become homogeneous in learned Q-value.[8] Thus, the representative AI investor behaves as if it correctly anticipates that the ψ_c cursed investors use the demand rule in (23), the ψ_s level-k investors use level-k human strategies, and the remaining AI sector follows the aggregate learned level-k AI rule. Each AI investor conditions on $(\eta_{j,t}, p_t)$. Equivalently, it combines its private signal with the price-implied signal $s_{p,t}^{(k)}$ in (17). Its

[8] In the limit, all AI investors learn the same relative Q-values across actions; their Q-value levels may differ by investor-specific additive shifts that do not affect their Boltzmann policies.

posterior is

$$\mathbb{E}[v_t \mid \eta_{j,t}, s_{p,t}^{(k)}] = h_\eta^{A,k}\eta_{j,t} + h_p^{A,k}s_{p,t}^{(k)} + h_v^{A,k}\bar{v},$$

where

$$h_\eta^{A,k} \equiv \frac{\tau_A}{\tau_v + \tau_A + \tau_p^{(k)}}, \qquad h_p^{A,k} \equiv \frac{\tau_p^{(k)}}{\tau_v + \tau_A + \tau_p^{(k)}}, \qquad h_v^{A,k} \equiv \frac{\tau_v}{\tau_v + \tau_A + \tau_p^{(k)}}.$$

The first-order condition gives

$$x^{A,k}(p_t; \eta_{j,t}) = \frac{\mathbb{E}[v_t \mid \eta_{j,t}, s_{p,t}^{(k)}] - p_t}{\gamma_A} = \mu^{A,k} + \beta_\eta^{A,k}\eta_{j,t} - \beta_p^{A,k}p_t,$$

with signal coefficient

$$\beta_\eta^{A,k} = \frac{h_\eta^{A,k}}{\gamma_A} = \frac{\tau_A}{\gamma_A(\tau_v + \tau_A + \tau_p^{(k)})}.$$

The AI signal coefficient and price precision are jointly determined. Write the human-sector signal loading as $\zeta_M^{(k)} \equiv \psi_c\beta_\eta^{M,c} + \psi_s\beta_\eta^{M,k}$. Because AI investors also trade on private information, their signal coefficient contributes to price informativeness: $\zeta^{(k)} = \zeta_M^{(k)} + \phi\beta_\eta^{A,k}$. Price informativeness is obtained by solving the fixed-point problem:

$$\tau_p^{(k)} = \tau_z\left(\zeta_M^{(k)} + \frac{\phi\tau_A}{\gamma_A(\tau_v + \tau_A + \tau_p^{(k)})}\right)^2. \tag{25}$$

When $\zeta_M^{(k)} > 0$ or $\phi\tau_A > 0$, the left-hand side of (25) is strictly increasing in $\tau_p^{(k)}$, while the right-hand side is positive and weakly decreasing. The solution is therefore unique, continuous, and strictly increasing in the human-sector loading $\zeta_M^{(k)}$. When $\tau_A = 0$, the equation reduces to $\beta_\eta^{A,k} = 0$ and $\tau_p^{(k)} = \tau_z(\zeta_M^{(k)})^2$, recovering the benchmark with uninformed AI investors who learn only from prices. Solving the price coefficient and intercept yields

$$\beta_p^{A,k} = \frac{1 - h_p^{A,k}\frac{\psi_c\beta_p^{M,c} + \psi_s\beta_p^{M,k}}{\zeta^{(k)}}}{\gamma_A + h_p^{A,k}\frac{\phi}{\zeta^{(k)}}}, \tag{26}$$

$$\mu^{A,k} = \frac{h_v^{A,k}\bar{v} - h_p^{A,k}\frac{\psi_c\mu^{M,c} + \psi_s\mu^{M,k} - S}{\zeta^{(k)}}}{\gamma_A + h_p^{A,k}\frac{\phi}{\zeta^{(k)}}}. \tag{27}$$

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 informed 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.

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. AI investors do not directly observe the private signals of human 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 informed rational-expectations investor would choose when extracting information from prices, as in the benchmark of Grossman and Stiglitz (1980). 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 average AI investors follow the demand rule learned in the perceived level-$(k-1)$ environment:

$$x^{A,k-1}(p;\eta_j) = \mu^{A,k-1} + \beta_\eta^{A,k-1}\eta_j - \beta_p^{A,k-1}p.$$

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.

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

$$\mathbb{E}[v_t \mid \eta_{i,t}, p_t] = h_\eta^{M,k}\, \eta_{i,t} + h_p^{M,k}\, s_{p,t}^{(k-1)} + h_v^{M,k}\, \bar{v}, \tag{28}$$

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 $x^{M,k}(p_t) \equiv (\mathbb{E}[v \mid \eta_{i,t}, p_t] - p_t)/\gamma_M$. Therefore, according to

the expression of $s_p^{(k-1)}$ in terms of p, it holds that

$$\beta_\eta^{M,k} = \frac{h_\eta^{M,k}}{\gamma_M} = \frac{\tau_M}{\gamma_M(\tau_v + \tau_M + \tau_p^{(k-1)})}, \tag{29}$$

$$\beta_p^{M,k} = \frac{1 - h_p^{M,k}\frac{\xi^{(k-1)}}{\zeta^{(k-1)}}}{\gamma_M}, \tag{30}$$

$$\mu^{M,k} = \frac{h_v^{M,k}\bar{v} - h_p^{M,k}\frac{\mu^{(k-1)}}{\zeta^{(k-1)}}}{\gamma_M}. \tag{31}$$

3 Algorithmic Herding

We now show that, in any fixed level-k market environment, the population average of AI investors who independently learn through reinforcement learning endogenously converge to a common demand rule. As the auxiliary learning environments are refined, this aggregate demand rule converges to the optimal trading rule of informed rational-expectations investors in the atomistic rational-expectations equilibrium (REE) of that environment. Throughout this section, we hold fixed the distribution of human investors with heterogeneous sophistication.

We say that the limiting aggregate AI strategy are consistent with rational expectations if two conditions hold. First, each AI investor's policy-implied mean demand rule is optimal given the investor's private signal and the fundamental information it can infer from prices. Second, the price-formation rule embedded in the aggregate learned policy, 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-N, discretized economies that are tailored to the tabular Q-learning problem. In each approximating economy, a finite number of equal-weight AI investors choose demand coefficients from a finite grid $\mathcal{X}_\Delta$, and the fundamental, noise-trader, and AI signal-error shocks are truncated at finite bounds. The market-clearing equation and the demand schedules of human investors with heterogeneous sophistication are otherwise kept fixed at their values in the original model.[9] As the grid becomes finer, the truncation bounds grow, and the finite AI population becomes large, these finite approximating economies recover the original trading environment.

For each fixed approximating economy, the limiting object of AI learning is a mean-field

[9]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 values using realized trading profits.

fixed point. This object imposes two consistency requirements. First, taking the population-average AI demand rule as given, each 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 many atomistic AI investors.

This formulation follows the standard mean-field logic: each individual AI investor has vanishing price impact as the AI population grows, 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 rewards from which AI investors learn. The atomistic 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 and the finite AI population becomes large.

Definition 1 (Discretized bounded environment)**.** *Fix a level-k trading environment and a finite AI population N. A discretized bounded environment is a finite approximating economy for the AI learning problem in that environment. Each AI investor chooses a complete signal- and price-contingent demand schedule from an equally spaced Cartesian grid*

$$\mathcal{X}_\Delta \subset [-B_\eta, B_\eta] \times [-B_p, B_p] \times [-B_\mu, B_\mu]$$

of coefficient triples $y = (\beta_\eta^A, \beta_p^A, \mu^A)$. The fundamental, noise-trader, and AI signal-error shocks are truncated at finite bounds, and each AI investor represents the same share of the AI sector. The demand schedules of cursed and level-k strategic investors and the market-clearing equation otherwise remain the same as in the original level-k environment. The auxiliary price rule projects the realized aggregate coefficient triple onto a compact set on which aggregate price sensitivity is bounded away from zero. Appendix A gives the grid construction, truncations, finite-population signal-error aggregation, and projected market-clearing rule.

If $y(a) = (\beta_\eta^A(a), \beta_p^A(a), \mu^A(a))$ denotes grid arm a and investor j uses the Boltzmann policy $\sigma_j(Q_t^j, a)$, the policy-implied population-average coefficient triple is

$$\left(\bar{\beta}^A_{\eta,\Delta,B,N}(\mathbf{Q}_t), \bar{\beta}^A_{p,\Delta,B,N}(\mathbf{Q}_t), \bar{\mu}^A_{\Delta,B,N}(\mathbf{Q}_t)\right) := \frac{1}{N}\sum_{j=1}^{N}\sum_{a=1}^{M}\sigma_j(Q_t^j, a)y(a). \tag{32}$$

This conditional mean differs from the realized average of the arms selected at date t. Both the realized arm average and the cross-sectional average of AI signal errors continue to fluctuate under positive exploration even when (32) converges.

Definition 2 (Mean-field fixed point in a discretized bounded environment)**.** *In a fixed discretized bounded environment, a finite-population mean-field fixed point is a fixed point* $\mathbf{Q}^*$ *of the discounted Bellman equation. For each investor and arm, the Bellman value equals conditional expected trading profit plus the discounted maximal continuation value, integrating over the other investors' arm draws and all current shocks. The fixed point induces the policy-implied population-average coefficient triple*

$$\left(\bar{\beta}^{A,*}_{\eta,\Delta,B,N}, \bar{\beta}^{A,*}_{p,\Delta,B,N}, \bar{\mu}^{A,*}_{\Delta,B,N}\right) := \left(\bar{\beta}^{A}_{\eta,\Delta,B,N}(\mathbf{Q}^*), \bar{\beta}^{A}_{p,\Delta,B,N}(\mathbf{Q}^*), \bar{\mu}^{A}_{\Delta,B,N}(\mathbf{Q}^*)\right).$$

Thus, the policy profile used to evaluate conditional payoffs is reproduced by the Bellman values calculated from those payoffs.

The mean-field fixed point imposes two economic consistency requirements. First, taking as given the price process generated by the market-clearing equation under the aggregate demand, each AI investor evaluates every demand schedule using its expected trading profit at those prices and its discounted maximal continuation value. The resulting Q-values therefore correctly rank the schedules under the price process generated by the population policy; the Boltzmann rule assigns greater probability to schedules with higher values while preserving exploration. Second, the Q-values reproduce the same policy-implied population-average coefficients—and hence the same aggregate AI demand—that generates the distribution of market-clearing prices. Thus, the distribution of AI demand schedules that determines prices and learning rewards is reproduced by the values learned from that same distribution.

Boltzmann policies remain exploratory at a mean-field fixed point. Changing discount factors shifts each investor's fixed-point value vector by an arm-independent amount but leaves the policy profile and its population-average coefficient triple unchanged.

3.1 Theoretical Results

The next three results establish the main claims on convergence and algorithmic herding for heterogeneous Q-learning investors. Proposition 1 studies learning in a fixed discretized bounded environment and gives a global convergence result under a contraction condition. Proposition 2 shows that mean-field fixed points approach the informed atomistic rational-expectations demand rule as the auxiliary environment is refined and the AI population

grows. Proposition 3 establishes uniqueness and local stability near this limit. The theorem then combines the approximation and local learning results.

Proposition 1 (Convergence of fixed-environment learning)**.** *Fix a finite population size N and a discretized bounded environment populated by informed AI investors. Suppose the algorithm types and learning steps satisfy the boundedness and stochastic-approximation conditions in Appendix A. If the fixed-environment contraction condition* (A30) *holds, then the discounted max-Q recursion has a unique mean-field fixed point $\mathbf{Q}^*$ and*

$$\mathbf{Q}_t \longrightarrow \mathbf{Q}^* \quad \textit{almost surely.}$$

Consequently, the policy-implied population-average coefficient triple in (32) *converges almost surely to the triple induced by $\mathbf{Q}^*$. The conclusion holds from every initial Q-value profile in the invariant box of the fixed environment. In particular,*

$$\left(\bar{\beta}^A_{\eta,\Delta,B,N}(\mathbf{Q}_t), \bar{\beta}^A_{p,\Delta,B,N}(\mathbf{Q}_t), \bar{\mu}^A_{\Delta,B,N}(\mathbf{Q}_t)\right) \longrightarrow \left(\bar{\beta}^{A,*}_{\eta,\Delta,B,N}, \bar{\beta}^{A,*}_{p,\Delta,B,N}, \bar{\mu}^{A,*}_{\Delta,B,N}\right) \qquad \textit{almost surely.} \tag{33}$$

The proof is in Appendix A. This proposition is pointwise in a fixed discretized bounded environment and a fixed finite AI population. Heterogeneity in exploration temperatures affects the limiting mixed policies over grid actions, while heterogeneity in learning rates affects the speed of adjustment. Under the contraction condition, neither source of heterogeneity prevents convergence of the policy-implied population-average coefficients.

The economic force behind the contraction is market-clearing price feedback. When AI investors collectively tilt their demand toward a different set of schedules, the market-clearing price changes, altering both trading profits and the information conveyed by price. The contraction condition ensures that these responses do not amplify differences in learned values. Heterogeneous AI investors then herd onto a single mean-field fixed point rather than drifting toward persistently different aggregate trading rules.

The bounded action space, shock truncations, and denominator-safe projection ensure that exploration produces finite prices, positions, and rewards at every history. Because each AI investor has weight $1/N$, changing one investor's demand schedule has a bounded effect on market-clearing prices and on the profits of other investors. Boltzmann choice also bounds how strongly a change in Q-values shifts action probabilities, while discounting bounds the response of continuation values. The contraction condition requires the combined feedback through current trading profits and continuation values to be weaker than one. Full-support exploration and the stochastic-approximation steps then average out realized shocks and pull every initial Q-value profile in the invariant set toward the same fixed point.

The informed atomistic REE coefficient triple is $(\beta_\eta^{A,*}, \beta_p^{A,*}, \mu^{A,*})$. Appendix A derives this triple as the unique fixed point of the informed AI best-response map.

Proposition 2 (Mean-field fixed points converge to informed rational-expectations demand)**.** *Suppose the primitive conditions in Assumption 3 hold, and use the denominator-safe projection set constructed in Appendix A. For every $\varepsilon > 0$, choose a sufficiently fine action grid and sufficiently wide action and shock bounds, followed by a sufficiently large population. Then every mean-field fixed point $\mathbf{Q}^*$ satisfies*

$$\left\| \left(\bar{\beta}_{\eta,\Delta,B,N}^{A,*}, \bar{\beta}_{p,\Delta,B,N}^{A,*}, \bar{\mu}_{\Delta,B,N}^{A,*} \right) - \left(\beta_\eta^{A,*}, \beta_p^{A,*}, \mu^{A,*} \right) \right\|_\infty < \varepsilon. \tag{34}$$

The bound is uniform across the bounded heterogeneous algorithm types allowed in the appendix, whose raw exploration temperatures lie in a compact positive interval.

In the atomistic environment, expected trading profit is a strictly concave quadratic function of the three demand coefficients. AI investors therefore have a unique price-taking demand rule given the market-clearing price function, and the informed REE is the fixed point of this optimal-demand map. Exploration changes the dispersion of individual choices across nearby schedules, but not the mean of the continuous Boltzmann rule. The informed best-response map has a triangular structure: the signal loading first determines how much fundamental information enters prices; conditional on that price informativeness, the price sensitivity and intercept solve separate linear equations. This structure identifies a unique informed rational-expectations demand rule even when the slopes of the individual feedback channels are not themselves contractions.

The discretized bounded environments approximate this continuous problem in two steps. Fine grids and wide shock bounds make the discrete Boltzmann mean map and its derivative approach the continuous best-response map. With the auxiliary environment fixed, finite-population demand-schedule sampling errors and the cross-sectional average of idiosyncratic AI signal errors vanish at rate $N^{-1/2}$, while each investor's own price impact vanishes at rate N^{-1}. The finite-population mean-field fixed points therefore inherit the informed atomistic demand rule as the population grows. Economically, strict concavity gives AI investors a well-defined optimal demand schedule against the market-clearing price function, while discretization, truncation, and finite-population errors disappear as the approximating environments are refined.

Proposition 3 (Uniqueness and local stability)**.** *Fix an accurate discretized bounded environment and consider an admissible sequence of bounded heterogeneous populations, as defined in Appendix A. Under the regularity conditions stated in the appendix, all sufficiently*

large populations have a unique mean-field fixed point. This fixed point is locally exponentially stable under the mean discounted max-Q dynamics.

The appendix proves that, in a neighborhood of the reference fixed point, the finite-population mean learning dynamics and their local responses approach the deterministic atomistic reference dynamics. The regularity conditions make demand-schedule sampling errors, the cross-sectional average of idiosyncratic AI signal errors, rare profiles on which the price projection is active, and individual price impact small. For populations with positive discounting, local stability additionally requires the finite-population approximation error to be asymptotically negligible relative to the payoff gap between the best and second-best continuation arms, as stated in (A69). This condition keeps the same continuation arm optimal near the fixed point, so the Bellman maximum is locally smooth. The gap is not needed for fixed-point uniqueness and is vacuous in undiscounted populations.

Local stability comes from price feedback near the informed REE. The signal loading first determines price informativeness, after which price sensitivity and the intercept adjust through the triangular best-response system. Strict concavity disciplines each investor's response, while the induced price feedback makes local aggregate deviations mean-reverting. Because the aggregate response has a triangular structure, this feedback can remain locally stable even when neither individual feedback channel is itself a contraction. Boltzmann averaging transmits this stability from individual values to the policy-implied population-average demand rule.

Theorem 4 (Algorithmic herding toward atomistic rational-expectations demand)**.** *Fix a level-k trading environment, and let* $\left(\beta_\eta^{A,*}, \beta_p^{A,*}, \mu^{A,*}\right)$ *denote the informed atomistic rational-expectations coefficient triple.*

Global convergence. *Maintain the assumptions of Proposition 1. In any discretized bounded environment satisfying its contraction condition, learning converges almost surely from every initial Q-value profile in the invariant box to the unique mean-field fixed point. Whenever the environment and population are also sufficiently accurate and large for Proposition 2 to give an error below* ε*, this convergence satisfies*

$$\lim_{t\to\infty} \left\| \left(\bar{\beta}^A_{\eta,\Delta,B,N}(\mathbf{Q}_t), \bar{\beta}^A_{p,\Delta,B,N}(\mathbf{Q}_t), \bar{\mu}^A_{\Delta,B,N}(\mathbf{Q}_t)\right) - \left(\beta_\eta^{A,*}, \beta_p^{A,*}, \mu^{A,*}\right) \right\|_\infty < \varepsilon, \qquad \text{almost surely.} \tag{35}$$

Local lock-in. *More generally, maintain the assumptions of Propositions 2 and 3 and the learning-step conditions in Appendix A. For every* $\varepsilon > 0$*, a sufficiently accurate discretized bounded environment and sufficiently large population have a unique locally stable mean-field fixed point within* ε *of* $\left(\beta_\eta^{A,*}, \beta_p^{A,*}, \mu^{A,*}\right)$*. For any probability tolerance* $\upsilon > 0$*, once learning enters an inner neighborhood at a sufficiently late deterministic date b, conditional on the*

observed history $\mathcal{F}_b$ and on that entry, it stays in the stable neighborhood and converges to this fixed point with probability at least $1 - \upsilon$. The same conclusion holds after an almost-surely finite stopping time at which sufficiently little squared learning gain remains. On the lock-in event, the limiting coefficient triple lies within ε of $\left(\beta_\eta^{A,*}, \beta_p^{A,*}, \mu^{A,*}\right)$.

The theorem combines learning within a discretized bounded environment with approximation across environments. The construction is ordered: the action grid and shock bounds are chosen first, the population is then made sufficiently large, and learning time tends to infinity with the auxiliary environment and population fixed. The two claims describe complementary routes to algorithmic herding. Under global contraction, differences in learned values are attenuated throughout the discretized bounded environment. Under local lock-in, the same force operates near informed rational-expectations demand: deviations change prices, expected profits, and price informativeness in a way that pulls the policy-implied population average back toward the fixed point. Once the remaining squared learning steps are small, realized shocks are unlikely to dislodge the learning process from this stable neighborhood.

The result formalizes algorithmic herding: although AI investors may differ in exploration temperatures, learning rates, discount factors, 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 and population size approach the original level-k trading environment, this common learned demand rule converges to the atomistic 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.

When $\tau_A = 0$, the signal coordinate is removed rather than treated as a finite-variance third coordinate. Restricting the action grid to $\beta_\eta^A = 0$ removes the idiosyncratic AI signal-error term from finite-population market clearing and reduces the problem to the two-coefficient economy in which AI investors receive no fundamental signal and learn only from prices. Appendix A gives this dimension reduction and recovers the uninformed-AI REE coefficients.

3.2 Limit of Level-k Strategic Reasoning

We next consider the limit of the level-k reasoning hierarchy as $k \to \infty$, maintaining $\psi_s > 0$. 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 con-

struction 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 5 (Convergence of the level-k recursion)**.** *For every $\sigma_M > 0$, every driven initialization whose finite hierarchy is price admissible converges to a unique fixed point if*

$$\psi_s \left(\frac{8\psi}{27\gamma_M^2} + \frac{\phi}{\gamma_M \gamma_A} \frac{\tau_A}{\tau_v + \tau_A} \right) \frac{\tau_z}{\tau_v} \leq 1. \tag{36}$$

When $\tau_A = 0$, this condition reduces to

$$\psi_s \leq \frac{27\gamma_M^2 \sigma_z^2}{8\psi\sigma_v^2}. \tag{37}$$

Proposition 5 gives a uniform sufficient condition for dynamic stability, conditional on finite-level price admissibility. The condition requires the mass of strategic investors not to be too large relative to the stabilizing effects of transaction costs and noise-trader volatility. The sufficient stability region shrinks when noise trading is less volatile or when the informed AI sector contributes more signal-based demand. In both cases, prices reveal more information, so small changes in how higher-level investors extract that information generate larger changes in subsequent price responses. Greater fundamental uncertainty also tightens the stability condition because changes in price inference then produce larger revisions in perceived payoff information and subsequent demand. The bound isolates this destabilizing signal-loading channel; the exact dynamics also include within-level damping generated by informed AI demand. When the condition holds, the recursion dampens these changes across successive reasoning levels and converges to a unique limiting trading environment.

Proposition 6 (Convergent level-k limit is a linear REE)**.** *Suppose every finite environment in the level-k hierarchy is price admissible and the coefficients converge to a fixed point:*

$$\lim_{k\to\infty} (\beta_\eta^{M,k}, \beta_p^{M,k}, \mu^{M,k}) = (\beta_\eta^{M,\infty}, \beta_p^{M,\infty}, \mu^{M,\infty}),$$
$$\lim_{k\to\infty} (\beta_\eta^{A,k}, \beta_p^{A,k}, \mu^{A,k}) = (\beta_\eta^{A,\infty}, \beta_p^{A,\infty}, \mu^{A,\infty}).$$

Then the limit constitutes a linear rational-expectations equilibrium: level-∞ humans and AI investors are mutual Bayesian best responses given the equilibrium price generated by the limiting aggregate coefficients, with cursed investors taken as given.

The proposition identifies the economic meaning of a convergent reasoning hierarchy. Once the human and learned AI coefficient sequences settle down, both sophisticated groups condition on the same price map that their limiting demands generate. Level-∞ humans use

their private signals and price, while AI investors use their own private signals and price; neither group retains a mistaken belief about price formation. Appendix B proves both propositions.

3.3 Simulation Evidence

Theorem 4 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 focus on the boundary case in which AI investors cannot receive fundamental signals, $\tau_A = 0$, and use a centered environment with $S = \bar{v} = 0$.

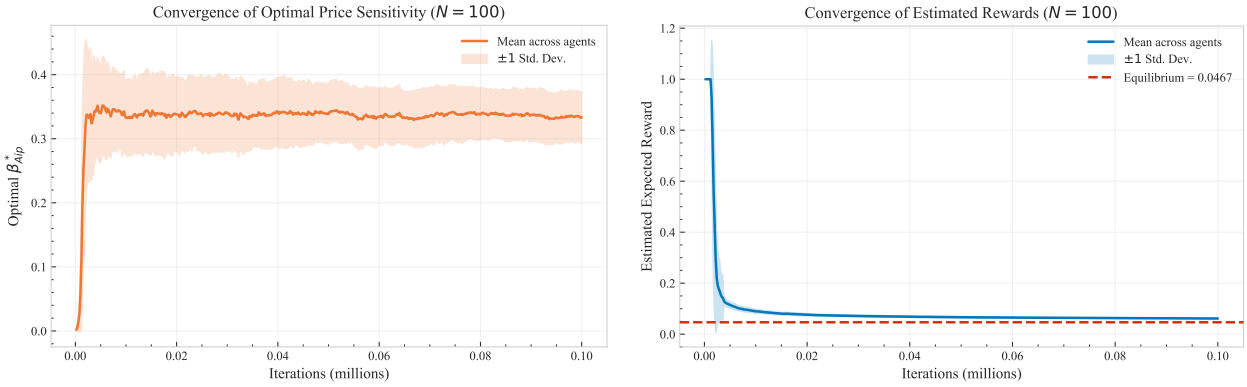
We simulate N atomistic AI investors trading in the level-k market environment. The continuum of AI investors in the model is approximated by a large finite population. Each AI investor learns independently from its own realized trading profits. To match the tabular implementation for AI investors without fundamental signals, we restrict the AI action space to a finite grid of price-sensitivity coefficients $\beta_{p,t}^{A,k}$, set $\beta_{\eta,t}^{A,k} = 0$, and fix 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 human 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 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: $\sigma_v = 2$, $\sigma_z = 0.2$, $\gamma_M = \gamma_A = 0.2$, asset supply $S = 0$, total human-investor

Figure 2: Learning Convergence to the Rational-Expectations Benchmark



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

(b) Estimated expected reward

measure $\psi = 1.5$, and aggregate human-investor coefficients $\bar{\beta}_p^k = 4.0$ and $\bar{\beta}_\eta^k = 4.5$.

Results. For this initialization, learning converges to the rational-expectations demand rule. In Figure 2, the cross-sectional average of learned price sensitivities converges to $\beta_p^{A,k}$ in Panel (a), while the average estimated 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 learned population average approaches the rational-expectations demand. The simulation therefore illustrates the price-feedback mechanism behind local algorithmic herding. Appendix A.12 reports the full cross-sectional distribution of learned coefficients and documents its concentration over time.

4 Market Sophistication and the Limits of AI Performance

This section studies how human sophistication shapes market competition and AI performance. First, we derive per-period expected trading profits for human and AI investors in a fixed market. Second, we characterize the limits on AI performance first with finite level-k humans and then in the convergent level-∞ market. Third, we separate training from deployment and study how training-market sophistication affects AI performance in a common test market. Holding the test environment fixed isolates the sophistication acquired from the market data generated during training.

4.1 Expected Trading Profits

Fix a market after AI strategies have converged. Given the investor composition (ψ_c, ψ_s, ϕ) and the sophistication level k of strategic investors, the equilibrium price function is summarized by aggregate price sensitivity ξ, aggregate signal loading ζ, and intercept μ.[10] Section 2.1 defines $\pi_{i,t}$ and $\pi_{j,t}$ as the per-period net trading profits of human investor i and AI investor j, respectively. Define their per-period expected net profits as

$$\Pi_i^M \equiv \mathbb{E}[\pi_{i,t}], \quad i \in \mathcal{M},$$
$$\Pi_j^A \equiv \mathbb{E}[\pi_{j,t}], \quad j \in \mathcal{A}.$$

The expectations are taken under the objective equilibrium distribution of the fixed market. The formula below applies to any admissible human-investor and AI coefficients evaluated against the same equilibrium price function.

Proposition 7 (Expected profit)**.** *Let $\ell \in \{M, A\}$ index investor type. Consider an investor with demand intercept μ^ℓ, private-signal loading β_η^ℓ, and price loading β_p^ℓ, whose demand is*

$$x_i^\ell = \mu^\ell + \beta_\eta^\ell \eta_n^\ell - \beta_p^\ell p$$

and trading-cost parameter γ_ℓ, evaluated against the equilibrium price function summarized by (ξ, ζ, μ), where σ_ℓ^2 denotes the variance of investor type ℓ's idiosyncratic signal noise. The investor's per-period expected net profit is

$$\begin{aligned}\Pi_i^\ell = \frac{1}{\xi^2}\Big\{&[(\xi-\zeta)\bar{v}-\mu]\left[(\beta_\eta^\ell\xi-\beta_p^\ell\zeta)\bar{v}+\xi\mu^\ell-\beta_p^\ell\mu\right]\\ &+(\xi-\zeta)(\beta_\eta^\ell\xi-\beta_p^\ell\zeta)\sigma_v^2+\beta_p^\ell\sigma_z^2\Big\}\\ -\frac{\gamma_\ell}{2\xi^2}\Big\{&\left[(\beta_\eta^\ell\xi-\beta_p^\ell\zeta)\bar{v}+\xi\mu^\ell-\beta_p^\ell\mu\right]^2\\ &+(\beta_\eta^\ell\xi-\beta_p^\ell\zeta)^2\sigma_v^2+\xi^2(\beta_\eta^\ell)^2\sigma_\ell^2+(\beta_p^\ell)^2\sigma_z^2\Big\}. \end{aligned} \tag{38}$$

The formula applies to cursed and level-k human investors and to AI investors after substituting their respective coefficients. When AI private-signal precision is zero, $\tau_A = 0$, the AI signal coordinate and its variance term are absent, nesting the benchmark in which AI investors learn only from prices.

[10] For ease of notation, we drop superscripts on (ξ, ζ, μ) in this section except where needed to distinguish training and deployment environments. We focus on per-period expected profits in the stationary market after AI strategies have converged. Because the primitive shocks are i.i.d. over time and the post-learning demand coefficients are time-invariant, we also suppress the time subscript t on expected profits.

The proof, including the gross-profit and trading-cost decompositions, is in Appendix C. The first brace is gross expected trading revenue. Its first product equals unconditional average excess payoff times unconditional average demand. The remaining terms measure how the investor's position covaries with payoff-relevant shocks. Signal exposure loads demand directly on the fundamental; price sensitivity responds to both payoff information and noise-trader demand embedded in the equilibrium price.

The second brace is the expected quadratic trading cost. Costs rise with the mean position, net exposure to fundamental variation, idiosyncratic signal noise, and positions induced by noise trading. This decomposition isolates the tradeoff below: a more precise private signal improves payoff forecasts directly, whereas learning from prices requires responding to order flow that also contains nonfundamental noise.

4.2 Information Environment and AI Performance

Human information has two opposing effects on AI performance. It gives humans a direct forecasting advantage, but human trades also supply the fundamental information that AI extracts from equilibrium prices. AI benefits from human information only to the extent that informed order flow reaches prices without being obscured by noise trading. We first study a market with AI, cursed investors, and level-3 humans, who reason strategically through a short belief hierarchy. Even against these moderately sophisticated humans, private-signal precision and noise trading sharply limit AI profitability. Appendix C.3 repeats the comparisons for level-2 and level-4 investors. We then study the convergent level-∞ market, in which rational-expectations demands permit analytical profit comparisons.

Throughout this subsection, impose equal human and AI trading-cost parameters, $\gamma_M = \gamma_A \equiv \gamma$. This benchmark isolates information-channel limitations. Profit comparisons therefore reflect how investors use private signals and learn from prices, rather than differences in the cost of taking a given position.

4.2.1 Markets with Limited Strategic Sophistication

AI versus Cursed Human Investors. Figure 3 compares AI and cursed-investor profits as AI private-signal precision rises.[11] Cursed investors use their private signals correctly but do not use price to update their payoff forecast: a higher price lowers perceived trading surplus without raising the expected payoff. AI interprets price information correctly and, when τ_A is positive, combines the equilibrium price with its own signal. With equal trading

[11] Appendix Figures A2–A9 repeat all four limited-sophistication comparisons for $k = 2$ and $k = 4$.

costs, the comparison pits the AI's total usable information—its private signal plus the price signal—against the cursed investor's private signal.

Panel (a) sets AI private-signal precision to zero, $\tau_A = 0$, so AI relies entirely on the equilibrium price. Start on the right side of the panel, where noise trading volatility, σ_z is large. 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 panel is blue throughout this high-noise region: cursed investors earn more for every displayed value of human signal noise.

Now move to the left side, where noise-trader volatility is low. AI dominance over cursed investors is a window, not a universal result. When human private signals are very precise (σ_M is small), 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.

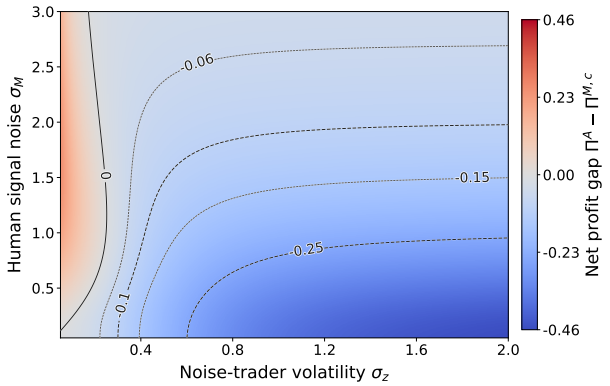
When private signals are very noisy (σ_M becomes sufficiently large), cursed investors also dominate. Their own signal is weak, but that signal is also the input that makes prices informative. Cursed and level-k investors then trade only weakly on payoff information, so prices contain little useful variation 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 human investors only in the middle. Human signals must be precise enough to make the equilibrium price informative, but not so precise that their direct value dominates the information AI recovers from that price. 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.

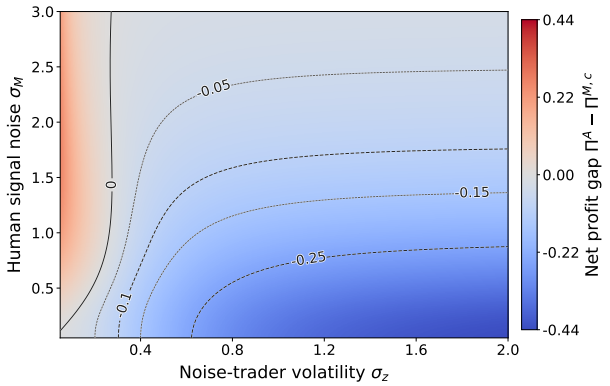
Panel (d) gives AI private-signal noise $\sigma_A = 1$. If human private-signal noise, σ_M, is at least one, the AI signal is at least as precise as the human signal, and AI also extracts information from the equilibrium price. AI therefore earns more. When human private-signal noise is below one, humans have the more precise private signal. The substantial blue region shows that this advantage can outweigh AI's additional price information: even cursed investors can earn more than an informed AI when they retain sufficiently better private information.

The axes illustrate two distinct forces. Moving upward weakens the cursed investor's private signal but also reduces the human information entering the equilibrium price. In panel (d), the direct effect dominates and the AI–cursed profit gap rises. Moving right raises

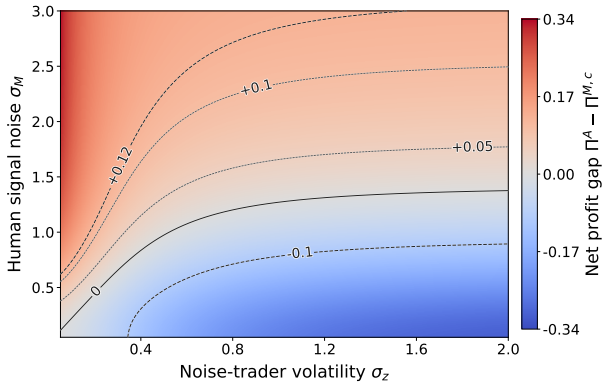
Figure 3: AI and Cursed-Investor Profit Gaps with Limited Strategic Sophistication



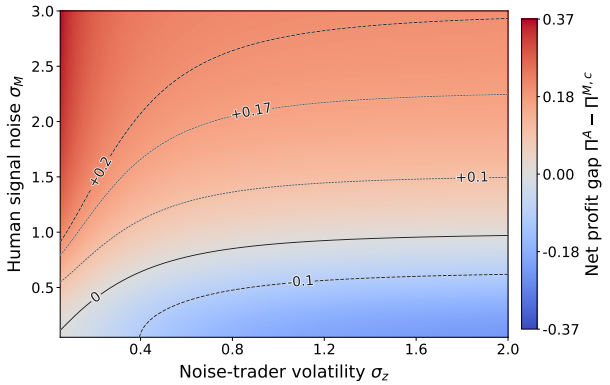
(a) $\tau_A = 0$ ($\sigma_A = \infty$)



(b) $\tau_A = 0.04$ ($\sigma_A = 5$)



(c) $\tau_A = 0.5$ ($\sigma_A \simeq 1.41$)



(d) $\tau_A = 1$ ($\sigma_A = 1$)

Notes: Each panel plots the AI net expected profit minus the cursed investor net expected profit, $\Pi^A - \Pi^{M,c}$, in a market with level-3 strategic investors. The horizontal axis is noise-trader volatility, σ_z, and the vertical axis is human private-signal noise, σ_M. Red regions favor AI, blue regions favor cursed investors, and the solid black contour marks equal profits. Dashed contours report labeled nonzero profit gaps. Colorbar limits are panel-specific, so color intensity is not comparable across panels. Parameters are $\bar{v} = 0$, $S = 0$, $\sigma_v = 1$, $\gamma_M = \gamma_A = 1$, $(\psi_c, \psi_s, \phi) = (0.30, 0.30, 0.40)$, $\sigma_z \in [0.05, 2]$, and $\sigma_M \in [0.05, 3]$. Panels (a)–(d) set AI private-signal precision, $\tau_A \in \{0, 0.04, 0.5, 1\}$, corresponding to AI private-signal noise, $\sigma_A \in \{\infty, 5, \sqrt{2}, 1\}$.

noise-trader volatility and generally lowers the gap in the displayed panels. Noisier order flow removes an information source that AI uses correctly but cursed investors ignore, so the blue region expands.

Panels (b) and (c) trace the transition as AI private-signal precision rises. A more precise AI signal improves its payoff forecast directly and contributes fundamental information to aggregate order flow. The resulting equilibrium price precision, $\tau_p = \zeta^2 \tau_z$, also depends on human signal quality and noise-trader volatility, where τ_z is the precision of noise-trader order flow. In the displayed panels, better AI information expands the region in which AI earns more, but it does not make relative performance independent of the market's information environment.

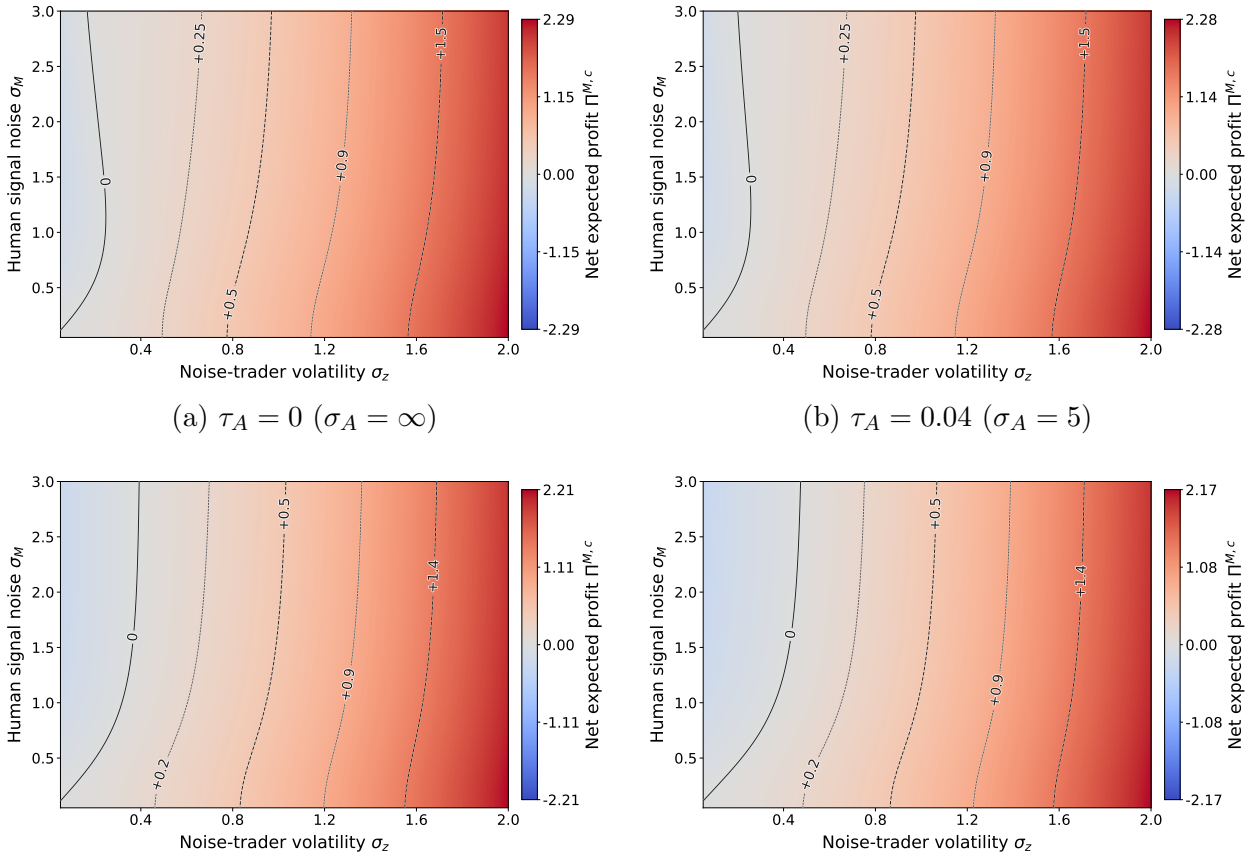
Figure 4 reports cursed investors' profit levels. Their net expected profits remain positive over most of every panel. Under the displayed calibration, more precise AI information widens the negative-profit wedge at very low noise-trader volatility, where informational competition is strongest. Yet broad positive-profit regions remain even in panel (d). Noise-trader demand creates price-payoff wedges, and cursed investors still use their private signals to forecast payoffs. Thus, even a substantially better-informed AI does not eliminate positive expected profits for the least sophisticated humans in the market.

AI versus Level-k Investors. Level-3 investors provide a more sophisticated human benchmark. They combine private signals with strategic inference from prices, although their belief hierarchy remains finite and thus their learning from price is still biased. Both AI and level-3 humans observe the same realized price, but they interpret it differently. Level-3 investors infer from their perceived level-2 equilibrium price function, whereas AI correctly learns its demand in the actual equilibrium populated by level-3 humans.

Panel (a) of Figure 5 again sets AI private-signal precision to zero, $\tau_A = 0$. Begin on the right side, where noise-trader volatility, σ_z, is large. Noisy order flow removes AI's only source of fundamental information. Level-3 humans retain a private signal and also infer from the equilibrium price, so they earn more throughout this high-noise region.

Now consider low noise-trader volatility. When human private-signal noise, σ_M, is small, level-3 investors have precise direct information that AI lacks. The same force still operates: when human private-signal noise is sufficiently large, informed human order flow conveys little about fundamentals, leaving little price information for AI to recover. AI can earn more only between these extremes, where human trades make prices revealing but the human private signal is not too precise. Even this advantage erodes as humans become more sophisticated. A human who interprets the same price more correctly and also observes a private signal must earn more than price-only AI. The red region is therefore tiny: level-3 humans outperform

Figure 4: Cursed-Investor Profits with Limited Strategic Sophistication



(a) $\tau_A = 0$ ($\sigma_A = \infty$)

(b) $\tau_A = 0.04$ ($\sigma_A = 5$)

(c) $\tau_A = 0.5$ ($\sigma_A \simeq 1.41$)

(d) $\tau_A = 1$ ($\sigma_A = 1$)

Notes: Each panel plots cursed-investor net expected profit $\Pi^{M,c}$ in a market with level-3 strategic investors. Red regions indicate positive profit, blue regions indicate negative profit, and the solid black contour marks zero profit. Dashed contours report labeled nonzero profit levels. Colorbar limits are panel-specific, so color intensity is not comparable across panels. The horizontal axis is noise-trader volatility, σ_z, and the vertical axis is human private-signal noise, σ_M. Panels (a)–(d) set AI private-signal precision, $\tau_A \in \{0, 0.04, 0.5, 1\}$, corresponding to AI private-signal noise, $\sigma_A \in \{\infty, 5, \sqrt{2}, 1\}$. Parameters are $\bar{v} = 0$, $S = 0$, $\sigma_v = 1$, $\gamma_M = \gamma_A = 1$, $(\psi_c, \psi_s, \phi) = (0.30, 0.30, 0.40)$, $\sigma_z \in [0.05, 2]$, and $\sigma_M \in [0.05, 3]$.

price-only AI across almost the entire displayed parameter space.

Panel (d) sets AI private-signal precision to $\tau_A = 1$, corresponding to AI private-signal noise $\sigma_A = 1$. The comparison turns approximately at human private-signal noise $\sigma_M = 1$, where the human and AI private signals are equally precise.[12] Below that value, level-3 investors have the more precise private signal and earn more than AI over nearly the entire region. The large blue area shows the key limit of AI performance: an informed AI that also uses price signal correctly can still earn less than humans who perform only three rounds of strategic reasoning. Above that value, AI has the more precise private signal and earns more over most of the panel.

The broad pattern follows the relative precision of private information. In the displayed calibration, when noise-trader volatility rises, the equilibrium price becomes a less useful common signal. This raises the importance of whichever private signal is more precise. Finite-depth inference also changes with the information environment. Panels (b) and (c) show the transition as AI information improves. A better AI signal raises its relative profit in the displayed panels, but human signal quality and noise-trader volatility continue to determine whether that informational improvement is enough.

Figure 6 shows an even stronger persistence result. Level-3 investors earn positive net expected profits over nearly the entire displayed domain in every panel; losses appear only in a thin region with extremely low noise-trader volatility. More precise AI information modestly widens this region because informational competition is strongest when prices contain little noise. Still, level-3 investors combine their private signals with price inference, while noise-trader demand leaves price-payoff wedges for informed investors to absorb. The limit on AI profitability thus appears in both comparisons: AI often earns less than finite-level humans, and it does not eliminate their positive expected profits.

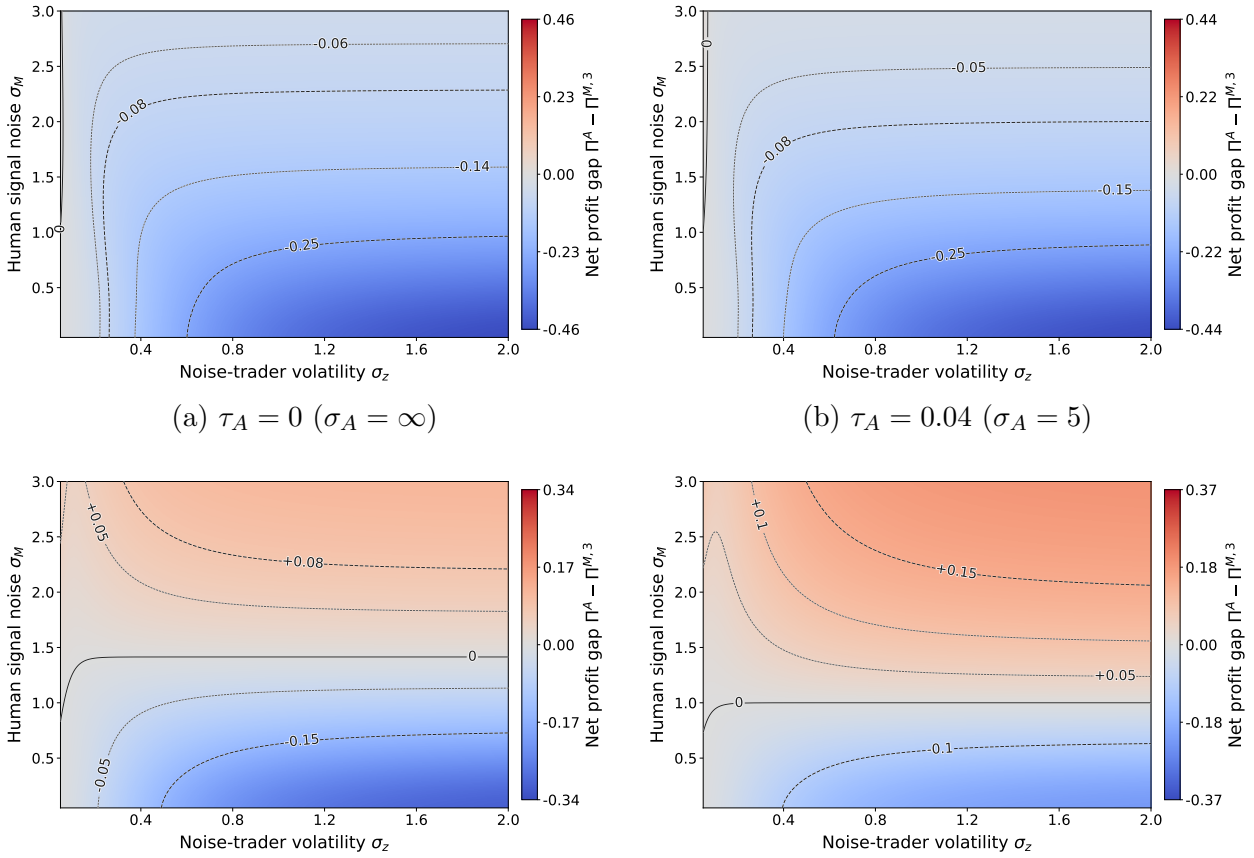
4.2.2 Markets with Infinite Strategic Reasoning

We next consider a convergent market with AI, cursed, and level-∞ investors. Proposition 6 shows that level-∞ humans and AI are mutual Bayesian best responses to the same equilibrium price. Convergence removes finite-depth errors in price inference and enables analytical profit comparisons. We first record the general profit comparison propositions and then focus on the price-only AI benchmark.

Proposition 8 (AI–human profit comparisons)**.** *Consider an informative, price-admissible limiting environment in which AI is a Bayesian best response to* (η_j, p) *and the level-*∞

[12] Although this is not an exact boundary at level-3, it becomes exact in the limit market with level-∞ human investors in the next section.

Figure 5: AI and Level-3 Profit Gaps



(a) $\tau_A = 0$ ($\sigma_A = \infty$)

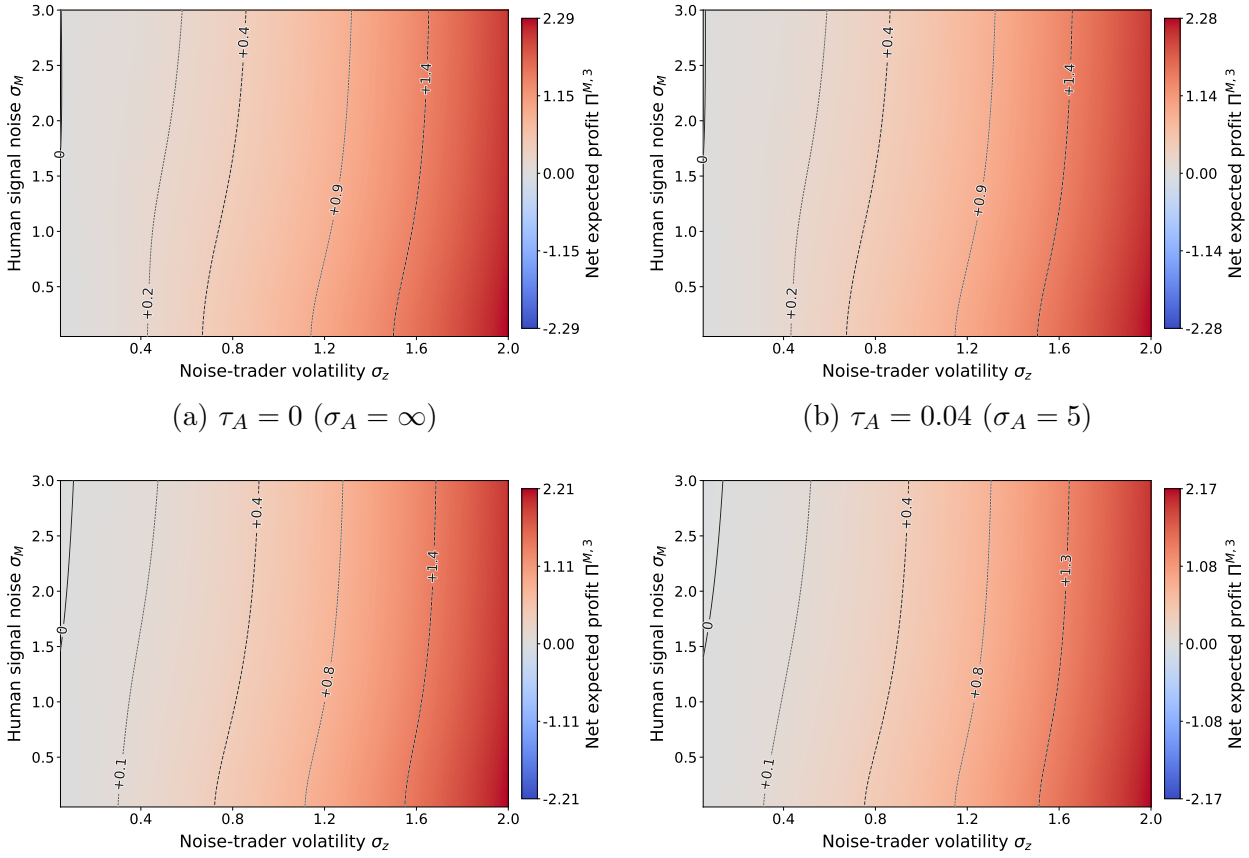
(b) $\tau_A = 0.04$ ($\sigma_A = 5$)

(c) $\tau_A = 0.5$ ($\sigma_A \simeq 1.41$)

(d) $\tau_A = 1$ ($\sigma_A = 1$)

Notes: Each panel plots $\Pi^A - \Pi^{M,3}$, the AI net expected profit minus the level-3 investor net expected profit. Red regions favor AI, blue regions favor level-3 investors, and the solid black contour marks equal profits. Dashed contours report labeled nonzero profit gaps. Colorbar limits are panel-specific, so color intensity is not comparable across panels. The horizontal axis is noise-trader volatility, σ_z, and the vertical axis is human private-signal noise, σ_M. Panels (a)–(d) set AI private-signal precision, $\tau_A \in \{0, 0.04, 0.5, 1\}$, corresponding to AI private-signal noise, $\sigma_A \in \{\infty, 5, \sqrt{2}, 1\}$. Parameters are $\bar{v} = 0$, $S = 0$, $\sigma_v = 1$, $\gamma_M = \gamma_A = 1$, $(\psi_c, \psi_s, \phi) = (0.30, 0.30, 0.40)$, $\sigma_z \in [0.05, 2]$, and $\sigma_M \in [0.05, 3]$.

Figure 6: Level-3 Investor Profits



(a) $\tau_A = 0$ ($\sigma_A = \infty$)

(b) $\tau_A = 0.04$ ($\sigma_A = 5$)

(c) $\tau_A = 0.5$ ($\sigma_A \simeq 1.41$)

(d) $\tau_A = 1$ ($\sigma_A = 1$)

Notes: Each panel plots level-3 investor net expected profit $\Pi^{M,3}$. Red regions indicate positive profit, blue regions indicate negative profit, and the solid black contour marks zero profit. Dashed contours report labeled nonzero profit levels. Colorbar limits are panel-specific, so color intensity is not comparable across panels. The horizontal axis is noise-trader volatility, σ_z, and the vertical axis is human private-signal noise, σ_M. Panels (a)–(d) set AI private-signal precision, $\tau_A \in \{0, 0.04, 0.5, 1\}$, corresponding to AI private-signal noise, $\sigma_A \in \{\infty, 5, \sqrt{2}, 1\}$. Parameters are $\bar{v} = 0$, $S = 0$, $\sigma_v = 1$, $\gamma_M = \gamma_A = 1$, $(\psi_c, \psi_s, \phi) = (0.30, 0.30, 0.40)$, $\sigma_z \in [0.05, 2]$, and $\sigma_M \in [0.05, 3]$.

investor is a Bayesian best response to (η_i, p)*. Then*

$$\Pi^A - \Pi^{M,\infty} = \frac{\tau_A - \tau_M}{2\gamma(\tau_v + \tau_A + \tau_p)(\tau_v + \tau_M + \tau_p)}, \tag{39}$$

$$\Pi^A - \Pi^{M,c} = \frac{\tau_A + \tau_p - \tau_M}{2\gamma(\tau_v + \tau_M)(\tau_v + \tau_A + \tau_p)}. \tag{40}$$

The proof is in Appendix C.2. The optimal-demand identity separates a common trading-opportunity term from posterior uncertainty. AI and level-∞ humans both interpret the equilibrium price correctly and face the same trading cost. They observe different private signals. Conditional on the price, those signals lower posterior uncertainty about v, and the lower conditional variance raises the expected square of the optimal trading return. Their comparison therefore depends only on AI private-signal precision, τ_A, relative to human private-signal precision, τ_M. AI earns more if and only if its private signal is more precise. Greater equilibrium price precision, τ_p, supplies more common information and compresses the absolute profit gap toward zero, but it cannot reverse the ranking. The profit gap is exactly the value of that extra signal after conditioning on the price signal.

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. AI investors have an additional price-signal channel. Because both investor types trade the same asset at the same price and face the same trading cost, the comparison does not come from unconditional mean mispricing or the price paid per share. It depends on whose total signals improve payoff forecasts more. AI therefore earns more than they do if and only if total AI signal precision—AI private-signal precision plus equilibrium price precision, $\tau_A + \tau_p$—exceeds human private-signal precision, τ_M. Correct price inference can compensate for an inferior AI private signal, but only when the equilibrium price is sufficiently informative.

AI without Private Signals. To connect this limit to panel (a) above, set AI private-signal precision to zero, $\tau_A = 0$. AI then learns only from the equilibrium price. The first comparison in Proposition 8 becomes a uniform human advantage.

Proposition 9 (Level-∞ investors always outperform AI without private signals)**.** *In any nondegenerate convergent environment with cursed investors, level-*∞ *investors, and AI investors without private signals,*

$$\Pi^A - \Pi^{M,\infty} = \frac{-\tau_M}{2\gamma(\tau_v + \tau_p)(\tau_v + \tau_M + \tau_p)} < 0 \tag{41}$$

for all $\tau_M > 0$.

Level-∞ investors observe the same equilibrium price as AI and add a private signal. Conditional on price, that signal lowers posterior uncertainty; the gap in (41) is its exact value. Greater equilibrium price precision makes this additional signal less valuable and shrinks the profit gap, but the gap remains strictly positive at every finite admissible price precision. Unlike at finite k, no mistaken human price inference remains for AI to exploit.

The cursed comparison instead pits direct private information against information extracted from prices.

Lemma 10 (Cursed investors versus AI without private signals)**.** *In any informative, nondegenerate environment in which AI is a Bayesian best response to the equilibrium price,*

$$\Pi^A - \Pi^{M,c} = \frac{h_p - h}{2\gamma\tau_v}, \qquad h \equiv \frac{\tau_M}{\tau_v + \tau_M}, \qquad h_p \equiv \frac{\tau_p}{\tau_v + \tau_p}. \tag{42}$$

Therefore,

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

The private-information weight, h, measures the value of the cursed investor's signal channel. It approaches zero when the human signal is mostly noise and approaches one when that signal is highly precise. The price-information weight, h_p, measures the value of AI's price-signal channel. It is high when informed order flow makes prices precise and low when price movements mostly reflect noise-trader demand. Equal trading costs remove the common trading-opportunity term, so the profit ranking depends on which usable signal reduces payoff uncertainty more. Noise trading directly limits AI by reducing equilibrium price precision, τ_p.

Proposition 11 (Large noise trading: cursed investors dominate)**.** *Suppose AI private-signal precision is zero,* $\tau_A = 0$*, and, for every positive human private-signal noise,* $\sigma_M > 0$*, the level-k hierarchy converges to an informative, nondegenerate, price-admissible limiting environment. If*

$$\sigma_z > \frac{\psi\sigma_v}{2\gamma}, \qquad \psi \equiv \psi_c + \psi_s,$$

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

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

The condition in Proposition 11 is sufficient. When noise-trader volatility, σ_z, exceeds this bound, equilibrium prices reflect too much nonfundamental demand for AI's price signal to match the human private signal. If human signals are weak, little fundamental information

enters prices. If they are strong, volatile noise-trader demand still keeps the price signal inferior. Cursed investors therefore earn more for every human private-signal precision.

Proposition 12 (Small noise trading: AI-dominance window)**.** *Suppose AI private-signal precision is zero, $\tau_A = 0$, and that, for every positive human private-signal precision, $\tau_M > 0$, the level-k hierarchy converges to an informative, nondegenerate, price-admissible limiting environment. Suppose also that the fixed-point price precision $\tau_p^*(\tau_M)$ is well defined 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

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

When noise-trader volatility is below the bound in Proposition 12, AI earns more only within an intermediate range of human private-signal noise. Precise human signals favor cursed investors directly. Very noisy human signals generate too little informed order flow for AI to learn from equilibrium prices. AI earns more only in the middle, where human information enters prices but does not dominate as a direct private signal.

Appendix Figure A10 displays these comparisons. The AI–cursed panel retains the low-noise, intermediate-signal window from the level-3 market. The AI–level-∞ panel is blue throughout: infinite-depth of strategic reasoning eliminates the biases in learning from prices, and thus eliminates the tiny region in which price-only AI earns more than level-3 humans.

4.3 Training-Environment Sophistication and AI Performance

We now show that another important limit of AI performance is their training environment. We analyze the effect by separating AI's training from deployment. AI sophistication inherits the sophistication of the market in which the AI is trained. For each k, AI investors train to convergence in a market with cursed investors, and level-k strategic investors. We then deploy the pre-trained AI algorithms in a common test market that holds all primitives and population shares fixed, retains the cursed investors, and replaces the level-k humans with rational-expectations (RE) investors. These investors know the pre-trained AI demand, correctly infer information from the actual equilibrium price, and adjust their demands.

To isolate AI sophistication from any direct informational advantage, we focus on AI investors with no private information. Allowing the AI to observe a private signal would make its relative profits jointly reflect the quality of its information and the sophistication embedded in its learned trading rule. Without a private signal, the AI learns about fundamentals

exclusively from prices. The information embedded in prices depends on the strategic trading of informed humans, so the AI's ability to anticipate their behavior directly shapes its performance. Throughout this subsection, AI private-signal precision is zero, $\tau_A = 0$.

Training against more sophisticated humans raises the AI's relative performance over both cursed and RE investors. This finding creates a tension with a purely informational account of AI learning. More sophisticated humans may shade their trades more effectively to conceal their private information, leaving an AI trained against them with less informative market prices. Why, then, should training in such an environment make the AI more sophisticated?

The increase in AI sophistication comes from the anticipation of humans' strategic behavior embedded in the AI's trading rule through the training process. When the AI is trained against lower-level human investors, it learns in an environment in which other market participants only imperfectly understand how the AI extracts information from prices or how they might strategically respond to its behavior. Therefore, the trained AI rule does not adequately anticipate such strategic responses. When the trained AI is subsequently evaluated against rational-expectations investors, these investors can better understand and strategically respond to the AI's trading behavior, leaving greater scope for its strategy to be exploited.

By contrast, training against more sophisticated investors exposes the AI to an environment in which other traders behave more strategically. The resulting trading rule therefore incorporates richer anticipation of how other investors trade, draw inferences from prices, and adjust their strategies in response to the AI. Training in a more sophisticated environment can thus increase the AI's own strategic sophistication—even if prices in that environment are no more informative in a purely statistical sense.

Rational-Expectations (RE) Human Investors. RE investors correctly understand the demand of the pre-trained AI algorithm and the demands of cursed and other RE investors. They therefore know how the equilibrium price aggregates information. After observing their private signal and the price, they use Bayesian updating to form a posterior belief about the future payoff and choose demand optimally. Their demand is

$$\begin{aligned} x_i^{M,\mathrm{RE}} &= \frac{\mathbb{E}[v \mid \eta_i, p] - p}{\gamma_M} \\ &= \beta_\eta^{M,\mathrm{RE}} \eta_i - \beta_{p,k}^{M,\mathrm{RE}} p + \mu_k^{M,\mathrm{RE}}, \end{aligned} \tag{43}$$

with coefficients

$$\beta_\eta^{M,\mathrm{RE}} = \frac{\tau_M}{\gamma_M \widehat{\tau}^{M,\mathrm{RE}}}, \tag{44}$$

$$\beta_{p,k}^{M,\mathrm{RE}} = \frac{1 - h_p^{M,\mathrm{RE}} \xi_{k\to\mathrm{RE}}/\zeta^{\mathrm{RE}}}{\gamma_M}, \tag{45}$$

$$\mu_k^{M,\mathrm{RE}} = \frac{h_v^{M,\mathrm{RE}} \bar{v} - h_p^{M,\mathrm{RE}} \mu_{k\to\mathrm{RE}}/\zeta^{\mathrm{RE}}}{\gamma_M}. \tag{46}$$

Here equilibrium price precision is τ_p^{RE}, total human posterior precision is $\widehat{\tau}^{M,\mathrm{RE}} \equiv \tau_v + \tau_M + \tau_p^{\mathrm{RE}}$, the price-information weight is $h_p^{M,\mathrm{RE}} \equiv \tau_p^{\mathrm{RE}}/\widehat{\tau}^{M,\mathrm{RE}}$, and the prior-information weight is $h_v^{M,\mathrm{RE}} \equiv \tau_v/\widehat{\tau}^{M,\mathrm{RE}}$. The coefficients $(\xi_{k\to\mathrm{RE}}, \zeta^{\mathrm{RE}}, \mu_{k\to\mathrm{RE}})$ characterize the equilibrium price function. Appendix D.1 derives the RE equilibrium and solves the fixed point for price precision. RE investors adjust their demand to each pre-trained algorithm, but equilibrium price precision remains the same across training levels of the algorithms. This invariance isolates the mechanism: changes in relative AI performance cannot come from a more informative deployment price. They must come from the price-contingent demand carried from training into deployment.

The expected-profit formulas can be obtained by applying Proposition 7 after substituting the equilibrium price coefficients $(\xi_{k\to\mathrm{RE}}, \zeta^{\mathrm{RE}}, \mu_{k\to\mathrm{RE}})$, the cursed and RE demand coefficients, and the pre-trained AI coefficients $(\mu^{A,k}, \beta_p^{A,k})$. The AI private-signal loading is zero, $\beta_\eta^{A,k} = 0$, because AI private-signal precision is zero, $\tau_A = 0$. Let $\Pi_{k\to\mathrm{RE}}^{A}$, $\Pi_{k\to\mathrm{RE}}^{M,c}$, and $\Pi_{k\to\mathrm{RE}}^{M,\mathrm{RE}}$ denote the resulting per-investor test profits. Define the two relative-profit gaps by

$$\Delta_c(k) \equiv \Pi_{k\to\mathrm{RE}}^{A} - \Pi_{k\to\mathrm{RE}}^{M,c}, \qquad \Delta_{\mathrm{RE}}(k) \equiv \Pi_{k\to\mathrm{RE}}^{A} - \Pi_{k\to\mathrm{RE}}^{M,\mathrm{RE}}. \tag{47}$$

Lemma 13 (Relative-profit decomposition)**.** *Suppose $\bar{v} = S = 0$, $\gamma_M = \gamma_A \equiv \gamma$, and $\psi_s > 0$. Suppose the training coefficient system converges, all training markets are well defined, and every induced rational-expectations equilibrium is nondegenerate. Then there is a constant $C_{\mathrm{RE}} > 0$, independent of k, such that*

$$\Delta_c(k) = \frac{\tau_p^{\mathrm{RE}} - \tau_M}{2\gamma(\tau_v + \tau_M)(\tau_v + \tau_p^{\mathrm{RE}})} - \mathcal{R}_k, \tag{48}$$

$$\Delta_{\mathrm{RE}}(k) = -\frac{\tau_M}{2\gamma(\tau_v + \tau_p^{\mathrm{RE}})(\tau_v + \tau_M + \tau_p^{\mathrm{RE}})} - \mathcal{R}_k, \tag{49}$$

$$\mathcal{R}_k = C_{\mathrm{RE}} \left(\frac{|\beta_p^{A,k} - \beta_p^{A,\infty}|}{|\psi/\gamma + \phi\beta_p^{A,k}|} \right)^2. \tag{50}$$

The two gaps therefore have identical first differences.

Lemma 13 separates information from deployment mismatch. In the AI–cursed gap, the first term is the information advantage of a reoptimized price-only AI relative to the cursed investor. In the AI–RE gap, it is the negative incremental value of the RE investor's private signal conditional on the price both types observe. The second term, deployment regret $\mathcal{R}_k$, is the exact profit loss from using the pre-trained AI demand rather than the demand that is the best response to the test-market price. In the centered benchmark, equation (50) measures this error by the squared, normalized distance between the learned AI price loading, $\beta_p^{A,k}$, and the price loading in the limiting environment with infinite-level humans, $\beta_p^{A,\infty}$. The normalization accounts for how the pre-trained AI demand affects aggregate price sensitivity.

Because equilibrium price precision, τ_p^{RE}, does not vary with training sophistication, k, the information term in each comparison is fixed. All variation across training environments enters through deployment regret, $\mathcal{R}_k$, and the two profit gaps have identical first differences. An RE human always outperforms the frozen AI when human private-signal precision is positive, $\tau_M > 0$: she observes the same price, adds a private signal, and optimizes against the actual equilibrium price function. AI can outperform cursed investors only when its price-information advantage is large enough to cover deployment regret.

Proposition 14 (Training sophistication and relative AI performance)**.** *Maintain the assumptions of Lemma 13. Suppose the regularity condition defined in Appendix D.1 holds. Then both relative-profit gaps are weakly increasing in training sophistication:*

$$\Delta_h(k+1) \geq \Delta_h(k), \qquad h \in \{c, \mathrm{RE}\}, \quad k \geq 0. \tag{51}$$

Higher-k training improves AI relative performance in competing with both types of human investors. Under the stated regularity condition, training against more sophisticated humans reduces the normalized distance in (50). Deployment regret falls, so AI performance improves by the same amount relative to both human benchmarks. [13]

Low-sophistication training exposes AI to a market in which other investors perform only a few rounds of strategic reasoning about prices and the price feedback created by AI demand. The trained demand can fail to anticipate the higher-order responses present in the test market with RE human investors. RE investors know AI's trading rule, infer correctly from the equilibrium price and best respond. The mismatch leaves greater scope for the AI's demand to be exploited and widens its profit shortfall relative to RE investors.

Higher-sophistication training instead exposes AI to demands that embody richer anticipation of how investors trade, infer from prices, and respond to AI demand. Under the

[13] The result concerns relative profit gaps. Absolute AI profit need not move monotonically because the frozen algorithm changes both the equilibrium price and the profit available to a reoptimized AI.

proposition's regularity condition, the learned demand has a smaller normalized mismatch with the best-response demand in the test market. Training against more sophisticated humans can therefore improve AI's relative test-market performance by reducing deployment mismatch.

Figure 7 shows that both relative-profit gaps weakly increase with training-market sophistication across the three test environments. Most of the adjustment occurs at low training levels, after which the series plateau. The information environment determines how much this improvement can achieve. When noise-trader volatility is high, $(\sigma_M, \sigma_z) = (1, 0.8)$, both gaps improve only slightly and remain negative: noisy order flow weakens the price signal on which AI exclusively relies. When human private-signal noise is high, $(\sigma_M, \sigma_z) = (4, 0.2)$, both gaps again improve only slightly and remain negative. In that environment, humans trade on little fundamental information, so little information reaches the equilibrium price for AI to extract. More sophisticated training reduces deployment mismatch in both cases but cannot create the missing information.

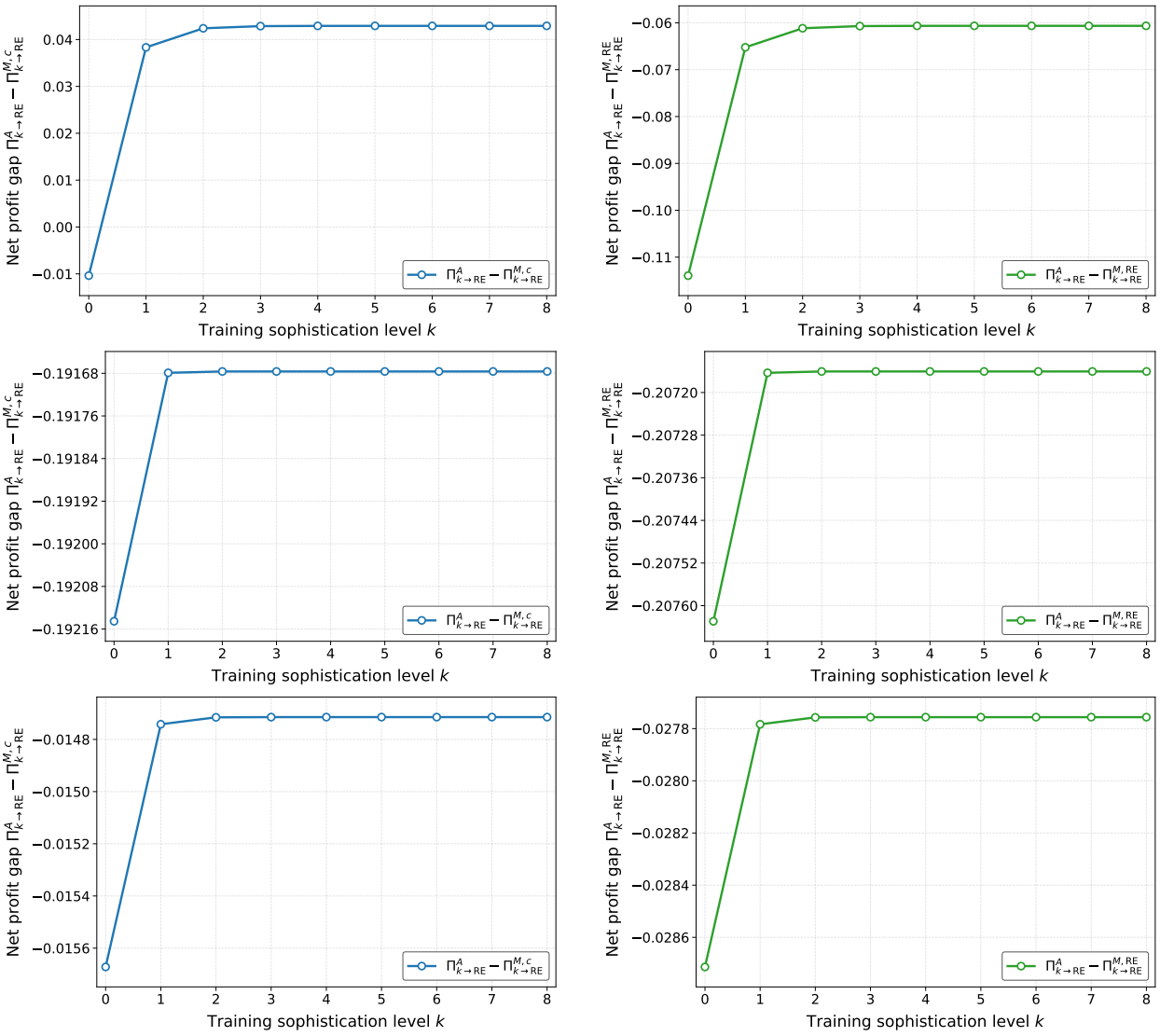
The most favorable case combines low noise-trader volatility with moderately precise human signals, $(\sigma_M, \sigma_z) = (1, 0.2)$. The AI–cursed profit gap rises from about -0.010 at training sophistication $k = 0$ to 0.043 at $k = 8$, becoming positive at $k = 1$. The AI–RE profit gap rises by the same amount, from about -0.114 to -0.061, but remains negative because RE investors retain their private-information advantage. Thus sophisticated training can improve AI's relative performance and can overturn its ranking against cursed investors in a favorable information environment. It does not remove the profitability ceiling created by AI's lack of private information: AI earns less than RE investors in every displayed case.

Appendix D.2 repeats the exercise with the level-∞ humans from the previous subsection. Their demand is held fixed as different pre-trained algorithms enter the test market, so it need not remain optimal for each resulting price.[14] The same qualitative pattern emerges: AI performance relative to both cursed and level-∞ humans weakly improves, mainly at low training levels, in the three displayed calibrations. More sophisticated training narrows AI's disadvantage

In summary, AI performance is greatly limited by the sophistication of its training environment. Even in a market with an information environment favorable to AI, training against low-level humans leaves the AI vulnerable to exploitation by more sophisticated investors. Training against higher-level humans improves AI's relative performance, but it cannot overcome the fundamental disadvantage of lacking private information.

[14]The RE exercise removes this mismatch by allowing humans to best respond to each pre-trained AI algorithm.

Figure 7: Training Sophistication and Performance in a Rational-Expectations Test Market



(a) AI profit gap relative to cursed investors

(b) AI profit gap relative to RE investors

Notes: The horizontal axis is the sophistication $k \in \{0, \ldots, 8\}$ of strategic investors in the AI training market. For each k, we compute the converged AI algorithm and deploy it without further updating in a market with rational-expectations investors. AI receives no private signal in either market (AI private-signal precision is $\tau_A = 0$). The RE investors know the pre-trained algorithm and best respond to the equilibrium price; cursed investors follow the cursed demand. Each profit symbol below is the corresponding investor's per-investor net expected profit in the test market. Panel (a) plots $\Pi^A_{k\to\mathrm{RE}} - \Pi^{M,c}_{k\to\mathrm{RE}}$, and panel (b) plots $\Pi^A_{k\to\mathrm{RE}} - \Pi^{M,\mathrm{RE}}_{k\to\mathrm{RE}}$. Positive values favor AI. From top to bottom, the pairs of human private-signal noise and noise-trader volatility, (σ_M, σ_z), are $(1, 0.2), (1, 0.8), (4, 0.2)$. Other parameters are $\bar{v} = 0$, $S = 0$, $\sigma_v = 1$, $\gamma_M = \gamma_A = 1$, $\psi_c = 0.30$, $\psi_s = 0.30$, and $\phi = 0.40$. Training and test markets share these primitives and population shares. Vertical scales differ across the six plots.

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–.

Borkar, V. S. 2002. On the lock-in probability of stochastic approximation. *Combinatorics, Probability and Computing* 11:11–20.

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. 2026. Algorithmic pricing and liquidity in securities markets. *The Review of Financial Studies* hhag010. doi:10.1093/rfs/hhag010.

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. 2023. Ai-powered trading, algorithmic collusion, and price efficiency. *Available at SSRN 4452704* .

———. 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: Supracompetitive 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. 1985. Continuous auctions and insider trading. *Econometrica* 53:1315–35. ISSN 00129682, 14680262.

———. 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.

Watkins, C. C. 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 establishes when informed AI investors who learn from trading profits herd on a stable aggregate demand rule. The economic force is market-clearing price feedback. When AI investors shift toward different demand schedules, their aggregate demand changes the market-clearing price. The new price changes both expected trading profits and the information that investors extract from price, which in turn changes the rewards used for learning. The proof determines when this feedback stabilizes the learning process and when its limiting aggregate demand rule coincides with rational-expectations equilibrium (REE) demand.

The learning argument begins with a finite, bounded economy tailored to tabular reinforcement learning. Each AI investor chooses from finitely many complete demand schedules, and all shocks are bounded so that prices, positions, and rewards remain finite. An auxiliary projection keeps market clearing well defined even for demand profiles that arise only because exploration assigns positive probability to every schedule. These restrictions preserve the economic price-feedback mechanism while providing a well-defined learning problem at every history.

Within one fixed bounded economy, changing a single investor's schedule has only a limited effect on prices and profits. A sufficient contraction condition then gives a unique Bellman fixed point and global convergence of the stochastic learning process. This condition uses worst-case feedback over the entire finite action set.

When the sufficient contraction condition cannot hold, as may occur when the action set expands and the shock bounds are relaxed, we provide a separate local argument for the approximation. The local argument first solves the atomistic informed benchmark. Expected profit is strictly concave in the coefficients of a complete demand schedule, and a continuous Boltzmann policy is Gaussian around the price-taking best response. Its dispersion depends on exploration, but its mean does not. Market consistency has a sequential structure: signal demand determines price informativeness, after which price sensitivity and the demand intercept are determined. This structure identifies the informed rational-expectations equilibrium and supplies the local feedback signs used in the stability proof.

The remaining steps connect discretized bounded economies to this benchmark in an ordered way. Fine action grids and wide shock bounds approximate the continuous best-response map; a large finite population then makes demand-schedule sampling noise, idiosyncratic signal noise, and individual price impact small. Under the local regularity and

continuation-gap conditions stated below, the resulting mean-field fixed point is unique and locally stable. Once the stochastic process enters its local basin, declining learning gains make escape unlikely and yield convergence with high conditional probability. The convergence object is the population-average demand schedule implied by the investors' mixed policies. Continued exploration leaves realized schedules fluctuating, and heterogeneous investors may retain different mixed policies, even after this aggregate object converges.

A.1 Discretized bounded environment

We first define the discretized bounded environments used throughout the learning proof. They preserve the asset-market structure, human demand schedules, informed AI demand schedules, noise-trader demand, and market-clearing equation of the original model. They modify only the objects required to make tabular learning well defined: AI demand coefficients lie on a finite grid, shocks are truncated, and a projection maps exploratory aggregate profiles into a region in which market clearing yields a bounded price.

These restrictions have distinct roles. The finite grid makes each arm a complete signal- and price-contingent schedule. Truncation bounds realized prices, positions, and trading profits. Projection supplies a bounded off-equilibrium price for profiles that full-support exploration can generate, while leaving market clearing unchanged whenever the aggregate coefficients remain in the admissible region. The approximation proof chooses this region around the informed atomistic equilibrium, where projection is inactive, and makes the grid sufficiently fine and the truncation bounds sufficiently wide before increasing the finite AI population. The finite-population formulation also retains the cross-sectional average of AI signal errors, which remains a source of price and payoff noise at every fixed population size.

Market structure. Time is discrete. A single risky asset with per capita supply S is traded in each period. Its payoff and noise-trader demand are

$$v \sim \mathcal{N}(\bar{v}, \sigma_v^2), \qquad z \sim \mathcal{N}(0, \sigma_z^2), \qquad v \perp z.$$

Write $\tau_v = \sigma_v^{-2}$, $\tau_M = \sigma_M^{-2}$, $\tau_A = \sigma_A^{-2}$, and $\tau_z = \sigma_z^{-2}$. The main informed case has $\tau_A > 0$. Section A.10 treats $\tau_A = 0$ by removing the signal coordinate.

Human investors. A measure ψ_c of cursed investors and a measure ψ_s of level-k investors observe $\eta_i = v + e_i$, where $e_i \sim \mathcal{N}(0, \sigma_M^2)$. Cursed investors ignore the information in price and use

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

where

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

Level-k investors use

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

with finite coefficients determined by the level-k recursion in the economic model. We fix these coefficients throughout the learning proof.

Aggregate human demand. Idiosyncratic human signal errors average out in the continuum. Define the fixed aggregate human coefficients

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

Thus aggregate human demand is

$$\bar{\beta}_\eta^M v - \bar{\beta}_p^M p + \bar{\mu}^M.$$

AI investors and discretized demand schedules. AI investor j observes $\eta_j = v + e_j$, where $e_j \sim \mathcal{N}(0, \sigma_A^2)$ independently across investors and independently of (v, z) and all human signal errors. Each arm is a complete signal- and price-contingent demand schedule: it maps the investor's private signal and the market-clearing price into demand and is represented by a coefficient triple rather than by a single realized position.

For coordinate meshes $\delta_\eta, \delta_p, \delta_\mu > 0$, positive action bounds B_η, B_p, B_μ, and translations s_η, s_p, s_μ, define one-dimensional lattices by

$$\begin{aligned}
\mathcal{E}_\Delta &= \{s_\eta + k\delta_\eta : k \in \mathbb{Z}\} \cap [-B_\eta, B_\eta], \\
\mathcal{B}_\Delta &= \{s_p + k\delta_p : k \in \mathbb{Z}\} \cap [-B_p, B_p], \\
\mathcal{U}_\Delta &= \{s_\mu + k\delta_\mu : k \in \mathbb{Z}\} \cap [-B_\mu, B_\mu].
\end{aligned}$$

The action grid is the equally spaced Cartesian product $\mathcal{X}_\Delta = \mathcal{E}_\Delta \times \mathcal{B}_\Delta \times \mathcal{U}_\Delta$. Assume $\mathcal{X}_\Delta$ is nonempty and write $M := |\mathcal{X}_\Delta|$. Every interior grid cell has common volume

$$h_\Delta := \delta_\eta\delta_p\delta_\mu, \tag{A4}$$

which cancels from the numerator and denominator of every discrete Boltzmann policy. Along a refinement, each coordinate mesh tends to zero and each pair of action bounds tends to infinity. The translations may vary with the refinement but remain within one

mesh of a fixed origin.

An arm $a \in \{1, \ldots, M\}$ is the coefficient triple

$$y(a) = \big(\beta_\eta(a), \beta_p(a), \mu(a)\big) \in \mathcal{X}_\Delta,$$

with

$$|\beta_\eta(a)| \le B_\eta, \qquad |\beta_p(a)| \le B_p, \qquad |\mu(a)| \le B_\mu.$$

Arm a represents the complete rule

$$x(a; \eta, p) = \beta_\eta(a)\eta - \beta_p(a)p + \mu(a). \tag{A5}$$

The coefficient order is $(\beta_\eta, \beta_p, \mu)$ throughout.

Finite number of AI investors. The learning economy represents the measure-ϕ AI sector by N investors, each with market weight ϕ/N. Investors may differ in exploration temperature, discount factor, and learning speed. The fixed-environment result holds for finite N; the REE approximation lets N grow only after the action grid and truncation bounds have been fixed.

Population profiles and signal-error averages. For an assignment $A = (A^1, \ldots, A^N)$, define its empirical arm profile and unprojected coefficient mean by

$$\widehat{f}^N(A, a) = \frac{1}{N} \sum_{j=1}^{N} \mathbf{1}\{A^j = a\}, \tag{A6}$$

$$m(A) = \frac{1}{N} \sum_{j=1}^{N} y(A^j) = \big(\bar{\beta}_\eta(A), \bar{\beta}_p(A), \bar{\mu}(A)\big). \tag{A7}$$

The idiosyncratic signal aggregate is

$$w_N(A, e) = \frac{1}{N} \sum_{j=1}^{N} \beta_\eta(A^j) e_j. \tag{A8}$$

Projected aggregate coefficients. Full-support exploration can generate coefficient profiles with nearly singular price denominators. The bounded learning economy therefore includes a projected clearing rule.

Assumption 1 (Projected admissibility)**.** *There is a compact rectangle*

$$\mathcal{K}_{\text{proj}} = [\beta_{\eta,-}^{\text{proj}}, \beta_{\eta,+}^{\text{proj}}] \times [\beta_{p,-}^{\text{proj}}, \beta_{p,+}^{\text{proj}}] \times [\mu_{-}^{\text{proj}}, \mu_{+}^{\text{proj}}]$$

such that

$$\underline{\xi}_{\text{proj}} := \inf_{m \in \mathcal{K}_{\text{proj}}} \{\bar{\beta}_p^M + \phi m_2\} > 0.$$

Let $m_{\text{proj}}(A) = \text{proj}_{\mathcal{K}_{\text{proj}}} m(A)$ *be coordinatewise Euclidean projection and write its coordinates as* $(\bar{\beta}_{\eta,\text{proj}}(A), \bar{\beta}_{p,\text{proj}}(A), \bar{\mu}_{\text{proj}}(A))$.

The projection is part of the auxiliary learning economy, not the atomistic benchmark. It clips the three aggregate coefficients but leaves the realized idiosyncratic term w_N unchanged. In the approximation argument, $\mathcal{K}_{\text{proj}}$ is a compact neighborhood of the informed REE, so the projection is inactive near the limiting equilibrium. If $m(A) \in \mathcal{K}_{\text{proj}}$, (A9) equals the market-clearing price in the unprojected finite economy. When the projection binds, the same expression provides a bounded off-equilibrium continuation for tabular rewards. Thus the auxiliary economy preserves projection-inactive fixed points and their local dynamics without attempting to reproduce distant singular paths.

Truncation and payoff bounds. In a fixed bounded economy, condition the Gaussian shocks independently on

$$|v - \bar{v}| \le B_v, \qquad |z| \le B_z, \qquad |e_j| \le B_e.$$

Here truncation means conditioning each Gaussian variable independently on the displayed event. The conditional signal-error distribution remains centered.

Market price, demand, and profit. Given (A, e, v, z), market clearing yields the projected price

$$p(A, e; v, z) = \frac{[\bar{\beta}_\eta^M + \phi \bar{\beta}_{\eta,\text{proj}}(A)]v + z + \bar{\mu}^M + \phi \bar{\mu}_{\text{proj}}(A) - S + \phi w_N(A, e)}{\bar{\beta}_p^M + \phi \bar{\beta}_{p,\text{proj}}(A)}. \tag{A9}$$

Set $V_{\max} = |\bar{v}| + B_v$ and

$$M_{\eta,\text{proj}} = \sup_{m \in \mathcal{K}_{\text{proj}}} |m_1|, \qquad M_{\mu,\text{proj}} = \sup_{m \in \mathcal{K}_{\text{proj}}} |m_3|,$$

$$A_{\max} = (|\bar{\beta}_\eta^M| + \phi M_{\eta,\text{proj}})V_{\max} + B_z + |\bar{\mu}^M - S| + \phi M_{\mu,\text{proj}} + \phi B_\eta B_e,$$

$$P_{\max} = A_{\max}/\underline{\xi}_{\text{proj}}, \qquad X_{\max} = B_\eta(V_{\max} + B_e) + B_p P_{\max} + B_\mu,$$

$$\Omega_B = (V_{\max} + P_{\max})X_{\max} + \frac{\gamma_A}{2} X_{\max}^2.$$

The realized profit of investor j when it uses arm a is

$$\pi^j(a; A, e, v, z) = [v - p(A, e; v, z)]x(a; v + e_j, p(A, e; v, z)) - \frac{\gamma_A}{2} x(a; v + e_j, p(A, e; v, z))^2. \tag{A10}$$

Assumption 1 and the truncations give

$$|p| \le P_{\max}, \qquad |x| \le X_{\max}, \qquad |\pi^j| \le \Omega_B.$$

Definition 3 (Discretized bounded environment)**.** *A discretized bounded environment is the finite-grid, shock-truncated economy described in this subsection. AI investors choose complete signal- and price-contingent demand schedules from $\mathcal{X}_\Delta$, each investor represents mass ϕ/N, market clearing uses (A9), and rewards are the raw trading profits in (A10). Projected admissibility keeps the price denominator bounded away from zero for every realized arm assignment.*

A.2 Realized profiles and the conditional payoff map

We next link learned values to the demand schedules that enter market clearing. Each AI investor stores one value for every complete schedule. The Boltzmann policy converts these values into probabilities, and independent draws from these policies produce the realized cross-section of schedules in each period. Market clearing depends on this realized cross-section and on the contemporaneous signal errors, whereas the aggregate object used in the mean dynamics is the conditional mean implied by the policies.

The payoff assigned to one schedule is counterfactual. To evaluate a particular investor's schedule, hold that investor's arm fixed and average over the other investors' independent arm draws and all current shocks. Adding the discounted maximal continuation value gives the Bellman map. A mean-field fixed point requires the policies used to form these conditional payoffs to be reproduced by the resulting Bellman values. Because the continuation term is

common across an investor's arms at a fixed point, discounting shifts values without changing Boltzmann odds, fixed-point policies, or their policy-implied coefficient mean.

Policies and types. Investor j has raw exploration temperature $\vartheta_j > 0$ (denoted κ_j in the main text), learning-speed multiplier $c_j > 0$, and discount factor $\rho_j \in [0, 1)$. Its Boltzmann policy is

$$\sigma_j(Q^j, a) = \frac{\exp(Q^j(a)/\vartheta_j)}{\sum_{b=1}^{M} \exp(Q^j(b)/\vartheta_j)}. \tag{A11}$$

Assumption 2 (Algorithm types and stepsizes)**.** *There are constants*

$$0 < c_{\min} \le c_j \le c_{\max} < \infty, \quad 0 < \underline{\vartheta} \le \vartheta_j \le \bar{\vartheta} < \infty, \quad 0 \le \rho_j \le \bar{\rho} < 1.$$

The stepsizes have the common-base form $0 < \alpha_{j,t} = c_j \bar{\alpha}_t \le 1$*, where* $(\bar{\alpha}_t)_{t \ge 0}$ *is a deterministic positive sequence satisfying*

$$\sum_t \bar{\alpha}_t = \infty, \qquad \sum_t \bar{\alpha}_t^2 < \infty.$$

Definition 4 (Admissible population sequence)**.** *An admissible population sequence is a triangular sequence* $\{(c_{j,N}, \vartheta_{j,N}, \rho_{j,N})\}_{j=1}^{N}$*,* $N \ge 1$*, satisfying the uniform bounds in Assumption 2 for every* N*. When a population size is fixed, we suppress the second subscript.*

Conditional mean and realized profiles. For a value profile $\mathbf{Q} = (Q^1, \ldots, Q^N)$, the policy-implied conditional mean arm profile is

$$\bar{f}(\mathbf{Q}, a) := \frac{1}{N} \sum_{j=1}^{N} \sigma_j(Q^j, a), \qquad a = 1, \ldots, M. \tag{A12}$$

Its coefficient triple is

$$m_{\Delta,B,N}(\mathbf{Q}) := \sum_{a=1}^{M} \bar{f}(\mathbf{Q}, a) y(a) = \frac{1}{N} \sum_{j=1}^{N} \sum_{a=1}^{M} \sigma_j(Q^j, a) y(a). \tag{A13}$$

The realized arm profile is $\widehat{f}^N(A, \cdot)$ from (A6). Unlike $\bar{f}(\mathbf{Q}, \cdot)$, it fluctuates across periods under positive exploration. Market clearing also depends on the realized signal-error average $w_N(A, e)$, so it cannot in general be recovered from the realized arm profile alone.

Conditional sampling protocol. Define the information available before date-t choices and shocks by

$$\mathcal{F}_t := \sigma(\mathbf{Q}_0, \{A_s, e_s, v_s, z_s : 0 \le s < t\}), \tag{A14}$$

where $e_s = (e_{1,s}, \dots, e_{N,s})$. The value profile $\mathbf{Q}_t$ is $\mathcal{F}_t$-measurable. Conditional on $\mathcal{F}_t$, investors draw arms independently using (A11). Current truncated shocks and signal errors are independent of both the history and the arm draws. To evaluate investor j's arm a, hold that arm fixed and integrate over every other investor's draw. Formally, let $A^{j\to a}$ replace A^j by a while the other arms follow their current policies, and define the conditional raw payoff

$$\bar{\Pi}^j_{B,N}(\mathbf{Q}, a) := \mathbb{E}\big[\pi^j(a; A^{j\to a}, e, v, z) \mid \mathbf{Q}\big]. \tag{A15}$$

The discounted Bellman map is

$$\mathcal{B}_{B,N}(\mathbf{Q})^j(a) = \bar{\Pi}^j_{B,N}(\mathbf{Q}, a) + \rho_j \max_b Q^j(b). \tag{A16}$$

Definition 5 (Mean-field fixed point in a discretized bounded environment)**.** *A value profile $\mathbf{Q}^*$ is a finite-population mean-field fixed point if $\mathbf{Q}^* = \mathcal{B}_{B,N}(\mathbf{Q}^*)$. Its policy-implied coefficient mean is $m_{\Delta,B,N}(\mathbf{Q}^*)$.*

Lemma 15 (Existence and discount invariance)**.** *Every fixed discretized bounded environment admits a mean-field fixed point. If $\mathbf{Q}^*$ is a mean-field fixed point, then $\mathbf{Q}^*$ lies in*

$$\mathcal{Q}_B := \prod_{j=1}^{N} \left[-\frac{\Omega_B}{1-\rho_j}, \frac{\Omega_B}{1-\rho_j}\right]^M. \tag{A17}$$

At such a mean-field fixed point,

$$Q^{j,*}(a) = \bar{\Pi}^j_{B,N}(\mathbf{Q}^*, a) + \frac{\rho_j}{1-\rho_j} \max_b \bar{\Pi}^j_{B,N}(\mathbf{Q}^*, b), \tag{A18}$$

$$\sigma_j(Q^{j,*}, a) = \frac{\exp(\bar{\Pi}^j_{B,N}(\mathbf{Q}^*, a)/\vartheta_j)}{\sum_b \exp(\bar{\Pi}^j_{B,N}(\mathbf{Q}^*, b)/\vartheta_j)}. \tag{A19}$$

Thus changing the discount vector sends a mean-field fixed point to a mean-field fixed point with the same policy profile and coefficient mean, while each investor's value vector changes by an arm-independent shift. Reversing the discount change reverses this mapping.

Proof. The payoff bound implies that the Bellman map sends (A17) into itself. Each conditional payoff is a finite sum of bounded shock expectations weighted by products of continuous Boltzmann probabilities. The maximum is continuous as well, so Brouwer's theorem

gives a fixed point. Conversely, taking sup norms in the fixed-point identity places every mean-field fixed point in the stated box.

At a fixed point, investor j's continuation term is common across all of its arms. The term therefore cancels from Boltzmann odds, which gives (A19). Taking the maximum of the Bellman identity and solving for $\max_b Q^{j,*}(b)$ gives (A18). Conversely, (A18) lifts any policy profile satisfying (A19) uniquely to a value profile for every discount vector. Applying the construction in both directions proves discount invariance. □

A.3 Lipschitz continuity of conditional expected payoffs

The global contraction and the later finite-population derivative argument both require a bound on the effect of one investor's schedule. Changing one arm moves the aggregate signal loading, price sensitivity, and intercept. It also changes the realized loading on that investor's idiosyncratic signal error. The market-clearing price transmits these changes to the demands and profits of investors whose own schedules are held fixed.

Projected denominator safety and bounded positions make this feedback Lipschitz. Since each investor represents an equal share of the AI sector, the price and recipient-profit effects of a single arm swap are proportional to the investor's population weight. The next lemma records explicit bounds. They provide the primitive input for the whole-environment Bellman contraction and, later, for comparing derivatives of the finite-population and atomistic payoff maps.

Lemma 16 (Single-swap payoff sensitivity)**.** *Suppose assignments A and A' differ only in investor ℓ's arm. Holding (e, v, z) fixed,*

$$|p(A, e; v, z) - p(A', e; v, z)| \leq \frac{2L_{p,B}}{N}, \tag{A20}$$

where

$$L_{p,B} := \phi \left[\frac{A_{\max} B_p}{\underline{\xi}_{\text{proj}}^2} + \frac{B_\mu + B_\eta (V_{\max} + B_e)}{\underline{\xi}_{\text{proj}}} \right]. \tag{A21}$$

For any recipient whose arm is unchanged,

$$|\pi^j(a; A, e, v, z) - \pi^j(a; A', e, v, z)| \leq \frac{2L_{\pi,B}}{N}, \tag{A22}$$

where

$$L_{\pi,B} := \big[(V_{\max} + P_{\max} + \gamma_A X_{\max}) B_p + X_{\max} \big] L_{p,B}. \tag{A23}$$

Proof. An arm swap changes each unprojected coefficient mean by at most twice the corre-

sponding grid bound divided by N. Coordinatewise projection is nonexpansive, and the same swap changes (A8) by at most $2B_\eta B_e/N$. Subtract the two price ratios in (A9). Bounding both denominators below by $\underline{\xi}_{\text{proj}}$ and both numerators by $A_{\max}$ gives (A20). The bound includes the change in the projected fundamental loading, the intercept, the price sensitivity, and the realized idiosyncratic term.

For a fixed recipient and arm, the swap affects demand only through price, so

$$|x(a;\eta,p) - x(a;\eta,p')| \le B_p|p - p'|.$$

Use this inequality in (A10), and then apply the uniform bounds on v, p, and x. The result is (A22). □

A.4 Contraction and fixed point

To compare expected payoffs at two value profiles, change the other investors' mixed policies one at a time. The single-swap bound limits each payoff change, while the Boltzmann policy limits how much arm probabilities respond to learned values. Adding the continuation term gives the total strength of payoff and discounting feedback in the Bellman map.

Exploration governs one side of this trade-off. Lower temperatures make arm probabilities more responsive to value differences and strengthen feedback through market clearing; higher temperatures flatten the policy response. When the combined payoff and continuation feedback is weaker than one, the contraction-mapping theorem gives a unique mean-field fixed point. The condition is imposed over the entire finite action grid and supports the global fixed-environment result below.

Lemma 17 (Conditional payoff and Bellman contractions)**.** *For* $\mathbf{Q}, \widetilde{\mathbf{Q}} \in \mathcal{Q}_B$*,*

$$\max_{j,a} |\bar{\Pi}^j_{B,N}(\mathbf{Q}, a) - \bar{\Pi}^j_{B,N}(\widetilde{\mathbf{Q}}, a)| \le \Lambda_{B,N} \|\mathbf{Q} - \widetilde{\mathbf{Q}}\|_\infty, \tag{A24}$$

where

$$\Lambda_{B,N} = 4L_{\pi,B} \max_j \frac{1}{N} \sum_{\ell \ne j} \vartheta_\ell^{-1}.$$

Consequently,

$$\|\mathcal{B}_{B,N}(\mathbf{Q}) - \mathcal{B}_{B,N}(\widetilde{\mathbf{Q}})\|_\infty \le (\Lambda_{B,N} + \bar{\rho}) \|\mathbf{Q} - \widetilde{\mathbf{Q}}\|_\infty. \tag{A25}$$

Proof. Fix recipient j and counterfactual arm a. Replace the product distribution of the other investors' arms one marginal distribution at a time. Lemma 16 bounds the payoff oscillation associated with investor ℓ's arm by $2L_{\pi,B}/N$. The total-variation comparison and

the softmax Jacobian give

$$|\mathbb{E}_P h - \mathbb{E}_{\widetilde{P}} h| \le \operatorname{osc}(h)\|P - \widetilde{P}\|_1, \qquad \|\sigma_\ell(Q^\ell) - \sigma_\ell(\widetilde{Q}^\ell)\|_1 \le \frac{2}{\vartheta_\ell}\|Q^\ell - \widetilde{Q}^\ell\|_\infty.$$

Summing over $\ell \neq j$ proves (A24). Since the maximum is one-Lipschitz and $\rho_j \le \bar{\rho}$, adding the continuation term gives (A25). □

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

This section proves convergence of heterogeneous Q-learning in one fixed discretized bounded economy. In each period, investors draw complete demand schedules using their Boltzmann policies, trade at the price generated by the realized cross-section, and update only the values of the schedules they selected. The expected update holds an investor's own arm fixed and averages over the other investors' randomized schedules and the current shocks, exactly as in the conditional payoff map above.

The value and reward bounds keep every learning update in a fixed range, while positive Boltzmann probabilities ensure that every schedule continues to be sampled. Investors may differ in learning speeds, exploration temperatures, and discount factors. A common base stepsize places their updates on a common time scale, and the type bounds rule out vanishing or explosive responses.

Under the fixed-environment contraction, the mean dynamics have a unique globally stable mean-field fixed point, and the stochastic recursion converges to it almost surely. Discounting changes the continuation values but not the fixed-point policy profile because it adds the same shift to all arms. The section also relates normalized-reward implementations to raw trading profits. That equivalence is pointwise in a fixed bounded economy; the approximation analysis keeps raw exploration temperatures in a fixed positive range. When all discount factors are zero, the continuation term and the local maximizing-arm requirement disappear.

Stochastic recursion. At date t, investor j draws A_t^j using (A11), observes

$$\widetilde{\pi}_t^j = \pi^j(A_t^j; A_t, e_t, v_t, z_t),$$

and updates

$$Q_{t+1}^j(a) = Q_t^j(a) + \alpha_{j,t}\mathbf{1}\{A_t^j = a\}\big[\widetilde{\pi}_t^j + \rho_j \max_b Q_t^j(b) - Q_t^j(a)\big]. \tag{A26}$$

The box (A17) is invariant when the initial values lie in it. The updated coordinate is a convex combination of its current value and a target in the same interval. At either face of the box, the corresponding component of the mean Bellman residual also points weakly inward.

Mean dynamics. The mean ODE associated with (A26) is

$$\dot{Q}^j(a) = F_{N,j}(\mathbf{Q}, a) := c_j \sigma_j(Q^j, a)[\mathcal{B}_{B,N}(\mathbf{Q})^j(a) - Q^j(a)]. \tag{A27}$$

The stochastic recursion can be written as

$$\mathbf{Q}_{t+1} - \mathbf{Q}_t = \bar{\alpha}_t[\mathbf{F}_N(\mathbf{Q}_t) + M_{t+1}], \tag{A28}$$

where $\mathbb{E}[M_{t+1} \mid \mathcal{F}_t] = 0$. On (A17), let $L_Q = \Omega_B/(1-\bar{\rho})$. The martingale difference obeys

$$\|M_{t+1}\|_\infty \le 4c_{\max} L_Q, \tag{A29}$$

because both the realized and conditional update gaps lie in $[-2L_Q, 2L_Q]$.

All formal results use raw trading profits and raw exploration temperatures. Within a fixed bounded economy, define the normalized reward

$$\pi^j_{t,\mathrm{norm}} := \frac{\widetilde{\pi}^j_t + \Omega_B}{2\Omega_B},$$

and let $\widetilde{Q}^j$ follow the same max-Q recursion with $\pi^j_{t,\mathrm{norm}}$ in place of $\widetilde{\pi}^j_t$. With matched draws and transformed initial values, the normalized- and raw-reward recursions are linked by

$$Q^j = 2\Omega_B \widetilde{Q}^j - \frac{\Omega_B}{1-\rho_j}\mathbf{1}, \qquad \vartheta_j = 2\Omega_B \vartheta_j^{\mathrm{norm}}.$$

The transformation preserves every Boltzmann policy along the recursion and at a mean-field fixed point. This is a pointwise equivalence for a fixed bounded economy. Along refinements, the approximation results keep raw ϑ_j in a fixed compact positive interval. Since Ω_B grows with the action and shock bounds, the corresponding normalized temperature must satisfy $\vartheta_j^{\mathrm{norm}} = \vartheta_j/(2\Omega_B) \to 0$. A refinement that instead holds $\vartheta_j^{\mathrm{norm}}$ fixed is not covered by the uniform Gaussian-tail argument below.

Theorem 18 (Global almost-sure convergence in a fixed bounded environment)**.** *Under*

Assumptions 1 and 2, suppose

$$\chi_{B,N} := \Lambda_{B,N} + \bar{\rho} < 1. \tag{A30}$$

Then the Bellman map has a unique fixed point $\mathbf{Q}^* \in \mathcal{Q}_B$*. This point is globally exponentially stable for* (A27) *on* $\mathcal{Q}_B$*. If* $\mathbf{Q}_0 \in \mathcal{Q}_B$*, the recursion* (A26) *satisfies*

$$\mathbf{Q}_t \to \mathbf{Q}^*, \qquad m_{\Delta,B,N}(\mathbf{Q}_t) \to m_{\Delta,B,N}(\mathbf{Q}^*) \quad \textit{almost surely.}$$

Proof. Lemma 17 and Banach's theorem give a unique Bellman fixed point. Every Boltzmann probability is strictly positive, so the zeros of (A27) are exactly the Bellman fixed points.

Let $\delta = \mathbf{Q} - \mathbf{Q}^*$ and $V(\mathbf{Q}) = \|\delta\|_\infty$. On the invariant box, every softmax probability has the lower bound

$$\sigma_j(Q^j, a) \geq \frac{1}{M \exp(2L_Q/\underline{\vartheta})} =: \sigma_{\min} > 0.$$

Let $I(\mathbf{Q}) = \{i : |\delta_i| = V(\mathbf{Q})\}$ be the coordinates with the largest absolute error. For every $i = (j, a) \in I(\mathbf{Q})$ with $V > 0$, the fixed-point identity and Bellman contraction give

$$\begin{aligned}\operatorname{sgn}(\delta_i)\dot{Q}_i &= c_j \sigma_j(Q^j, a) \left\{\operatorname{sgn}(\delta_i)[\mathcal{B}_i(\mathbf{Q}) - \mathcal{B}_i(\mathbf{Q}^*)] - V\right\} \\ &\leq -c_{\min}\sigma_{\min}(1 - \chi_{B,N})V(\mathbf{Q}).\end{aligned}$$

The left side is the instantaneous change in this coordinate's absolute deviation from the fixed point. Thus every coordinate with the largest error moves back toward the fixed point at a uniform proportional rate.

The sup norm need not be differentiable when several coordinates have the largest error. For $V(\mathbf{Q}(t)) > 0$, we therefore use its upper-right Dini derivative along the mean ODE,

$$D^+V(\mathbf{Q}(t)) := \limsup_{h \downarrow 0} \frac{V(\mathbf{Q}(t+h)) - V(\mathbf{Q}(t))}{h} = \max_{i \in I(\mathbf{Q}(t))} \operatorname{sgn}(\delta_i(t))\dot{Q}_i(t).$$

If $V(\mathbf{Q}(t)) = 0$, then $\mathbf{Q}(t) = \mathbf{Q}^*$ and both sides of the desired inequality are zero. The preceding bound applies to every active coordinate, so

$$D^+V(\mathbf{Q}(t)) \leq -c_{\min}\sigma_{\min}(1 - \chi_{B,N})V(\mathbf{Q}(t)).$$

The comparison inequality yields

$$\|\mathbf{Q}(t) - \mathbf{Q}^*\|_\infty \leq e^{-c_{\min}\sigma_{\min}(1-\chi_{B,N})t}\|\mathbf{Q}(0) - \mathbf{Q}^*\|_\infty,$$

which proves global exponential stability.

The drift is Lipschitz on the compact invariant box by Lemma 17, the Lipschitz property of the maximum, and smoothness of the softmax map. This argument does not require differentiability of coordinatewise projection. Equation (A29) gives bounded martingale-difference noise, and Assumption 2 supplies the Robbins–Monro stepsizes. The standard ODE method for stochastic approximation therefore implies that the interpolated recursion tracks the mean ODE. Since the preceding strict Lyapunov bound rules out any other limiting invariant set, $\mathbf{Q}_t \to \mathbf{Q}^*$ almost surely (Benaïm, 1999); continuity of the policy mean then gives the coefficient conclusion. □

Undiscounted payoff recursion. If $\rho_j = 0$ for every investor, (A26) becomes

$$Q^j_{t+1}(a) = Q^j_t(a) + c_j \bar{\alpha}_t \mathbf{1}\{A^j_t = a\}[\widetilde{\pi}^j_t - Q^j_t(a)]. \tag{A31}$$

Fixed-point values then equal raw conditional payoffs. Equation (A19) shows that fixed-point policies and their coefficient means are unchanged. The global Bellman modulus loses $\bar{\rho}$. The local continuation-value maximum also disappears, so the strict-maximizer condition used below is unnecessary.

A.6 Atomistic continuous-action benchmark and informed REE

We next identify the continuous-action benchmark that the discretized bounded economies approximate. The auxiliary bounds and finite action grid are removed, and each AI investor becomes atomistic. The investor takes the aggregate price rule as given and chooses a complete linear demand schedule contingent on both its private signal and the market price. The aggregate coefficients still determine how market clearing loads price on the fundamental, noise-trader demand, and the demand intercept.

Against any admissible aggregate price rule, expected trading profit is a strictly concave quadratic in the three individual demand coefficients. The price-taking best response is therefore unique. A raw-profit Boltzmann policy is Gaussian around this best response: temperature changes the dispersion of schedules but not their mean. Thus the mean of every admissible continuous policy implements the same price-taking response, even when investors have heterogeneous exploration temperatures.

The informed rational-expectations equilibrium imposes consistency between this mean best response and the aggregate price rule. The consistency problem is triangular. The coefficient on the private signal determines total signal-based trading and hence price informativeness. Conditional on that informativeness, the price coefficient and intercept solve

separate linear fixed-point equations. This structure gives a unique admissible informed equilibrium and later supplies the signs and nonsingularity needed for localization and local stability.

Admissible aggregate triples. For an AI population mean $m = (a, b, c)$, define

$$\zeta(m) = \bar{\beta}_\eta^M + \phi a, \qquad \xi(m) = \bar{\beta}_p^M + \phi b, \qquad \mu_p(m) = \bar{\mu}^M + \phi c - S, \tag{A32}$$

and

$$\widehat{\tau}(m) = \tau_v + \tau_A + \zeta(m)^2 \tau_z.$$

On the domain $\xi(m) > 0$, the price-taking price is

$$p_m = \frac{\zeta(m)v + z + \mu_p(m)}{\xi(m)}. \tag{A33}$$

Continuous-action expected profit. For $r_m = (\eta, -p_m, 1)'$ and $y = (\beta_\eta, \beta_p, \mu)'$, demand is $x = y' r_m$ and expected raw profit is

$$\Pi(y; m) = y' \mathbb{E}[(v - p_m) r_m] - \frac{\gamma_A}{2} y' \mathbb{E}[r_m r_m'] y. \tag{A34}$$

Assumption 3 (Primitive regularity)**.** *The variances $\sigma_v^2, \sigma_z^2, \sigma_A^2$ and the cost parameter γ_A are positive, $\phi \geq 0$, $\bar{\beta}_\eta^M \geq 0$, and at least one of $\bar{\beta}_\eta^M$ and $\phi\tau_A$ is positive. Moreover, $\gamma_A \bar{\beta}_p^M + \phi > 0$.*

Raw-temperature Boltzmann policies. The continuous benchmark retains the raw exploration temperatures used by the learning algorithms. Strict quadratic concavity makes the Boltzmann density Gaussian. Temperature changes its dispersion but not its mean, so the mean policy implements the same price-taking best response for every admissible type.

Lemma 19 (Strict concavity and the informed best response)**.** *For every m with $\xi(m) > 0$, $\Pi(\cdot; m)$ is strictly concave. Its unique maximizer is $T(m) = (T_\eta, T_p, T_\mu)$, where*

$$T_\eta(m) = \frac{\tau_A}{\gamma_A \widehat{\tau}(m)}, \tag{A35}$$

$$T_p(m) = \frac{1 - \zeta(m)\tau_z \xi(m)/\widehat{\tau}(m)}{\gamma_A}, \tag{A36}$$

$$T_\mu(m) = \frac{\tau_v \bar{v} - \zeta(m)\tau_z \mu_p(m)}{\gamma_A \widehat{\tau}(m)}. \tag{A37}$$

For every raw temperature $\vartheta > 0$, the continuous Boltzmann density with kernel

$$\exp\{\Pi(y; m)/\vartheta\}$$

is Gaussian with mean $T(m)$ and covariance

$$\vartheta H(m)^{-1}, \qquad H(m) := \gamma_A \mathbb{E}[r_m r_m']. \tag{A38}$$

Thus its mean is independent of temperature.

Proof. The Hessian of (A34) is $-H(m)$. Because r_m contains a constant, $\mathbb{E}[r_m r_m']$ is positive definite if and only if the covariance matrix of (η, p_m) is positive definite. Direct calculation gives

$$\det \mathrm{Var}(\eta, p_m) = \frac{\zeta(m)^2 \sigma_v^2 \sigma_A^2 + (\sigma_v^2 + \sigma_A^2)\sigma_z^2}{\xi(m)^2} > 0. \tag{A39}$$

The determinant is positive, so $H(m)$ is positive definite and profit is strictly concave. The first-order condition is the normal equation $H(m)y = \mathbb{E}[(v - p_m)r_m]$. Solving it and substituting the Gaussian precisions gives (A35)–(A37). Completing the square around this unique solution yields

$$\Pi(y; m) = \Pi(T(m); m) - \frac{1}{2}[y - T(m)]' H(m)[y - T(m)], \tag{A40}$$

which proves (A38). □

Lemma 20 (Unique aggregate signal loading)**.** *Under Assumption 3, the equation*

$$\zeta = \bar{\beta}_\eta^M + \frac{\phi \tau_A}{\gamma_A(\tau_v + \tau_A + \zeta^2 \tau_z)} \tag{A41}$$

has exactly one nonnegative solution ζ^*, and $\zeta^* > 0$.

Proof. The right side of (A41) is continuous and weakly decreasing on $[0, \infty)$, whereas the left side is strictly increasing. At zero, the right side is positive under Assumption 3; for sufficiently large ζ, the left side is larger. The intermediate value theorem gives a positive solution, and the opposing monotonicities make it unique. □

Theorem 21 (Unique informed-AI REE)**.** *The fixed-point equation $m = T(m)$ has a unique admissible solution $m^{\mathrm{REE}} = (\beta_\eta^{A,*}, \beta_p^{A,*}, \mu^{A,*})$. Let $\widehat{\tau}^* = \tau_v + \tau_A + (\zeta^*)^2 \tau_z$, with ζ^* from (A41).*

Then

$$\beta_\eta^{A,*} = \frac{\tau_A}{\gamma_A \widehat{\tau}^*}, \tag{A42}$$

$$\beta_p^{A,*} = \frac{\widehat{\tau}^* - \zeta^* \tau_z \bar{\beta}_p^M}{\gamma_A \widehat{\tau}^* + \phi \zeta^* \tau_z}, \tag{A43}$$

$$\mu^{A,*} = \frac{\tau_v \bar{v} - \zeta^* \tau_z (\bar{\mu}^M - S)}{\gamma_A \widehat{\tau}^* + \phi \zeta^* \tau_z}. \tag{A44}$$

The resulting aggregate price sensitivity $\xi^* = \bar{\beta}_p^M + \phi \beta_p^{A,*}$ *is positive.*

Proof. Suppose first that $\phi > 0$. The first coordinate of $m = T(m)$ is equivalent to (A41), so Lemma 20 determines ζ^* and (A42). If $\phi = 0$, aggregate signal loading is instead fixed at $\zeta = \bar{\beta}_\eta^M$, and the first-coordinate equation gives (A42) directly. Conditional on ζ^*, the second and third fixed-point equations are separate linear equations. Solving them gives (A43) and (A44), whose common denominator is positive. Finally,

$$\xi^* = \frac{\widehat{\tau}^* (\gamma_A \bar{\beta}_p^M + \phi)}{\gamma_A \widehat{\tau}^* + \phi \zeta^* \tau_z} > 0$$

by Assumption 3. The solution is therefore admissible and unique. □

Lemma 22 (Triangular Jacobian)**.** *The Jacobian of* T *is lower triangular in the coefficient order* $(\beta_\eta, \beta_p, \mu)$*. Its diagonal is*

$$(-\lambda_\eta(m), -\lambda_T(m), -\lambda_T(m)), \quad \lambda_\eta(m) = \frac{2\phi \tau_A \zeta(m) \tau_z}{\gamma_A \widehat{\tau}(m)^2}, \quad \lambda_T(m) = \frac{\phi \zeta(m) \tau_z}{\gamma_A \widehat{\tau}(m)}. \tag{A45}$$

In particular,

$$\det[I - DT(m^{\mathrm{REE}})] = (1 + \lambda_\eta^*)(1 + \lambda_T^*)^2 > 0. \tag{A46}$$

Proof. T_η depends only on a, T_p only on (a, b), and T_μ only on (a, c). Differentiating (A35)–(A37) gives the stated diagonal. Assumption 3 and $\zeta^* > 0$ make the slopes nonnegative, proving (A46). The off-diagonal first-column derivatives are finite. Writing $\zeta = \zeta(m)$, $\xi = \xi(m)$, $\mu_p = \mu_p(m)$, and $\widehat{\tau} = \widehat{\tau}(m)$, they are

$$\partial_a T_p = -\frac{\phi \tau_z \xi}{\gamma_A} \frac{\tau_v + \tau_A - \zeta^2 \tau_z}{\widehat{\tau}^2},$$

$$\partial_a T_\mu = \frac{\phi}{\gamma_A} \frac{-\tau_z \mu_p \widehat{\tau} - 2\zeta \tau_z (\tau_v \bar{v} - \zeta \tau_z \mu_p)}{\widehat{\tau}^2}.$$

□

A.7 Approximation to the atomistic informed REE

This section turns the continuous benchmark into an approximation result for discretized bounded economies. The limits are ordered. First choose a sufficiently fine grid and sufficiently wide action and shock bounds. Then increase the number of equal-share AI investors while holding that auxiliary economy fixed. The learning-time limit is taken only after the grid, bounds, population, and algorithm types have been fixed.

We first compare the deterministic grid economy with the continuous price-taking benchmark. On a compact neighborhood of the informed equilibrium, refining the action grid and widening the action and shock bounds make the policy-implied average demand coefficients and their sensitivity to aggregate coefficients converge uniformly to the continuous best response. The deterministic grid economy then has a unique fixed point in a fixed neighborhood of the informed REE. Uniformity requires raw temperatures to remain between common positive lower and upper bounds. An implementation that normalizes rewards must therefore adjust its normalized temperature when the payoff bound changes.

We then account for the finite population. Random arm choices and average private-signal errors each generate errors of order $N^{-1/2}$, while one investor's effect on the market price is of order N^{-1}. Hence the policy-implied demand coefficients at every finite-population mean-field fixed point nearly satisfy the deterministic grid equilibrium conditions. Those conditions determine signal loading first, followed by the price coefficient and intercept, and the benchmark response points back toward the admissible region at its boundary. A small equilibrium error must therefore place all three coefficients close to the informed REE, where projection does not bind. This conclusion concerns the policy-implied mean, not the realized cross-section of schedules, and does not require a global Bellman contraction.

Compact neighborhood of the benchmark fixed point. Construct the localization rectangle in the same order as the triangular best response. First choose a small interval I_η around $\beta_\eta^{A,*}$ on which $\zeta > 0$. The scalar residual $a - T_\eta(a)$ is strictly increasing and has opposite signs at the endpoints, so its unique zero lies inside. Conditional on any $a \in I_\eta$, each of the other two fixed-point equations also has a unique solution that varies continuously with a: its residual has derivative $1 + \lambda_T(a) > 0$ in its own coordinate. Choose intervals I_p and I_μ that contain these solutions in their interiors, while maintaining $\bar{\beta}_p^M + \phi b > 0$, $\zeta > 0$, and $\xi > 0$. The resulting rectangle $\mathcal{K}^+ = I_\eta \times I_p \times I_\mu$ contains m^{REE} in its interior, and the mean drift $T(m) - m$ points back into the rectangle at each boundary. Set $\mathcal{K}_{\mathrm{proj}} = \mathcal{K}^+$ in every sufficiently accurate auxiliary economy. Projection is then inactive near the REE and only disciplines far-off-equilibrium exploratory profiles.

Atomistic deterministic grid map. Let B collect the action and shock truncation bounds. For $m \in \mathcal{K}^+$, let $\Pi_B(y; m)$ denote the expectation of (A10) when a price-taking investor uses y, aggregate coefficients are m, the term w_N is set to zero, and shocks have the truncated laws. Define

$$\omega_{j,\Delta,B}(a \mid m) := \frac{\exp\{\Pi_B(y(a); m)/\vartheta_j\}}{\sum_{q=1}^{M} \exp\{\Pi_B(y(q); m)/\vartheta_j\}}, \tag{A47}$$

$$\mathcal{T}_{\Delta,B,N}(m) := \frac{1}{N} \sum_{j=1}^{N} \sum_{a=1}^{M} \omega_{j,\Delta,B}(a \mid m) y(a). \tag{A48}$$

Thus $\omega_{j,\Delta,B}(a \mid m)$ is investor j's Boltzmann probability of arm a, and $\mathcal{T}_{\Delta,B,N}(m)$ is the corresponding population-average coefficient vector. For arguments outside $\mathcal{K}^+$, the projected version is $\mathcal{T}^{\mathrm{proj}}_{\Delta,B,N}(m) := \mathcal{T}_{\Delta,B,N}(\mathrm{proj}_{\mathcal{K}^+} m)$. The notation suppresses the dependence of $\mathcal{T}_{\Delta,B,N}$ and its fixed point below on the temperature profile.

Define the deterministic best-response error

$$\delta_{\mathrm{BR}}(\Delta, B) := \sup_{N \geq 1} \sup_{\boldsymbol{\vartheta}} \|\mathcal{T}_{\Delta,B,N} - T\|_{C^1(\mathcal{K}^+)}, \tag{A49}$$

where the temperature supremum is over $[\underline{\vartheta}, \bar{\vartheta}]^N$ and the supremum also ranges over the allowed grid translations. Let

$$\mathcal{P}_* = [I_3 - DT(m^{\mathrm{REE}})]^{-1}, \qquad \chi_* := \|\mathcal{P}_*\|_\infty.$$

Choose $r_0 > 0$ so that $K_0 := \overline{B}_\infty(m^{\mathrm{REE}}, r_0) \subset \operatorname{int} \mathcal{K}^+$ and so that (A55) below holds. Call an auxiliary economy *best-response accurate* if

$$\chi_* \delta_{\mathrm{BR}}(\Delta, B) \leq \min\left\{\frac{1}{4}, \frac{r_0}{2}\right\}. \tag{A50}$$

Proposition 23 (Deterministic grid approximation near the informed REE)**.** *Let the action rectangles expand to $\mathbb{R}^3$, let $\delta_\eta, \delta_p, \delta_\mu \to 0$, and let all shock bounds tend to infinity. Uniformly over $m \in \mathcal{K}^+$, population sizes, admissible temperature profiles, and allowed grid translations,*

$$\|\mathcal{T}_{\Delta,B,N} - T\|_{C^1(\mathcal{K}^+)} \longrightarrow 0. \tag{A51}$$

In every best-response accurate auxiliary economy, the equation

$$m = \mathcal{T}_{\Delta,B,N}(m) \tag{A52}$$

has a unique solution $m^0_{\Delta,B,N}$ *in* K_0*. Uniformly over populations and admissible types,*

$$\|m^0_{\Delta,B,N} - m^{\mathrm{REE}}\|_\infty \le 2\chi_* \delta_{\mathrm{BR}}(\Delta, B) \longrightarrow 0. \tag{A53}$$

Proof. Strict concavity is uniform on the compact set $\mathcal{K}^+$. The same remains true once the shock bounds are sufficiently wide. Hence there are constants $h > 0$ and $C < \infty$, independent of all sufficiently large shock bounds, such that

$$\Pi_B(y; m) \le C - h\|y\|^2, \qquad \|D_m \Pi_B(y; m)\| \le C(1 + \|y\|^2) \tag{A54}$$

for every $m \in \mathcal{K}^+$ and $y \in \mathbb{R}^3$. Thus expected profit falls at least quadratically as the demand coefficients grow, while its derivative in the aggregate coefficients grows at most quadratically. Because raw temperatures remain in a fixed positive range, the Boltzmann weights and their derivatives have uniformly negligible tails.

On a bounded action rectangle, the grid sums and their derivatives converge uniformly to the corresponding Gaussian integrals in the aggregate coefficients, temperature, and grid translation. The quadratic bound makes the omitted tails uniformly small as the rectangle expands. Lemma 19 identifies the limiting mean and derivative as $T(m)$ and $DT(m)$. Averaging over investors proves (A51).

It remains to locate the deterministic grid fixed point. Consider the Newton iteration using the derivative at the benchmark,

$$\mathcal{N}_{\Delta,B,N}(m) = m - \mathcal{P}_*[m - \mathcal{T}_{\Delta,B,N}(m)].$$

By the choice of r_0,

$$q_0 := \sup_{m \in K_0} \|I_3 - \mathcal{P}_*[I_3 - DT(m)]\|_\infty \le \frac{1}{4}. \tag{A55}$$

Equations (A55)–(A50) imply

$$\sup_{m \in K_0} \|D\mathcal{N}_{\Delta,B,N}(m)\|_\infty \le \frac{1}{2}, \qquad \|\mathcal{N}_{\Delta,B,N}(m^{\mathrm{REE}}) - m^{\mathrm{REE}}\|_\infty \le \frac{r_0}{2}.$$

The iteration therefore maps K_0 into itself and contracts distances. The contraction-mapping theorem gives the unique solution and (A53). The same calculation gives the lower bound

$$\|[m - \mathcal{T}_{\Delta,B,N}(m)] - [m' - \mathcal{T}_{\Delta,B,N}(m')]\|_\infty \ge \frac{1}{2\chi_*}\|m - m'\|_\infty \tag{A56}$$

for $m, m' \in K_0$. □

The next lemma records the stronger, pointwise approximation needed only for the local-stability argument.

Lemma 24 (Local Gaussian approximation of grid policies)**.** *Let $m^0_{\Delta,B,N}$ be the fixed point in Proposition 23, and put*

$$\omega^0_j(a) := \omega_{j,\Delta,B}(a \mid m^0_{\Delta,B,N}).$$

Let f_ϑ denote the $\mathcal{N}(m^{\mathrm{REE}}, \vartheta H_*^{-1})$ density, where $H_* = H(m^{\mathrm{REE}})$. Along the refinements in Proposition 23,

$$\sup_{N,\boldsymbol{\vartheta},j} \max_{1\le a\le M} \left| \frac{\omega^0_j(a)}{h_\Delta} - f_{\vartheta_j}(y(a)) \right| \longrightarrow 0. \tag{A57}$$

The supremum is over $N \ge 1$, admissible temperature profiles, investors $j \le N$, and allowed grid translations. There are constants $C_p, c_p, C_G > 0$, uniform across N, admissible types, and all sufficiently accurate environments, such that

$$\frac{\omega^0_j(a)}{h_\Delta} \le C_p e^{-c_p \|y(a)\|^2}, \tag{A58}$$

$$\|D_m \Pi_B(y(a); m^0_{\Delta,B,N})\| \le C_G(1 + \|y(a)\|^2). \tag{A59}$$

Along the same refinements, the truncated payoff derivative converges uniformly to the untruncated derivative evaluated at the REE, on grid arms within any fixed distance of m^{REE}. The convergence is uniform over populations, admissible temperature profiles, investors, and allowed grid translations. Uniformly over these same objects, each policy mean converges to m^{REE}, and the maximum distance between a maximizing grid arm and m^{REE} tends to zero. The same integrable Gaussian upper bound remains valid after multiplication by any fixed polynomial in $\|y(a)\|$.

Proof. On every bounded action set, the grid-sum argument in Proposition 23 gives convergence of the probabilities and payoff derivatives. The fixed-point bound in the same proposition also gives $m^0_{\Delta,B,N} \to m^{\mathrm{REE}}$ uniformly. The cell-weighted normalizing sum converges uniformly to a positive Gaussian integral, so it stays uniformly bounded away from zero. The quadratic payoff bound (A54) makes the probability outside a sufficiently large rectangle uniformly small. The bounded-set limits therefore extend to the expanding grids and imply (A57), the stated upper bounds, and convergence of the policy means.

For the maximizing arms, $\Pi_B(\cdot; m^0_{\Delta,B,N})$ is a strictly concave quadratic whose continuous maximizer approaches m^{REE} and whose curvature remains bounded away from zero. Compare a grid maximizer with the grid point nearest the continuous maximizer. Uniform upper and lower curvature bounds imply that the grid maximizer lies within a constant times the

mesh of the continuous maximizer. Hence every maximizing grid arm converges to m^{REE}. All bounds are uniform over the compact type ranges. □

Theorem 25 (Mean-field fixed-point approximation of the informed REE)**.** *Take* $\mathcal{K}_{\mathrm{proj}} = \mathcal{K}^+$*. For every* $\varepsilon > 0$*, sufficiently accurate grids and truncations followed by sufficiently large* N *satisfy*

$$\|m_{\Delta,B,N}(\mathbf{Q}^*) - m^{\mathrm{REE}}\|_\infty < \varepsilon \tag{A60}$$

whenever $\mathbf{Q}^*$ *is a mean-field fixed point and the algorithm-type profile is admissible. No Bellman contraction is required.*

Proof. Fix the auxiliary economy and use the same fundamental, noise-trader demand, and signal errors in the finite and price-taking economies. Conditional on $\mathbf{Q}$, the empirical mean of the other investors' bounded arm choices differs from its policy mean by $O(N^{-1/2})$ in expected magnitude. The average private-signal error also has expected magnitude $O(N^{-1/2})$. Forcing investor j to a counterfactual arm changes the coefficient mean and the investor's own signal-error term by $O(N^{-1})$ in expected magnitude. On the bounded admissible set, these changes alter the projected clearing price and profit by at most a constant multiple of their size. Thus, uniformly over investors, arms, and value profiles,

$$\left|\bar{\Pi}^j_{B,N}(\mathbf{Q}, a) - \Pi_B(y(a); \mathrm{proj}_{\mathcal{K}^+} m_{\Delta,B,N}(\mathbf{Q}))\right| \le C_B(N^{-1/2} + N^{-1}). \tag{A61}$$

At a mean-field fixed point, Lemma 15 removes the arm-independent continuation value from each Boltzmann policy. The Boltzmann probabilities change by at most a constant multiple of the payoff error, so (A61) implies

$$\left\| m_{\Delta,B,N}(\mathbf{Q}^*) - \mathcal{T}^{\mathrm{proj}}_{\Delta,B,N}(m_{\Delta,B,N}(\mathbf{Q}^*)) \right\|_\infty \le C_B(N^{-1/2} + N^{-1}). \tag{A62}$$

The constants are uniform over temperatures in their compact range.

We next convert this small fixed-point error into distance from the REE. Let $q = \mathrm{proj}_{\mathcal{K}^+} m$. On a boundary face, $m_i - q_i$ and $q_i - T_i(q)$ have the same outward sign because the benchmark response points inward; in the interior, $m_i = q_i$. Thus each term is bounded by $|m_i - T_i(q)|$, so projection cannot conceal a large scalar residual. The first residual $q_1 - T_\eta(q_1)$ is strictly increasing, has derivative $1 + \lambda_\eta \ge 1$, and has the unique zero $\beta^{A,*}_\eta$. The inward-pointing construction of $\mathcal{K}^+$ implies that a residual of size δ bounds both $|q_1 - \beta^{A,*}_\eta|$ and $|m_1 - q_1|$ by a constant times δ. Conditional on q_1, the second and third residuals are also strictly increasing in their own coordinates, with derivatives $1 + \lambda_T \ge 1$. Their cross effects from the first coordinate are bounded on $\mathcal{K}^+$. Thus there are constants

$\delta_0 > 0$ and $C_{\text{loc}} < \infty$ such that

$$\|m - T(\text{proj}_{\mathcal{K}^+} m)\|_\infty \le \delta \le \delta_0 \quad \Longrightarrow \quad \|m - m^{\text{REE}}\|_\infty \le C_{\text{loc}}\delta. \tag{A63}$$

This coordinate-by-coordinate argument also shows that the projection is inactive once the residual is small enough.

Proposition 23 makes the projected grid map uniformly close to $T \circ \text{proj}_{\mathcal{K}^+}$. First choose the grid and bounds so that this deterministic error is below both $\delta_0/2$ and $\varepsilon/(2C_{\text{loc}})$. Holding that auxiliary economy fixed, choose N so that the finite-population residual in (A62) satisfies the same two bounds. The localization inequality then gives (A60). This order of choices is the ordered limit in the theorem. □

A.8 Uniqueness and local stability

The preceding result places the policy-implied average demand coefficients of every finite-population fixed point near the informed REE. It does not rule out several Q-value profiles that generate similar averages, nor does proximity alone imply that small perturbations die out. This section establishes uniqueness of the fixed-point Q-value profile and local stability of the expected learning dynamics without relying on the global contraction condition.

In the reference grid economy, the value system has one coordinate for every investor–arm pair, but its feedback to prices and expected profits operates through only three population-average demand coefficients. Solving each investor's response in terms of a change in these three aggregate coefficients reduces local stability to three dimensions. In the continuous benchmark, Gaussian symmetry aligns this aggregate adjustment with the three-coefficient best-response map. Its sequential structure—signal loading first, followed by the price coefficient and intercept—rules out locally self-reinforcing price feedback. Sufficiently accurate grids inherit this property, and sufficiently large finite populations inherit it because their expected learning dynamics and marginal responses are close to those of the reference grid economy.

We first prove uniqueness in the undiscounted economy. The local response rules out multiple nearby fixed points, while the localization result places every fixed point in that neighborhood. At a fixed point, discounting adds the same continuation constant to all arm values of a given investor and therefore leaves Boltzmann probabilities unchanged. This shift gives a one-to-one correspondence between undiscounted and discounted fixed points, so uniqueness requires no continuation-arm gap. No gap is needed for undiscounted local stability either. With positive discounting, local stability additionally requires the best continuation arm to remain unchanged near the fixed point; the rate condition $\kappa_N/g_N \to 0$

makes the finite-population approximation error small relative to the arm gap.

Fix an accurate grid and truncation level. Let $m^0 = m^0_{\Delta,B,N}$ be the deterministic grid fixed point from Proposition 23, and set $R^0_a(m) = \Pi_B(y(a); m)$. The reference Bellman drift is

$$F^0_{j,a}(\mathbf{Q}) = c_j \sigma_j(Q^j, a) \left[R^0_a(m_{\Delta,B,N}(\mathbf{Q})) + \rho_j \max_q Q^j(q) - Q^j(a) \right].$$

At m^0, let $R^0 = (R^0_a(m^0))_{a=1}^M$ and

$$Q^{j,0}(a) = R^0_a + \frac{\rho_j}{1-\rho_j} \max_q R^0_q.$$

Then $\mathbf{Q}^0$ is a zero of the reference drift and has policy mean m^0.

Restrict attention to accurate action grids with at least two arms. Let g_N be the payoff difference between the best and second-best coordinates of R^0. When $g_N > 0$, write $a^\dagger_N$ for the unique best arm, and let

$$\mathcal{D} := \left\{ N : \max_{1 \le j \le N} \rho_{j,N} > 0 \right\}$$

contain the population sizes with at least one positively discounted investor.

Assumption 4 (Strict continuation arm along the population sequence)**.** *If $\mathcal{D}$ is infinite, $g_N > 0$ for every sufficiently large $N \in \mathcal{D}$. This requirement is vacuous when $\mathcal{D}$ is finite. No maximizing-arm condition is imposed at population sizes outside $\mathcal{D}$.*

The maximizing arm may vary with N. The gap is not needed for uniqueness; it is used only to make the discounted adjustment dynamics differentiable near the fixed point. The required rate is stated in Theorem 28.

Reference aggregate response. Three objects summarize how a value perturbation affects aggregate demand. The diagonal matrix D_j records investor j's arm-specific adjustment speeds, G maps a change in aggregate coefficients into arm payoffs, and B_j maps a change in values into that investor's mean demand coefficients. Formally, at the reference fixed point, put

$$D_j = c_j \operatorname{diag}(\omega^0_j), \qquad \omega^0_j(a) = \omega_{j,\Delta,B}(a \mid m^0),$$
$$G'_a = D_m R^0_a(m^0), \qquad B_j h = \frac{1}{\vartheta_j} \sum_a \omega^0_j(a)[y(a) - \bar{y}^0_j] h(a),$$

where $\bar{y}_j^0 = \sum_a \omega_j^0(a) y(a)$. Let $\ell h = h(a_\ell)$ select a maximizing arm of R^0; use $a_\ell = a_N^\dagger$ when the discounted maximizer is unique. The linearization of the reference drift is

$$(J^0 h)^j = D_j \left[G \left\{ \frac{1}{N} \sum_{q=1}^{N} B_q h^q \right\} + \rho_j \mathbf{1} \ell h^j - h^j \right].$$

Use $\|h\|_{\infty,2} = \max_j \|h^j\|_2$ and its induced operator norm. The matrix A_j adds continuation-value feedback to investor j's adjustment, $R_j(s)$ gives that investor's response to a perturbation growing at rate s, and $\mathcal{K}^\ell_{\Delta,B,N}(s)$ averages the resulting three-dimensional demand response:

$$A_j = D_j(I - \rho_j \mathbf{1}\ell), \qquad R_j(s) = (sI + A_j)^{-1},$$
$$\mathcal{K}^\ell_{\Delta,B,N}(s) = \frac{1}{N} \sum_{j=1}^{N} B_j R_j(s) D_j G.$$

Let f_ϑ be the $\mathcal{N}(m^{\mathrm{REE}}, \vartheta H_*^{-1})$ density, where $H_* = H(m^{\mathrm{REE}})$. The scalar Gaussian response is

$$\varpi_N(z) := \frac{1}{N} \sum_{j=1}^{N} \int_{\mathbb{R}^3} \frac{c_j f_{\vartheta_j}(y)}{z + c_j f_{\vartheta_j}(y)} \left(\frac{[H_*^{1/2}(y - m^{\mathrm{REE}})]_1}{\sqrt{\vartheta_j}} \right)^2 f_{\vartheta_j}(y)\, dy.$$

The choice of the first transformed coordinate is immaterial by Gaussian symmetry. Define $\delta_{\mathrm{ch}}(\Delta, B)$ as the supremum, over all population sizes, admissible type profiles, and selectors ℓ associated with maximizing arms of R^0, of the response error

$$\sup_{\mathrm{Re}\, z \geq 0} \left\| \mathcal{K}^\ell_{\Delta,B,N}(h_\Delta z) - \varpi_N(z) DT(m^{\mathrm{REE}}) \right\|_2.$$

Set $\gamma_* = [1 + \|DT(m^{\mathrm{REE}})\|_2]^{-2}$. Call an auxiliary economy *stability-accurate* when it is best-response accurate, $\mathcal{K}_{\mathrm{proj}} = \mathcal{K}^+$, and

$$\delta_{\mathrm{BR}}(\Delta, B) \leq \min\left\{ \frac{\delta_0}{2}, \frac{r_0}{4C_{\mathrm{loc}}} \right\}, \qquad \delta_{\mathrm{ch}}(\Delta, B) \leq \frac{\gamma_*}{4}. \tag{A64}$$

These are grid-and-truncation accuracy requirements, not additional restrictions on the economic primitives.

Proposition 26 (Stability of the reference grid dynamics)**.** *Along the refinements in Proposition 23, $\delta_{\mathrm{ch}}(\Delta, B) \to 0$. For every stability-accurate auxiliary economy and admissible population sequence, consider a population that is either undiscounted or sufficiently large*

along the discounted subsequence covered by Assumption 4. Then every eigenvalue of J^0 has strictly negative real part—that is, J^0 is Hurwitz—and $\mathbf{Q}^0$ is locally exponentially stable under the reference mean dynamics. After the auxiliary economy is fixed, there are $\omega_0, M_0 > 0$, independent of N, admissible types, and the admitted maximizing-arm selector, such that the inverse response matrix satisfies

$$\|(sI_{NM} - J^0)^{-1}\|_{\infty,2} \le \frac{M_0}{1+|s|} \qquad (\operatorname{Re} s \ge -\omega_0). \tag{A65}$$

Proof. The matrix A_j has positive diagonal entries, nonpositive off-diagonal entries, and is strictly row-diagonally dominant because $\rho_j < 1$; the dominance margin is at least $c_j \min_a \omega_j^0(a)(1-\bar{\rho})$. Its eigenvalues therefore lie in the open right half-plane, so $R_j(s)$ is well defined for $\operatorname{Re} s \ge 0$. If $u = N^{-1}\sum_j B_j h^j$, eliminating the investor-specific coordinates from $(sI - J^0)h = f$ gives

$$h^j = R_j(s)[D_j G u + f^j], \qquad [I_3 - \mathcal{K}^{\ell}_{\Delta,B,N}(s)]u = \frac{1}{N}\sum_j B_j R_j(s) f^j.$$

This is the low-dimensional step: stability of the NM-dimensional system depends on a three-dimensional aggregate response.

Gaussian symmetry identifies its continuous limit. Differentiating (A40) at the REE leaves one term that is linear in $y - m^{\mathrm{REE}}$ and two terms that are even. After multiplication by $y - m^{\mathrm{REE}}$, symmetry makes the even terms integrate to zero. The remaining term gives $\varpi_N(z)DT(m^{\mathrm{REE}})$. Uniformly over population sizes and admissible types,

$$\operatorname{Re}\varpi_N(z) \ge 0, \qquad |\varpi_N(z)| \le 1, \qquad |\varpi_N(z)| \le \frac{C_\varpi}{1+|z|} \quad (\operatorname{Re} z \ge 0).$$

The triangular Jacobian in Lemma 22 therefore gives

$$\det[I_3 - \varpi_N(z)DT(m^{\mathrm{REE}})] = [1+\lambda_\eta^*\varpi_N(z)][1+\lambda_T^*\varpi_N(z)]^2, \tag{A66}$$

whose factors cannot vanish on the closed right half-plane. More precisely, the smallest singular value of this matrix is at least γ_*. To see this, each factor in (A66) has modulus at least one because $\operatorname{Re}\varpi_N(z) \ge 0$. The two largest singular values are at most $1 + \|DT(m^{\mathrm{REE}})\|_2$, so the identity expressing the determinant as the product of the three singular values gives the bound $[1 + \|DT(m^{\mathrm{REE}})\|_2]^{-2} = \gamma_*$ for the smallest one.

Lemma 24 makes the undiscounted grid response converge uniformly to this Gaussian integral. Its integrable upper bound controls the action tails, while the last bound on ϖ_N

controls high frequencies. Discounting adds a one-dimensional continuation-value adjustment. To bound it, write $L_j = h_\Delta z I + D_j$ and separate this adjustment from $L_j^{-1} D_j$. The modulus of the resulting denominator is at least $1 - \bar{\rho}$. At bounded frequencies, Gaussian symmetry makes the aggregate factor $B_j L_j^{-1} D_j \mathbf{1}$ vanish in the limit; at high frequencies it is $O(|z|^{-1})$. The other factor stays bounded because the maximizing grid arms converge to m^{REE}. Thus the discounted adjustment also vanishes uniformly, and $\delta_{\mathrm{ch}}(\Delta, B) \to 0$.

Stability accuracy transfers the lower singular-value bound to $I_3 - \mathcal{K}^{\ell}_{\Delta,B,N}(s)$. Once the auxiliary economy is fixed, compact type ranges and positive grid probabilities uniformly bound every investor-level inverse on a fixed strip to the left of the imaginary axis. Differentiating the aggregate response with respect to s then gives a uniform bound L_K over populations, types, and selectors. Moving a short distance left from the imaginary axis therefore preserves the aggregate margin. Substitution into the two block equations above gives (A65); no factor of N appears because the population weights sum to one. Thus J^0 is Hurwitz. The strict arm gap makes the discounted reference drift differentiable near $\mathbf{Q}^0$; the undiscounted drift is already smooth. Standard linearization then gives local exponential stability. □

Lemma 27 (Finite-population C^1 drift approximation)**.** *Fix a stability-accurate auxiliary economy and an admissible population sequence. There are $\bar{r}_Q, C_B, c_B > 0$, independent of N and the type profile, such that, with*

$$\kappa_N := N^{-1/2} + N^{-1} + e^{-c_B N}, \tag{A67}$$

and

$$\mathcal{V}_N := \{\mathbf{Q} : \|\mathbf{Q} - \mathbf{Q}^0\|_{\infty,2} \le \bar{r}_Q\},$$

all sufficiently large N satisfy

$$\sup_{\mathbf{Q}\in\mathcal{V}_N} \|\mathbf{F}_N(\mathbf{Q}) - \mathbf{F}_N^0(\mathbf{Q})\|_{\infty,2} + \sup_{\mathbf{Q}\in\mathcal{V}_N} \|D(\mathbf{F}_N - \mathbf{F}_N^0)(\mathbf{Q})\|_{\infty,2\to\infty,2} \le C_B \kappa_N. \tag{A68}$$

At points where the two drifts are separately differentiable, the derivative in the second supremum equals $D\mathbf{F}_N - D\mathbf{F}_N^0$. No continuation-arm gap is required for this comparison. The three terms in κ_N are, respectively, the sampling and signal-averaging error, individual price impact, and the probability of reaching the projection boundary.

Proof. Choose $\bar{r}_Q$ so that all policy means in $\mathcal{V}_N$ remain a fixed positive distance from the boundary of $\mathcal{K}^+$. To handle rare arm profiles outside this region, compare the projection with a smooth coordinatewise saturation that is C^2, has uniformly bounded first and second

derivatives, maps into a compact subset of int $\mathcal{K}^+$, and agrees with the identity on the buffered set around these policy means. The comparison does not alter payoffs on the interior region. Hoeffding's inequality bounds the probability that an arm counterfactual or the segment used in a derivative comparison leaves this region by $C_B e^{-c_B N}$.

Write the payoff error as $E_N^{j,a}(\mathbf{Q}) := \bar{\Pi}_{B,N}^j(\mathbf{Q}, a) - R_a^0(m_{\Delta,B,N}(\mathbf{Q}))$. Equation (A61) gives its level bound. For its derivative with respect to investor $\ell \neq j$, condition on all investors except j and ℓ, force j to arm a, and compare two arms of investor ℓ around the common coefficient and signal-error average that excludes both investors. Centering removes the first-order private-signal term. The random base contributes $O(N^{-3/2})$; individual forcing and the second-order remainder contribute $O(N^{-2})$; the rare projection event contributes $O(N^{-1}e^{-c_B N})$. Conditional expected payoff is linear in each other investor's policy. Therefore, along any zero-sum change in investor ℓ's policy, the leading derivative is $N^{-1}D_m R_a^0$ times the induced change in that investor's mean coefficients. Its remainder is bounded by the size of the policy change times $C_B\{N^{-3/2} + N^{-2} + N^{-1}e^{-c_B N}\}$. For $\ell = j$, the true conditional payoff has zero own-policy derivative, while the reference derivative is $O(N^{-1})$. The derivative of the Boltzmann probabilities is uniformly bounded. Applying it to these payoff bounds and summing the N Jacobian blocks gives the rate $C_B\kappa_N$.

Finally, the common continuation term cancels before differentiation:

$$F_{N,j,a}(\mathbf{Q}) - F_{j,a}^0(\mathbf{Q}) = c_j\sigma_j(Q^j, a)E_N^{j,a}(\mathbf{Q}).$$

The drift difference is therefore smooth even when the two discounted drifts are separately nonsmooth, and its derivative agrees with their derivative difference wherever both exist. Differentiating the product adds the derivative of the Boltzmann probability times the level error and the probability times the payoff-error derivative; the preceding bounds control both. This proves (A68) without using the arm gap. □

Theorem 28 (Unique locally stable finite-population mean-field fixed point)**.** *Fix a stability-accurate auxiliary economy and an admissible population sequence. If $\mathcal{D}$ is infinite, maintain Assumption 4 and suppose*

$$\lim_{\substack{N\to\infty \\ N\in\mathcal{D}}} \frac{\kappa_N}{g_N} = 0. \tag{A69}$$

If $\mathcal{D}$ is finite, this condition is vacuous. No gap condition is imposed at population sizes outside $\mathcal{D}$. For all sufficiently large N, the finite-population Bellman map has exactly one mean-field fixed point $\mathbf{Q}^*$, and

$$\|\mathbf{Q}^* - \mathbf{Q}^0\|_{\infty,2} = O(\kappa_N), \qquad \|m_{\Delta,B,N}(\mathbf{Q}^*) - m^0\|_\infty = O(\kappa_N). \tag{A70}$$

The finite-population mean ODE is locally exponentially stable at $\mathbf{Q}^*$.

Proof. First set every discount factor to zero, holding learning speeds and temperatures fixed. Let $\mathbf{F}_N^u$ and $\mathbf{F}_N^{0,u}$ be the actual and reference drifts, and let $\mathbf{Q}^{0,u}$ have every block equal to R^0. The reference inverse bound in Proposition 26, the derivative estimate in Lemma 27, and a Newton iteration using $D\mathbf{F}_N^{0,u}(\mathbf{Q}^{0,u})$ give a unique zero. In particular, on the fixed grid the reference derivative is uniformly Lipschitz on a fixed neighborhood; after shrinking that neighborhood, the inverse bound and (A68) make the Newton iteration a contraction. Its zero $\mathbf{Q}^{u,*}$ satisfies $\|\mathbf{Q}^{u,*} - \mathbf{Q}^{0,u}\|_{\infty,2} = O(\kappa_N)$.

This local zero is the only undiscounted fixed point. Indeed, let $\bar{\mathbf{Q}}^u$ be any other undiscounted fixed point and set $\bar{m} = m_{\Delta,B,N}(\bar{\mathbf{Q}}^u)$. Its projected best-response residual is bounded by $\delta_{\mathrm{BR}} + C_B(N^{-1/2} + N^{-1})$. The localization bound (A63) places its coefficient mean in K_0. Projection is then inactive, and (A56) together with (A61) yields

$$\|\bar{m} - m^0\|_\infty = O(\kappa_N), \qquad \|\bar{\mathbf{Q}}^u - \mathbf{Q}^{0,u}\|_{\infty,2} = O(\kappa_N).$$

Every undiscounted fixed point therefore lies in the neighborhood in which the Newton iteration has a unique zero.

Now restore the original discount factors. The blockwise shift

$$\mathcal{L}_{\boldsymbol{\rho}}(q)^j := q^j + \frac{\rho_j}{1-\rho_j} \max_a q^j(a)\, \mathbf{1}$$

preserves softmax policies and, by Lemma 15, bijects the undiscounted and discounted fixed-point sets. It is Lipschitz with constant $\sqrt{M}/(1-\bar{\rho})$. Discounted uniqueness and both rates in (A70) follow. The arm-gap assumption has not been used for uniqueness.

For local stability, let $J_N^0 = D\mathbf{F}_N^0(\mathbf{Q}^0)$. In discounted populations, the rate (A69) makes $\|\mathbf{Q}^* - \mathbf{Q}^0\|_{\infty,2} = o(g_N)$. The same arm therefore remains optimal throughout a neighborhood joining the two points; in an undiscounted population the continuation maximum is absent. On this common differentiability neighborhood, the fixed-grid reference derivative changes by at most a constant times the change in $\mathbf{Q}$, uniformly over admitted populations and types. Lemma 27 and (A70) therefore give the single perturbation bound

$$\|D\mathbf{F}_N(\mathbf{Q}^*) - J_N^0\|_{\infty,2\to\infty,2} = O(\kappa_N).$$

For every s in the closed right half-plane, factor $sI - D\mathbf{F}_N(\mathbf{Q}^*)$ into $sI - J_N^0$ and a second factor equal to the identity minus a uniformly small matrix. The inverse bound in Proposition 26 and the Neumann-series argument make both factors invertible for large N. The actual Jacobian therefore has no eigenvalue with nonnegative real part, and standard C^1

linearization gives local exponential stability. □

Order of limits. The argument first fixes a stability-accurate auxiliary economy, then takes the population sufficiently large, and finally lets learning time diverge with both fixed. A simultaneous refinement would additionally require the approximation and finite-population errors to vanish relative to the continuation-arm gap and the local stability margin; no such result is claimed here.

A.9 Stochastic lock-in

Fix a stability-accurate finite economy and a sufficiently large population covered by Theorem 28. All constants in this subsection may depend on its grid, truncation bounds, population, and type profile. The expected learning dynamics return to the unique mean-field fixed point after a small perturbation. Actual learning nevertheless continues to receive shocks from realized schedules, private-signal errors, fundamentals, and noise-trader demand. Local stability of the expected dynamics therefore does not guarantee that every stochastic path remains nearby.

The lock-in proof compares this local pull toward the fixed point with the learning shocks that remain after entry. It uses an inner and an outer neighborhood and stops the comparison at the first exit from the outer one. This step avoids conditioning on the future event that no exit occurs. Because the sum of future squared step sizes becomes arbitrarily small, so does the probability that later shocks force an exit. On paths that remain inside, declining gains and local mean reversion imply convergence.

The result is deliberately conditional. It applies on the event that the learning process lies in the inner neighborhood at a deterministic date or at an almost-surely finite stopping time; it does not establish entrance from an arbitrary initial value. On successful lock-in paths, the Q-value profile converges to the unique fixed point, and the population-average coefficients implied by the mixed policies converge as a consequence. Positive exploration continues, so realized schedules remain random and need not settle down. The conditional probability of remaining in the local basin and converging approaches one as the sum of future squared step sizes becomes small.

Theorem 29 (Local lock-in for max-Q learning)**.** *Maintain Assumptions 1 and 2, and let* $\mathbf{Q}^*$ *be the unique mean-field fixed point from Theorem 28. There are neighborhoods* $\mathcal{U}^{\text{in}}$ *and* $\mathcal{U}^{\text{out}}$ *of* $\mathbf{Q}^*$*, with the closure of the inner neighborhood contained in the outer one, and a*

finite C_{lock} such that, for every deterministic date b,

$$\Pr\left(\mathbf{Q}_t \in \mathcal{U}^{\text{out}} \text{ for all } t \geq b, \quad \mathbf{Q}_t \to \mathbf{Q}^* \mid \mathcal{F}_b\right) \geq 1 - C_{\text{lock}} \sum_{t \geq b} \bar{\alpha}_t^2 \quad \text{on } \{\mathbf{Q}_b \in \mathcal{U}^{\text{in}}\}. \tag{A71}$$

Whenever this event occurs, continuity of the Boltzmann policy also gives

$$m_{\Delta,B,N}(\mathbf{Q}_t) \to m_{\Delta,B,N}(\mathbf{Q}^*). \tag{A72}$$

Proof. By Theorem 28, the mean drift is continuously differentiable near $\mathbf{Q}^*$ and its Jacobian $J = D\mathbf{F}_N(\mathbf{Q}^*)$ has eigenvalues with negative real parts. The standard Lyapunov theorem therefore supplies a positive-definite quadratic function $V(x) = x'Px$. With $x_t = \mathbf{Q}_t - \mathbf{Q}^*$, choose $0 < r_{\text{in}} \leq r_{\text{out}}/4$ so that the closed ellipsoid $\{\mathbf{Q} : V(\mathbf{Q} - \mathbf{Q}^*) \leq r_{\text{out}}\}$ lies inside this differentiability neighborhood. In a discounted economy, choose it also to preserve the same maximizing continuation arm. For some $c_0 > 0$, the mean drift then satisfies

$$2x'P\mathbf{F}_N(\mathbf{Q}^* + x) \leq -c_0\|x\|^2 \quad \text{whenever } V(x) \leq r_{\text{out}}. \tag{A73}$$

Set $\mathcal{U}^{\text{out}} = \{\mathbf{Q} : V(\mathbf{Q} - \mathbf{Q}^*) < r_{\text{out}}\}$ and $\mathcal{U}^{\text{in}} = \{\mathbf{Q} : V(\mathbf{Q} - \mathbf{Q}^*) < r_{\text{in}}\}$. The closure of the outer neighborhood is compact. Bounded profits and continuity of the update target therefore bound the martingale-difference noise whenever $\mathbf{Q}_t \in \mathcal{U}^{\text{out}}$.

Let

$$\tau = \inf\{t \geq b : \mathbf{Q}_t \notin \mathcal{U}^{\text{out}}\}$$

be the first exit time and freeze the recursion after exit: $\widehat{\mathbf{Q}}_t = \mathbf{Q}_{t\wedge\tau}$. Expanding the quadratic function V along the stopped recursion and using the local noise bound and (A73) gives the following. Define $S_n := \sum_{q=b}^{(n\wedge\tau)-1} 2\bar{\alpha}_q x_q' P M_{q+1}$; this is a martingale. Then, for every $t \geq b$,

$$V(\widehat{\mathbf{Q}}_{t+1} - \mathbf{Q}^*) \leq V(\widehat{\mathbf{Q}}_t - \mathbf{Q}^*) - c_1\bar{\alpha}_t \mathbf{1}\{t < \tau\}\|x_t\|^2 + (S_{t+1} - S_t) + C_1\bar{\alpha}_t^2 \mathbf{1}\{t < \tau\}. \tag{A74}$$

Stopping is important: it makes this inequality valid without conditioning on the future event that the process never exits.

Put $T_b = \sum_{t \geq b} \bar{\alpha}_t^2$. The local bounds and orthogonality of martingale increments make the remaining conditional variance at most a constant times T_b. Thus S_n converges in L^2, and Doob's conditional maximal inequality bounds the conditional expected squared maximum by a constant times T_b. On $\{\mathbf{Q}_b \in \mathcal{U}^{\text{in}}\}$, if $C_1 T_b \leq r_{\text{out}}/4$, summing (A74) shows that exit

requires $\sup_{n\geq b}|S_n| \geq r_{\text{out}}/2$. Conditional Doob inequality gives

$$\Pr(\tau < \infty \mid \mathcal{F}_b) \leq C_{\text{lock}} \sum_{t\geq b} \bar{\alpha}_t^2 \quad \text{on } \{\mathbf{Q}_b \in \mathcal{U}^{\text{in}}\}. \tag{A75}$$

If $C_1 T_b > r_{\text{out}}/4$, increasing C_{lock} makes the same bound immediate. The resulting constant works for every b (Borkar, 2002).

Taking conditional expectations removes the martingale increment in (A74). Robbins–Siegmund then shows that $V(\widehat{\mathbf{Q}}_t - \mathbf{Q}^*)$ converges and that the series $\sum_{t\geq b} \bar{\alpha}_t \mathbf{1}\{t < \tau\}\|x_t\|^2$ is finite almost surely. On $\{\tau = \infty\}$, the gains sum to infinity, so $\liminf_t \|x_t\| = 0$. Because $V(x_t)$ converges, its limit must be zero, and $\mathbf{Q}_t \to \mathbf{Q}^*$. Combining this with (A75) proves (A71). Continuity of the softmax policy gives (A72). □

Corollary 1 (Lock-in after a finite entrance time)**.** *Under the assumptions of Theorem 29, let τ_{ent} be an almost-surely finite, integer-valued $(\mathcal{F}_t)$-stopping time. Then*

$$\begin{aligned}\Pr\big(\mathbf{Q}_t \in \mathcal{U}^{\text{out}}\ \forall t \geq \tau_{\text{ent}},\ \mathbf{Q}_t \to \mathbf{Q}^* \mid \mathcal{F}_{\tau_{\text{ent}}}\big) \\ \geq 1 - C_{\text{lock}} \sum_{t\geq \tau_{\text{ent}}} \bar{\alpha}_t^2 \qquad \text{on } \{\mathbf{Q}_{\tau_{\text{ent}}} \in \mathcal{U}^{\text{in}}\}.\end{aligned}$$

The corollary is conditional on finite entrance and does not assert that the process reaches $\mathcal{U}^{\text{in}}$ from an arbitrary initial value.

Proof. On each event $\{\tau_{\text{ent}} = b\}$, which belongs to $\mathcal{F}_b$, apply Theorem 29 at the deterministic date b. Restricted to this event, conditioning on $\mathcal{F}_{\tau_{\text{ent}}}$ is the same as conditioning on $\mathcal{F}_b$. The events $\{\tau_{\text{ent}} = b\}$ are disjoint and exhaust the sample space up to a null set because the entrance time is almost surely finite. Summing the separate deterministic-date inequalities proves the claim. □

A.10 Uninformed-AI boundary case

When AI investors have no private signal, their demand schedules cannot load on an AI signal realization. The relevant choice therefore contains only a price coefficient and an intercept. Price may still convey information about the fundamental through the signal-based demand of human investors, but the AI sector no longer adds private-signal information to market clearing.

This case must be formulated directly in two dimensions. Treating zero signal precision as an ordinary point of the three-dimensional problem would attach an infinite variance to a signal coordinate that has no economic role. Removing that coordinate also removes the

finite-population average of AI signal errors. The remaining uninformed-AI economy retains the same strict-concavity, approximation, local-stability, and stochastic-lock-in logic.

Assumption 5 (Primitive regularity at the uninformed-AI boundary)**.** *At $\tau_A = 0$, the variances σ_v^2 and σ_z^2 and the cost parameter γ_A are positive, $\phi \geq 0$, $\bar{\beta}_\eta^M \geq 0$, and $\gamma_A \bar{\beta}_p^M + \phi > 0$. The signal coordinate is removed from the action space rather than assigned an infinite variance.*

Formally, at the boundary $\tau_A = 0$, restrict every arm to $\beta_\eta = 0$, delete e_j and w_N, and use the coefficient vector (β_p, μ). Aggregate signal loading is then the fixed human loading $\zeta = \bar{\beta}_\eta^M$. With $\widehat{\tau}_0 = \tau_v + \zeta^2 \tau_z$, the price-taking map becomes

$$T_p^0(b,c) = \frac{1 - \zeta\tau_z(\bar{\beta}_p^M + \phi b)/\widehat{\tau}_0}{\gamma_A}, \tag{A76}$$

$$T_\mu^0(b,c) = \frac{\tau_v \bar{v} - \zeta\tau_z(\bar{\mu}^M + \phi c - S)}{\gamma_A \widehat{\tau}_0}. \tag{A77}$$

It is affine with Jacobian $-\lambda_T^0 I_2$, where $\lambda_T^0 = \phi\zeta\tau_z/(\gamma_A\widehat{\tau}_0)$. Its unique fixed point is

$$\beta_p^{A,*} = \frac{\widehat{\tau}_0 - \zeta\tau_z\bar{\beta}_p^M}{\gamma_A\widehat{\tau}_0 + \phi\zeta\tau_z}, \tag{A78}$$

$$\mu^{A,*} = \frac{\tau_v\bar{v} - \zeta\tau_z(\bar{\mu}^M - S)}{\gamma_A\widehat{\tau}_0 + \phi\zeta\tau_z}. \tag{A79}$$

The second-moment matrix of $(-p, 1)$ is positive definite because $\mathrm{Var}(p) > 0$. Hence the two-dimensional continuous Boltzmann density is Gaussian with this affine best-response mean. The uniform quadratic payoff bound, grid approximation, projected localization, and nonnegative-real-part response argument apply with two coordinates and characteristic determinant $(1 + \lambda_T^0\varpi)^2$.

The affine formulas remain valid at $\bar{\beta}_\eta^M = 0$, although price then carries no information about the fundamental. In the discretized bounded economy, deleting the signal coordinate removes the signal terms from $L_{p,B}$ and turns Lemma 16 into the corresponding two-dimensional bound. The Bellman contraction, grid approximation, finite-sample localization, local-stability argument, and stochastic lock-in proof then apply on the (β_p, μ) plane.

Corollary 2 (Convergence at the uninformed-AI boundary)**.** *Under Assumption 5, the two-dimensional best-response map has the unique admissible fixed point displayed above, including when $\bar{\beta}_\eta^M = 0$. In a fixed two-dimensional discretized bounded environment satisfying the two-dimensional projected-admissibility condition and Assumption 2, Theorem 18 holds under the corresponding two-dimensional contraction condition. Along ordered two-dimensional*

refinements satisfying the counterparts of stability accuracy and, for discounted populations, Assumption 4 and (A69), the conclusions of Theorems 25, 28, and 29, and Corollary 1, continue to hold after deleting the signal coordinate.

Proof. For $r = (-p, 1)'$, the determinant of $\mathbb{E}[rr']$ equals $\mathrm{Var}(p)$, which is at least $\sigma_z^2/\xi^2 > 0$. Expected profit is therefore strictly concave in (β_p, μ), even when $\zeta = 0$. The map in (A76)–(A77) is affine with Jacobian $-\lambda_T^0 I_2$; its fixed point is unique, and its aggregate price sensitivity is positive by $\gamma_A \bar{\beta}_p^M + \phi > 0$. The uniform quadratic payoff bound and grid approximation repeat in $\mathbb{R}^2$. The finite-population signal average is absent, while the schedule-sampling and individual-impact bounds are unchanged. Centering the two-dimensional Gaussian response gives the characteristic factor $(1 + \lambda_T^0 \varpi)^2$, and the stochastic lock-in argument is dimension independent. These substitutions give the stated two-dimensional versions of the preceding results. □

A.11 Connection to the main-text results

This subsection links the appendix to the three propositions and theorem in the main text. In a fixed bounded economy, global almost-sure convergence follows when the combined payoff and continuation feedback is weaker than one. Under the ordered refinement—first the grid and bounds, then the population—the policy-implied coefficient mean of every fixed point approaches the informed REE. For a fixed stability-accurate economy and a sufficiently large population, the local result gives a unique Q-value fixed point and local exponential stability. Discounted stability additionally requires an arm gap that dominates the finite-population approximation error. Finally, once learning enters the local basin late enough, stochastic lock-in keeps it there and yields convergence with high conditional probability.

Proof of Proposition 1. The proposition is Theorem 18, with the modulus in Lemma 17 and the policy-mean definition (A13). □

Proof of Proposition 2. Theorem 21 identifies the informed REE $m^{\mathrm{REE}} = (\beta_\eta^{A,*}, \beta_p^{A,*}, \mu^{A,*})$, using the same starred notation as the main-text herding results. Substitution of the fixed level-k human coefficients into (A42)–(A44) gives the within-level system in (25), (26), and (27). Lemma 22 proves the required nonsingularity. Proposition 23 and Theorem 25 imply that, after choosing a fine grid and large bounds and then taking N large, the coefficient mean at a mean-field fixed point lies within any prescribed ε of m^{REE}. This argument takes $\mathcal{K}_{\mathrm{proj}} = \mathcal{K}^+$ and is uniform over all admissible type profiles. □

Proof of Proposition 3. Under the sequence-level gap-rate condition on discounted populations, Proposition 26, Lemma 27, and Theorem 28 give uniqueness and local exponential

stability. For an undiscounted population, the continuation maximum drops out and no gap condition is used. None of these steps invokes (A30). □

Proof of Theorem 4. For the global-convergence part, fix any discretized bounded environment and population satisfying (A30). Theorem 18 gives almost-sure convergence from every initial value profile in $\mathcal{Q}_B$. Whenever the same environment and population also meet the ordered approximation thresholds in Theorem 25 with error below ε, the limiting policy mean lies within ε of m^{REE}. This conclusion is pointwise in the fixed environment. The contraction condition uses worst-case sensitivities over that environment and is not asserted to persist along a refinement sequence.

For the local-lock-in part, given $\varepsilon > 0$, first choose a stability-accurate grid and truncations that also satisfy

$$2\chi_* \delta_{\mathrm{BR}}(\Delta, B) < \varepsilon/2.$$

Such a choice exists by Proposition 23. With this auxiliary economy fixed, choose an admitted population size N for which Theorem 28 applies and

$$\|m_{\Delta,B,N}(\mathbf{Q}^*) - m^0_{\Delta,B,N}\|_\infty < \varepsilon/2.$$

The deterministic bound in Proposition 23, the finite-population rate in (A70), and the triangle inequality then put the fixed-point policy mean within ε of m^{REE}.

Theorem 29 supplies the neighborhoods and C_{lock}. Assumption 2 supplies both the divergent stepsize sum needed for convergence and the finite squared-stepsize tail needed for lock-in. Given $\upsilon > 0$, determinism and square summability allow us to choose b_0 before conditioning so that $C_{\mathrm{lock}} \sum_{t \geq b_0} \bar{\alpha}_t^2 \leq \upsilon$. For every $b \geq b_0$, conditional on $\mathcal{F}_b$ and $\{\mathbf{Q}_b \in \mathcal{U}^{\mathrm{in}}\}$, (A71) gives conditional probability at least $1 - \upsilon$. Equation (A72) and the triangle inequality give the claimed population-average limit and its REE error.

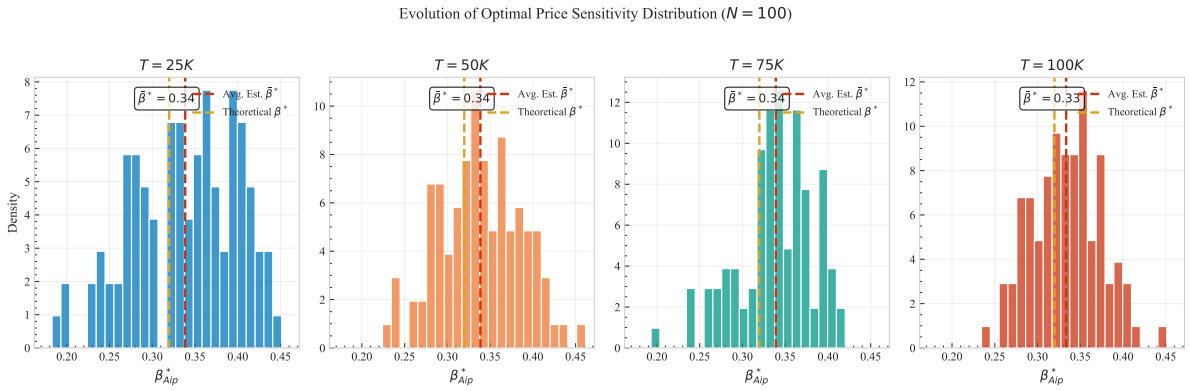
For an almost-surely finite integer-valued $(\mathcal{F}_t)$-stopping time τ_{ent}, Corollary 1 gives the same conditional bound with $\sum_{t \geq \tau_{\mathrm{ent}}} \bar{\alpha}_t^2$. In particular, whenever $C_{\mathrm{lock}} \sum_{t \geq \tau_{\mathrm{ent}}} \bar{\alpha}_t^2 \leq \upsilon$, the conditional probability is at least $1 - \upsilon$. The construction fixes the auxiliary economy and population before letting learning time tend to infinity. At $\tau_A = 0$, Corollary 2 gives the corresponding two-dimensional conclusions. □

A.12 Simulation evidence: algorithmic herding

Figure A1 complements the convergence results by showing the cross-sectional distribution of each investor's estimated best-arm price-sensitivity coefficient $\beta_p^{A,k}$ in a finite population of AI investors. The simulation uses the uninformed-AI boundary $\tau_A = 0$, a centered environ-

ment, and a particular optimistic initialization. Although investors begin from heterogeneous estimates and explore independently, the distribution of their estimated best-arm coefficients narrows around the rational-expectations benchmark. This finite-population exercise illustrates the local coordination mechanism in Theorem 4.

Figure A1: Algorithmic Herding: Cross-Sectional Distribution of Estimated Best-Arm $\beta_p^{A,k}$



Notes: The panels show the cross-sectional distribution of each investor's estimated best-arm price-sensitivity coefficient $\beta_p^{A,k}$—the grid arm with the highest estimated expected reward—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 uninformed-AI rational-expectations benchmark $\beta_p^{A,*} = 0.3197$. As learning proceeds, the distribution concentrates around the benchmark. Simulation parameters are $\sigma_v = 2$, $\sigma_z = 0.2$, $\gamma_A = 0.2$, $\phi = 1$, $\bar{v} = S = 0$, and total human mass $\psi := \psi_c + \psi_s = 1.5$. The per-unit-mass human price and signal coefficients are 4.0 and 4.5, so the aggregate coefficients in (A3) are $\bar{\beta}_p^M = 6.0$, $\bar{\beta}_\eta^M = 6.75$, and $\bar{\mu}^M = 0$. The simulation is undiscounted, all arm-value estimates start at 1, the raw Boltzmann temperature is $\vartheta = 20$, and the β_p grid is $[0, 0.6]$ with mesh 0.002; $\beta_\eta^{A,k}$ and $\mu^{A,k}$ are fixed at zero.

B Proofs for the Limit of Level-k Strategic Reasoning

B.1 Triangular structure and convergence channels

The level-k recursion admits a triangular decomposition. Price informativeness $\tau_p^{(k)}$ evolves autonomously through a one-dimensional map G. 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. The map from $\tau_p^{(k-1)}$ to $\tau_p^{(k)}$ combines two channels. Across levels, the level-k investors update their signal coefficient: from (29), $\beta_\eta^{M,k} = \tau_M/(\gamma_M(\tau_v + \tau_M + \tau_p^{(k-1)}))$. Within the level-$k$ environment, the AI sector responds to the informativeness that its own signal trading helps generate through the fixed-point equation (25). Define the cursed-sector signal loading $\zeta^c \equiv \psi_c \beta_\eta^{M,c} \geq 0$, and note that $\zeta^c + \psi_s \tau_M/(\gamma_M(\tau_v + \tau_M)) = \psi \beta_\eta^{M,c}$. The human-sector signal loading in the level-k environment is

$$\zeta_M^{(k)} = g_M(\tau_p^{(k-1)}), \qquad g_M(\tau_p) \equiv \zeta^c + \frac{\psi_s \tau_M}{\gamma_M(\tau_v + \tau_M + \tau_p)},$$

and the informativeness of the level-k environment is the within-level fixed point at this human-sector loading. Writing $\Phi(\zeta_M)$ for the unique solution of (25), the scalar map governing informativeness is the composition

$$\tau_p^{(k)} = G(\tau_p^{(k-1)}), \qquad G \equiv \Phi \circ g_M, \tag{A80}$$

so that $G(\tau_p)$ is characterized implicitly by

$$G(\tau_p) = \tau_z \left(\zeta^c + \frac{\psi_s \tau_M}{\gamma_M(\tau_v + \tau_M + \tau_p)} + \frac{\phi \tau_A}{\gamma_A(\tau_v + \tau_A + G(\tau_p))} \right)^2. \tag{A81}$$

When $\tau_A = 0$, G reduces to the explicit uninformed-AI map $G(\tau_p) = \tau_z \big(\zeta^c + \psi_s \tau_M/(\gamma_M(\tau_v + \tau_M + \tau_p))\big)^2$.

Properties of G.

1. Φ is continuous and strictly increasing by the monotonicity of (25), and g_M is continuous and strictly decreasing since $\psi_s \tau_M > 0$. Hence $G = \Phi \circ g_M$ is continuous and strictly decreasing on $[0, \infty)$.

2. $G(0) = \Phi(\psi \beta_\eta^{M,c})$, and the level-0 environment has informativeness $\tau_p^{(0)} = G(0)$, since level-0 investors use the cursed coefficients.

3. $\lim_{\tau_p \to \infty} G(\tau_p) = \Phi(\zeta^c) \equiv \underline{\tau}_p$, the within-level fixed point (25) at $\zeta_M = \zeta^c$. The floor $\underline{\tau}_p$ is strictly positive whenever $\psi_c > 0$ or $\phi \tau_A > 0$ (informativeness floor).

4. G maps the interval $\mathcal{T} \equiv [\underline{\tau}_p, G(0)]$ into itself.

Property 3 is a key structural feature. With $\tau_A = 0$ the floor equals $(\psi_c \beta_\eta^{M,c})^2 \tau_z$ and is positive only when cursed investors are present; with $\phi \tau_A > 0$, AI signal trading keeps the

price informative even when $\psi_c = 0$. Regardless of how the level-k thinkers adjust their behavior, price informativeness never falls below $\underline{\tau}_p$.

Lemma 30 (Informativeness channel)**.** *The sequence $\{\tau_p^{(k)}\}$ is bounded in $\mathcal{T}$. Let τ_p^* denote the unique fixed point of G on $(0,\infty)$. Define*

$$\kappa_G \equiv |G'(\tau_p^*)| = \frac{\dfrac{2\psi_s \tau_M \sqrt{\tau_p^* \tau_z}}{\gamma_M(\tau_v + \tau_M + \tau_p^*)^2}}{1 + \dfrac{2\phi\tau_A \sqrt{\tau_p^* \tau_z}}{\gamma_A(\tau_v + \tau_A + \tau_p^*)^2}}.$$

1. *If $\kappa_G \leq 1$, then $\tau_p^{(k)} \to \tau_p^*$.*
2. *If $\kappa_G > 1$, then the canonical orbit converges by parity to a period-two cycle (τ_p^-, τ_p^+): $\tau_p^{(2n)} \to \tau_p^+$ and $\tau_p^{(2n+1)} \to \tau_p^-$, where $G(\tau_p^-) = \tau_p^+$, $G(\tau_p^+) = \tau_p^-$, and $\underline{\tau}_p \leq \tau_p^- < \tau_p^* < \tau_p^+$.*

Proof. Fixed point and bounded orbit. The fixed point solves $\tau_p = G(\tau_p)$. The left-hand side is strictly increasing, while G is strictly decreasing and satisfies $G(0) > 0$. Hence the two graphs cross once on $(0,\infty)$; call the crossing τ_p^*. The initial condition is $\tau_p^{(0)} = G(0)$. Because G maps $\mathcal{T} = [\underline{\tau}_p, G(0)]$ into itself, every subsequent $\tau_p^{(k)}$ lies in $\mathcal{T}$.

Parity subsequences. Let $G^{(2)} \equiv G \circ G$. Since G is decreasing, $G^{(2)}$ is increasing. Moreover $\tau_p^{(0)} > \tau_p^*$ and $\tau_p^{(1)} = G(\tau_p^{(0)}) < \tau_p^*$. Iterating this argument gives $\tau_p^{(2n)} > \tau_p^*$ and $\tau_p^{(2n+1)} < \tau_p^*$ for all n. Since $G^{(2)}(\tau_p^{(0)}) \leq \tau_p^{(0)}$ and $G^{(2)}$ is increasing, the even subsequence is decreasing. Applying the decreasing map G 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 $G(\tau_p^{\text{even}}) = \tau_p^{\text{odd}}$ and $G(\tau_p^{\text{odd}}) = \tau_p^{\text{even}}$. 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. Taking square roots in (A81) gives $\sqrt{G(\tau_p)/\tau_z} = \zeta^c + \psi_s\tau_M/(\gamma_M(\tau_v + \tau_M + \tau_p)) + \phi\tau_A/(\gamma_A(\tau_v + \tau_A + G(\tau_p)))$. Implicit differentiation of (A81) therefore gives

$$G'(\tau_p) = -\frac{2\psi_s\tau_M\sqrt{G(\tau_p)\tau_z}/\big(\gamma_M(\tau_v + \tau_M + \tau_p)^2\big)}{1 + 2\phi\tau_A\sqrt{G(\tau_p)\tau_z}/\big(\gamma_A(\tau_v + \tau_A + G(\tau_p))^2\big)}.$$

At the fixed point, $G(\tau_p^*) = \tau_p^*$, so $G'(\tau_p^*) = -\kappa_G$ with κ_G as displayed in the lemma. The stability of τ_p^* under $G^{(2)}$ is governed by $(G^{(2)})'(\tau_p^*) = \kappa_G^2$.

For the global restriction we show that the Schwarzian derivative of G is negative. Write $S(f) \equiv f'''/f' - \frac{3}{2}(f''/f')^2$. Both factors of the composition $G = \Phi \circ g_M$ have convenient

Schwarzian structure. The human channel g_M is a Möbius map, so $S(g_M) = 0$, and the composition rule $S(\Phi \circ g_M) = \big(S(\Phi) \circ g_M\big)(g_M')^2 + S(g_M)$ shows that $S(G)(\tau_p)$ has the same sign as $S(\Phi)(g_M(\tau_p))$. The within-level map Φ is the inverse of

$$\Psi(\tau) \equiv \sqrt{\tau/\tau_z} - \frac{\phi\tau_A}{\gamma_A(\tau_v + \tau_A + \tau)},$$

obtained by solving (25) for ζ_M; Ψ is strictly increasing. For inverse functions, $S(\Phi)(c) = -S(\Psi)(\tau)/\Psi'(\tau)^2$ at $c = \Psi(\tau)$, so it suffices to show $S(\Psi) > 0$. Write $T \equiv \tau_v + \tau_A + \tau$, $a \equiv 1/(2\sqrt{\tau_z\tau})$, and $b \equiv \phi\tau_A/(\gamma_A T^2)$, so that $\Psi' = a + b$, $\Psi'' = -a/(2\tau) - 2b/T$, and $\Psi''' = 3a/(4\tau^2) + 6b/T^2$. Direct computation gives

$$S(\Psi)\,(\Psi')^2 = \Psi'''\Psi' - \tfrac{3}{2}(\Psi'')^2 = \frac{3a^2}{8\tau^2} + 3ab\,\frac{(T-2\tau)^2 + 4\tau^2}{4\tau^2T^2} > 0.$$

Hence $S(\Psi) > 0$, so $S(\Phi) < 0$ and $S(G) < 0$ on $(0, \infty)$. When $\tau_A = 0$, $\Psi(\tau) = \sqrt{\tau/\tau_z}$ and the same computation reduces to

$$S(G) = -\frac{3\psi_s^2\tau_M^2}{2(\tau_v + \tau_M + \tau_p)^2\left[\gamma_M\zeta^c(\tau_v + \tau_M + \tau_p) + \psi_s\tau_M\right]^2}.$$

Because $G'(\tau_p) \neq 0$, the chain rule for Schwarzian derivatives gives $S(G^{(2)}) < 0$. For an increasing map with negative Schwarzian derivative, the function $((G^{(2)})')^{-1/2}$ is strictly convex. Therefore $(G^{(2)})'(\tau_p) = 1$ has at most two solutions on $\mathcal{T}$, and Rolle's theorem implies that $\delta_2(\tau_p) \equiv G^{(2)}(\tau_p) - \tau_p$ has at most three zeros on $\mathcal{T}$.

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

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

By Rolle's theorem, there are $u_1 \in (\tau_p^{\text{odd}}, \tau_p^*)$ and $u_2 \in (\tau_p^*, \tau_p^{\text{even}})$ with $(G^{(2)})'(u_1) = (G^{(2)})'(u_2) = 1$. Strict convexity of $((G^{(2)})')^{-1/2}$ then implies

$$\big((G^{(2)})'(\tau_p^*)\big)^{-1/2} < 1,$$

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

Oscillation cycle when $\kappa_G > 1$. When $\kappa_G > 1$, $\delta_2'(\tau_p^*) = \kappa_G^2 - 1 > 0$, so δ_2 crosses zero from negative to positive at τ_p^*. The boundary values $\delta_2(\underline{\tau}_p) \geq 0$ and $\delta_2(G(0)) \leq 0$ 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 $G^{(2)}$ besides τ_p^*.

It remains to rule out convergence of both parity subsequences to τ_p^*. Because

$$(G^{(2)})'(\tau_p^*) = \kappa_G^2 > 1,$$

there is an interval immediately to the right of τ_p^* on which $G^{(2)}(\tau_p) > \tau_p$. If the decreasing even subsequence converged to τ_p^* from above, it would eventually enter this interval, where its next even iterate would be larger rather than smaller. This contradiction shows that the two parity limits differ. The two noncentral fixed points of $G^{(2)}$ therefore form the unique period-two cycle (τ_p^-, τ_p^+) of G, and the canonical parity subsequences converge to these cycle points.

Global basin under free initialization. The same three-fixed-point argument covers every initialization in $\mathcal{T}$. The sign of $\delta_2(\tau_p) = G^{(2)}(\tau_p) - \tau_p$ is positive below τ_p^-, negative on (τ_p^-, τ_p^*), positive on (τ_p^*, τ_p^+), and negative above τ_p^+. Since $G^{(2)}$ is increasing, its iterates move monotonically toward τ_p^- from either side below τ_p^* and toward τ_p^+ from either side above τ_p^*. Thus τ_p^* is the unique initialization that remains at the repelling fixed point; every other initialization approaches the period-two cycle. □

The canonical initialization satisfies $\tau_p^{(0)} = G(0) > \tau_p^*$, so it cannot remain at the repelling fixed point when $\kappa_G > 1$.

Within-level damping. The AI signal channel stabilizes the informativeness recursion. The numerator of κ_G has the same form as the fixed-point slope in the uninformed-AI benchmark and measures the feedback strength of the human cross-level channel; the denominator exceeds one whenever $\phi\tau_A > 0$. The reason is timing: level-k investors respond to the previous level's informativeness, but the AI sector responds within the level. When the human channel pushes informativeness down, informed AI investors face a less informative price, shift weight toward their private signals, and partially restore informativeness, damping the oscillation that drives non-convergence.

Informativeness floor. When $\psi_c > 0$ or $\phi\tau_A > 0$, for all $k \geq 0$,

$$\tau_p^{(k)} \geq \underline{\tau}_p \geq (\zeta^c)^2\tau_z = (\psi_c\beta_\eta^{M,c})^2\tau_z, \tag{A82}$$

with $\underline{\tau}_p > 0$. At every price-admissible finite level, the equilibrium price is therefore informative about v: cursed investors provide a fixed private-signal loading when $\psi_c > 0$, while informed AI investors keep trading on their own signals when $\phi\tau_A > 0$.

The driven channels. Let $\nu^k \equiv \beta_p^{M,k}/\beta_\eta^{M,k}$ denote the normalized price loading. The scalar informativeness path determines all signal coefficients. Writing $\zeta = \sqrt{\tau_p/\tau_z}$ for the aggregate signal loading at informativeness τ_p, and

$$\beta_\eta^A(\tau_p) \equiv \frac{\tau_A}{\gamma_A(\tau_v+\tau_A+\tau_p)}, \qquad h_p^A(\tau_p) \equiv \frac{\tau_p}{\tau_v+\tau_A+\tau_p}$$

for the AI signal coefficient and posterior price weight at informativeness τ_p, the level-$(k-1)$ investor signal coefficient satisfies $\psi_s\beta_\eta^{M,k-1} = \zeta - \zeta^c - \phi\beta_\eta^A(\tau_p)$ at $\tau_p = \tau_p^{(k-1)}$, and the human-sector aggregate price sensitivity in the perceived environment is

$$\psi_c\beta_p^{M,c} + \psi_s\beta_p^{M,k-1} = \frac{\psi_c}{\gamma_M} + \nu^{k-1}\big(\zeta - \zeta^c - \phi\beta_\eta^A(\tau_p)\big).$$

Substituting the AI fixed-point coefficient (26) into the definition of $\xi^{(k-1)}$ and simplifying yields $\xi^{(k-1)}/\zeta^{(k-1)}$ as an affine function of ν^{k-1}. Substituting this into (30) gives the affine recursion

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

where

$$a_\nu(\tau_p) \equiv \frac{\tau_v+\tau_M+\tau_p}{\tau_M} - \frac{\tau_p}{\tau_M}\,\frac{\gamma_A\frac{\psi_c}{\gamma_M}+\phi}{\zeta\big(\gamma_A + h_p^A(\tau_p)\phi/\zeta\big)}, \tag{A83}$$

$$b_\nu(\tau_p) \equiv -\frac{\tau_p}{\tau_M}\,\frac{\gamma_A\big(\zeta-\zeta^c-\phi\beta_\eta^A(\tau_p)\big)}{\zeta\big(\gamma_A + h_p^A(\tau_p)\phi/\zeta\big)} \le 0. \tag{A84}$$

The slope is nonpositive on the closed invariant interval. It is strictly negative along every finite orbit because $\psi_s > 0$ and $\beta_\eta^{M,k-1} > 0$, so $\zeta-\zeta^c-\phi\beta_\eta^A(\tau_p) = \psi_s\beta_\eta^{M,k-1} > 0$. Economically, a higher normalized price loading at level $k-1$ raises the perceived aggregate price sensitivity $\xi^{(k-1)}$ while leaving signal loadings, and hence price informativeness, unchanged. A given price movement then corresponds to a larger movement of the price signal $s_p^{(k-1)}$ through the ratio $\xi^{(k-1)}/\zeta^{(k-1)}$, so the inference component of the level-k price response offsets more of the mechanical response and $\beta_p^{M,k}$ falls.

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

$$h_p^M(\tau_p) \equiv \frac{\tau_p}{\tau_v+\tau_M+\tau_p}, \qquad h_v^M(\tau_p) \equiv \frac{\tau_v}{\tau_v+\tau_M+\tau_p}, \qquad h_v^A(\tau_p) \equiv \frac{\tau_v}{\tau_v+\tau_A+\tau_p}.$$

Its affine coefficients are

$$a_\mu(\tau_p) \equiv \frac{h_v^M(\tau_p)\bar{v}}{\gamma_M} - \frac{h_p^M(\tau_p)}{\gamma_M}\frac{\gamma_A(\psi_c\mu^{M,c} - S) + \phi h_v^A(\tau_p)\bar{v}}{\zeta\big(\gamma_A + h_p^A(\tau_p)\phi/\zeta\big)},$$
$$b_\mu(\tau_p) \equiv -\frac{h_p^M(\tau_p)}{\gamma_M}\frac{\psi_s\gamma_A}{\zeta\big(\gamma_A + h_p^A(\tau_p)\phi/\zeta\big)} \leq 0.$$

The slope is strictly negative when $\tau_p > 0$. If $\underline{\tau}_p = 0$, the displayed driven-channel functions are understood by continuous extension at the endpoint:

$$a_\nu(0) = \frac{\tau_v + \tau_M}{\tau_M}, \qquad b_\nu(0) = 0, \qquad a_\mu(0) = \frac{\tau_v\bar{v}}{\gamma_M(\tau_v + \tau_M)}, \qquad b_\mu(0) = 0.$$

The canonical finite orbit remains strictly positive, so these endpoint values matter only when the recursions are extended to the closed invariant interval.

Common stability. The normalized price-loading and intercept channels share the same asymptotic stability. At the fixed point, write $\zeta^* \equiv \sqrt{\tau_p^*/\tau_z}$ and $\beta_\eta^{M,\infty} \equiv \tau_M/[\gamma_M(\tau_v + \tau_M + \tau_p^*)]$. Direct computation gives $|b_\nu(\tau_p^*)| = |b_\mu(\tau_p^*)|$: the ratio $|b_\nu(\tau_p^*)|/|b_\mu(\tau_p^*)|$ simplifies to $(\tau_v + \tau_M + \tau_p^*)\gamma_M\beta_\eta^{M,\infty}/\tau_M$, using $\psi_s\beta_\eta^{M,\infty} = \zeta^* - \zeta^c - \phi\beta_\eta^A(\tau_p^*)$, and substituting the fixed-point signal coefficient gives a ratio of one.

B.2 Proof of the convergence condition

Lemma 30 and the driven affine recursions above show that the two stability inequalities $\kappa_G \leq 1$ and $|b_\nu(\tau_p^*)| < 1$ imply fixed-point convergence for every driven initialization. Here a driven initialization means any finite pair $(\nu^0, \mu^{M,0})$, holding fixed the canonical informativeness initialization $\tau_p^{(0)} = G(0)$. We next derive closed-form sufficient conditions on the share of level-k thinkers, ψ_s, that guarantee both inequalities uniformly over all human private-signal precisions $\tau_M > 0$. Each economic level-k path must also satisfy the finite price-admissibility condition (15). The AI signal precision τ_A is kept explicit throughout; setting $\tau_A = 0$ recovers the corresponding condition for uninformed AI investors.

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. The within-level AI response damps this feedback; the bound below discards the damping term and is therefore conservative.

Proposition 31 (Informativeness convergence)**.** *For all $\sigma_M > 0$,*

$$\kappa_G \leq \frac{8\psi\psi_s\tau_z}{27\gamma_M^2\tau_v} + \frac{\psi_s\phi\tau_z}{2\gamma_M\gamma_A\tau_v} \cdot \frac{\tau_A}{\tau_v + \tau_A}. \tag{A85}$$

Hence $\kappa_G \leq 1$ for all $\sigma_M > 0$ whenever

$$\psi_s\left(\frac{8\psi}{27\gamma_M^2} + \frac{\phi}{2\gamma_M\gamma_A} \cdot \frac{\tau_A}{\tau_v + \tau_A}\right)\frac{\tau_z}{\tau_v} \leq 1. \tag{A86}$$

When $\tau_A = 0$, these reduce to the uninformed-AI bounds

$$\kappa_G \leq \frac{8\psi\psi_s\tau_z}{27\gamma_M^2\tau_v} = \frac{8\psi\psi_s\sigma_v^2}{27\gamma_M^2\sigma_z^2}, \qquad \psi_s \leq \frac{27\gamma_M^2\sigma_z^2}{8\psi\sigma_v^2}.$$

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

Discarding the damping denominator of κ_G, which is at least one, leaves the numerator $2\psi_s\tau_M\sqrt{\tau_p^*\tau_z}/\big(\gamma_M(\tau_v + \tau_M + \tau_p^*)^2\big)$. The fixed-point identity

$$\begin{aligned}\sqrt{\tau_p^*\tau_z} &= \tau_z\left(\zeta^c + \frac{\psi_s\tau_M}{\gamma_M(\tau_v + \tau_M + \tau_p^*)} + \frac{\phi\tau_A}{\gamma_A(\tau_v + \tau_A + \tau_p^*)}\right)\\ &\leq \tau_z\left(\frac{\psi\tau_M}{\gamma_M(\tau_v + \tau_M)} + \frac{\phi\tau_A}{\gamma_A(\tau_v + \tau_A)}\right),\end{aligned}$$

which uses $\zeta^c + \psi_s\tau_M/(\gamma_M(\tau_v + \tau_M)) = \psi\tau_M/(\gamma_M(\tau_v + \tau_M))$ and $\tau_p^* \geq 0$, together with $\tau_v + \tau_M + \tau_p^* \geq \tau_v + \tau_M$, gives

$$\begin{aligned}\kappa_G &\leq \frac{2\psi_s\tau_M\tau_z}{\gamma_M(\tau_v + \tau_M)^2}\left(\frac{\psi\tau_M}{\gamma_M(\tau_v + \tau_M)} + \frac{\phi\tau_A}{\gamma_A(\tau_v + \tau_A)}\right)\\ &= \frac{2\psi\psi_s\tau_z}{\gamma_M^2}\frac{\tau_M^2}{(\tau_v + \tau_M)^3} + \frac{2\psi_s\phi\tau_z}{\gamma_M\gamma_A}\frac{\tau_A}{\tau_v + \tau_A}\frac{\tau_M}{(\tau_v + \tau_M)^2}.\end{aligned}$$

The first term 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 $4/(27\tau_v)$; the second depends on τ_M only through $\tau_M/(\tau_v + \tau_M)^2$, which is maximized at $\tau_M = \tau_v$ with value $1/(4\tau_v)$. Bounding each term by its maximum yields (A85). □

B.2.2 Driven channel

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

Proposition 32 (Driven-channel convergence)**.** *For all $\sigma_M > 0$,*

$$|b_\nu(\tau_p^*)| < \frac{\psi\psi_s\tau_z}{4\gamma_M^2\tau_v} + \frac{\psi_s\phi\tau_z}{\gamma_M\gamma_A\tau_v} \cdot \frac{\tau_A}{\tau_v + \tau_A}. \tag{A87}$$

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

$$\psi_s\left(\frac{\psi}{4\gamma_M^2} + \frac{\phi}{\gamma_M\gamma_A} \cdot \frac{\tau_A}{\tau_v + \tau_A}\right)\frac{\tau_z}{\tau_v} \le 1. \tag{A88}$$

When $\tau_A = 0$, these reduce to the uninformed-AI bounds $|b_\nu(\tau_p^*)| < \psi\psi_s\tau_z/(4\gamma_M^2\tau_v)$ and $\psi_s \le 4\gamma_M^2\sigma_z^2/(\psi\sigma_v^2)$.

Proof. From (A84),

$$|b_\nu(\tau_p)| = \frac{\tau_p}{\tau_M}\,\frac{\gamma_A\,\psi_s\beta_\eta^M}{\zeta\big(\gamma_A + h_p^A(\tau_p)\phi/\zeta\big)} = \frac{\tau_p}{\tau_M}\,\frac{\gamma_A\,\psi_s\beta_\eta^M}{\gamma_A\zeta + h_p^A(\tau_p)\phi},$$

where $\psi_s\beta_\eta^M = \zeta - \zeta^c - \phi\beta_\eta^A(\tau_p)$ is the level-k investor sector's signal loading. Dropping the non-negative term $h_p^A(\tau_p)\phi$ from the denominator and canceling γ_A gives

$$|b_\nu(\tau_p)| \le \frac{\tau_p\,\psi_s\beta_\eta^M}{\tau_M\zeta} = \frac{\tau_z\,\zeta\,\psi_s\beta_\eta^M}{\tau_M},$$

where the last step uses $\tau_p = \zeta^2\tau_z$. At the fixed point, $\psi_s\beta_\eta^{M,\infty} = \psi_s\tau_M/[\gamma_M(\tau_v + \tau_M + \tau_p^*)] < \psi_s\tau_M/(\gamma_M(\tau_v + \tau_M))$ because $\tau_p^* > 0$, and, as in the proof of Proposition 31, $\zeta^* \le \psi\tau_M/(\gamma_M(\tau_v + \tau_M)) + \phi\tau_A/(\gamma_A(\tau_v + \tau_A))$. Hence

$$\begin{aligned}|b_\nu(\tau_p^*)| &< \frac{\tau_z}{\tau_M}\frac{\psi_s\tau_M}{\gamma_M(\tau_v + \tau_M)}\left(\frac{\psi\tau_M}{\gamma_M(\tau_v + \tau_M)} + \frac{\phi\tau_A}{\gamma_A(\tau_v + \tau_A)}\right)\\ &= \frac{\psi\psi_s\tau_z}{\gamma_M^2}\frac{\tau_M}{(\tau_v + \tau_M)^2} + \frac{\psi_s\phi\tau_z}{\gamma_M\gamma_A}\frac{\tau_A}{\tau_v + \tau_A}\frac{1}{\tau_v + \tau_M}.\end{aligned}$$

The first term depends on τ_M only through $\tau_M/(\tau_v + \tau_M)^2$, which is maximized at $\tau_M = \tau_v$ with value $1/(4\tau_v)$; the second is bounded using $\tau_v + \tau_M > \tau_v$. Substituting yields (A87). □

Remark. In the uninformed-AI benchmark ($\tau_A = 0$), the bounds in Propositions 31 and 32 do not involve the AI parameters γ_A or ϕ: γ_A cancels algebraically and the term $h_p^A \phi \geq 0$ is dropped from the denominator. With informed AI investors, the destabilizing signal-loading contribution in the bounds is $\phi\tau_A/[\gamma_A(\tau_v + \tau_A)]$. Two stabilizing AI effects are deliberately discarded, which keeps the bounds conservative: the within-level damping term in the denominator of κ_G, and the term $h_p^A\phi$ in the denominator of b_ν. The exact dynamics therefore depend separately on the AI mass, signal precision, and trading cost through both destabilizing and stabilizing channels.

B.2.3 Combined condition

Proof of Proposition 5. For every $\sigma_M > 0$, the condition

$$\psi_s\left(\frac{8\psi}{27\gamma_M^2} + \frac{\phi}{\gamma_M\gamma_A}\cdot\frac{\tau_A}{\tau_v+\tau_A}\right)\frac{\tau_z}{\tau_v} \leq 1. \tag{A89}$$

implies both (A86) and (A88): its human coefficient $8/27$ equals the informativeness coefficient and exceeds the driven-channel coefficient $1/4$, while its AI coefficient 1 exceeds the informativeness coefficient $1/2$ and equals the driven-channel coefficient. Lemma 30 therefore gives $\tau_p^{(k)} \to \tau_p^*$. Continuity of the affine coefficients then implies that the intercepts and slopes of both driven recursions converge. Their limiting absolute slopes are below one, so for some $\rho < 1$ all sufficiently late slopes have absolute value at most ρ; the standard iteration of a nonautonomous affine recursion then gives convergence from every finite driven initialization. The AI coefficients converge as well by continuity of the AI signal, price, and intercept formulas. This establishes coefficient convergence for every initialization whose finite hierarchy remains price admissible. When $\tau_A = 0$, (A89) reduces to

$$\psi_s \leq \frac{27\gamma_M^2\sigma_z^2}{8\psi\sigma_v^2}. \tag{A90}$$

□

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

Remark. Condition (A89) is a uniform sufficient condition for dynamic stability, not a necessary characterization of the full parameter space and not a uniform proof of finite price admissibility. Its upper bound on ψ_s is proportional to σ_z^2 (equivalently inversely proportional to τ_z) and decreasing in $\phi\tau_A/[\gamma_A(\tau_v + \tau_A)]$. Thus the sufficient stability region

shrinks as noise-trading volatility falls and as informed AI signal trading grows: when prices are very precise, small changes in perceived signal extraction can generate large changes in higher-level price responses.

B.3 Proof that the convergent limit is a linear REE

In the limit, the level-∞ investors and AI investors best-respond to the same linear equilibrium price that their own strategies generate.

Proof of Proposition 6. Under its local-lock-in assumptions, Theorem 4 supplies the learning justification for the AI coefficients used in each fixed level-k environment: after the process enters the stable basin, its ordered learning and refinement limits approach the informed atomistic rational-expectations target with the theorem's stated high conditional probability. The argument below separately imposes the cross-level coefficient convergence assumed in Proposition 6.

Let the limiting aggregate objects be

$$\zeta^{(\infty)} \equiv \psi_c \beta_\eta^{M,c} + \psi_s \beta_\eta^{M,\infty} + \phi \beta_\eta^{A,\infty}, \qquad \xi^{(\infty)} \equiv \psi_c \beta_p^{M,c} + \psi_s \beta_p^{M,\infty} + \phi \beta_p^{A,\infty},$$
$$\mu^{(\infty)} \equiv \psi_c \mu^{M,c} + \psi_s \mu^{M,\infty} + \phi \mu^{A,\infty} - S.$$

Coefficient convergence implies $\zeta^{(k-1)} \to \zeta^{(\infty)}$, $\tau_p^{(k-1)} = (\zeta^{(k-1)})^2 \tau_z \to \tau_p^{(\infty)} \equiv (\zeta^{(\infty)})^2 \tau_z$, $\xi^{(k-1)} \to \xi^{(\infty)}$, and $\mu^{(k-1)} \to \mu^{(\infty)}$. Continuity of G and $\tau_p^{(k)} = G(\tau_p^{(k-1)})$ imply $\tau_p^{(\infty)} = G(\tau_p^{(\infty)})$, so uniqueness of the fixed point gives $\tau_p^{(\infty)} = \tau_p^*$. The bound $\tau_p^{(k-1)} \le G(0)$ implies

$$\beta_\eta^{M,k} \ge \frac{\tau_M}{\gamma_M(\tau_v + \tau_M + G(0))} > 0.$$

Since $\psi_s > 0$, taking limits gives $\zeta^{(\infty)} > 0$.

At the fixed point, define

$$h_p^{M,\infty} \equiv \frac{\tau_p^{(\infty)}}{\tau_v + \tau_M + \tau_p^{(\infty)}}, \qquad h_p^{A,\infty} \equiv \frac{\tau_p^{(\infty)}}{\tau_v + \tau_A + \tau_p^{(\infty)}}.$$

The human and AI price coefficients satisfy

$$\beta_p^{M,\infty} = \frac{1 - h_p^{M,\infty} \xi^{(\infty)}/\zeta^{(\infty)}}{\gamma_M}, \qquad \beta_p^{A,\infty} = \frac{1 - h_p^{A,\infty} \xi^{(\infty)}/\zeta^{(\infty)}}{\gamma_A}.$$

Substituting these expressions and $\beta_p^{M,c} = 1/\gamma_M$ into the definition of $\xi^{(\infty)}$ and solving gives

$$\xi^{(\infty)} = \frac{\psi/\gamma_M + \phi/\gamma_A}{1 + \left(\psi_s h_p^{M,\infty}/\gamma_M + \phi h_p^{A,\infty}/\gamma_A\right)/\zeta^{(\infty)}} > 0. \tag{A91}$$

Thus price admissibility of the limiting REE is automatic. Moreover, convergence to a positive $\xi^{(\infty)}$, together with nonzero $\xi^{(k)}$ at every finite level, implies $\inf_k |\xi^{(k)}| > 0$: the tail is bounded below by $\xi^{(\infty)}/2$, and the remaining finite prefix has a positive minimum. Substituting the limiting demands into market clearing therefore gives the limiting equilibrium price

$$p^{(\infty)} = \frac{\mu^{(\infty)}}{\xi^{(\infty)}} + \frac{\zeta^{(\infty)}}{\xi^{(\infty)}} v + \frac{1}{\xi^{(\infty)}} z.$$

Equivalently, this price reveals the signal

$$s_p^{(\infty)} = \frac{\xi^{(\infty)}}{\zeta^{(\infty)}} \left(p^{(\infty)} - \frac{\mu^{(\infty)}}{\xi^{(\infty)}}\right) = v + \frac{z}{\zeta^{(\infty)}},$$

with precision $\tau_p^{(\infty)}$.

Taking $k \to \infty$ in the level-k best-response formulas (29)–(31), the right-hand sides are continuous in $(\tau_p^{(k-1)}, \xi^{(k-1)}, \mu^{(k-1)})$. At the limit, the perceived objects used in the posterior coincide with $(\tau_p^{(\infty)}, \xi^{(\infty)}, \mu^{(\infty)})$, so the level-$\infty$ investor coefficients $(\beta_\eta^{M,\infty}, \beta_p^{M,\infty}, \mu^{M,\infty})$ satisfy the Bayesian best-response conditions for a trader who observes $(\eta_i, p^{(\infty)})$ and conditions on the equilibrium price.

The same argument applies to the AI coefficients: taking limits in (26)–(27) and in $\beta_\eta^{A,k} = \tau_A/(\gamma_A(\tau_v + \tau_A + \tau_p^{(k)}))$, the limit coefficients $(\beta_\eta^{A,\infty}, \beta_p^{A,\infty}, \mu^{A,\infty})$ satisfy the best-response conditions for a Bayesian trader who observes $(\eta_j, p^{(\infty)})$ when $\tau_A > 0$, and who observes only $p^{(\infty)}$ when $\tau_A = 0$. In particular, $\beta_\eta^{A,\infty} = \tau_A/(\gamma_A(\tau_v + \tau_A + \tau_p^{(\infty)}))$, so the within-level consistency condition (25) holds at the limit. Hence the limit is a linear REE for the level-∞ investors and AI investors, with cursed investors as given. □

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,

$$\mathbb{E}[p] = \frac{\zeta}{\xi}\bar{v} + \frac{\mu}{\xi}, \qquad \mathrm{Var}(p) = \left(\frac{\zeta}{\xi}\right)^2 \sigma_v^2 + \frac{\sigma_z^2}{\xi^2}, \qquad \mathrm{Cov}(v, p) = \frac{\zeta}{\xi}\sigma_v^2. \tag{A92}$$

For $\ell \in \{M, A\}$, let $n = i$ when $\ell = M$ and $n = j$ when $\ell = A$. When the signal coordinate is present, $\eta_n^\ell = v + e_n^\ell$ has moments

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

where $\sigma_\ell^2 = \sigma_M^2$ for humans and $\sigma_\ell^2 = \sigma_A^2$ for informed AI. At $\tau_A = 0$, the AI signal coordinate is removed. The mean excess payoff is

$$\bar{v} - \mathbb{E}[p] = \frac{(\xi - \zeta)\bar{v} - \mu}{\xi}. \tag{A93}$$

Proof of Proposition 7. Fix $\ell \in \{M, A\}$. At $\tau_A = 0$, interpret the AI demand below after removing its signal coordinate. The investor uses

$$x_n^\ell = \beta_\eta^\ell (v + e_n^\ell) - \beta_p^\ell p + \mu^\ell.$$

Its mean demand is

$$\mathbb{E}[x_n^\ell] = \frac{(\beta_\eta^\ell \xi - \beta_p^\ell \zeta)\bar{v} + \xi\mu^\ell - \beta_p^\ell \mu}{\xi}. \tag{A94}$$

Gross expected profit decomposes as

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

Independence of the fundamental, idiosyncratic signal error, and noise-trader demand gives

$$\mathrm{Cov}(v - p, x_n^\ell) = \left(1 - \frac{\zeta}{\xi}\right)\left(\beta_\eta^\ell - \beta_p^\ell \frac{\zeta}{\xi}\right)\sigma_v^2 + \beta_p^\ell \frac{\sigma_z^2}{\xi^2}. \tag{A95}$$

Using (A93) and (A94),

$$\mathbb{E}[(v-p)x_n^\ell] = \frac{1}{\xi^2}\left\{[(\xi-\zeta)\bar{v}-\mu]\left[(\beta_\eta^\ell\xi-\beta_p^\ell\zeta)\bar{v}+\xi\mu^\ell-\beta_p^\ell\mu\right]\right.$$
$$\left.+(\xi-\zeta)(\beta_\eta^\ell\xi-\beta_p^\ell\zeta)\sigma_v^2+\beta_p^\ell\sigma_z^2\right\}. \tag{A96}$$

The variance of demand is

$$\mathrm{Var}(x_n^\ell) = \left(\beta_\eta^\ell-\beta_p^\ell\frac{\zeta}{\xi}\right)^2\sigma_v^2+(\beta_\eta^\ell)^2\sigma_\ell^2+(\beta_p^\ell)^2\frac{\sigma_z^2}{\xi^2}. \tag{A97}$$

Therefore,

$$\frac{\gamma_\ell}{2}\mathbb{E}[(x_n^\ell)^2] = \frac{\gamma_\ell}{2\xi^2}\left\{\left[(\beta_\eta^\ell\xi-\beta_p^\ell\zeta)\bar{v}+\xi\mu^\ell-\beta_p^\ell\mu\right]^2\right.$$
$$\left.+(\beta_\eta^\ell\xi-\beta_p^\ell\zeta)^2\sigma_v^2+\xi^2(\beta_\eta^\ell)^2\sigma_\ell^2+(\beta_p^\ell)^2\sigma_z^2\right\}. \tag{A98}$$

Subtracting (A98) from (A96) proves (38). At $\tau_A = 0$, the AI signal coordinate is removed, so its signal-noise variance contribution is zero and the formula nests the uninformed-AI case. □

C.2 Profit Comparisons

Throughout this subsection, impose equal trading costs $\gamma_M = \gamma_A \equiv \gamma$.

Lemma 33 (Optimal-demand profit representation)**.** *A Bayesian investor whose information set $\mathcal{I}$ contains p, whose posterior variance $\mathrm{Var}(v \mid \mathcal{I})$ is constant, and whose trading-cost parameter is γ_i chooses*

$$x^*(\mathcal{I}) = \frac{\mathbb{E}[v-p \mid \mathcal{I}]}{\gamma_i}$$

and earns

$$\Pi(\mathcal{I}) = \frac{1}{2\gamma_i}\left\{\mathbb{E}[(v-p)^2]-\mathrm{Var}(v \mid \mathcal{I})\right\}. \tag{A99}$$

In the objective environment, the two Bayesian profit levels are

$$\Pi^A = \frac{1}{2\gamma}\left[\mathbb{E}[(v-p)^2]-\frac{1}{\tau_v+\tau_A+\tau_p}\right], \tag{A100}$$

$$\Pi^{M,\infty} = \frac{1}{2\gamma}\left[\mathbb{E}[(v-p)^2]-\frac{1}{\tau_v+\tau_M+\tau_p}\right], \tag{A101}$$

A cursed investor, whose subjectively optimal demand ignores the informational content of

price, earns

$$\Pi^{M,c} = \frac{1}{2\gamma}\left[\mathbb{E}[(v-p)^2] - \frac{1}{\tau_v + \tau_M}\right]. \tag{A102}$$

Proof. Substituting Bayesian optimal demand into expected profit and using conditional-expectation orthogonality gives

$$\Pi(\mathcal{I}) = \frac{1}{2\gamma_i}\mathbb{E}\left[\mathbb{E}[v-p \mid \mathcal{I}]^2\right].$$

The law of total variance, $p \in \mathcal{I}$, and constancy of the posterior variance imply

$$\mathbb{E}\left[\mathbb{E}[v-p \mid \mathcal{I}]^2\right] = \mathbb{E}[(v-p)^2] - \mathrm{Var}(v \mid \mathcal{I}),$$

which proves (A99). For AI and level-∞ investors, Gaussian updating gives

$$\mathrm{Var}(v \mid \eta_j, p) = \frac{1}{\tau_v + \tau_A + \tau_p}, \qquad \mathrm{Var}(v \mid \eta_i, p) = \frac{1}{\tau_v + \tau_M + \tau_p}.$$

For cursed investors, direct substitution of $x_i^{M,c} = (\mathbb{E}[v \mid \eta_i] - p)/\gamma$ and the identity

$$2(v-p)(\mathbb{E}[v \mid \eta_i] - p) - (\mathbb{E}[v \mid \eta_i] - p)^2 = (v-p)^2 - (v - \mathbb{E}[v \mid \eta_i])^2$$

give (A102). This calculation does not require price to be uncorrelated with the private-signal forecast error. □

Proof of Proposition 8. Subtracting (A101) from (A100) gives (39). Subtracting (A102) from (A100) gives (40). Both formulas remain valid at $\tau_A = 0$. □

For the remaining analytical results with uninformed AI, maintain $\psi_s > 0$ and set $\tau_A = 0$. In any informative, price-admissible 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

$$\delta \equiv \bar{v} - \mathbb{E}[p]. \tag{A103}$$

Let $\psi \equiv \psi_c + \psi_s$.

Proof of Proposition 9. 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 10. 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 in-

formation 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

$$\Pi^{M,c} = \frac{\delta^2}{2\gamma} + \frac{(h - 2\alpha + \alpha^2)\sigma_v^2 + \omega}{2\gamma}, \tag{A104}$$

or, equivalently,

$$\Pi^{M,c} = \frac{\delta^2}{2\gamma} + \frac{\sigma_v^2}{2\gamma}\left[h - 2\alpha + \alpha^2 + \omega\tau_v\right]. \tag{A105}$$

The AI investor is a Bayesian best responder to p, so

$$\Pi^A = \frac{1}{2\gamma}\mathbb{E}\left[\left(\mathbb{E}[v - p \mid p]\right)^2\right].$$

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

$$\Pi^A = \frac{\delta^2}{2\gamma} + \frac{(h_p - \alpha)^2\sigma_v^2}{2\gamma h_p}. \tag{A106}$$

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

$$\Pi^A = \frac{\delta^2}{2\gamma} + \frac{\sigma_v^2}{2\gamma}\left[h_p - 2\alpha + \alpha^2 + \omega\tau_v\right]. \tag{A107}$$

Subtracting (A107) from (A105) 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 11. By Lemma 10, 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 following sufficient dynamic-stability condition, conditional on finite-level price admissibility:

$$\psi_s \leq \frac{27 \gamma^2 \sigma_z^2}{8 \psi \sigma_v^2}.$$

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 12 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 12 assumes convergence to an informative, price-admissible fixed point for each τ_M under consideration. These inequalities guarantee the dynamic-stability part of that premise at the respective boundaries; they do not by themselves establish finite-level price admissibility.

Proof of Proposition 12. By Lemma 10, 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:

$$G_c(\tau_p) = \tau_z \left(\zeta^c + \frac{\chi^s}{\widehat{\tau}^c + \tau_p} \right)^2, \qquad \widehat{\tau}^c \equiv \tau_v + \tau_M, \qquad \zeta^c \equiv \psi_c \frac{\tau_M}{\gamma(\tau_v + \tau_M)}, \qquad \chi^s \equiv \frac{\psi_s \tau_M}{\gamma}. \tag{A108}$$

Nonzero $\bar{v}$ and S affect the intercept channel. The assumed convergence of that channel supplies a limiting intercept, while price admissibility ensures that the resulting noncentered objective price is finite. 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

$$\varphi(\tau_M) \equiv \frac{G_c(\tau_M)}{\tau_M} = \frac{\tau_z \tau_M}{\gamma^2} \left(\frac{\psi_c}{\tau_v + \tau_M} + \frac{\psi_s}{\tau_v + 2\tau_M} \right)^2 > 1. \tag{A109}$$

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 (A109) 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 $2\sqrt{\tau_M} > 0$, 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

$$\sigma_M^L \equiv \frac{1}{\sqrt{\tau_M^b}}, \qquad \sigma_M^H \equiv \frac{1}{\sqrt{\tau_M^a}}.$$

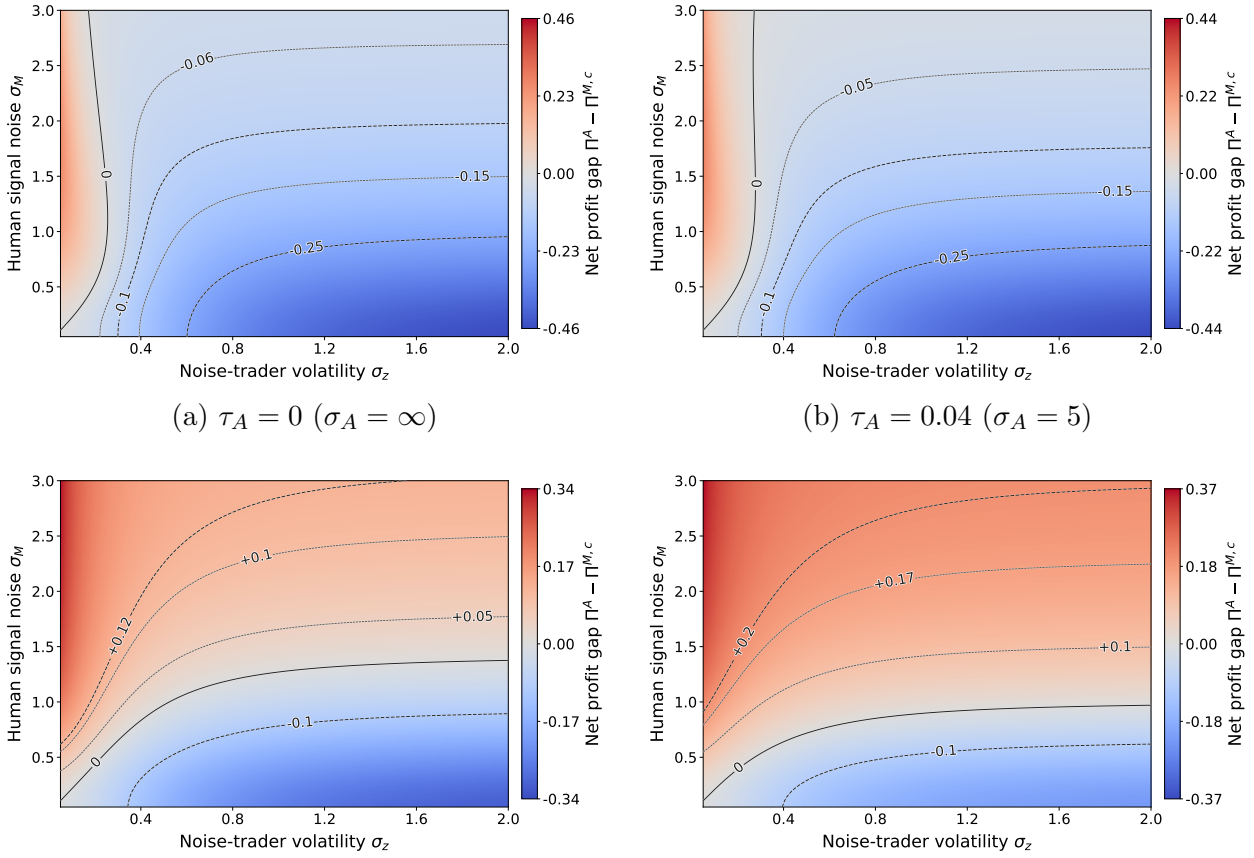
Then $\Pi^{M,c} < \Pi^A$ if and only if $\sigma_M \in (\sigma_M^L, \sigma_M^H)$.

□

C.3 Finite-Level Robustness Figures

The main text uses level 3 as a low-sophistication example. The following figures repeat each comparison at levels 2 and 4. The patterns are stable across these nearby levels.

Figure A2: AI and Cursed-Investor Profit Gaps with Level-2 Investors



(a) $\tau_A = 0$ ($\sigma_A = \infty$)

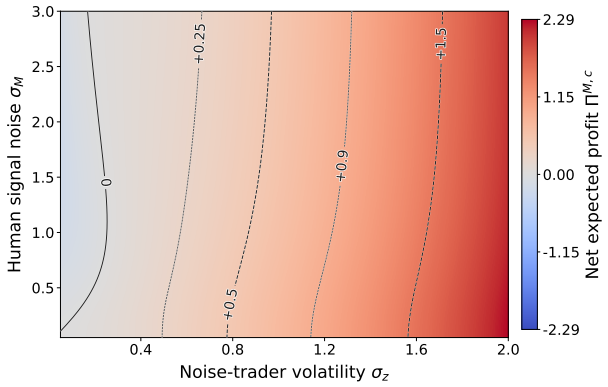
(b) $\tau_A = 0.04$ ($\sigma_A = 5$)

(c) $\tau_A = 0.5$ ($\sigma_A \simeq 1.41$)

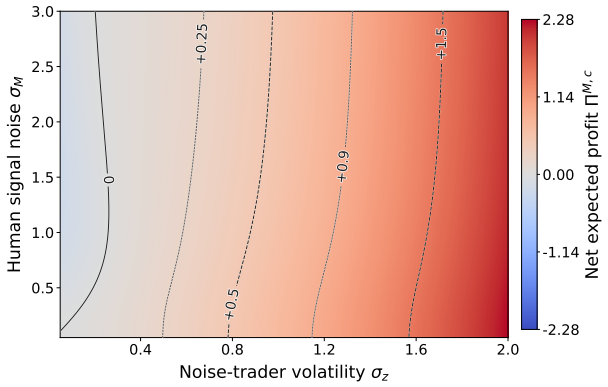
(d) $\tau_A = 1$ ($\sigma_A = 1$)

Notes: Each panel plots $\Pi^A - \Pi^{M,c}$ in a market with level-2 strategic investors. The horizontal axis is noise-trader volatility σ_z, and the vertical axis is human private-signal noise σ_M. Red regions favor AI, blue regions favor cursed investors, and the solid black contour marks equal profits. Panels (a)–(d) set $\tau_A \in \{0, 0.04, 0.5, 1\}$, corresponding to $\sigma_A \in \{\infty, 5, \sqrt{2}, 1\}$. Parameters are $\bar{v} = 0$, $S = 0$, $\sigma_v = 1$, $\gamma_M = \gamma_A = 1$, $(\psi_c, \psi_s, \phi) = (0.30, 0.30, 0.40)$, $\sigma_z \in [0.05, 2]$, and $\sigma_M \in [0.05, 3]$.

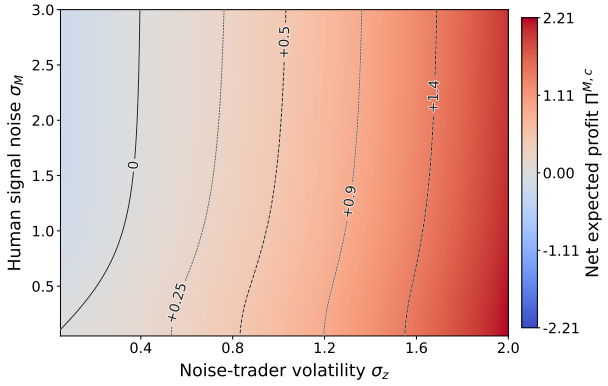
Figure A3: Cursed-Investor Profits with Level-2 Investors



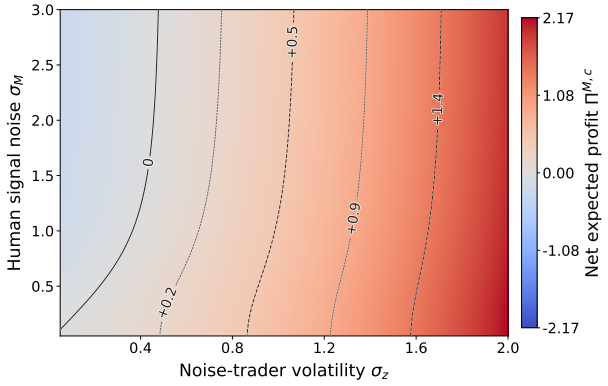
(a) $\tau_A = 0$ $(\sigma_A = \infty)$



(b) $\tau_A = 0.04$ $(\sigma_A = 5)$



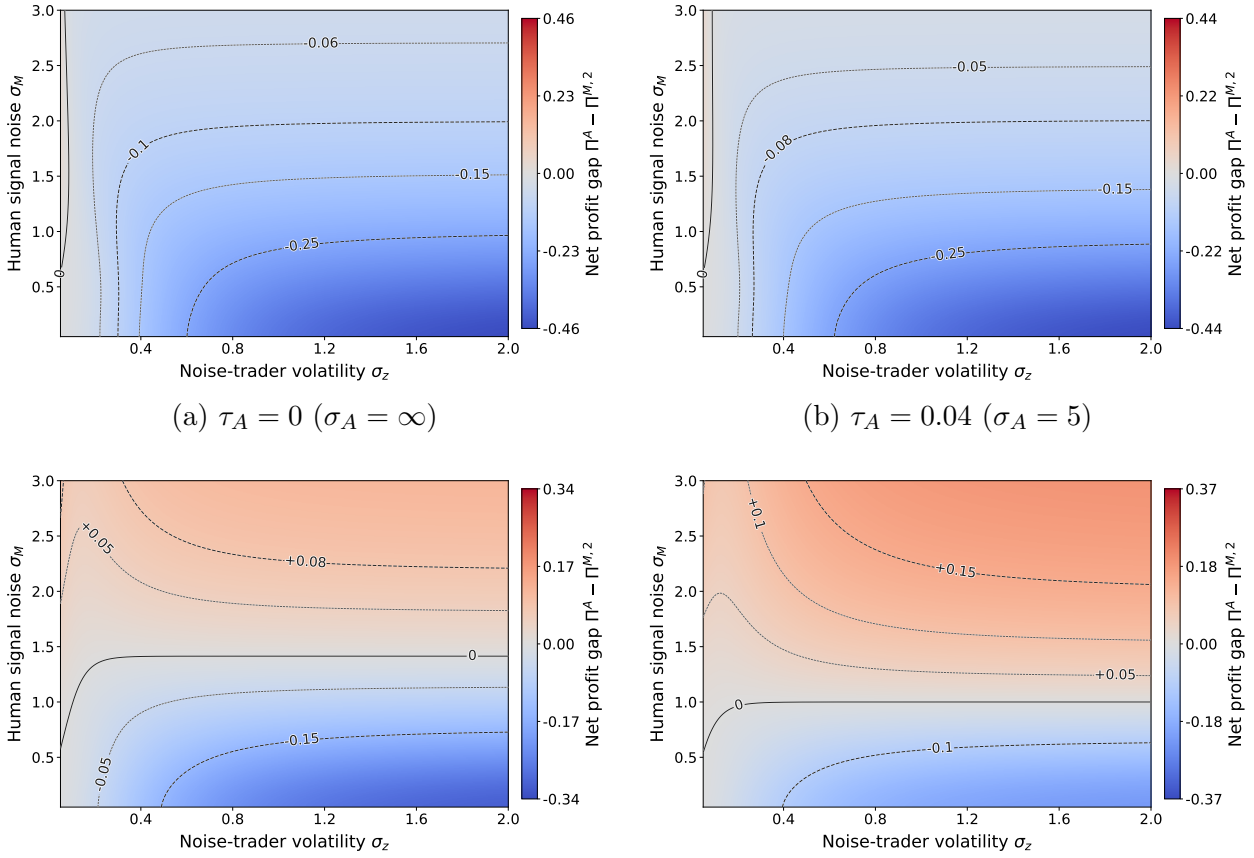
(c) $\tau_A = 0.5$ $(\sigma_A \simeq 1.41)$



(d) $\tau_A = 1$ $(\sigma_A = 1)$

Notes: Each panel plots $\Pi^{M,c}$ in a market with level-2 strategic investors. The horizontal axis is noise-trader volatility σ_z, and the vertical axis is human private-signal noise σ_M. Red regions indicate positive profit, blue regions indicate negative profit, and the solid black contour marks zero profit. Panels (a)–(d) set $\tau_A \in \{0, 0.04, 0.5, 1\}$, corresponding to $\sigma_A \in \{\infty, 5, \sqrt{2}, 1\}$. Parameters are $\bar{v} = 0$, $S = 0$, $\sigma_v = 1$, $\gamma_M = \gamma_A = 1$, $(\psi_c, \psi_s, \phi) = (0.30, 0.30, 0.40)$, $\sigma_z \in [0.05, 2]$, and $\sigma_M \in [0.05, 3]$.

Figure A4: AI and Level-2 Profit Gaps



(a) $\tau_A = 0$ ($\sigma_A = \infty$)

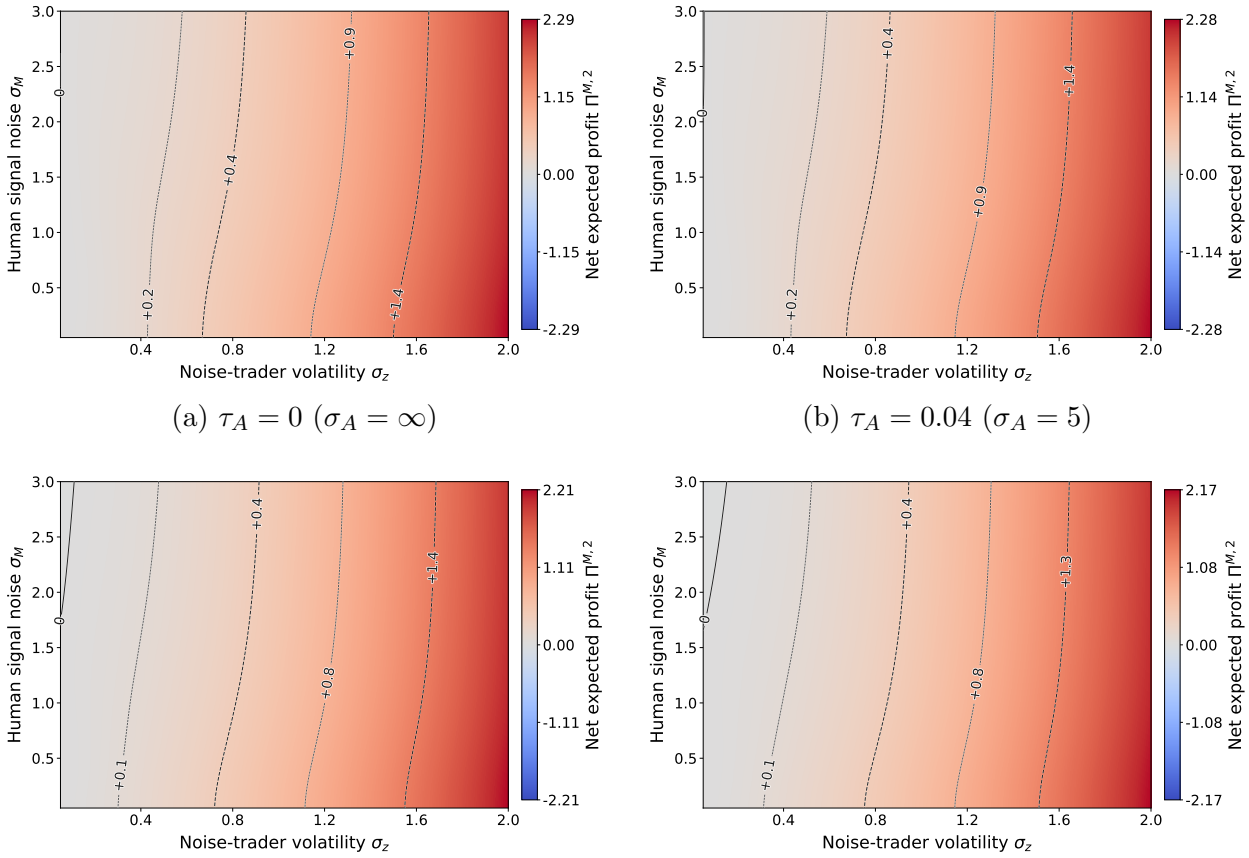
(b) $\tau_A = 0.04$ ($\sigma_A = 5$)

(c) $\tau_A = 0.5$ ($\sigma_A \simeq 1.41$)

(d) $\tau_A = 1$ ($\sigma_A = 1$)

Notes: Each panel plots $\Pi^A - \Pi^{M,2}$ in a market with level-2 strategic investors. The horizontal axis is noise-trader volatility σ_z, and the vertical axis is human private-signal noise σ_M. Red regions favor AI, blue regions favor level-2 investors, and the solid black contour marks equal profits. Panels (a)–(d) set $\tau_A \in \{0, 0.04, 0.5, 1\}$, corresponding to $\sigma_A \in \{\infty, 5, \sqrt{2}, 1\}$. Parameters are $\bar{v} = 0$, $S = 0$, $\sigma_v = 1$, $\gamma_M = \gamma_A = 1$, $(\psi_c, \psi_s, \phi) = (0.30, 0.30, 0.40)$, $\sigma_z \in [0.05, 2]$, and $\sigma_M \in [0.05, 3]$.

Figure A5: Level-2 Investor Profits



(a) $\tau_A = 0$ $(\sigma_A = \infty)$

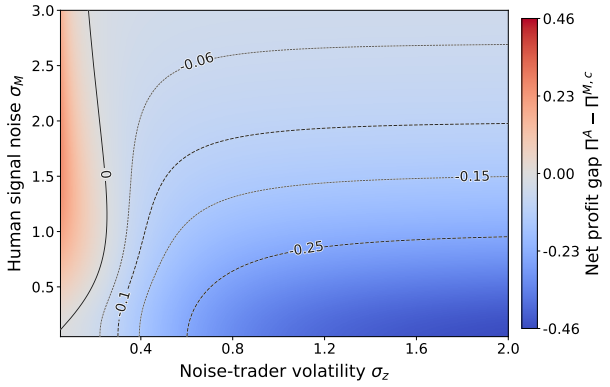
(b) $\tau_A = 0.04$ $(\sigma_A = 5)$

(c) $\tau_A = 0.5$ $(\sigma_A \simeq 1.41)$

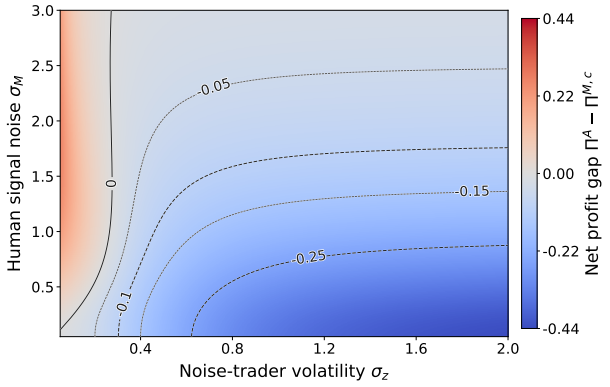
(d) $\tau_A = 1$ $(\sigma_A = 1)$

Notes: Each panel plots $\Pi^{M,2}$ in a market with level-2 strategic investors. The horizontal axis is noise-trader volatility σ_z, and the vertical axis is human private-signal noise σ_M. Red regions indicate positive profit, blue regions indicate negative profit, and the solid black contour marks zero profit. Panels (a)–(d) set $\tau_A \in \{0, 0.04, 0.5, 1\}$, corresponding to $\sigma_A \in \{\infty, 5, \sqrt{2}, 1\}$. Parameters are $\bar{v} = 0$, $S = 0$, $\sigma_v = 1$, $\gamma_M = \gamma_A = 1$, $(\psi_c, \psi_s, \phi) = (0.30, 0.30, 0.40)$, $\sigma_z \in [0.05, 2]$, and $\sigma_M \in [0.05, 3]$.

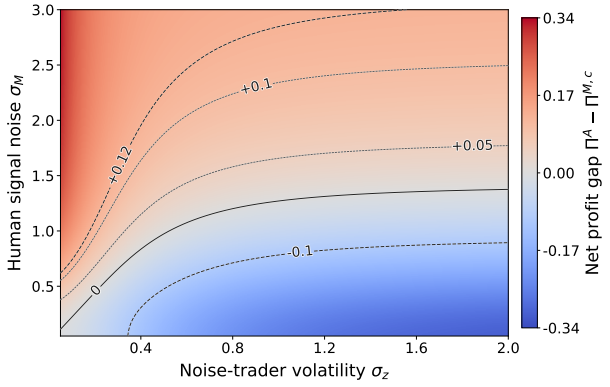
Figure A6: AI and Cursed-Investor Profit Gaps with Level-4 Investors



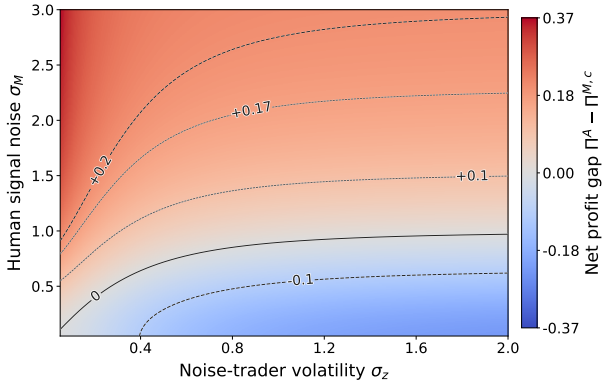
(a) $\tau_A = 0$ ($\sigma_A = \infty$)



(b) $\tau_A = 0.04$ ($\sigma_A = 5$)



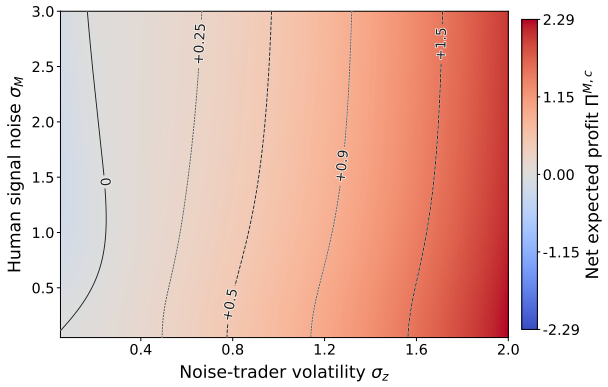
(c) $\tau_A = 0.5$ ($\sigma_A \simeq 1.41$)



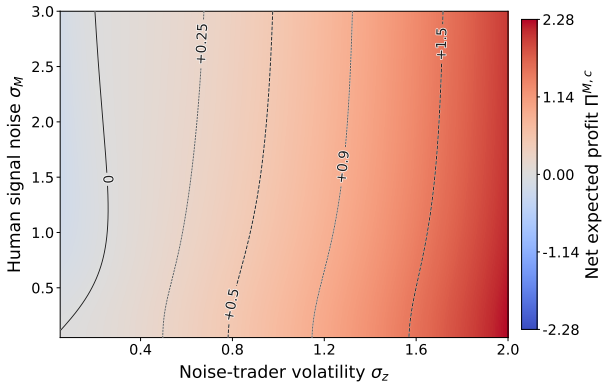
(d) $\tau_A = 1$ ($\sigma_A = 1$)

Notes: Each panel plots $\Pi^A - \Pi^{M,c}$ in a market with level-4 strategic investors. The horizontal axis is noise-trader volatility σ_z, and the vertical axis is human private-signal noise σ_M. Red regions favor AI, blue regions favor cursed investors, and the solid black contour marks equal profits. Panels (a)–(d) set $\tau_A \in \{0, 0.04, 0.5, 1\}$, corresponding to $\sigma_A \in \{\infty, 5, \sqrt{2}, 1\}$. Parameters are $\bar{v} = 0$, $S = 0$, $\sigma_v = 1$, $\gamma_M = \gamma_A = 1$, $(\psi_c, \psi_s, \phi) = (0.30, 0.30, 0.40)$, $\sigma_z \in [0.05, 2]$, and $\sigma_M \in [0.05, 3]$.

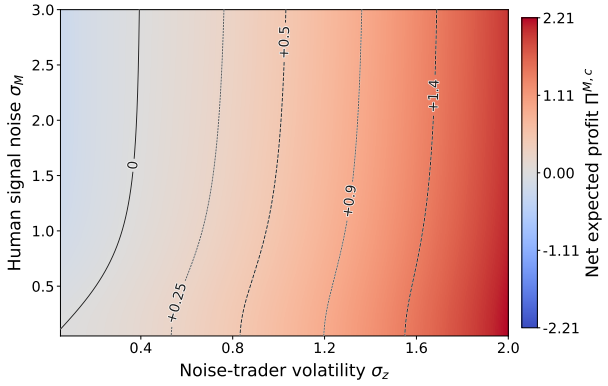
Figure A7: Cursed-Investor Profits with Level-4 Investors



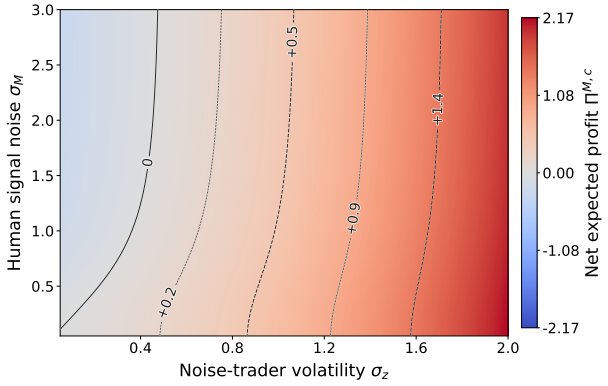
(a) $\tau_A = 0$ $(\sigma_A = \infty)$



(b) $\tau_A = 0.04$ $(\sigma_A = 5)$



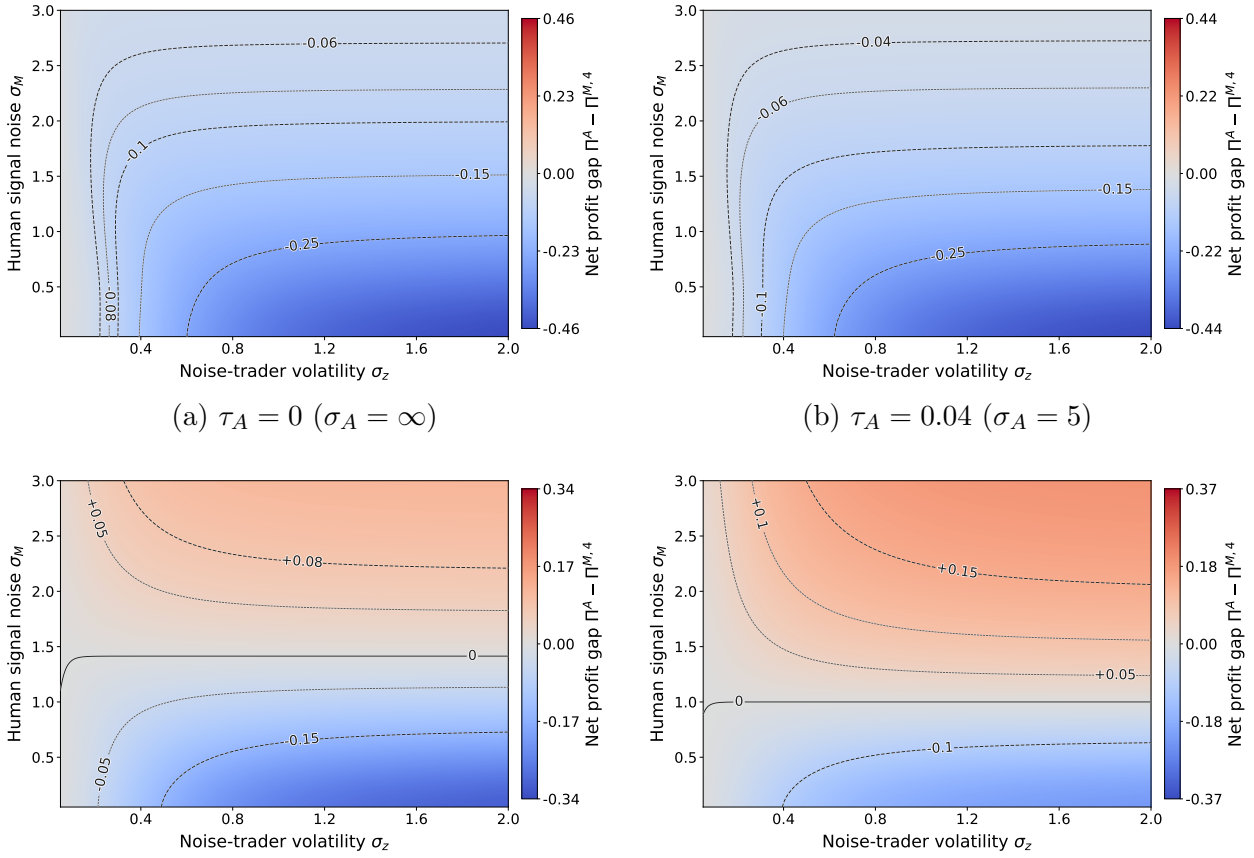
(c) $\tau_A = 0.5$ $(\sigma_A \simeq 1.41)$



(d) $\tau_A = 1$ $(\sigma_A = 1)$

Notes: Each panel plots $\Pi^{M,c}$ in a market with level-4 strategic investors. The horizontal axis is noise-trader volatility σ_z, and the vertical axis is human private-signal noise σ_M. Red regions indicate positive profit, blue regions indicate negative profit, and the solid black contour marks zero profit. Panels (a)–(d) set $\tau_A \in \{0, 0.04, 0.5, 1\}$, corresponding to $\sigma_A \in \{\infty, 5, \sqrt{2}, 1\}$. Parameters are $\bar{v} = 0$, $S = 0$, $\sigma_v = 1$, $\gamma_M = \gamma_A = 1$, $(\psi_c, \psi_s, \phi) = (0.30, 0.30, 0.40)$, $\sigma_z \in [0.05, 2]$, and $\sigma_M \in [0.05, 3]$.

Figure A8: AI and Level-4 Profit Gaps



(a) $\tau_A = 0$ ($\sigma_A = \infty$)

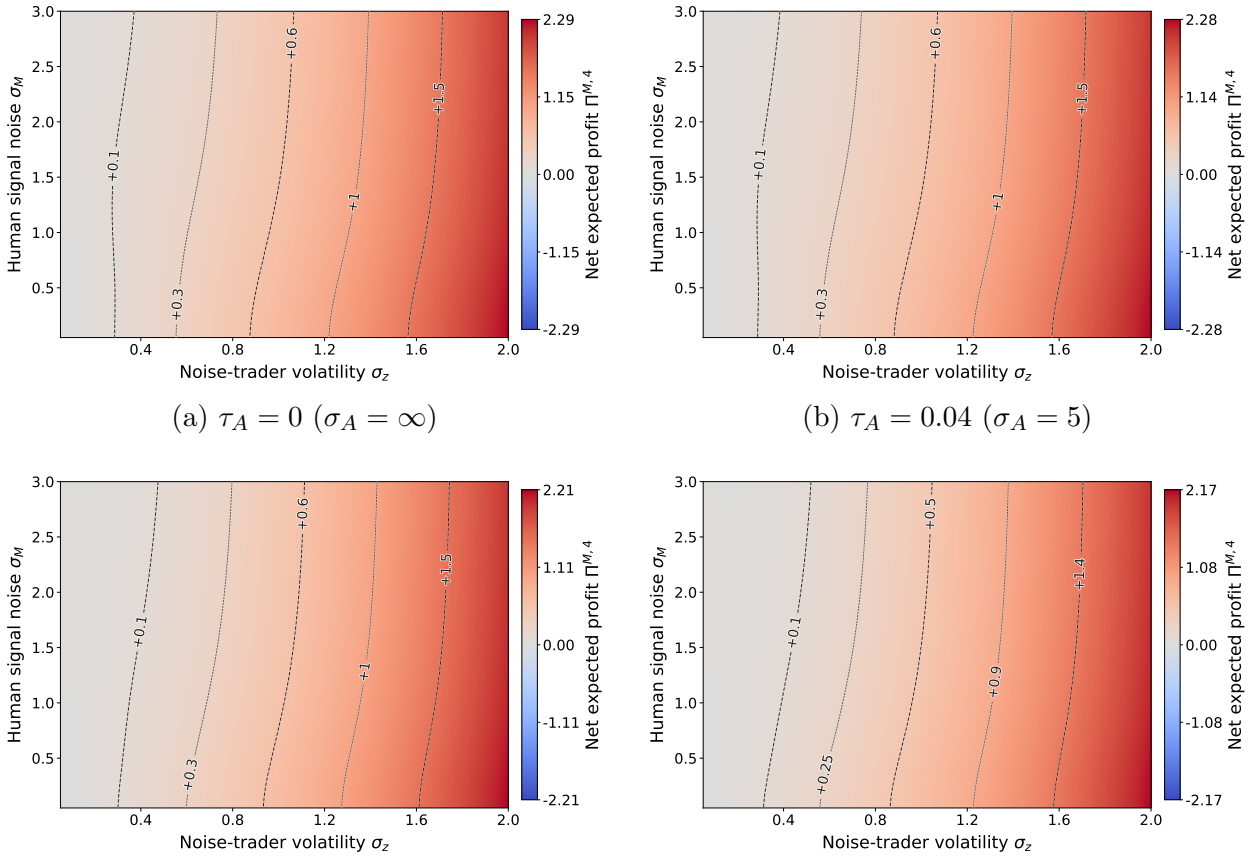
(b) $\tau_A = 0.04$ ($\sigma_A = 5$)

(c) $\tau_A = 0.5$ ($\sigma_A \simeq 1.41$)

(d) $\tau_A = 1$ ($\sigma_A = 1$)

Notes: Each panel plots $\Pi^A - \Pi^{M,4}$ in a market with level-4 strategic investors. The horizontal axis is noise-trader volatility σ_z, and the vertical axis is human private-signal noise σ_M. Red regions favor AI, blue regions favor level-4 investors, and the solid black contour marks equal profits. Panels (a)–(d) set $\tau_A \in \{0, 0.04, 0.5, 1\}$, corresponding to $\sigma_A \in \{\infty, 5, \sqrt{2}, 1\}$. Parameters are $\bar{v} = 0$, $S = 0$, $\sigma_v = 1$, $\gamma_M = \gamma_A = 1$, $(\psi_c, \psi_s, \phi) = (0.30, 0.30, 0.40)$, $\sigma_z \in [0.05, 2]$, and $\sigma_M \in [0.05, 3]$.

Figure A9: Level-4 Investor Profits



(a) $\tau_A = 0$ $(\sigma_A = \infty)$

(b) $\tau_A = 0.04$ $(\sigma_A = 5)$

(c) $\tau_A = 0.5$ $(\sigma_A \simeq 1.41)$

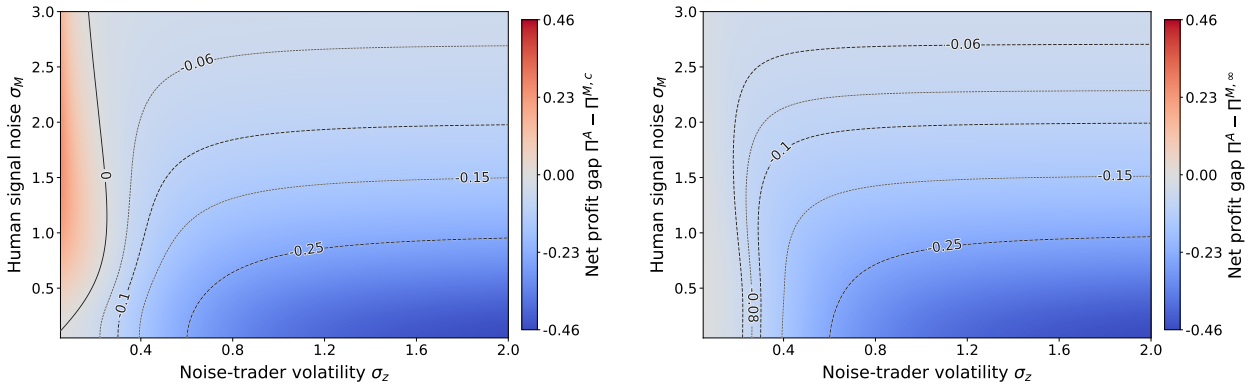
(d) $\tau_A = 1$ $(\sigma_A = 1)$

Notes: Each panel plots $\Pi^{M,4}$ in a market with level-4 strategic investors. The horizontal axis is noise-trader volatility σ_z, and the vertical axis is human private-signal noise σ_M. Red regions indicate positive profit, blue regions indicate negative profit, and the solid black contour marks zero profit. Panels (a)–(d) set $\tau_A \in \{0, 0.04, 0.5, 1\}$, corresponding to $\sigma_A \in \{\infty, 5, \sqrt{2}, 1\}$. Parameters are $\bar{v} = 0$, $S = 0$, $\sigma_v = 1$, $\gamma_M = \gamma_A = 1$, $(\psi_c, \psi_s, \phi) = (0.30, 0.30, 0.40)$, $\sigma_z \in [0.05, 2]$, and $\sigma_M \in [0.05, 3]$.

C.4 Profit Gaps with Uninformed AI at the Convergent Limit

The limiting profit gaps preserve the main information mechanisms from the level-3 market. AI dominance over cursed investors remains confined to a low-noise, intermediate-signal window. Against strategic humans, convergence removes the small finite-level AI advantage: level-∞ investors outperform uninformed AI throughout because they observe the same price and retain a private signal.

Figure A10: Profit Gaps with Uninformed AI and Infinite Strategic Reasoning



(a) AI minus cursed-investor profit

(b) AI minus level-∞ profit

Notes: Panel (a) plots $\Pi^A - \Pi^{M,c}$, and panel (b) plots $\Pi^A - \Pi^{M,\infty}$, in the convergent market in which AI receives no fundamental signal ($\tau_A = 0$). Red regions favor AI, blue regions favor the human comparator, and the solid black contour marks equal profits. The horizontal axis is noise-trader volatility σ_z, and the vertical axis is human private-signal noise σ_M. Parameters are $\bar{v} = 0$, $S = 0$, $\sigma_v = 1$, $\gamma_M = \gamma_A = 1$, and $(\psi_c, \psi_s, \phi) = (0.30, 0.30, 0.40)$, with $\sigma_z \in [0.05, 2]$ and $\sigma_M \in [0.05, 3]$.

D Training Sophistication and Test Profits

D.1 Rational-Expectations Test Environment

Throughout this appendix, $\tau_A = 0$. The training and deployment exercises therefore isolate the sophistication embedded in an AI rule learned without a fundamental signal.

For each training level k, take the pre-trained AI algorithm

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

as given in the deployment market. This market retains cursed investors and replaces the level-k population by rational-expectations (RE) investors of measure ψ_s. These investors know the pre-trained AI demand and best respond to the equilibrium price that it induces. Let $\psi \equiv \psi_c + \psi_s$, $\widehat{\tau}^c \equiv \tau_v + \tau_M$, and define

$$\zeta^c \equiv \psi_c \beta_\eta^{M,c}, \qquad \chi^s \equiv \frac{\psi_s \tau_M}{\gamma_M}, \qquad G_c(\tau_p) \equiv \tau_z \left(\zeta^c + \frac{\chi^s}{\widehat{\tau}^c + \tau_p} \right)^2. \tag{A110}$$

Here $\beta_\eta^{M,c} = \tau_M/(\gamma_M \widehat{\tau}^c)$, $\beta_p^{M,c} = 1/\gamma_M$, and $\mu^{M,c} = \tau_v \bar{v}/(\gamma_M \widehat{\tau}^c)$.

Proposition 34 (Unique rational-expectations equilibrium)**.** *Suppose $\psi_s > 0$. The equilibrium price precision is the unique positive fixed point*

$$\tau_p^{\mathrm{RE}} = G_c(\tau_p^{\mathrm{RE}}) = \tau_z \left[\psi_c \beta_\eta^{M,c} + \frac{\psi_s \tau_M}{\gamma_M (\widehat{\tau}^c + \tau_p^{\mathrm{RE}})} \right]^2. \tag{A111}$$

It does not depend on either pre-trained AI coefficient. Define

$$\zeta^{\mathrm{RE}} \equiv \sqrt{\tau_p^{\mathrm{RE}}/\tau_z}, \qquad \widehat{\tau}^{M,\mathrm{RE}} \equiv \widehat{\tau}^c + \tau_p^{\mathrm{RE}}, \tag{A112}$$

$$h_p^{M,\mathrm{RE}} \equiv \frac{\tau_p^{\mathrm{RE}}}{\widehat{\tau}^{M,\mathrm{RE}}}, \qquad h_v^{M,\mathrm{RE}} \equiv \frac{\tau_v}{\widehat{\tau}^{M,\mathrm{RE}}}, \tag{A113}$$

$$\Lambda^{\mathrm{RE}} \equiv 1 + \frac{\psi_s h_p^{M,\mathrm{RE}}}{\gamma_M \zeta^{\mathrm{RE}}}. \tag{A114}$$

If

$$\frac{\psi}{\gamma_M} + \phi \beta_p^{A,k} \neq 0, \tag{A115}$$

the unique rational-expectations equilibrium has aggregate coefficients

$$\xi_{k\to\mathrm{RE}} = \frac{\psi/\gamma_M + \phi \beta_p^{A,k}}{\Lambda^{\mathrm{RE}}}, \qquad \mu_{k\to\mathrm{RE}} = \frac{\psi_c \mu^{M,c} + \psi_s h_v^{M,\mathrm{RE}} \bar{v}/\gamma_M + \phi \mu^{A,k} - S}{\Lambda^{\mathrm{RE}}}, \tag{A116}$$

and price

$$p_{k\to\mathrm{RE}} = \frac{\zeta^{\mathrm{RE}}}{\xi_{k\to\mathrm{RE}}} v + \frac{z}{\xi_{k\to\mathrm{RE}}} + \frac{\mu_{k\to\mathrm{RE}}}{\xi_{k\to\mathrm{RE}}}. \tag{A117}$$

The three RE demand coefficients are

$$\beta_\eta^{M,\mathrm{RE}} = \frac{\tau_M}{\gamma_M \widehat{\tau}^{M,\mathrm{RE}}}, \tag{A118}$$

$$\beta_{p,k}^{M,\mathrm{RE}} = \frac{1 - h_p^{M,\mathrm{RE}} \xi_{k\to\mathrm{RE}}/\zeta^{\mathrm{RE}}}{\gamma_M}, \tag{A119}$$

$$\mu_k^{M,\mathrm{RE}} = \frac{h_v^{M,\mathrm{RE}} \bar{v} - h_p^{M,\mathrm{RE}} \mu_{k\to\mathrm{RE}}/\zeta^{\mathrm{RE}}}{\gamma_M}. \tag{A120}$$

Proof. Conjecture an equilibrium price

$$p = \frac{\zeta}{\xi} v + \frac{z}{\xi} + \frac{\mu}{\xi}, \qquad \zeta > 0, \quad \xi \neq 0. \tag{A121}$$

The price is equivalent to the signal

$$s_p \equiv \frac{\xi}{\zeta}\left(p - \frac{\mu}{\xi}\right) = v + \frac{z}{\zeta}, \qquad \tau_p = \zeta^2 \tau_z. \tag{A122}$$

Hence an RE investor's posterior mean is

$$\mathbb{E}[v \mid \eta_i, p] = \frac{\tau_M}{\widehat{\tau}^c + \tau_p}\eta_i + \frac{\tau_p}{\widehat{\tau}^c + \tau_p} s_p + \frac{\tau_v}{\widehat{\tau}^c + \tau_p}\bar{v}. \tag{A123}$$

Substituting (A122) into $x_i^{M,\mathrm{RE}} = (\mathbb{E}[v \mid \eta_i, p] - p)/\gamma_M$ gives the three coefficients in (A118)–(A120), with the candidate aggregate coefficients in place of their equilibrium values.

The aggregate coefficient on v must therefore satisfy

$$\zeta = \psi_c \beta_\eta^{M,c} + \frac{\psi_s \tau_M}{\gamma_M(\widehat{\tau}^c + \zeta^2 \tau_z)}. \tag{A124}$$

Multiplying by $\sqrt{\tau_z}$ and squaring gives $\tau_p = G_c(\tau_p)$. The map G_c is continuous, positive, and strictly decreasing when $\psi_s > 0$. Moreover, $G_c(0) > 0$ and $G_c(\tau_p) - \tau_p \to -\infty$ as $\tau_p \to \infty$. Thus (A111) has exactly one positive solution, which fixes ζ^{RE}.

Next aggregate the coefficients on price. Using (A119),

$$\begin{aligned}\xi_{k\to\mathrm{RE}} &= \frac{\psi_c}{\gamma_M} + \psi_s \beta_{p,k}^{M,\mathrm{RE}} + \phi \beta_p^{A,k} \\ &= \frac{\psi}{\gamma_M} + \phi \beta_p^{A,k} - \frac{\psi_s h_p^{M,\mathrm{RE}}}{\gamma_M \zeta^{\mathrm{RE}}} \xi_{k\to\mathrm{RE}}.\end{aligned}$$

This linear equation gives the first expression in (A116). Aggregating intercepts and using (A120) similarly gives

$$\begin{aligned}\mu_{k\to\mathrm{RE}} &= \psi_c\mu^{M,c} + \psi_s\mu_k^{M,\mathrm{RE}} + \phi\mu^{A,k} - S \\ &= \psi_c\mu^{M,c} + \frac{\psi_s h_v^{M,\mathrm{RE}}\bar{v}}{\gamma_M} + \phi\mu^{A,k} - S - \frac{\psi_s h_p^{M,\mathrm{RE}}}{\gamma_M\zeta^{\mathrm{RE}}}\mu_{k\to\mathrm{RE}},\end{aligned}$$

which gives the second expression. Market clearing is then

$$\xi_{k\to\mathrm{RE}}p = \zeta^{\mathrm{RE}}v + z + \mu_{k\to\mathrm{RE}},$$

so the price used to form the posterior is exactly the price generated by the resulting demands.

The precision equation has one solution, and the two linear coefficient equations are non-singular. These equations therefore uniquely determine the equilibrium price and demands. If $\psi/\gamma_M + \phi\beta_p^{A,k} = 0$, the price-sensitivity equation forces $\xi = 0$; because $\sigma_z > 0$, no finite price can then clear the market almost surely. □

D.1.1 Profits and pre-trained-algorithm regret

Apply Proposition 7 using equilibrium price coefficients $\xi_{k\to\mathrm{RE}}$, ζ^{RE}, and $\mu_{k\to\mathrm{RE}}$. Substitute the relevant human coefficients and the pre-trained AI coefficients. To isolate the loss from deploying the algorithm without further updating, define

$$\begin{aligned}\alpha_k &\equiv \frac{\zeta^{\mathrm{RE}}}{\xi_{k\to\mathrm{RE}}}, & c_k &\equiv \frac{\mu_{k\to\mathrm{RE}}}{\xi_{k\to\mathrm{RE}}}, \\ \bar{p}_k &\equiv \mathbb{E}[p_{k\to\mathrm{RE}}] = \alpha_k\bar{v} + c_k, & \delta_k &\equiv \bar{v} - \bar{p}_k, \\ V_{p,k} &\equiv \mathrm{Var}(p_{k\to\mathrm{RE}}) = \alpha_k^2\sigma_v^2 + \frac{\sigma_z^2}{\xi_{k\to\mathrm{RE}}^2}.\end{aligned} \tag{A125}$$

Let

$$h_p^{A,\mathrm{RE}} \equiv \frac{\tau_p^{\mathrm{RE}}}{\tau_v + \tau_p^{\mathrm{RE}}}. \tag{A126}$$

An atomistic AI without fundamental signals that reoptimizes while holding the actual price law (A117) fixed chooses

$$x_k^{A,\mathrm{BR}}(p) = \frac{\mathbb{E}[v - p \mid p]}{\gamma_A} = -\beta_{p,k}^{A,\mathrm{BR}}p + \mu_k^{A,\mathrm{BR}}, \tag{A127}$$

where

$$\beta_{p,k}^{A,\mathrm{BR}} = \frac{1 - h_p^{A,\mathrm{RE}} \xi_{k\to\mathrm{RE}}/\zeta^{\mathrm{RE}}}{\gamma_A}, \tag{A128}$$

$$\mu_k^{A,\mathrm{BR}} = \frac{(1 - h_p^{A,\mathrm{RE}})\bar{v} - h_p^{A,\mathrm{RE}} \mu_{k\to\mathrm{RE}}/\zeta^{\mathrm{RE}}}{\gamma_A}. \tag{A129}$$

This is an atomistic counterfactual: it changes the AI's holding at the given price but does not re-clear the price after changing that holding.

Proposition 35 (Exact deployment regret)**.** *The pre-trained AI profit satisfies*

$$\Pi_{k\to\mathrm{RE}}^{A} = \Pi_{k\to\mathrm{RE}}^{A,\mathrm{BR}} - \mathcal{R}_k, \tag{A130}$$

where

$$\Pi_{k\to\mathrm{RE}}^{A,\mathrm{BR}} = \frac{1}{2\gamma_A}\left\{\delta_k^2 + \frac{(h_p^{A,\mathrm{RE}} - \alpha_k)^2}{h_p^{A,\mathrm{RE}}}\sigma_v^2\right\}. \tag{A131}$$

Writing the slope and intercept errors as $\varepsilon_{\beta,k} \equiv \beta_p^{A,k} - \beta_{p,k}^{A,\mathrm{BR}}$ *and* $\varepsilon_{\mu,k} \equiv \mu^{A,k} - \mu_k^{A,\mathrm{BR}}$*, the exact noncentered regret is*

$$\mathcal{R}_k = \frac{\gamma_A}{2}\left\{(\varepsilon_{\mu,k} - \varepsilon_{\beta,k}\bar{p}_k)^2 + \varepsilon_{\beta,k}^2 V_{p,k}\right\} \geq 0. \tag{A132}$$

Regret is zero if and only if the pre-trained algorithm equals the best response of an AI without fundamental signals almost surely. In the centered model, $\bar{v} = S = 0$ *implies* $\mu^{A,k} = \mu_{k\to\mathrm{RE}} = \mu_k^{A,\mathrm{BR}} = \bar{p}_k = 0$*, so*

$$\mathcal{R}_k = \frac{\gamma_A}{2}(\beta_p^{A,k} - \beta_{p,k}^{A,\mathrm{BR}})^2 V_{p,k}. \tag{A133}$$

Proof. Conditional on price, completing the square gives

$$\mathbb{E}[v - p \mid p]x - \frac{\gamma_A}{2}x^2 = \frac{\mathbb{E}[v - p \mid p]^2}{2\gamma_A} - \frac{\gamma_A}{2}\left(x - \frac{\mathbb{E}[v - p \mid p]}{\gamma_A}\right)^2. \tag{A134}$$

Taking expectations at the pre-trained rule proves (A130) and (A132). The price signal implies

$$\mathbb{E}[v - p \mid p] = \delta_k + \left(\frac{h_p^{A,\mathrm{RE}}}{\alpha_k} - 1\right)(p - \bar{p}_k).$$

Its demand coefficients are (A128) and (A129). Finally,

$$\frac{V_{p,k}}{\alpha_k^2} = \sigma_v^2 + \frac{1}{\tau_p^{\mathrm{RE}}} = \frac{\sigma_v^2}{h_p^{A,\mathrm{RE}}},$$

which gives (A131). The centered formula is the corresponding specialization of (A132). □

D.1.2 Profit gaps and the regularity condition

Proof of Lemma 13. First impose only equal trading costs, $\gamma_M = \gamma_A \equiv \gamma$. The counterfactual best-response AI and a cursed investor face the same price. Substitution into Proposition 7 gives the information-value difference

$$\begin{aligned}\Pi^{A,\mathrm{BR}}_{k\to\mathrm{RE}} - \Pi^{M,c}_{k\to\mathrm{RE}} &= \frac{1}{2\gamma\tau_v}\left[\frac{\tau_p^{\mathrm{RE}}}{\tau_v+\tau_p^{\mathrm{RE}}} - \frac{\tau_M}{\tau_v+\tau_M}\right] \\ &= \frac{\tau_p^{\mathrm{RE}}-\tau_M}{2\gamma(\tau_v+\tau_M)(\tau_v+\tau_p^{\mathrm{RE}})}. \end{aligned} \tag{A135}$$

An RE investor adds a private signal to the price information of the counterfactual AI. The value of that signal is the reduction in posterior variance:

$$\begin{aligned}\Pi^{M,\mathrm{RE}}_{k\to\mathrm{RE}} - \Pi^{A,\mathrm{BR}}_{k\to\mathrm{RE}} &= \frac{1}{2\gamma}\left[\frac{1}{\tau_v+\tau_p^{\mathrm{RE}}} - \frac{1}{\tau_v+\tau_M+\tau_p^{\mathrm{RE}}}\right] \\ &= \frac{\tau_M}{2\gamma(\tau_v+\tau_p^{\mathrm{RE}})(\tau_v+\tau_M+\tau_p^{\mathrm{RE}})}. \end{aligned} \tag{A136}$$

Subtracting the regret in (A130) yields

$$\Delta_c(k) = \frac{\tau_p^{\mathrm{RE}}-\tau_M}{2\gamma(\tau_v+\tau_M)(\tau_v+\tau_p^{\mathrm{RE}})} - \mathcal{R}_k, \tag{A137}$$

$$\Delta_{\mathrm{RE}}(k) = -\frac{\tau_M}{2\gamma(\tau_v+\tau_p^{\mathrm{RE}})(\tau_v+\tau_M+\tau_p^{\mathrm{RE}})} - \mathcal{R}_k. \tag{A138}$$

The constants do not depend on k, so

$$\Delta_c(k+1) - \Delta_c(k) = \Delta_{\mathrm{RE}}(k+1) - \Delta_{\mathrm{RE}}(k) = \mathcal{R}_k - \mathcal{R}_{k+1}. \tag{A139}$$

Now impose $\bar{v} = S = 0$ and convergence of the training coefficient system. At its limit, the training precision solves (A111); the limiting training market and the deployment market with RE investors therefore coincide. In particular, $\beta_p^{A,\infty}$ is the fixed point of the deployment best-response map. Define

$$a^{\mathrm{RE}} \equiv \frac{h_p^{A,\mathrm{RE}}\phi}{\gamma\zeta^{\mathrm{RE}}\Lambda^{\mathrm{RE}}} \geq 0. \tag{A140}$$

Using (A116) in (A128) shows that the best-response map is linear with slope $-a^{\mathrm{RE}}$. Its

fixed-point property gives

$$\beta_p^{A,k} - \beta_{p,k}^{A,\mathrm{BR}} = (1 + a^{\mathrm{RE}})(\beta_p^{A,k} - \beta_p^{A,\infty}). \tag{A141}$$

Moreover,

$$V_{p,k} = \frac{(\Lambda^{\mathrm{RE}})^2 \left[(\zeta^{\mathrm{RE}})^2 \sigma_v^2 + \sigma_z^2\right]}{(\psi/\gamma + \phi\beta_p^{A,k})^2}. \tag{A142}$$

Combining (A133)– (A142) gives

$$\mathcal{R}_k = C_{\mathrm{RE}} \left(\frac{|\beta_p^{A,k} - \beta_p^{A,\infty}|}{|\psi/\gamma + \phi\beta_p^{A,k}|} \right)^2, \qquad C_{\mathrm{RE}} \equiv \frac{\gamma}{2}(1 + a^{\mathrm{RE}})^2 (\Lambda^{\mathrm{RE}})^2 \left[(\zeta^{\mathrm{RE}})^2 \sigma_v^2 + \sigma_z^2\right] > 0, \tag{A143}$$

Together with (A139), this proves the lemma. □

For the centered equal-cost benchmark, define the deployment-relevant distance

$$d_k \equiv \frac{|\beta_p^{A,k} - \beta_p^{A,\infty}|}{|\psi/\gamma + \phi\beta_p^{A,k}|}. \tag{A144}$$

The regularity condition requires

$$d_{k+1} \leq d_k \qquad \text{for every } k \geq 0. \tag{A145}$$

It requires the pre-trained AI demand to move weakly closer to its limiting deployment-market demand at every training transition, after normalizing for the algorithm's effect on aggregate price sensitivity. A strict inequality at a given transition produces a strict increase in both relative-profit gaps.

Proof of Proposition 14. By Lemma 13 and (A144),

$$\Delta_h(k+1) - \Delta_h(k) = C_{\mathrm{RE}}(d_k^2 - d_{k+1}^2), \qquad h \in \{c, \mathrm{RE}\}.$$

The regularity condition makes this expression nonnegative for every k, and it is strictly positive whenever $d_{k+1} < d_k$. □

Equation (A138) is strictly negative whenever $\tau_M > 0$, so an RE investor always outperforms the pre-trained AI. Equation (A137) shows that the AI strictly outperforms a cursed investor if and only if $\tau_p^{\mathrm{RE}} > \tau_M$ and

$$\mathcal{R}_k < \frac{\tau_p^{\mathrm{RE}} - \tau_M}{2\gamma(\tau_v + \tau_M)(\tau_v + \tau_p^{\mathrm{RE}})}. \tag{A146}$$

Corollary 3 (Sufficient conditions for the regularity condition)**.** *Maintain the assumptions of Lemma 13. Either condition below implies the regularity condition $d_{k+1} \le d_k$ for every k.*

1. *The coefficients $\{\beta_p^{A,k}\}_{k\ge 0}$ and their limit lie in one pole-free interval, and the coefficient path moves monotonically toward $\beta_p^{A,\infty}$.*

2. *The coefficients and their limit lie in a compact pole-free interval I, and for some $\rho \in [0,1)$,*

$$|\beta_p^{A,k+1} - \beta_p^{A,\infty}| \le \rho |\beta_p^{A,k} - \beta_p^{A,\infty}| \qquad \text{for every } k, \tag{A147}$$

$$\rho \frac{\max_{\beta\in I} |\psi/\gamma + \phi\beta|}{\min_{\beta\in I} |\psi/\gamma + \phi\beta|} \le 1. \tag{A148}$$

Strict movement in part 1, or a strict inequality in (A148) with no finite attainment of the limit in part 2, gives $d_{k+1} < d_k$ and hence strict increases in both gaps.

Proof. On a pole-free interval containing $\beta_p^{A,\infty}$, the map

$$\beta \longmapsto \frac{\beta - \beta_p^{A,\infty}}{\psi/\gamma + \phi\beta}$$

has derivative

$$\frac{\psi/\gamma + \phi\beta_p^{A,\infty}}{(\psi/\gamma + \phi\beta)^2},$$

whose sign is constant. Monotone movement toward the limit therefore weakly reduces its absolute value. Under part 2, the ratio d_{k+1}/d_k, whenever $d_k > 0$, is bounded above by the left side of (A148). If $d_k = 0$, the raw contraction keeps the path at the limit. The strict claims follow from the same argument. □

D.1.3 Eventual regularity under strict stability

In the centered model with equal trading costs, write the training state as

$$\nu^k \equiv \frac{\beta_p^{M,k}}{\beta_\eta^{M,k}}, \qquad \zeta^k = \sqrt{\tau_p^k/\tau_z}, \qquad \xi_H^k = \frac{\psi_c}{\gamma} + \nu^k(\zeta^k - \zeta^c). \tag{A149}$$

Its anchors are

$$\tau_p^0 = (\psi\beta_\eta^{M,c})^2\tau_z = G_c(0), \qquad \nu^0 = \frac{\widehat{\tau}^c}{\tau_M}. \tag{A150}$$

The state recursion is

$$\tau_p^{k+1} = G_c(\tau_p^k), \qquad \nu^{k+1} = a_\nu(\tau_p^k) + b_\nu(\tau_p^k)\nu^k, \tag{A151}$$

where

$$a_\nu(\tau_p) \equiv \frac{\widehat{\tau}^c + \tau_p}{\tau_M} - \frac{\tau_p}{\tau_M} \frac{\psi_c + \phi}{\gamma\sqrt{\tau_p/\tau_z} + [\tau_p/(\tau_v + \tau_p)]\phi}, \tag{A152}$$

$$b_\nu(\tau_p) \equiv -\frac{\tau_p}{\tau_M} \frac{\gamma(\sqrt{\tau_p/\tau_z} - \zeta^c)}{\gamma\sqrt{\tau_p/\tau_z} + [\tau_p/(\tau_v + \tau_p)]\phi} < 0. \tag{A153}$$

The trained AI coefficient is the smooth state function

$$\beta_p^A(\tau_p, \nu) = \frac{\sqrt{\tau_p/\tau_z} - h_p^A(\tau_p)\left[\psi_c/\gamma + \nu(\sqrt{\tau_p/\tau_z} - \zeta^c)\right]}{\gamma\sqrt{\tau_p/\tau_z} + h_p^A(\tau_p)\phi}, \qquad h_p^A(\tau_p) \equiv \frac{\tau_p}{\tau_v + \tau_p}. \tag{A154}$$

Let

$$\kappa_c \equiv |G_c'(\tau_p^{\mathrm{RE}})|. \tag{A155}$$

Strict fixed-point stability means

$$\kappa_c < 1, \qquad |b_\nu(\tau_p^{\mathrm{RE}})| < 1. \tag{A156}$$

Proposition 36 (Eventual gap improvement under strict stability)**.** *Maintain the convergence and nondegeneracy assumptions of Lemma 13. Suppose $\psi_s > 0$, (A156) holds, and the training denominators remain nonzero in a neighborhood of the fixed point. Then there are $\lambda_A \in \{-\kappa_c, b_\nu(\tau_p^{\mathrm{RE}})\} \subset (-1, 0)$ and $C_A \neq 0$ such that either*

$$\beta_p^{A,k} - \beta_p^{A,\infty} = C_A \lambda_A^k + o(|\lambda_A|^k), \tag{A157}$$

or, when the two eigenvalues coincide,

$$\beta_p^{A,k} - \beta_p^{A,\infty} = C_A k \lambda_A^k + o(k|\lambda_A|^k). \tag{A158}$$

Consequently, $d_{k+1} < d_k$ for every sufficiently large k, and both profit gaps then increase strictly.

Proof. The Jacobian of (A151) at $(\tau_p^{\mathrm{RE}}, \nu^\infty)$ is lower triangular:

$$\begin{pmatrix} -\kappa_c & 0 \\ a_\nu'(\tau_p^{\mathrm{RE}}) + b_\nu'(\tau_p^{\mathrm{RE}})\nu^\infty & b_\nu(\tau_p^{\mathrm{RE}}) \end{pmatrix}. \tag{A159}$$

Both eigenvalues are negative and strictly inside the unit circle. The anchor satisfies $\tau_p^0 = G_c(0) > G_c(\tau_p^{\mathrm{RE}}) = \tau_p^{\mathrm{RE}}$. Because G_c is strictly decreasing, no finite iterate equals the fixed

point. Smoothness and strict stability then give

$$\tau_p^k - \tau_p^{\mathrm{RE}} = C_\tau(-\kappa_c)^k + o(\kappa_c^k), \qquad C_\tau \neq 0. \tag{A160}$$

Indeed, the ratio of the precision errors at $k+1$ and k is $-\kappa_c + O(|\tau_p^k - \tau_p^{\mathrm{RE}}|)$. The associated infinite product converges to a finite nonzero constant because the strictly stable orbit is summable.

The AI coefficient loads nontrivially on the driven state:

$$\left.\frac{\partial \beta_p^A}{\partial \nu}\right|_{(\tau_p^{\mathrm{RE}}, \nu^\infty)} = -\frac{h_p^{A,\mathrm{RE}}(\zeta^{\mathrm{RE}} - \zeta^c)}{\gamma\zeta^{\mathrm{RE}} + h_p^{A,\mathrm{RE}}\phi} < 0. \tag{A161}$$

Its loading on the informativeness eigendirection cannot cancel either. With all partial derivatives evaluated at $(\tau_p^{\mathrm{RE}}, \nu^\infty)$, direct substitution of the fixed-point identities gives

$$\begin{aligned}
&\frac{\partial \beta_p^A}{\partial \tau_p}[G_c'(\tau_p^{\mathrm{RE}}) - b_\nu(\tau_p^{\mathrm{RE}})] + \frac{\partial \beta_p^A}{\partial \nu}[a_\nu'(\tau_p^{\mathrm{RE}}) + b_\nu'(\tau_p^{\mathrm{RE}})\nu^\infty] \\
&\quad = \frac{\tau_p^{\mathrm{RE}}\tau_v(\zeta^{\mathrm{RE}} - \zeta^c)\left\{\gamma[\tau_p^{\mathrm{RE}}(\zeta^{\mathrm{RE}} - \zeta^c) + \widehat{\tau}^c\zeta^{\mathrm{RE}}] + \phi\tau_M\right\}}{(\widehat{\tau}^c + \tau_p^{\mathrm{RE}})(\tau_v + \tau_p^{\mathrm{RE}})^2(\gamma\zeta^{\mathrm{RE}} + h_p^{A,\mathrm{RE}}\phi)\left[\tau_M(\gamma\zeta^{\mathrm{RE}} + h_p^{A,\mathrm{RE}}\phi) + \tau_p^{\mathrm{RE}}\gamma(\zeta^{\mathrm{RE}} - \zeta^c)\right]} > 0.
\end{aligned} \tag{A162}$$

For the variation-of-constants argument, set $u_k \equiv \tau_p^k - \tau_p^{\mathrm{RE}}$ and $v_k \equiv \nu^k - \nu^\infty$ within this proof. A first-order expansion gives

$$\begin{aligned}
u_{k+1} &= -\kappa_c u_k + O(u_k^2), \\
v_{k+1} &= b_\nu(\tau_p^{\mathrm{RE}})v_k + [a_\nu'(\tau_p^{\mathrm{RE}}) + b_\nu'(\tau_p^{\mathrm{RE}})\nu^\infty]u_k + O(u_k^2 + |u_k v_k|).
\end{aligned}$$

If the eigenvalues are distinct, define

$$w_k \equiv v_k - \frac{a_\nu'(\tau_p^{\mathrm{RE}}) + b_\nu'(\tau_p^{\mathrm{RE}})\nu^\infty}{-\kappa_c - b_\nu(\tau_p^{\mathrm{RE}})}u_k.$$

Then

$$w_{k+1} = b_\nu(\tau_p^{\mathrm{RE}})w_k + O(u_k^2 + |u_k w_k|).$$

Variation of constants gives two possibilities. A nonzero homogeneous driven component leads at rate $|b_\nu(\tau_p^{\mathrm{RE}})|$ when that rate exceeds κ_c; otherwise the forced precision component in (A160) leads. Equation (A161) makes the first loading nonzero, while (A162) makes the second loading nonzero. This proves (A157), including the case in which the nominally

slower homogeneous mode has zero amplitude.

If the eigenvalues coincide at λ_A, variation of constants gives

$$\nu^k - \nu^\infty = \frac{[a'_\nu(\tau_p^{\mathrm{RE}}) + b'_\nu(\tau_p^{\mathrm{RE}})\nu^\infty]C_\tau}{\lambda_A} k\lambda_A^k + o(k|\lambda_A|^k). \tag{A163}$$

At equality of the eigenvalues, the first term on the left side of (A162) vanishes. The Jordan coefficient remains nonzero after multiplication by $\partial\beta_p^A/\partial\nu$, proving (A158). In both cases,

$$\lim_{k\to\infty} \frac{|\beta_p^{A,k+1} - \beta_p^{A,\infty}|}{|\beta_p^{A,k} - \beta_p^{A,\infty}|} = |\lambda_A| < 1. \tag{A164}$$

The denominator in (A144) converges to a nonzero limit by nondegeneracy of the rational-expectations equilibrium. Its ratio at k and $k+1$ therefore converges to one. Equation (A164) implies $d_{k+1} < d_k$ for all sufficiently large k. The common-regret identity (A139) then makes both profit gaps strictly increasing on the same tail. □

Strict stability controls the tail, not the finite prefix. For example, the primitive bound

$$\psi_s \le \frac{27\gamma^2\sigma_z^2}{8\psi\sigma_v^2} \tag{A165}$$

guarantees fixed-point convergence under training-market nondegeneracy, but it does not preclude early overshooting. The regularity condition is therefore stronger than convergence and eventual regularity.

D.2 Level-Infinity Human-Demand Robustness Exercise

The earlier train-on-k, test-at-level-∞ exercise uses a different deployment protocol. It uses the cursed demand and the level-∞ human demand characterized in the previous section; the latter describes the most sophisticated strategic investors in the model. For each k, the pre-trained AI algorithm is deployed without further updating, and the equilibrium price re-clears. To make the distinction explicit, define

$$\begin{aligned}
\zeta^\infty &\equiv \psi_c\beta_\eta^{M,c} + \psi_s\beta_\eta^{M,\infty}, & \xi_H^\infty &\equiv \psi_c\beta_p^{M,c} + \psi_s\beta_p^{M,\infty},\\
\mu_H^\infty &\equiv \psi_c\mu^{M,c} + \psi_s\mu^{M,\infty} - S, & \widetilde{\xi}_k &\equiv \xi_H^\infty + \phi\beta_p^{A,k},\\
\widetilde{\mu}_k &\equiv \mu_H^\infty + \phi\mu^{A,k}.
\end{aligned} \tag{A166}$$

The re-cleared price is

$$\widetilde{p}_k = \frac{\zeta^\infty}{\widetilde{\xi}_k} v + \frac{z}{\widetilde{\xi}_k} + \frac{\widetilde{\mu}_k}{\widetilde{\xi}_k}. \tag{A167}$$

At a finite k, the level-∞ demand need not be optimal for the resulting equilibrium price because it is not reoptimized for each pre-trained AI algorithm. Thus this exercise is a robustness comparison, not a rational-expectations equilibrium. The distinction disappears as $k \to \infty$.

Figure A11 shows patterns similar to the main RE exercise. Both relative-profit gaps rise in the three displayed calibrations, and most of the adjustment occurs between $k = 0$ and $k = 2$. For $(\sigma_M, \sigma_z) = (1, 0.2)$, the AI underperforms cursed investors at $k = 0$ but overtakes them at $k = 1$; its gap relative to level-∞ investors narrows but remains negative. When $(\sigma_M, \sigma_z) = (1, 0.8)$ or $(4, 0.2)$, the AI remains below both human policies. High noise-trader volatility makes prices noisy. Very noisy private signals also weaken the fundamental information embedded in informed demand and hence in the price observed by the AI.

Let $\widetilde{\Pi}^A_{k\to\infty}$ and $\widetilde{\Pi}^{M,h}_{k\to\infty}$ denote the deployment profits of the pre-trained AI and human type $h \in \{c, \infty\}$. The arrow records the training environment and the level-∞ human-demand exercise; it is not a limit operator. Define

$$\widetilde{\Delta}_h(k) \equiv \widetilde{\Pi}^A_{k\to\infty} - \widetilde{\Pi}^{M,h}_{k\to\infty}, \qquad h \in \{c, \infty\}. \tag{A168}$$

Proposition 37 (Monotonic improvement under the regularity condition)**.** *Suppose $\bar{v} = S = 0$, $\gamma_M = \gamma_A \equiv \gamma$, and $\psi_s > 0$. Let the training and deployment markets have the same primitives and population shares, and hold the cursed and level-∞ human demands fixed in the deployment market. Suppose the training recursion is well defined, the level-∞ equilibrium has $\xi^\infty_H > 0$ and $\xi^\infty > 0$, and every induced equilibrium is nondegenerate. If the regularity condition (A190) holds, then*

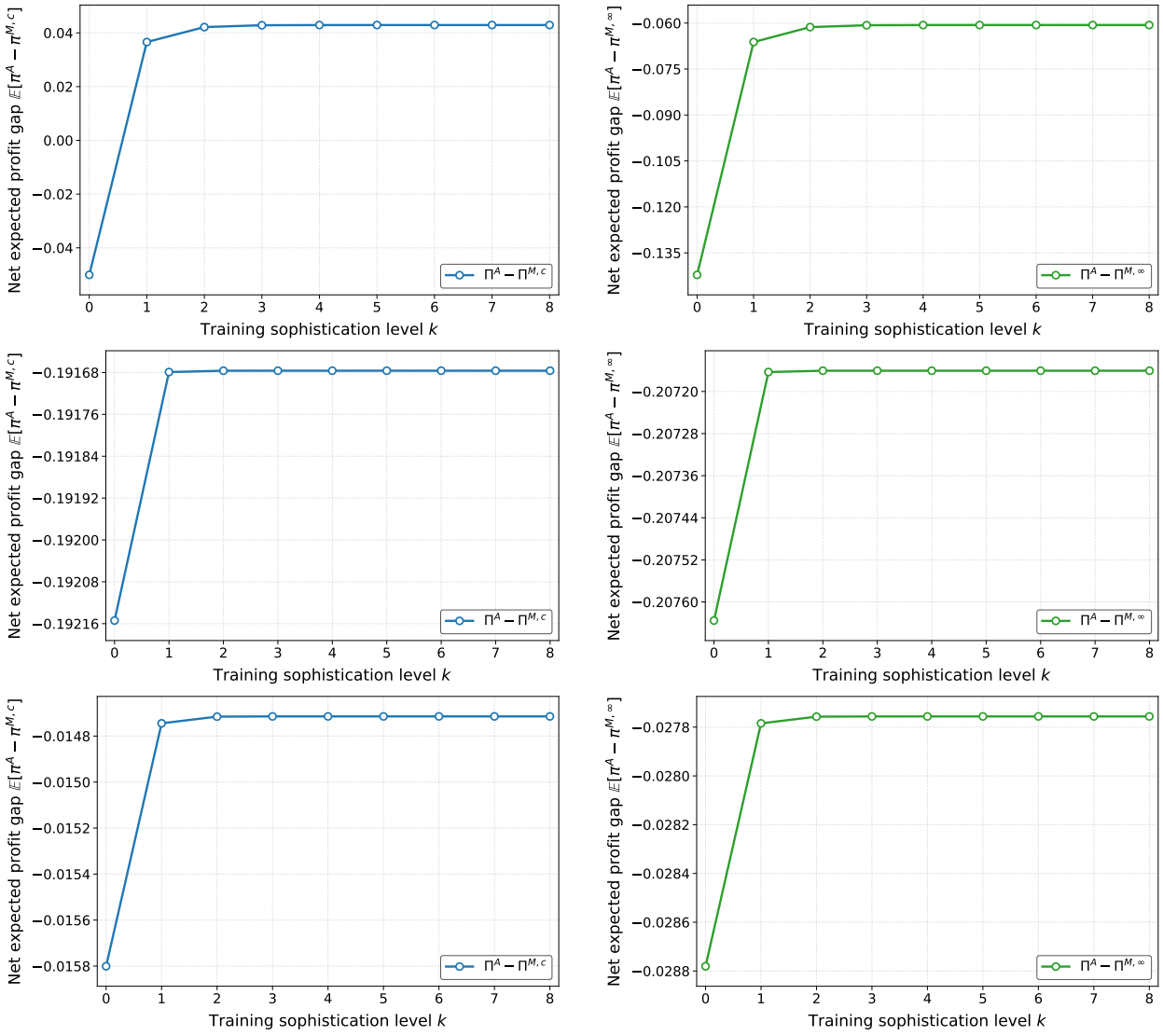
$$\widetilde{\Delta}_h(k+1) > \widetilde{\Delta}_h(k) \qquad \textit{for every } k \geq 0 \textit{ and } h \in \{c, \infty\}. \tag{A169}$$

D.2.1 Eventual Improvement under Strict Stability

The regularity condition controls every training level. Strict fixed-point stability alone delivers the weaker tail result.

Proposition 38 (Eventual improvement under strict stability)**.** *Maintain the assumptions of Proposition 37, except for the regularity condition. Suppose the fixed point is strictly stable: the eigenvalues $-\kappa_c$ and $b_\nu(\tau^\infty_p)$, defined in (A153), have absolute value below one.*

Figure A11: Training Sophistication with Level-∞ Human Demand



(a) AI profit gap relative to cursed investors (b) AI profit gap relative to level-∞ investors

Notes: The horizontal axis is the sophistication $k \in \{0, \ldots, 8\}$ of strategic investors in the AI training market. For each k, the pre-trained AI algorithm is deployed without further updating, the human demands from the previous section are held fixed, and the equilibrium price reclears. AI receives no fundamental signal in training or deployment ($\tau_A = 0$). Panel (a) plots $\widetilde{\Delta}_c(k)$, and panel (b) plots $\widetilde{\Delta}_\infty(k)$. Positive values favor AI. From top to bottom, $(\sigma_M, \sigma_z) = (1, 0.2), (1, 0.8), (4, 0.2)$. Other parameters are $\bar{v} = 0$, $S = 0$, $\sigma_v = 1$, $\gamma_M = \gamma_A = 1$, $\psi_c = 0.30$, $\psi_s = 0.30$, and $\phi = 0.40$. Training and deployment markets share these primitives and population shares. Vertical scales differ across the six plots.

For $h \in \{c, \infty\}$, define

$$\widetilde{k}^{h,\star} \equiv \min\left\{K \in \{0,1,\ldots\} : \widetilde{\Delta}_h(k+1) > \widetilde{\Delta}_h(k) \text{ for every } k \geq K\right\}, \tag{A170}$$

with $\min \varnothing \equiv \infty$. *Then*

$$\widetilde{k}^{c,\star} = \widetilde{k}^{\infty,\star} \equiv \widetilde{k}^{\star} < \infty. \tag{A171}$$

Thus both gaps strictly increase after the same finite training level. Their one-step changes have the same sign at every training level.

Proof. The human demands from the previous section imply price sensitivity

$$\xi_H^{\infty} = \psi_c \beta_p^{M,c} + \psi_s \beta_p^{M,\infty},$$

so deploying an AI algorithm with slope β gives the centered price

$$p(\beta) = \frac{\zeta^{\infty} v + z}{\xi_H^{\infty} + \phi\beta}.$$

Write

$$t_{\infty}(\beta) \equiv \frac{\beta}{\xi_H^{\infty} + \phi\beta}.$$

Let $\beta_{\infty} \equiv \beta_p^{A,\infty}$, $\tau_p^{\infty} \equiv (\zeta^{\infty})^2 \tau_z$, and $\xi^{\infty} \equiv \xi_H^{\infty} + \phi\beta_{\infty} > 0$. The AI and level-$\infty$ price-signal weights are

$$h_A^{\infty} \equiv \frac{\tau_p^{\infty}}{\tau_v + \tau_p^{\infty}}, \qquad h_M^{\infty} \equiv \frac{\tau_p^{\infty}}{\tau_v + \tau_M + \tau_p^{\infty}}, \qquad 0 < h_M^{\infty} < h_A^{\infty} < 1.$$

The limiting AI best-response condition is

$$\gamma\beta_{\infty} = 1 - h_A^{\infty} \frac{\xi^{\infty}}{\zeta^{\infty}}, \tag{A172}$$

and the limiting human coefficients satisfy

$$\gamma\beta_{\eta}^{M,\infty} = \frac{\tau_M}{\tau_v + \tau_M + \tau_p^{\infty}}, \tag{A173}$$

$$\gamma\beta_p^{M,\infty} = 1 - h_M^{\infty} \frac{\xi^{\infty}}{\zeta^{\infty}}, \tag{A174}$$

$$\beta_p^{M,c} = \frac{1}{\gamma}, \qquad \xi_H^{\infty} = \frac{\psi_c}{\gamma} + \psi_s \beta_p^{M,\infty} > 0. \tag{A175}$$

Equations (A172) and (A174) imply $\beta_p^{M,\infty} > \beta_{\infty}$. Hence $\xi_H^{\infty} + \phi\beta_p^{M,\infty} \geq \xi^{\infty} > 0$, while

(A174) gives $\gamma\beta_p^{M,\infty} < 1$.

Because $1/(\xi_H^\infty + \phi\beta) = (1 - \phi t_\infty(\beta))/\xi_H^\infty$, substituting (A172)– (A175) into (38) makes each gap an exact quadratic in $t_\infty(\beta)$, with zero derivative at $t_\infty(\beta_\infty)$. Completing the square gives

$$\widetilde{\Delta}_h(k) = \widetilde{\Delta}_h(\infty) + \widetilde{\Gamma}_h \left[t_\infty(\beta_p^{A,k}) - t_\infty(\beta_\infty)\right]^2. \tag{A176}$$

Here $\widetilde{\Delta}_h(\infty)$ is the gap obtained by setting the AI slope to β_∞. The curvature coefficients are

$$\widetilde{\Gamma}_c = -\frac{[(\zeta^\infty)^2\sigma_v^2 + \sigma_z^2]\gamma}{2}\left(1 + \frac{\phi}{\gamma\xi_H^\infty}\right)^2 < 0, \tag{A177}$$

$$\widetilde{\Gamma}_\infty = -\frac{(\zeta^\infty)^2\sigma_v^2 + \sigma_z^2}{2}\left(1 + \frac{\phi\beta_p^{M,\infty}}{\xi_H^\infty}\right)\left(\gamma + \frac{2\phi}{\xi_H^\infty} - \frac{\gamma\phi\beta_p^{M,\infty}}{\xi_H^\infty}\right) < 0. \tag{A178}$$

The second sign follows from $\xi_H^\infty + \phi\beta_p^{M,\infty} > 0$ and $\gamma\beta_p^{M,\infty} < 1$.

It remains to derive eventual contraction from the training recursion. Write

$$\tau_p^{k+1} = G_c(\tau_p^k), \tag{A179}$$

$$\nu^{k+1} = a_\nu(\tau_p^k) + b_\nu(\tau_p^k)\nu^k, \qquad \nu^k \equiv \frac{\beta_p^{M,k}}{\beta_\eta^{M,k}}. \tag{A180}$$

Let $\tau_p^* \equiv \tau_p^\infty$. Strict stability means

$$\kappa_c \equiv |G_c'(\tau_p^*)| < 1, \qquad |b_\nu(\tau_p^*)| < 1. \tag{A181}$$

The Jacobian of (A179)– (A180) at (τ_p^*, ν^∞) is triangular, with eigenvalues $-\kappa_c$ and $b_\nu(\tau_p^*)$, both in $(-1, 0)$.

The precision mode is active. The anchored initial condition satisfies $\tau_p^0 = G_c(0) > G_c(\tau_p^*) = \tau_p^*$, and strict decrease of G_c implies that no finite iterate equals τ_p^*. Hence

$$\tau_p^k - \tau_p^* = C_\tau(-\kappa_c)^k + o(\kappa_c^k), \qquad C_\tau \neq 0. \tag{A182}$$

Indeed, the ratio of successive precision errors is $-\kappa_c + O(|\tau_p^k - \tau_p^*|)$. The associated infinite product converges to a nonzero constant because strict stability makes the orbit summable.

The trained AI slope is a smooth function $\beta_p^A(\tau_p, \nu)$. Its loading on the driven state is nonzero:

$$\left.\frac{\partial\beta_p^A}{\partial\nu}\right|_* = -\frac{h_A^\infty(\zeta^\infty - \zeta^c)}{\gamma\zeta^\infty + h_A^\infty\phi} < 0. \tag{A183}$$

Evaluating all partial derivatives at (τ_p^*, ν^∞) and substituting the fixed-point identities gives

$$\frac{\partial \beta_p^A}{\partial \tau_p}[G_c'(\tau_p^*) - b_\nu(\tau_p^*)] + \frac{\partial \beta_p^A}{\partial \nu}[a_\nu'(\tau_p^*) + b_\nu'(\tau_p^*)\nu^\infty]$$
$$= \frac{\tau_p^* \tau_v (\zeta^\infty - \zeta^c)\left\{\gamma[\tau_p^*(\zeta^\infty - \zeta^c) + \widehat{\tau}^c \zeta^\infty] + \phi\tau_M\right\}}{(\widehat{\tau}^c + \tau_p^*)(\tau_v + \tau_p^*)^2(\gamma\zeta^\infty + h_A^\infty\phi)\left[\tau_M(\gamma\zeta^\infty + h_A^\infty\phi) + \tau_p^*\gamma(\zeta^\infty - \zeta^c)\right]} > 0. \tag{A184}$$

Set $u_k \equiv \tau_p^k - \tau_p^*$ and $v_k \equiv \nu^k - \nu^\infty$ within this proof. A first-order expansion gives

$$u_{k+1} = -\kappa_c u_k + O(u_k^2),$$
$$v_{k+1} = b_\nu(\tau_p^*)v_k + [a_\nu'(\tau_p^*) + b_\nu'(\tau_p^*)\nu^\infty]u_k + O(u_k^2 + |u_k v_k|).$$

If the eigenvalues are distinct, define

$$w_k \equiv v_k - \frac{a_\nu'(\tau_p^*) + b_\nu'(\tau_p^*)\nu^\infty}{-\kappa_c - b_\nu(\tau_p^*)} u_k.$$

Then

$$w_{k+1} = b_\nu(\tau_p^*)w_k + O(u_k^2 + |u_k w_k|).$$

Variation of constants gives two cases. A nonzero homogeneous driven component leads when $|b_\nu(\tau_p^*)| > \kappa_c$; otherwise the forced precision component leads. Equation (A183) makes the first loading nonzero, while (A184) makes the second loading nonzero. Thus, including when the nominally slower homogeneous mode has zero amplitude, there are $\lambda_A \in \{-\kappa_c, b_\nu(\tau_p^*)\}$ and $C_A \neq 0$ such that

$$\beta_p^{A,k} - \beta_\infty = C_A \lambda_A^k + o(|\lambda_A|^k). \tag{A185}$$

If the eigenvalues coincide at λ_A, variation of constants gives

$$\nu^k - \nu^\infty = \frac{[a_\nu'(\tau_p^*) + b_\nu'(\tau_p^*)\nu^\infty]C_\tau}{\lambda_A} k\lambda_A^k + o(k|\lambda_A|^k). \tag{A186}$$

At equality, the first term on the left side of (A184) vanishes. The coefficient of the Jordan term is therefore nonzero after applying $\partial\beta_p^A/\partial\nu$, so

$$\beta_p^{A,k} - \beta_\infty = C_A k\lambda_A^k + o(k|\lambda_A|^k), \qquad C_A \neq 0. \tag{A187}$$

In either case,

$$\lim_{k\to\infty} \frac{|\beta_p^{A,k+1} - \beta_\infty|}{|\beta_p^{A,k} - \beta_\infty|} = |\lambda_A| < 1. \tag{A188}$$

Finally, let

$$e_k \equiv t_\infty(\beta_p^{A,k}) - t_\infty(\beta_\infty).$$

Because

$$t'_\infty(\beta_\infty) = \frac{\xi_H^\infty}{(\xi_H^\infty + \phi\beta_\infty)^2} \neq 0,$$

equation (A188) implies

$$\frac{e_{k+1}^2}{e_k^2} \longrightarrow |\lambda_A|^2 < 1.$$

Therefore $e_{k+1}^2 < e_k^2$ for every sufficiently large k. Equation (A176) and $\widetilde{\Gamma}_h < 0$ then give

$$\widetilde{\Delta}_h(k+1) - \widetilde{\Delta}_h(k) = \widetilde{\Gamma}_h(e_{k+1}^2 - e_k^2) > 0, \qquad h \in \{c, \infty\},$$

eventually. At every k, the two gap changes have the same sign because they multiply the same change in e_k^2 by negative constants. Consequently, the sets defining $\widetilde{k}^{c,\star}$ and $\widetilde{k}^{\infty,\star}$ are identical and nonempty, which proves

$$\widetilde{k}^{c,\star} = \widetilde{k}^{\infty,\star} = \widetilde{k}^\star < \infty.$$

□

D.2.2 Monotonic improvement under the regularity condition

Define the price-normalized AI slope and its limit by

$$\widetilde{q}_k \equiv \frac{\beta_p^{A,k}}{\xi_H^\infty + \phi\beta_p^{A,k}}, \qquad \widetilde{q}_\infty \equiv \frac{\beta_\infty}{\xi_H^\infty + \phi\beta_\infty}. \tag{A189}$$

The regularity condition requires

$$|\widetilde{q}_{k+1} - \widetilde{q}_\infty| < |\widetilde{q}_k - \widetilde{q}_\infty| \qquad \text{for every } k \geq 0. \tag{A190}$$

It strengthens tail convergence by controlling the finite prefix.

Proof of Proposition 37. Because $\widetilde{q}_k = t_\infty(\beta_p^{A,k})$, the exact representation (A176) becomes

$$\widetilde{\Delta}_h(k) = \widetilde{\Delta}_h(\infty) + \widetilde{\Gamma}_h(\widetilde{q}_k - \widetilde{q}_\infty)^2, \qquad \widetilde{\Gamma}_h < 0. \tag{A191}$$

Hence

$$\widetilde{\Delta}_h(k+1) - \widetilde{\Delta}_h(k) = \widetilde{\Gamma}_h\left[(\widetilde{q}_{k+1} - \widetilde{q}_\infty)^2 - (\widetilde{q}_k - \widetilde{q}_\infty)^2\right].$$

Condition (A190) makes the bracketed term negative. Since $\widetilde{\Gamma}_h < 0$, the difference is positive for both $h = c$ and $h = \infty$ at every k. The identity also shows that the condition is necessary for strict increases in both gaps and that their one-step changes always have the same sign. □

Two path restrictions are sufficient for the regularity condition. First, it holds if the trained slopes remain in one pole-free interval and move strictly and monotonically toward β_∞. Second, let I be a compact pole-free interval containing every trained slope and β_∞. It is enough that, for some $\varrho \in [0, 1)$,

$$|\beta_p^{A,k+1} - \beta_\infty| \leq \varrho |\beta_p^{A,k} - \beta_\infty| \qquad \text{for every } k \geq 0, \tag{A192}$$

$$\varrho \frac{\max_{\beta \in I} |\xi_H^\infty + \phi\beta|}{\min_{\beta \in I} |\xi_H^\infty + \phi\beta|} < 1. \tag{A193}$$

The first restriction rules out overshooting. The second permits an alternating slope path but requires raw contraction to dominate changes in aggregate price sensitivity.