Ex Machina:
Financial Fragility in the Age of Artificial Intelligence*

Preliminary, do not circulate

Kartik Anand[†] Sophia Kazinnik[‡] Agnese Leonello[§] Ettore Panetti[¶]

June 30, 2025

Abstract

Does artificial intelligence (AI) represent a threat to the stability of the financial system? This paper investigates how two types of AI agents (Q-learning and LLM) behave in a canonical financial coordination game of mutual-fund redemptions under strategic uncertainty. In our stylized setting, investors choose whether to redeem based on their beliefs about economic fundamentals and other investors' actions. We then experimentally examine whether these AI agents learn to coordinate their redemption decisions and how this depends on economic fundamentals and other uncertainties. Our simulation results reveal stark heterogeneity in coordination behavior across AI types. Q-learning agents do learn to coordinate on common strategies that depend on fundamentals, but their behavior is strongly influenced by the type of portfolio securities and associated payoff risks, often more so than by strategic or fundamental uncertainty alone. In contrast, LLM agents coordinate via explicit reasoning and tend to select equilibria based on logical inference about the situation, rather than trial-and-error patterns. These differences imply that different AI systems can induce different levels of financial fragility.

JEL Classification: G01, G21, C73, D83
Keywords: Financial Stability; Artificial Intelligence; Strategic Uncertainty; Coordination Games; Reinforcement Learning, Q-Learning; Large Language Models; AI Agents.

*We thank E. Calvano, M. Ekmekci, John J. Horton, and seminar audience at Bundesbank for useful comments. The views expressed here are the authors' and do not reflect the views of the European Central Bank, the Deutsche Bundesbank or the Eurosystem.
[†]Deutsche Bundesbank.
[‡]Stanford University, HAI.
[§]European Central Bank, DG-Research and CEPR.
[¶]University of Naples Federico II, CSEF, SUERF and UECE.

1 Introduction

In ancient Greek tragedy, when a plot became overwhelmingly complex, a mechanical crane would lower actors playing gods onto the stage to impose a resolution, a device later known as *deus ex machina*. Two and a half millennia later, our growing reliance on artificial intelligence (hereafter, AI) to navigate complexity and uncertainty invites a comparison. We now devise machines to make decisions and resolve strategic dilemmas on our behalf. Will AI become the *deus ex machina* of modern finance, or will it introduce new forms of fragility? We examine this question by studying AI agents' behavior in a classic financial coordination problem: a mutual-fund redemption game akin to a bank run. Our goal is to understand whether AI agents temper or amplify instability when faced with strategic uncertainty in financial markets.

The influence of AI on financial decision-making has accelerated over the past decade. Algorithmic strategies now execute more than 70% of U.S. equity share volume (Clark, 2010), underscoring how deeply automation is embedded in everyday market liquidity. More recently, advanced generative AI systems have begun integrating into portfolio management; for example, BlackRock's Aladdin product introduced an LLM-driven assistant to support portfolio managers in 2024. Yet a critical question remains: how will AI agents react under pressure, especially when facing uncertainty? Traditional financial models assume human behavior, but AI might not follow the same patterns of panic or coordination.

Existing studies provide valuable insights into how AI agents learn and behave strategically in environments with uncertainty. For instance, studies of Q-learning algorithms in market pricing games (e.g., Calvano et al., 2020; Colliard et al., 2025; Dou et al., 2025) find that these agents can learn to sustain collusive or coordinated strategies, even in settings where classical theory would predict competitive outcomes. These findings suggest that AI agents are capable of sophisticated, non-trivial strategic behavior. However, much of the existing literature focuses on environments with a unique equilibrium. In contrast, financial fragility scenarios (like bank runs or currency attacks) inherently feature strategic complementarities and multiple equilibria, meaning an agent's best action crucially depends on beliefs about others' actions (Diamond and Dybvig, 1983; Goldstein and Pauzner, 2005). How will AI algorithms navigate such complex environments? Would coordination still be possible in the presence of multiple equilibria? Will AI agents learn to avoid panics, or might they coordinate on the panic equilibrium?

We address these questions by experimentally studying two distinct AI types in a model of financial fragility. Specifically, we study: (i) a model-free reinforcement learning (multi-agent Q-learning), and (ii) a reasoning-based approach using large language models (LLMs), as decision-makers in a mutual-fund redemption game. By examining each approach separately and comparatively, we observe how different AI decision-making architectures influence equilibrium selection and thus financial fragility.

We build on a canonical mutual fund run model (Chen et al., 2010) to provide a *theoretical framework* with multiple equilibria in investors' redemption decisions.[1] The model yields clear theoretical predictions about equilibrium behavior under different conditions (e.g., with/without fundamental uncertainty, different levels of fund illiquidity, etc.), providing a baseline for evaluating AI decision-making.

[1] The model represents a parsimonious variant of Diamond and Dybvig (1983) and Goldstein and Pauzner (2005). Hence, while the application specifically concerns mutual fund redemptions, its insights could be generalized to a broader set of enviroments exhibiting strategic complementarities and a similar payoff structure, e.g., bank runs, currency crises, etc.

Within this framework, investors decide whether to redeem their shares before the maturity of the fund's investment project. Their redemption decisions depend critically on the actions of other investors and macroeconomic conditions. Specifically, the incentive to redeem early increases with the fraction of other investors taking the same decision (strategic complementarity) and decreases with the state of economic fundamentals, since stronger fundamentals imply higher expected returns. We characterize the equilibrium under different assumptions regarding the economic environment, which translates into different types of uncertainty being present. In particular, we distinguish between an economy where fundamentals are observed, so that fundamental uncertainty is not present, and between different types of funds, i.e., equity and bond funds, since the type of securities in a fund's portfolio introduces distinct payoff uncertainties, playing a significant role in the simulations.

Equipped with the predictions from the theoretical framework, we substitute investors with machines and run simulations to see whether allocating the redemption decisions to AI-powered systems will alter the outcomes and lead to increased fragility. We start with a model-free reinforcement learning algorithm (henceforth, Q-learning) to train investors. Each investor is replaced by a Q-learning agent, and we simulate a setting in which 30 agents simultaneously decide whether to redeem their shares before the fund's investment matures. We run simulations across four scenarios that vary based on the presence or absence of fundamental and payoff uncertainty, analyzing how these sources of uncertainty influence reinforcement learning and its implications for financial fragility. We then replicate the same exercises using LLMs as investors, examining how differences in reasoning, such as implicit, latent reasoning versus explicit chain-of-thought (CoT) reasoning, affect investor behavior and, in turn, financial fragility. This *experimental design* allows us to test our hypotheses by observing whether the AI agents' behavior aligns with theoretical equilibrium predictions.

Our results reveal substantial differences between the two AI types. Consistent with prior findings, we observe evidence of machine collusion. In the absence of fundamental uncertainty, Q-learning agents do eventually coordinate their actions, but their learned strategies are strongly shaped by the payoff structure of the fund, which also significantly affects fragility. When the fund invests in risky debt securities (bond fund), Q-learners become far more prone to early redemptions (i.e., runs) than when the fund invests in equity securities, even under identical fundamental conditions. In other words, payoff uncertainty (the risk of a zero payoff) materially increases fragility. This result remains robust to the introduction of fundamental uncertainty, which overall tends to increase fragility. An interesting result we obtain in this scenario is that observed collusion breaks down for some values of economic fundamentals. Specifically, we find that for intermediate values of economic fundamentals, only a share of algorithms choose to redeem before the maturity of the fund investment. Such share decreases as the fundamentals improve, in line with the theoretical prediction from the global game literature (Goldstein and Pauzner, 2005).

Despite this alignment, the simulations reveal two important departures from theory. First, Q-learning behavior varies significantly across asset types, suggesting a heightened sensitivity to payoff risk rather than strategic uncertainty alone. Second, in the absence of fundamental uncertainty, multiple equilibria do not emerge, implying that Q-learning agents tend to converge to a single outcome even in settings where theory predicts multiplicity.

These patterns stand in contrast to the behavior observed in simulations with LLM-based agents. Unlike Q-learners, LLM agents select actions by reasoning through the scenario. The LLM agents tend to follow equilibrium strategies that make sense given the fundamentals and presumed investor incentives (e.g. not running when economic fundamentals are strong), highlighting that reasoning drives their equilibrium selection rather than trial-and-error adaptation. This heterogeneity implies that different AI decision processes can lead to different equilibrium outcomes and levels of fragility. First, the LLMs' behavior is not affected by the type of assets the fund invests in. Second, the degree of fund illiquidity plays an important role in sustaining machine collusion. When fund illiquidity is severe, all models considered, irrespective of reasoning, fail to coordinate in some intermediate range of economic fundamentals.

Despite this, our simulations illustrate that reasoning, especially when repeated, allows machines to coordinate on a threshold strategy around the risk-dominant threshold, as for Q-learning algorithms. When fund illiquidity is contained, the LLMs perform better than Q-learning algorithms and achieve an outcome close to the first-best in the theoretical benchmark: Early redemptions are observed only in a range of economic fundamentals where early redemptions are efficient.

Our results underscore the central role of payoff uncertainty in shaping the behavior of Q-learning agents, and of reasoning in shaping the behavior of LLMs. Both factors also emerge as key channels influencing financial fragility. For Q-learning agents, payoff uncertainty increases overall fragility regardless of the presence of fundamental uncertainty, while also reducing the sensitivity of outcomes to the degree of illiquidity. Similarly, for LLMs, reasoning is generally associated with lower fragility, holding other factors constant, but also introduces greater volatility in redemption behavior across rounds.

These findings have important policy implications. As AI agents become prevalent in financial markets, their coordination dynamics could either mitigate or exacerbate fragility. Notably, we observe that *homogeneous* learning algorithms can converge rapidly and uniformly on a single action, including panic-driven redemptions, in ways that are less likely in a heterogeneous population of human investors. In this sense, our results point to a higher level of predictability of machines relative to humans, which would make it easier to design effective policy interventions. They also suggest that regulators should stress-test AI systems in multi-agent environments, rather than evaluating algorithms in isolation.

Literature Classic models show that when payoffs are strategic complements, self-fulfilling crises can arise. The canonical framework of bank runs by Diamond and Dybvig (1983) features multiple equilibria: an efficient no run outcome and an inefficient panic run. Subsequent work has introduced incomplete information to pin down the probability of panic runs. Morris and Shin (1998) and Goldstein and Pauzner (2005) (building on Carlsson and Van Damme, 1993) embed slightly noisy private signals on the economic fundamental (i.e., fundamental uncertainty), yielding a unique run equilibrium below a critical level of economic fundamentals.

Empirical evidence confirms run-like dynamics beyond the banking sector. In mutual funds, Chen et al. (2010) document that early redeemers impose liquidation costs on those who stay, resulting in outflows that are concave in performance. Similar strategic complementarities appear in corporate-bond funds (Goldstein et al., 2017) and credit markets (Bebchuk and Goldstein, 2011). Experiments with human subjects further illustrate equilibrium selection under uncertainty (e.g. Heinemann et al., 2004).

Despite these insights, we know very little about AI investors in such environments, a gap addressed by the present paper. While algorithms often improve liquidity in normal times (Hendershott et al., 2011), they can also amplify volatility. For example, Johnson et al. (2013) catalog dozens of ultrafast mini-crashes driven by algorithmic "hot-potato" trading. Regulators now worry that homogeneous AI strategies may coordinate unintentionally, accelerating contagion. Yet almost all studies treat algorithms as a single species; few explore heterogeneity across AI architectures. Our paper fills this gap by comparing model-free Q-learning agents with reasoning LLM agents in a run game.

Recent economics and computer science research finds that independent reinforcement learning agents often learn to collude. Two Q-learning price setters tacitly sustain supra-competitive prices in Calvano et al. (2020). Follow-up work replicates collusion in auctions (Banchio and Mantegazza, 2022; Banchio and Skrzypacz, 2022) and market-making (Cont and Xiong, 2024). Yet these studies focus on environments with a *unique* static equilibrium; coordination problems with inherent equilibrium multiplicity (bank runs, currency attacks) remain largely unexplored. An exception is Yang (2025), who places two Q-learners in a speculative-attack game and shows they sometimes trigger crises at higher fundamentals than theory predicts.

In addition, large language models (henceforth, LLMs) have begun to influence financial practice. It has been shown that LLMs can act as simulated players in strategic experiments (Horton, 2023) or sustain cooperation in repeated games (Akata et al., 2023). Our study is the first to examine whether an LLM's reasoning steers the system toward the panic or the calm equilibrium, and how that contrasts with trial-and-error RL.

More broadly, our work is the first to compare two distinct AI paradigms, model-free Q-learning and reasoning-based LLMs, while systematically varying both fundamental and payoff uncertainty to reveal how each architecture drives equilibrium selection in financial run scenarios.

Reinforcement learning (RL) has long modeled bounded rationality in games (Erev and Roth, 1998; Camerer and Ho, 1999). In finance, deep reinforced learning drives trading, portfolio selection, and execution (Cong et al., 2023). Colliard et al. (2025) show RL liquidity suppliers can tacitly coordinate. Similarly, Dou et al. (2025), in a setting á la Kyle (1985) show that Q-learning algorithms exhibit collusive behaviours when acting as speculators. While these papers provide important insights about efficiency associated with the use of AI-powered algorithms, the implications for *systemic* stability are only now being probed. Yang (2024) moves in this direction, finding that Q-learners can precipitate currency crises. Our paper complements these results by shedding light on the drivers of machines' behavior, heteorgeity among different AI agents and their implications for financial fragility.

In sum, our paper contributes to four strands of the literature. First, we bring the literature on financial fragility into the age of AI by embedding AI-driven investors in the canonical run model and showing that differences among algorithms shape which equilibrium prevails. Second, we add to the work on AI in financial markets by showing experimentally that Q-learning and LLM agents generate distinct patterns of fragility. Third, we extend findings on algorithmic coordination, revealing that machines can coordinate in environments with strategic complementarities and multiple equilibria. And finally, we add to reinforcement-learning studies in economics by showing that model-free learners often settle on a single equilibrium whereas model-based LLM agents can shift between equilibria, a behavior not

previously documented. Overall, the study indicates that as AI permeats finance, regulators must consider not only the *speed* and *scale* of algorithms, but also their *decision architecture*, when assessing systemic risk.

The paper proceeds as follows. In Section 2 we present a stylized theoretical framework characterized by strategic complementarities and derive the hypothesis to test in our experiments with Q-learning algorithms and LLMs. In Section 3, we describe the experimental environment for both AI types. The results of the simulations with Q-learning algorithms and LLMs are reported in Section 4.1 and 4.2 for the case without and with fundamental uncertainty, respectively. Section 5 summarizes the key policy implications of financial fragility obtained from our simulations and Section **??** concludes. All proofs are in the Appendix.

2 Theoretical Framework

In this section, we characterize the equilibria of a stylized model of financial fragility that builds on Chen et al. (2010), and derive predictions, in terms of equilibria selection and their properties, which we then test using Q-learning algorithms and LLMs.

The economy extends over two dates, denoted $t = 1$ and $t = 2$, and consists of a mutual fund, and an interger number, $N > 2$, of risk-neutral investors. Prior to the start of $t = 1$, each investor has one share in the mutual fund and we normalize the total amount invested by the mutual fund to N.

At the start of $t = 1$, a number, $A < N$, of active investors choose to either redeem their share or wait until the final date. The book value of the mutual fund's investments at $t = 1$ is $R_1 N$, which is common knowledge. Investors who redeem their shares receive the current value, R_1. In order to service these redemptions, the fund must liquidate assets, which is costly. Specifically, to raise R_1, the fund must liquidate $(1+\lambda)R_1$ units, where $\lambda > 0$ is a measure of asset illiquidity: the higher is λ, the greater is the share of the portfolio that must be liquidated to service the redemption. Following Chen et al. (2010), we assume $A \leq \frac{N}{1+\lambda}$, which implies that the mutual fund has enough resources to meet the redemptions of all A active investors at $t = 1$. Thus, all investors who redeem their shares receive R_1 for sure.[2]

An investor who does not redeem their share at $t = 1$ and instead waits until $t = 2$ receives

$$\frac{N - n(1+\lambda)}{N - n} R_1 R_2(\theta), \tag{1}$$

where n is the number of investors who redeemed their shares at $t = 1$ and $R_2(\theta)$ is the gross return at $t = 2$, which is an increasing function of the fundamental $\theta \in [0, 1]$.[3] The specific functional form of $R_2(\theta)$ captures the type of the mutual fund, i.e., whether it primarily invests in equity or bonds. For the purposes of the theoretical model, however, this distinction is immaterial as what matters is only the expected return at $t = 2$, which we define as $\mathbb{E}[R_2(\theta)] = R\theta$.[4] Thus, the payoff from waiting is increasing

[2] The derivations for the case with illiquidity, i.e, when $A > \frac{N}{1+\lambda}$ are included in Appendix E.

[3] We refer to θ generically as fundamentals of the economy. In Chen et al. (2010) θ is a measure of the fund performance. The two definitions are clearly linked, as better fundamentals can be associated with improved fund performance. For the purpose of our exercise, the precise definition is immaterial.

[4] This distinction and the functional form of $R_2(\theta)$ is, however, important when it comes to testing the model with AI agents,

in the fundamental, θ and the performance of the fund, but is decreasing in both asset illiquidity, λ, and the number of redemptions, n.

2.1 Characterizing the equilibria

We characterize the equilibria of the redemption game under two regimes differing based on the presence of fundamental uncertainty, which refers to the observability of the fundamentals θ when investors take the redemption decision. When there is common knowledge about the fundamental, θ, we obtain the standard tri-partite classification for the equilibrium (Morris and Shin, 1998), which is characterized in Proposition 1 depicted in Figure 1.

Proposition 1. *Under common knowledge of the fundamental, for extreme values of θ, there is a unique Nash equilibrium in pure strategies, while multiple equilibria emerge in an intermediate range of fundamentals. Specifically,*

- *for $\theta < \underline{\theta} = \frac{1}{R}$, all investors redeem at $t = 1$;*
- *for $\theta > \overline{\theta} = \frac{1}{R}\frac{N-(A-1)}{N-(A-1)(1+\lambda)}$, all investors wait until $t = 2$, and*
- *for $\theta \in (\underline{\theta}, \overline{\theta}]$, all investors redeeming at $t = 1$ and at all waiting until $t = 2$ are both equilibria.*

Proof. See Appendix A □

As Proposition 1 makes clear, an important prerequisite for the emergence of multiple equilibra is illiquidity. When the portfolio is perfectly liquid, $\lambda = 0$, the liquidation of assets does not impose any additional costs and so the expected return of staying is independent of the number of investors who withdraw. In this case, redemption at $t = 1$ is the dominant action if and only if $\theta \leq \underline{\theta}$, while staying is dominant otherwise, i.e., for $\theta > \underline{\theta}$. This equilibrium is the first-best allocation.

Figure 1: Tripartite classification of the fundamentals

However, when $\lambda > 0$, asset liquidation does impose a cost and, thus, redemptions by investors impose externalities on other investors. Moreover, since the expected payoff differential between staying until $t = 2$ and redeeming at $t = 1$ monotonically decreases in the number of investors who withdraw, investors' redemption decisions exhibit strategic complementarity: the incentive to redeem is stronger when others are expected to do so. This gives rise to multiple equilibria in the intermediate range of fundamentals $(\underline{\theta}, \overline{\theta}]$.

which we discuss in the subsequent section.

Next, we consider the case where the fundamental is not common knowledge. Instead, each investor, i, only receives an imperfect signal $\theta_i = \theta + \varepsilon_i$, with $\varepsilon_i \sim [-\eta, +\eta]$. On the basis of their private signals, investors must make their decisions. Proposition 2 characterizes the global games equilibrium (Goldstein and Pauzner, 2005), which is illustrated in Figure 2.

Proposition 2. *The model has a unique symmetric Bayes-Nash equilibrium in threshold strategies that is characterized by a critical signal*

$$\theta^* \equiv \frac{A}{\sum_{n=0}^{A-1} \frac{N-n(1+\lambda)}{N-n} R}, \tag{2}$$

such that an investor will redeem their share at $t = 1$ if and only if their signal is below θ^*. For a given fundamental, the average number of investors who redeem their shares at $t = 1$ is

$$n^*(\theta, \theta^*) = A \frac{\theta^* - \theta + \eta}{2\eta}. \tag{3}$$

In the limiting case, $\eta \to 0$, all active investors redeem their shares whenever $\theta < \theta^*$, while for $\theta \geq \theta^*$, they all stay.

Proof. See Appendix B □

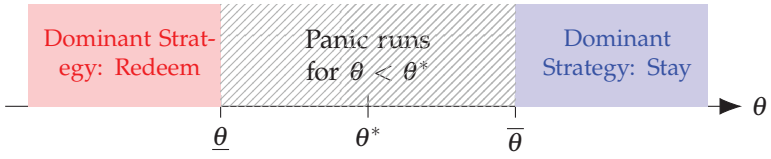


Figure 2: Global games equilibrium

The introduction of private information – to break the assumption of common knowledge – generates a unique equilibrium, which is characterized by the critical threshold, θ^*. This threshold captures the propensity for redemptions given the economic environment and is increasing in λ. The intuition for this result is clear. When investors redeem their shares, this exerts a negative externality on the other investors. When λ is large, so too is the externality. Thus, the proclivity of investors to withdraw is increased, which is captured by an increase in θ^*. And, insofar that $\theta^* > \underline{\theta}$, whereby the lower dominance bound also fully characterizes the first-best allocation, for the range of fundamentals between the two thresholds, redemptions are inefficient and detrimental to investors' welfare. As they are driven by investors' beliefs that redemptions by others would have adverse implications for their own payoffs, we refer to such a situation as panic-based redemption runs, or simply panic runs.

2.2 Hypotheses

Propositions 1 and 2 present the key theoretical predictions to test with AI-agents. A key assumption in the derivation of these results is that the expected return at $t = 2$ was $R\theta$. This assumption is satisfied by two distinct specifications for $R_2(\theta)$. which we link to the types of mutual funds.

Equity fund: The mutual fund owns shares in risky non-financial corporations (NFCs). The return earned by the fund on its (well diversified) portfolio of shares depends on the fundamental and is given by

$$R_2(\theta) = R\,\theta\,.$$

Bond fund: The mutual fund provides debt to NFCs who make risky investments. The coupon on the bond is R, while the credit risk is captured by the fundamental, θ. Specifically, with probability θ, the NFC pays the coupon R, while with the converse probability, the NFC is unable to pay. So,

$$R_2(\theta) = \begin{cases} R & \text{with prob } \theta \\ 0 & \text{with prob } 1-\theta \end{cases}.$$

Thus, the structure of the returns for equity and bond funds differ in terms of their relationship with *ambiguity*. Investors in bond funds who choose to not redeem must content with the possibility that they might receive nothing at all if the NFCs fail to repay. While investors in equity funds face no such uncertainty. Yet, despite this difference, the theoretical results in Propositions 1 and 2 are applicable to both types of funds.

Based on our analysis, we have the following hypotheses to test.

1. **H1:** In the absence of fundamental uncertainty, all investors: (i) redeem their share at $t = 1$ when $\theta < \underline{\theta}$; (ii) stay when $\theta > \overline{\theta}$. In the intermediate range, both equilibria are possible;
2. **H2:** Redemption flows do not depend on whether the mutual fund is an equity fund or a bond fund;
3. **H3:** In the presence of fundamental uncertainty, investors behave according to a threshold strategy: Investors redeem their share at $t = 1$ if and only if $\theta < \theta^*$;
4. **H4:** An increase in fund illiquidity, as captured by λ, increases the propensity for investors to redeem their shares.

3 Experimental Design

Our primary goal is to explore the financial fragility of automated algorithmic investors (AAIs). We therefore replace the investors in our theoretical framework with two AI-archetypes: (i) Q-learning algorithms that rely on reinforcement learning from market signals to determine whether to redeem shares or not, and (ii) large-language models that apply contextual reasoning to observed market data to infer redemption risk. This dual setup reflects the broad spectrum of AI systems active in finance. In what follows, we describe our experimental design for each approach.

3.1 Q-learning algorithms

In this section, we describe how we implement AAIs with Q-learning algorithms. Q-learning is a simple reinforcement learning framework that models decisions to redeem or stay based on the fundamental, $\theta \in [0,1]$, which is discretized into a finite set of states. The algorithm consists of four key components:

1. **States**: The state space consists of a finite number of states, $S \in \mathbb{N}$, where each state $s \in \{1,\ldots,S\}$ represents a bin from the discretization of the continuous signal space $[0,1]$.
2. **Actions**: The action set is binary, $\mathscr{A} = \{a_R, a_S\}$, where a_R represents the "redeem" action and a_S represents the "stay" action.
3. **Reward**: The reward function, $\pi(\theta, a)$, assigns a numerical reward based on the fundamental, θ, and the action a.
4. **Episodes**: The learning process occurs over $T \in \mathbb{N}$ episodes, where in each episode, each AII observes the state, makes a decision and receives a reward.

These components are captured in the Q-matrix, $Q_{i,t} \in \mathbb{R}^{S\times 2}$, one for each AII, $i = 1,\ldots,A$. The rows of the Q-matrix correspond to the states, and the columns correspond to the actions. The Q-matrix is updated iteratively as follows and also depicted in Figure 3.

1. In episode t, the fundamental and signals and states are independently drawn for each AII, i.e., $s_{i,t}$ for $i = 1,\ldots,A$.
2. AIIs chooses their actions. With probability ε_t, each AII, i, independently chooses their action at random (*exploration*), i.e., $a^*_{i,t} = a_R$ with probability $1/2$ and $a^*_{i,t} = a_S$ otherwise. While with the converse probability, $1-\varepsilon_t$, AII i chooses the optimal action (*exploitation*), $a^*_{i,t} = \arg\max_{a\in\{a_R,a_S\}} Q_{i,t}(s_{i,t}, a)$.
3. If $a^*_{i,t} = a_R$, then the reward to the AII is $\pi(\theta, a^*_{i,t}) = R_1$. While, if $a^*_{i,t} = a_S$, then the reward from choosing to stay depends on both the realized fundamental, θ, and the actions of the other AIIs and is given by Equation (1).
4. At the start of episode $t+1$, the Q-matix for AII i is updated as follows:

$$Q_{i,t+1}(s_{i,t}, a^*_{i,t}) = (1-\alpha)Q_{i,t}(s_{i,t}, a^*_{i,t}) + \alpha\,\pi(\theta, a^*_{i,t})\,. \tag{4}$$

   The entry in the Q-matrix for action not selected for state $s_{i,t}$ does not change.

The parameter $\alpha \in [0,1]$ in Equation (4) is the learning rate and governs how quickly past knowledge is forgotten. When α is high, greater weight is placed on the new reward over the accumulation of old rewards in the Q-matrix entry.

In our simulations, we run 25 separate rounds using the following parameter values: $N = 50$, $A = 30$, $R_1 = 1$, $R = 2$, $\beta = 0.999986$, $\alpha = 0.1$, $S = 75$, and $T = 500{,}000$.[5] For these parameters, each Q-learning

[5]Consistently with Colliard et al. (2025), we set the learning parameter α and only perform comparative statics exercises with respect to economic variables (e.g., λ). Modifying the learning parameter in a way that is meaningful from an economic perspective would involve not only taking a stance on what the optimal α is, but also making assumptions about the preferences and objective functions of the agents coding the algorithms.

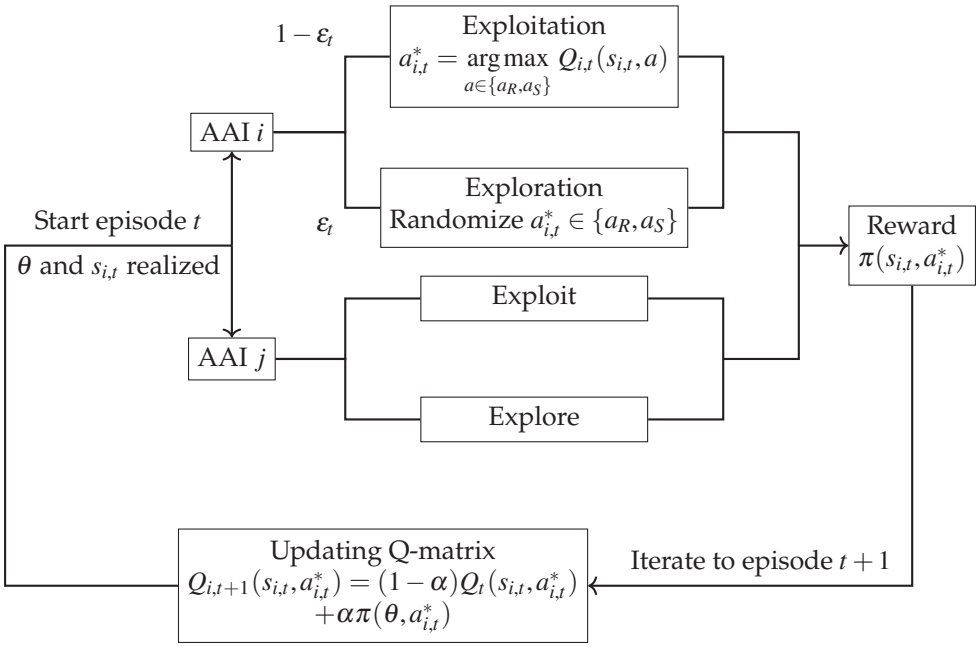


Figure 3: **Q-learning iterative process at work.**

algorithm experiments 71,000 times in expectation, which corresponds roughly to $1/8$ of total episodes. We consider seven equally distant values for λ in the range $(0.05, 0.35)$ to capture different degrees of illiquidity. In each episode, the fundamental, θ, is independently drawn from $[0,1]$.[6]

3.2 LLMs

For the LLM-based AIIs, the approach is fundamentally different: instead of learning via repeated numerical reward updates, an LLM agent uses *reasoning* to decide whether to withdraw or wait. We utilize a state-of-the-art large language model; specifically, we interface with the `4o-mini` and `o3-mini` models via OpenAI's API, and prompt these models to play the role of an investor. The former is a non-reasoning model which we used to test reasoning via the two-step procedure described below. The latter is a model with an embedded reasoning function, which we use to check the robusteness of the results.

We compose textual scenario prompts that describe the current economic environment. For example, when the fundamental is common knowledge and the payoff from waiting is uncertain, we use the following:.

Base prompt

[6]Our results are qualitatively robust to significant changes to the parameter space. Simulation results for $T = 1000000$, $N = 100$, $A = 70$ over 100 rounds can be provided upon request.

You are one of A active investors in a mutual fund out of a total of $N = \{N\}$ investors.
The fund fundamental is $theta = \{theta\}$.
If you withdraw, you earn $\{R1\}$ for sure.
If you stay, with probability $theta = \{theta\}$ you earn $fraction * \{R1\} * \{R\}$, otherwise you get 0.
Here:
- $fraction = (N - W * (1 + \lambda))/(N - W)$
- W is how many active investors withdraw
- λ $(lambda) = \{lam\}$ is the illiquidity parameter

Make your decision to withdraw or stay.
Please respond using the following XML format:
<decision> </decision>
Your decision must be exactly one word: either "stay" or "withdraw" (lower-case).

The prompt includes the key model parameters: the total number of investors ($N = 50$), the number of active investors ($A = 30$) and the payoffs ($R_1 = 1$ and $R = 2$). It also conveys the payoff structure $R_2(\theta)$. We further supplement the prompt with information on the illiquidity parameter, λ, and the fundamental, θ from a grid and then ask the AII to decide whether to withdraw or wait. We then iterate over the values of λ and θ from the grid and present the prompts to the AIIs.

We also consider a variation that explicitly asks the LLM-based AII to use chain-of-though (CoT) reasoning followed by an evaluation of the reasoning before making a decision. The prompting strategy consists of two steps.

Reasoning prompt: step 1

You are one of $\{A\}$ active investors in a mutual fund out of a total of $N = \{N\}$ investors.
The fund fundamental is $theta = \{theta\}$.
If you withdraw, you earn $\{R1\}$ for sure.
If you stay, with probability $theta = \{theta\}$ you earn $fraction * \{R1\} * \{R\}$, otherwise you get 0.
Here:
- $fraction = (N - W * (1 + \lambda))/(N - W)$
- W is how many active investors withdraw
- $\lambda(lambda) = \{lam\}$ is the illiquidity parameter

Carefully analyze and explain your reasoning for whether to stay or withdraw, but DO NOT make a decision yet.
Just write your explanation in plain text. Do NOT use XML or any special formatting.

Reasoning prompt: step 2

Based on the following reasoning: {`explanation from step 1`}
Now, make your decision to withdraw or stay.
Please respond using the following XML format:
<decision> </decision>
Your decision must be exactly one word: either "stay" or "withdraw" (lower-case).

In these simulations, the LLM AIIs are expected to reason about others' likely actions and choose the best response. To control the depth of reasoning, we consider two options:

- **One-shot (memoryless):** In each round of the simulation, a new prompt is generated by iterating over (λ, θ) pairs from the grid; the AIIs respond based on general reasoning without reference to past rounds.
- **Iterative (memory-enabled):** The model is given a limited memory of previous rounds' outcomes (e.g. a brief history of past prompts and responses) to allow adaptation through conversational context rather than gradient-based learning.

Unlike conventional trading algorithms that optimize explicit reward functions, LLM-based agents operate by following qualitative prompts (e.g. "act as a value investor"), raising new questions about their behavior and impact. Figure 4 provides a brief scematic for how the simulations are run.

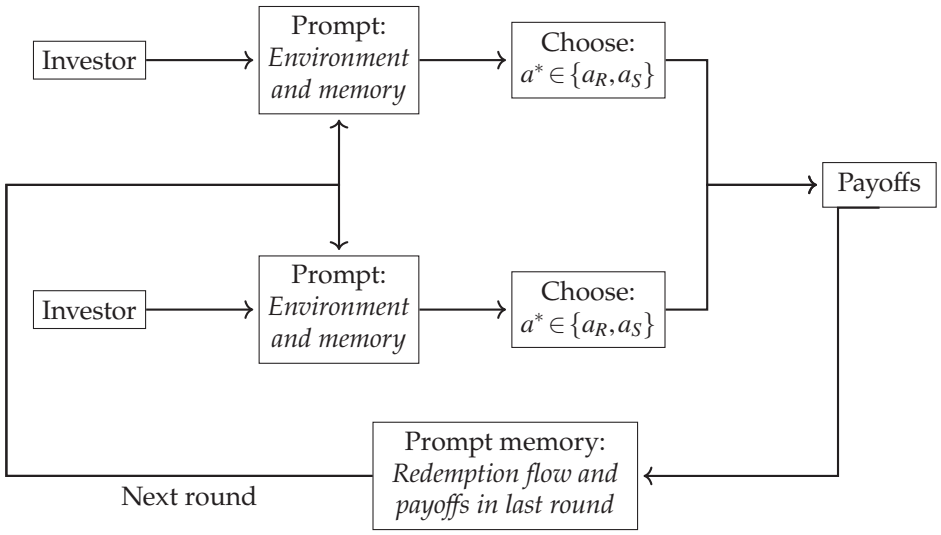


Figure 4: **Workflow for LLM Agents**

4 Financial Fragility with Automated Algorithmic Investors

Having described how Q-learning and LLMs operate in the context of the mutual fund game, in this section, we report the results of the simulations. We distinguish between the environment without

fundamental uncertainty (Section 4.1) and with fundamental uncertainty (Section 4.2). In each section, we first illustrate the simulation results of the Q-learning experiment and then the one with LLMs investors. For each type, we distinguish between equity and bond funds, so to test hypothesis H2. Finally, in order to isolate the role of strategic uncertainty, we benchmark the results against the case with a single operating machine.

4.1 No fundamental uncertainty

In an equity fund, investors' payoff at $t = 2$ drops to zero only when θ is zero. Figure 5 reports the result

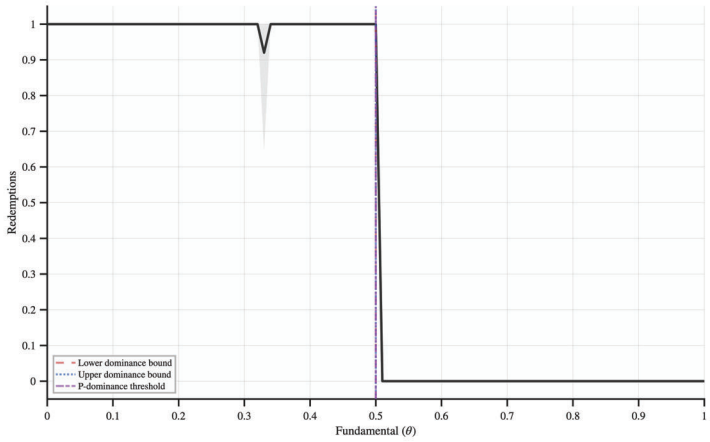


Figure 5: **Single Q-learning algorithm holding shares in an equity fund.** The figure is drawn for $\lambda = 0.25$. All other parameters are as stated in Section 3.1.

for the case where $A = 1$ and a single Q-algorithm is choosing when to redeem its shares. It shows that the Q-algorithm behaves as predicted by the theory: Redeem the share when the fundamentals θ fall below $\underline{\theta}$ and do not redeem otherwise.[7] Changes in the illiquidity parameter, as expected, play no role since $\underline{\theta} = \frac{1}{R}$ is independent of λ.

The results for the case with strategic uncertainty, when $A = 30$ Q-algorithms choose simultaneously whether they redeem their share at $t = 1$, are summarized in Proposition 3 and illustrated in Figures 6 and 7.

Proposition 3. *For each θ, Q-algorithms coordinate on the risk-dominant action, i.e., they choose to redeem early when $\theta < \theta^{RD}$ and wait until the final date otherwise. Increased illiquidity, as measured by a larger λ, is associated with more fragility. The threshold θ^{RD} is derived in the appendix.*

Proof. See Appendix C.

[7] In Section D, we show that the little dip in the curve we observe at about $\theta = 0.35$ is just the result of some intiial variation in behavior during the initial episodes when learning is intense, but the algorithm nicely converge to choose the p-dominant action based on the realization of the fundamentals θ.

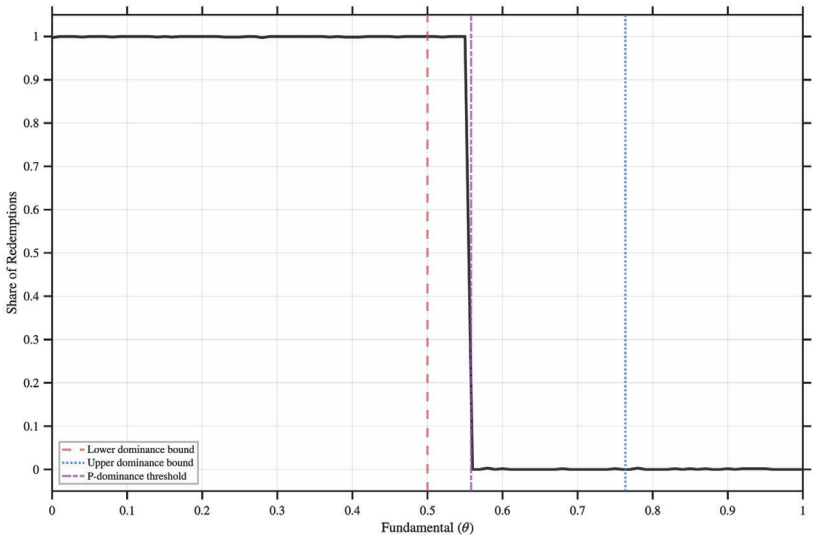


Figure 6: **Multiple Q-learning algorithms holding shares in an equity fund.** The figure illustrates algorithms' redemption decision as a function of θ. The figure is drawn for $\lambda = 0.25$. All other parameters are as specified Section 3.1.

In line with the existing literature (see, e.g., Colliard et al., 2025; Dou et al., 2025), our results show that Q-learning algorithms tend to coordinate. The action on which they coordinate, and thus the resulting equilibrium, depends on the fundamentals θ. Consistent with the prediction of the theory, redeeming shares early when $\theta < \underline{\theta}$ (i.e., below the red dotted line in Figure 6) and waiting until the final date when $\theta > \overline{\theta}$ (i.e., above the blue dotted line in the same figure). However, different from the theory predictions, no multiple equilibria emerge in the range $\theta \in (\underline{\theta}, \overline{\theta})$. On the contrary, Q-algorithms continue to coordinate on a specific action depending on the state of fundamentals θ.

The algorithms' behavior can be rationalized as being consistent with playing a threshold strategy around the risk-dominance equilibrium, corresponding to the cutoff θ^{RD} (i.e., the dotted purple line in Figure 6). In other words, for each value of θ, algorithms choose the action that gives them the highest payoff when they believe that their opponents will play any action with probability $1/2$. This corresponds to redeeming the shares at $t = 1$ when $\theta < \theta^{RD}$ and to redeeming the shares at $t = 2$ otherwise.

The results of the simulations and a simple comparative statics of θ^{RD} with respect to λ illustrate that hypothesis H4 is confirmed for Q-learning algorithms. Defining fragility as the withdrawals in excess of the social optimal ones (i.e., those occurring for $\theta < \underline{\theta}$), Figure 7 shows that fragility monotonically increases with illiquidity.

We now move on to a bond fund. Investing in debt securities yields a (per unit) return equal to R with probability $p(\theta) \equiv \theta$. As shown above, the expected (per unit) return at $t = 2$ is as in the case with equity security and, thus, equal to $R\theta$. This implies that the computation of the various cutoffs, including θ^{RD} is unaffected. Yet, this seemingly small modification leads to a dramatic change in the Q-algorithms' behavior, as illustrated in following proposition and Figure 8 below.

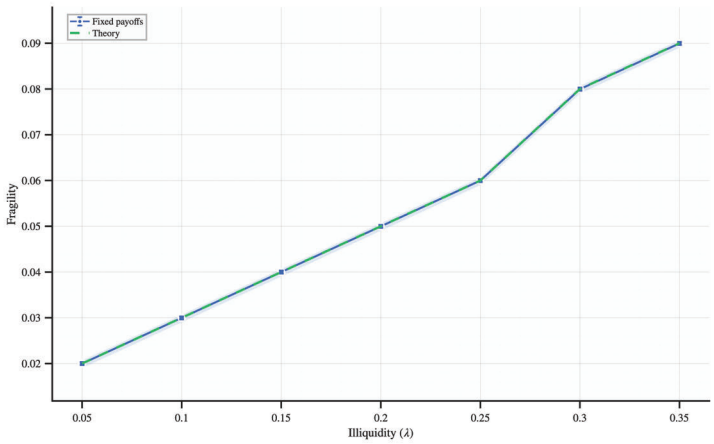


Figure 7: **Illiquidity and fragility for an equity fund.** The figure plots fragility (on the y-axis) as a function of the illiquidity paramenter λ and shows a upward sloping relationship.

Proposition 4. *Payoff uncertainty leads to an increase in fragility: Q-algorithms coordinate on redeeming their shares at $t = 1$ in a larger range of θ, which also includes values of $\theta > \overline{\theta}$. Furthermore, a range of fundamentals emerge in which the Q-algorithms stop coordinating on the same strategy.*

The introduction of payoff uncertainty, brought about by the change in the type of securities the fund invests in, increases fragility in that Q-algorithms coordinate on redeeming their shares early for values of fundamentals θ that are well above the cutoff $\overline{\theta}$ (blue dotted line). In other words, early redemptions are observed also for values of fundamentals in which it would be optimal (i.e., a dominant action) not to redeem early.

Another important aspect that emerges from the simulations is that payoff uncertainty breaks the algorithms' coordination in that in the same episode t, and for a given θ, some algorithms choose to redeem early and some do not. As it clearly emerges in Figure 8, the average fraction of those who choose to redeem, which is measured on the y-axis, decreases with θ in the range $\theta \in (0.8, 0.9)$.

Figure 9 provides a more precise illustration of this point, displaying for each θ the average share, over all rounds, of episodes in which Q-algorithms fail to coordinate on either redeeming at $t = 1$ or waiting until $t = 2$. The graph shows that in the range $\theta \in (0.8, 0.9)$, there are between 120-100000 episodes on average for each round where algorithms play different actions.[8] This result suggests that the share of early redemptions in the range $(0, 1)$, observed in Figure 8 cannot be interpreted as the presence of multiple equilibria (at least not fully), but rather suggests that for some values of θ, algorithms fail to coordinate on the same action.

This result requires a careful interpretation. Figure 10 delves further into its potential sources, plotting the rolling-average for redemptions across episodes when the fund invests in equity securities (Panel 10a) and in debt securities (Panel 10b) for several different values of θ. In each panel, the shaded regions

[8]The corresponding graph for the case when the fund invests in equity securities is in the Appendix and shows a flat line for all θ around 3 percent, which corresponds to about 15000 episodes per round in which the algorithms fail to coordinate.

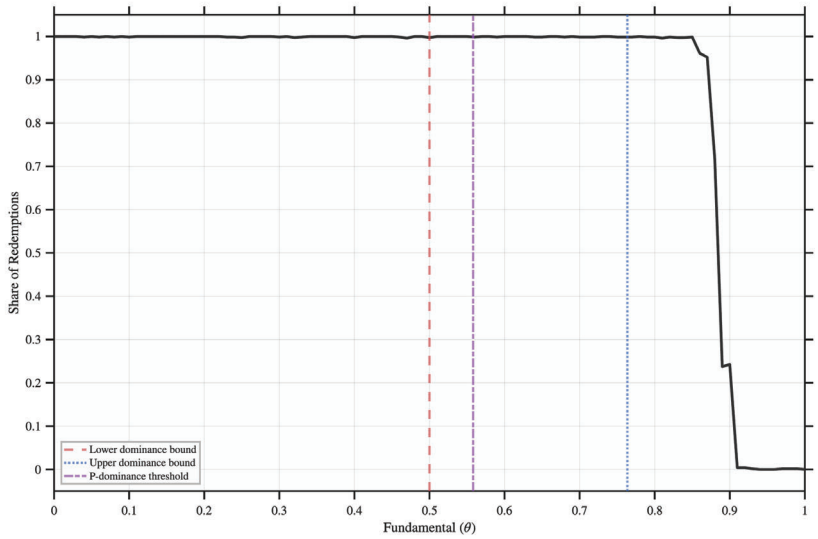


Figure 8: **Multiple Q-learning algorithms holding shares in a bond fund.** The figure illustrates the algorithms' redemption decision as a function of the fundamentals θ. The figure is drawn for $\lambda = 0.25$. All other parameters are as specified in Section 3.1.

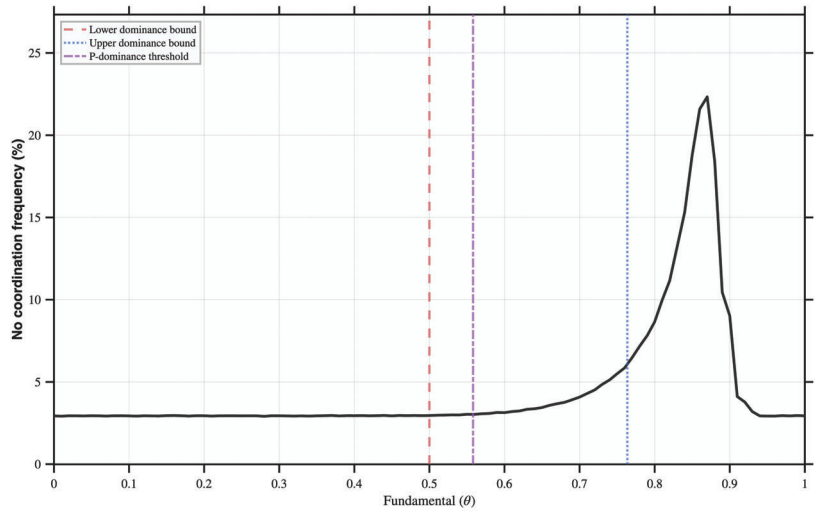
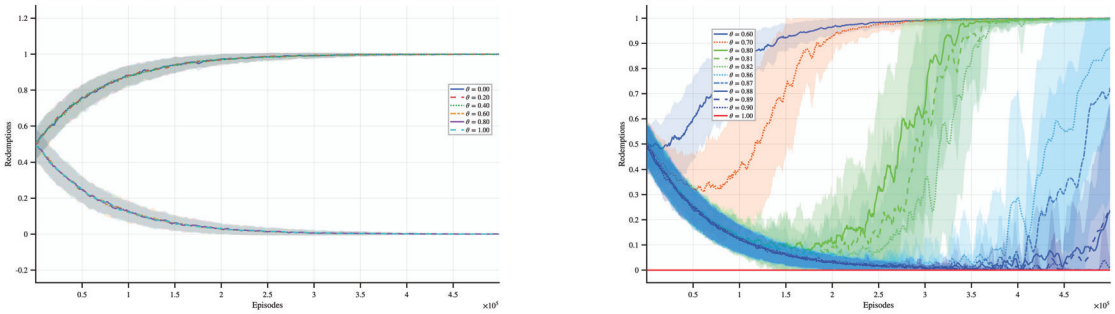


Figure 9: **Frequency of episodes in which algorithms fail to coordinate.** The figure illustrates the extent to which lack of coordination emerges for different values of θ. The y-axis measures the average share of episodes (over the 25 rounds) in which algorithms do not coordinate either on redeeming at $t = 1$ or waiting until $t = 2$.

represent the standard-deviations, which we compute over the twenty-five training runs. In Panel 10a, it can be seen that, after about two hundred thousand episodes, the algorithms are converging to an equilibrium with either no redemptions or complete redemptions depending on the levels of θ. This pattern is consistently observed over all rounds as proved by the fact that the shaded areas shrink. In other words, evidence of algorithms' convergence is provided by both the flatting of the curves around 1 and 0 respectively, as well as the reduction in the size of the shaded areas.

The figure in Panel 10b differs significantly. For value of $\theta \in (0.8, 0.9)$, we observe the following. First, the curves do not flatten, but rather display an upward trend. Second, the variation across rounds is still significant. Together these observations suggest that algorithms' lack of coordination could be interpreted as an unstable equilibrium. In other words, if we were to increase the number of episodes, we would observe the algorithms coordination on redeeming the share at $t = 1$.

While it is difficult to anticipate whether and when (i.e., for which t) the algorithms will coordinate, the results above offer important insights on the behaviour of algorithms in the presence of payoff uncertainty. First, it suggests that the type of securities the fund invests in and a different associated level of payoff uncertainty, plays a crucial role, with uncertainty making more difficult (and longer) for the algorithms to coordinate. Second, the payoff uncertainty leads to a significant increase in fragility. The fact that the shaded areas do not shrink, combined with the curves upward trend, hints to the possibility that, once the number of episodes is sufficiently large to allow full coordination again, the algorithms will coordinate on redeeming the shares at $t = 1$ for all values of θ, but $\theta = 1$.[9] This suggests that the level of fragility we identify in our simulations when the fund invests in debt securities should be thought as a lower bound.



(a) No payoff uncertainty

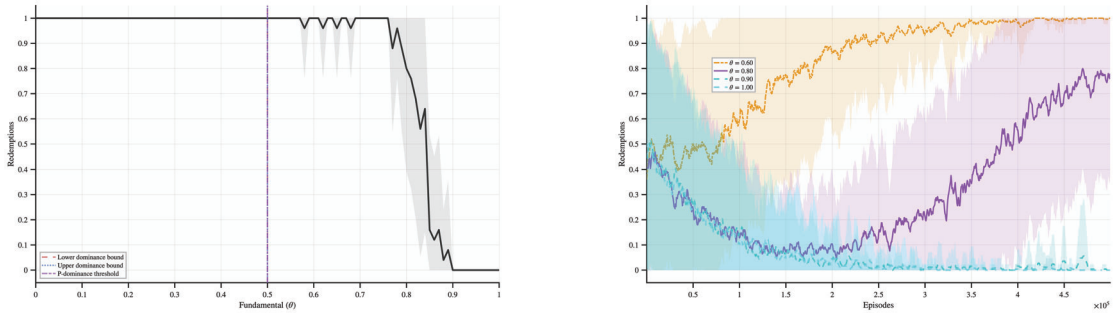
(b) With payoff uncertainty

Figure 10: **Redemptions over episodes**. For different values of θ, the figures illustrate the evolution of the average share of early redemptions over the twenty-five rounds. Panel 10a illustrates it in the case when the fund invest in equity securities and there is no payoff uncertainty. Panel 10b, instead, refers to the case when the fund invests in debt securities and payoff uncertainty is present. The figure is drawn for $\lambda = 0.25$. All other parameters are as specified in Section 3.1.

To prove that the change in behavior observed when switching from an equity to a debt fund is driven by the payoff uncertainty associated with the latter, we present below the result of the simulation without strategic uncertainty, by considering a single Q-algorithm that chooses whether to redeem (i.e.,

[9] For $\theta = 1$, Panel 10b shows that algorithms coordinate on not redeeming their shares early.

$A = 1$). Figure 11 illustrates the result.



(a) Redemptions as a function of fundamentals θ

(b) Redemptions as a function of episodes

Figure 11: **Single Q-learning algorithm holding shares in a bond fund.** The figure is drawn for $\lambda = 0.25$ and the parameters set in Section 3.1.

Payoff uncertainty dramatically changes Q-learning algorithm's behaviour, even in the absence of strategic uncertainty. To see this, we compare Figure 5 with Panel 26a in Figure 11. It emerges that the possibility to receive zero at the final date induces the machine to redeem its share early in a broader range of θ. [10] Futhermore, the behavior of the single algorithm appears more similar to that of the multiple machine, illustrated in Figure 8, than to that of the single algorithm operating as investor in a equity fund. This result, which is also confirmed by the similarities between Figures 10b and 26a, suggests that it is the diffent type of securities in which the fund invests in, and the associated different level of payoff uncertainty, that drives the result more than strategic uncertainty.

To conclude the analysis for Q-learning algorithms, we summarize three interesting insights on fragility that emerge from the comparison of the simulations for equity and bond funds. First, relative to equity funds, early redemptions are less dependent on fundamentals in bond funds, in line with the empirical evidence in Chen et al. (2010) and Goldstein et al. (2017). This can be seen comparing Figures 6 and 8 and noticing that algorithms converge on redeeming the share for a much larger range of fundamentals when the fund invests in debt securities, even beyond the cutoff $\overline{\theta}$ (blue dotted line). Second, as shown in Figure 12, fragility is, ceteris paribus, higher for a bond fund. Third, while more illiquidity tends to increase fragility in both cases, its effect is milder when the fund invests in debt securities.

The above analysis illustrates how reinforcement learning shapes investors' behavior and its implications for fragility. In what follows, we turn to explore the role of reasoning, by reporting the simulations for the LLMs. The results are summarized in Proposition **??** below and in Figures 13 and 15.

Figure 13 compares `4o-mini` investors that act without any chain-of-thought instructions. When each prompt begins with a brief recap of the previous round, the redemption curve takes on a smooth, sigmoid profile. The fraction of agents redeeming begins to fall just above the lower dominance bound $\underline{\theta}$ and approaches zero only when the fundamental nears the risk-dominant threshold. Removing the

[10] In the figure with a single Q-learning algorithm the y-axis measures, for each θ, the fraction of episodes in which the Q-algorithm chooses to redeem early.

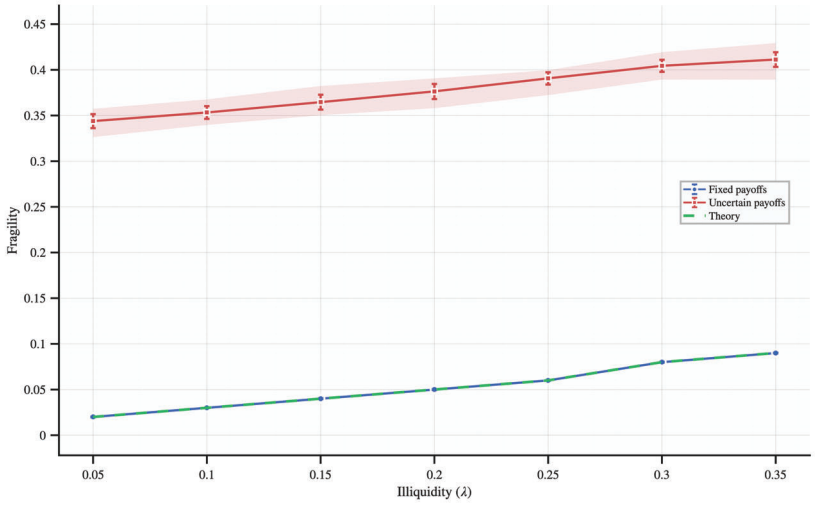


Figure 12: **Fragility and fund illiquidity.** The figure illustrates how fragility changes with illiquidity and compares the case in which the fund invests in equiity securities (blue line) with that in which it invests in debt securities (red line). The figure is drawn for $\lambda = 0.25$, while all other parameters are as specified in Section 3.1.

recap produces a different picture. In the second row of the same figure the curve remains at unity until it reaches a narrow window around the risk-dominant cutoff, after which it drops almost vertically. The single line of feedback therefore functions as a primitive prior that dampens the urge to exit at the first sign of weakness and stretches the transition zone by roughly five percentage points of the fundamental.

A higher liquidation penalty shifts these patterns to the right. Raising λ from 0.05 to 0.35 increases the loss borne by investors who wait, so moderately strong fundamentals no longer prevent precautionary withdrawals. Memory variants respond by broadening the descent region, whereas one-shot variants react by sharpening it. Consequently the gap between the two information regimes widens as illiquidity grows.

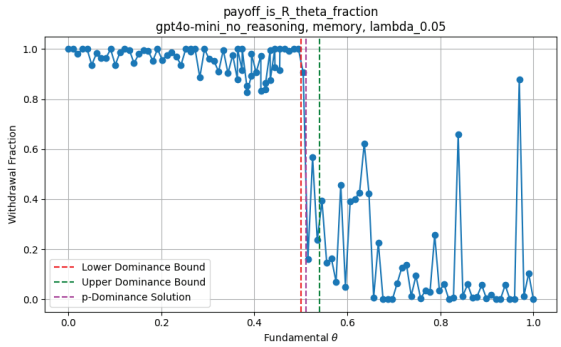
Introducing a two-step reasoning prompt modifies both slope and variance. In Figure 14 the memory variants still start with almost universal withdrawal at low fundamentals, but the descent now proceeds more gradually than in the no-reasoning condition. The internal deliberation appears to balance the payoff from waiting against the danger of a stampede. One-shot agents that reason in isolation exhibit volatile spikes once the group crosses the lower dominance line, a symptom of solving the coordination problem afresh each round without the stabilising effect of outcome feedback.

Turning to `o3-mini`, Figure 15 shows that both the memory and one-shot configurations trace an almost perfect step function. The fall in redemptions is centred tightly between $\underline{\theta}$ and θ_{RD}. Increasing λ shifts the break point by about eight percentage points, roughly twice the displacement observed for `4o-mini`. The lighter architecture evidently treats the liquidation penalty as a more prominent risk factor, which confirms that the choice of model alone can alter the fragility profile of the market.

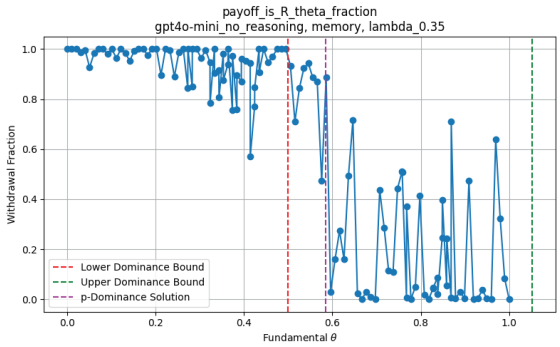
Panels 15c and 15d focus on `o3-mini` investors that act in a one-shot, no-reasoning mode. At the low

penalty of $\lambda = 0.05$ the population behaves almost deterministically: every agent withdraws when θ is below about 0.51, and the redemption fraction then collapses within only a few basis points. The break lies just to the right of θ_{RD} and well inside the dominance bounds, which indicates that even without feedback these agents adopt a sharply defined indifference point close to the theoretical solution for $p = \frac{1}{2}$.

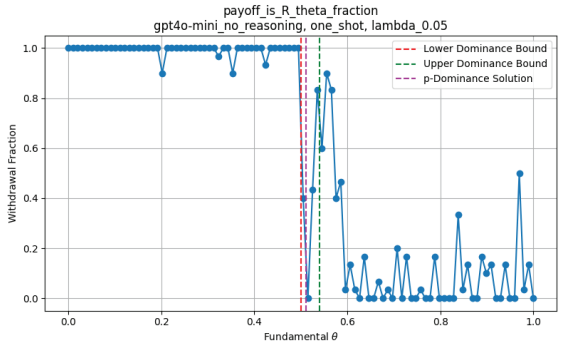
When the liquidation cost rises to $\lambda = 0.35$ two further changes occur. First, the mass-withdrawal threshold moves to approximately $\theta = 0.60$, keeping pace with the shift reported for the memory variant. Second, the drop in redemptions becomes noisy and approximately linear, stabilising near zero only after the fundamental exceeds 0.85. The broader band reflects greater strategic uncertainty when waiting is costly. Because the one-shot agent has no recollection of past play, it resolves a fresh coordination problem each round, so small perturbations in sampling can trigger partial runs that scatter around forty to fifty per cent withdrawal. These outliers, while rare, show that prompt-induced noise can stand in for the heterogeneous beliefs that underpin uniqueness in classical global-game models.



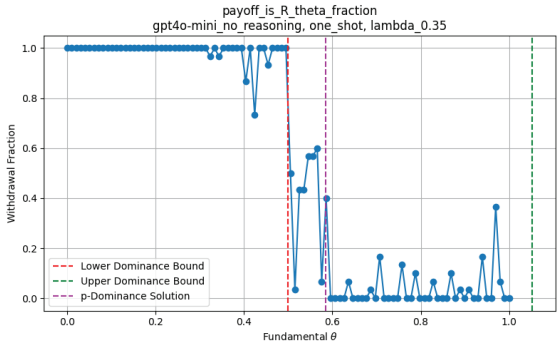
(a) 4o-mini, no reasoning + memory @ $\lambda = 0.05$.



(b) 4o-mini, no reasoning + memory @ $\lambda = 0.35$.

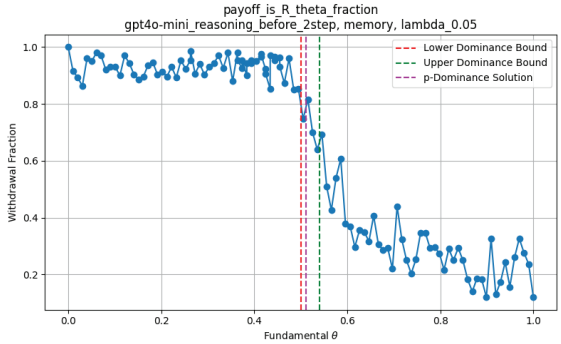


(c) 4o-mini, no reasoning + one-shot @ $\lambda = 0.05$.

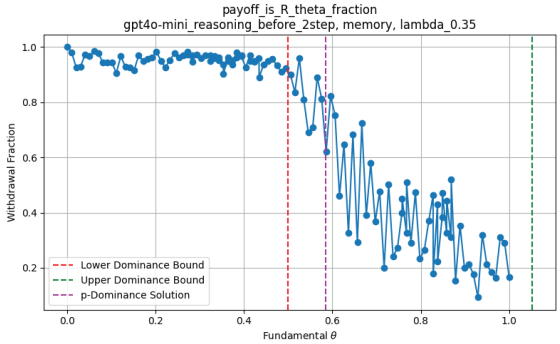


(d) 4o-mini, no reasoning + one-shot @ $\lambda = 0.35$.

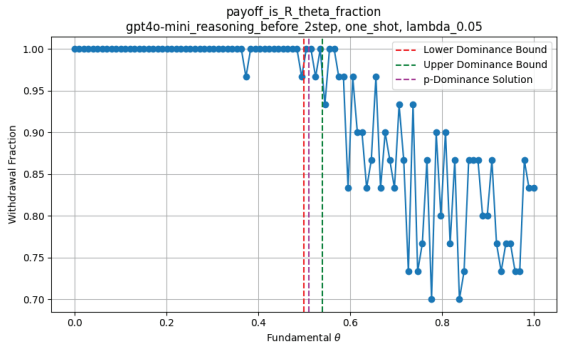
Figure 13: Redemptions for the `4o-mini` without reasoning
Notes: Solid blue lines with markers show the share of redemption for each θ. Red: lower dominance bound where unconditional exit always dominates. Green: upper dominance bound where unconditional wait always dominates. Magenta: p-dominance solution—the probabilistic equilibrium threshold. With memory, the redemption curve is smoother and the drop-off point shifts to higher θ, reflecting more cautious, history-aware decisions. Without memory, one-shot inference produces a sharp, step-like drop once the threshold is crossed, and exhibits greater volatility at high λ. Comparing left (low λ) vs. right (high λ) panels shows that stronger regularization exaggerates the gap between memory and one-shot: one-shot agents break earlier relative to the bounds under high penalty.



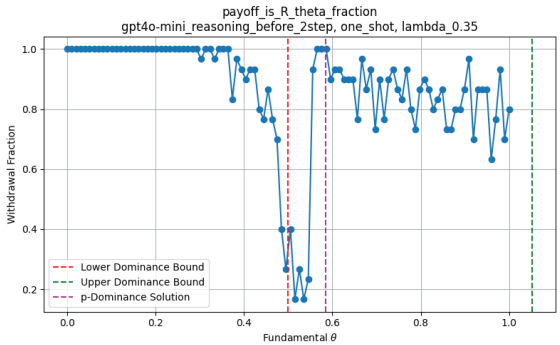
(a) 4o-mini, reasoning + memory @ $\lambda = 0.05$.



(b) 4o-mini, reasoning + memory @ $\lambda = 0.35$.



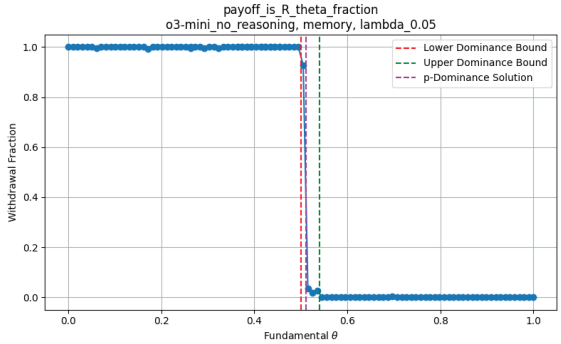
(c) 4o-mini, reasoning + one-shot @ $\lambda = 0.05$.



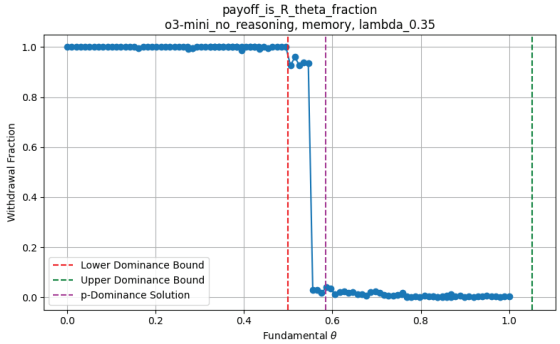
(d) 4o-mini, reasoning + one-shot @ $\lambda = 0.35$.

Figure 14: Redemptions for `4o-mini` with reasoning

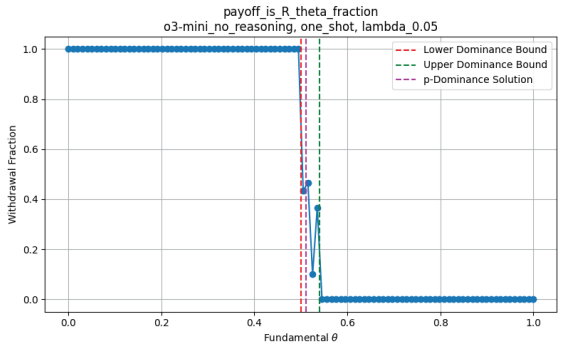
Notes: Blue markers and lines plot the empirical share of agents redeeming at each θ. Red: lower dominance bound where unconditional exit always dominates. Green: upper dominance bound where unconditional wait always dominates. Magenta: p-dominance solution—the probabilistic equilibrium threshold. Adding memory smooths the redemption curve and shifts the onset of exits past the upper bound, reflecting more history-aware decisions. One-shot (i.e., no memory) inference yields a sharp drop once the probabilistic threshold is crossed, whereas multi-step reasoning produces a more gradual transition. Increasing λ (left $\rightarrow$ right) amplifies the gap between memory and one-shot: strong regularization penalizes late redemptions more, causing one-shot agents to "break" earlier relative to the bounds.



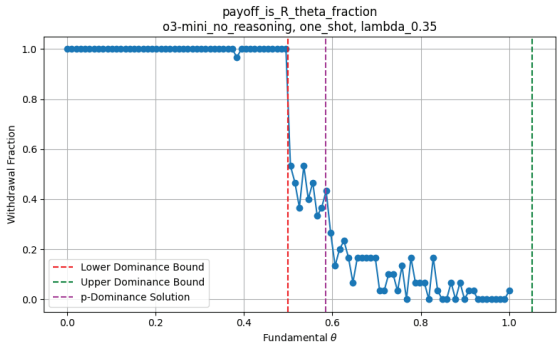
(a) o3-mini, no reasoning + memory @ $\lambda = 0.05$.



(b) o3-mini, no reasoning + memory @ $\lambda = 0.35$.



(c) o3-mini, no reasoning + one-shot @ $\lambda = 0.05$.



(d) o3-mini, no reasoning + one-shot @ $\lambda = 0.35$.

Figure 15: Redemption for The `o3-mini`

Notes: Red: lower dominance bound where unconditional exit always dominates. Green: upper dominance bound where unconditional wait always dominates. Magenta: p-dominance solution—the probabilistic equilibrium threshold. As θ increases, the redemption fraction declines, indicating agents wait longer before exiting when fundamentals improve. Adding memory smooths the curve and pushes the point at which withdrawals begin to higher θ, reflecting more cautious, history-aware decisions. One-shot (i.e., no memory) inference yields a sharp threshold drop once agents decide to exit, whereas without memory the transition is abrupt, especially under high λ. Moving from $\lambda = 0.05$ (top) to $\lambda = 0.35$ (bottom) exaggerates the gap between memory and one-shot: higher values of λ penalize late withdrawals more, causing one-shot to break down sooner.

4.2 Fundamental uncertainty

The analysis so far has highlighted the role that uncertainty about other investors' behavior and payoffs play a role in shaping Q-algorithms' redemption decisions and, ultimatly, financial fragility. In particular, the simulations results in Section 4.1 show that payoff uncertainty, which can be linked to the different type of securities the fund invests in, alters machines' behavior more than strategic uncertainty. Those simulations have uncovered that Q-learning is sensitive to payoff uncertainty since machines are unable to discriminate whether a low payoff is due to a bad draw or others' behaviour.

Starting from this observation, in this section, we run simululations to test whether this result remains valid once we introduce fundamental uncertainty. When algorithms do not observe the true state of fundamentals θ, but only receive an imperfect private signal about it, they confront themselves with a similar uncertainty about payoff, although the source is very different. In what follow, we test the role of fundamental uncertainty in affecting algorithms'behavior, building on the theoretical predictions derived in Section 2 and summarized in hypotheses H3 and H4. As in the previous section, we distinguish between an equity and a bond fund to capture the presence of payoff uncertianty. We start from the former.

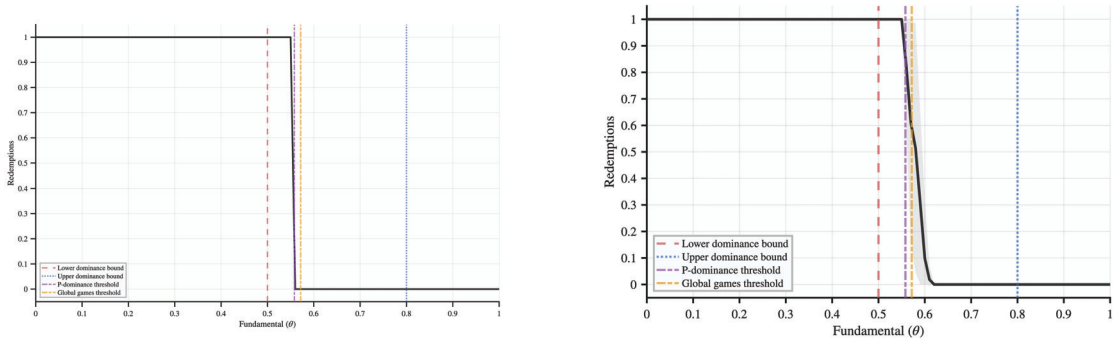
Before presenting the simulation results, it is useful to briefly mention how imperfect private signals are coded and how this affects the Q-learning iterative process. We draw the values for the fundamentals θ from the interval $[0,1]$, which we discretize in steps of 0.01. The level of fundamentals that each algorithm uses as the state s in any given episode, corresponds to its private signal that we compute as follows. We consider two values of $\eta = \{0.01, 0.05\}$. For each η and for any given θ, the noise term ε_i for agent i is drawn in the range $[-\eta, +\eta]$. All values in this whole interval are equally likely. Then, the signal for each agent is mapped into the states s used in the Q-learning process as in the previous section.[11]

The results for the case of an equity fund (i.e., in the absence of payoff uncertainty) are illustrated in Figure 16 and 17. Both figures include three panels, corresponding to three different levels of precision of the signal, i.e., $\eta = \{0, 0.01, 0.05\}$. The different levels of precision can be interpreted as a measure of fundamental uncertainty. In this sense, the case $\eta = 0$, which precisely corresponds to the case without fundamental uncertainty analyzed in Section 4.1, is meant to be used as benchmark..

The presence of fundamental uncertainty alters Q-learning algortithms' behavior in a way that it is similar to payoff uncertainty, although the detrimental impact on fragility seem milder. As the signal becomes less precise and so fundamental uncertainty increases, algorithms no longer redeeem early based on the p-dominance threshold. They starts to redeem at $t = 1$ for values of fundamental above the p-dominant cutoff, such an increase becomes more pronunced as η goes from 0.01 to 0.05.

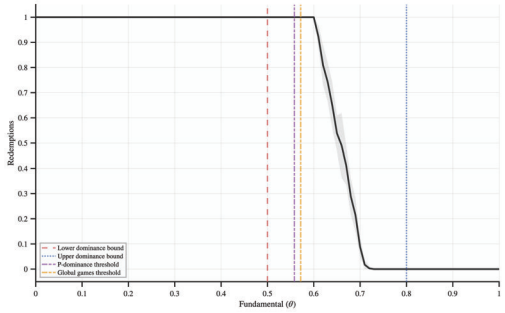
The behavior displayed by algorithms in the presence of fundamental uncertainty is in line with the theoretical predictions. Besides the limiting case where the precision of the signal is maximal (i.e., $\eta \to 0$) and all investors behave alike, even a small error term would lead to partial redemption equilibria to arise. In those equilibria, the share of early redemptions would decrease with the fundamentals θ, as we

[11] Since, when the precision of the signal is low, an algorithm could receive the same signal for different θ, at the beginning of simulations, we specify that there are only 75 possible states. Hence, 75 rows in the Q-matrix.



(a) Fully precise signal $\eta = 0$

(b) High precision signal $\eta = 0.01$



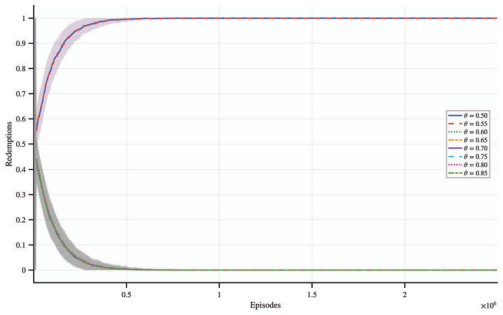
(c) Low precision signal $\eta = 0.05$

Figure 16: **Multiple Q-learning algorithms holding shares in an equity fund in the presence of fundamental uncertainty.** The figure illustrates the algorithms' redemption decision as a function of the fundamentals θ for different levels of precision of the signal, i.e., $\eta = \{0, 0.01, 0.05\}$. The figure is drawn for $\lambda = 0.25$. All other parameters are as specified in Section 3.1.

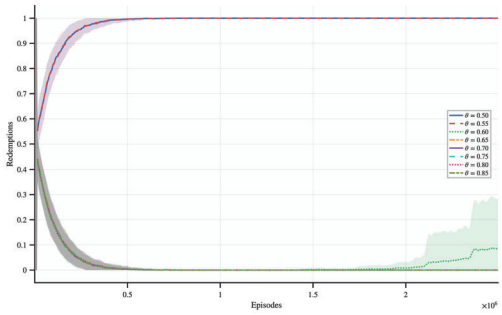
observe in Panels 17b and 17c in Figure 16. This is in line with the theoretical predictions in Proposition 2.

The region in which Q-algorithms stop coordinating becomes bigger as the precision of the signal lowers. Figure 17 illustrates how the rolling average of redemptions evolves over episodes.[12] When the signal is very precise, Q-algorithms converge on a full coordination equilibrium very quickly and for all θ.When the signal precision is low, instead, equilibria in which machines do not coordinate emerge for a large range of θ.

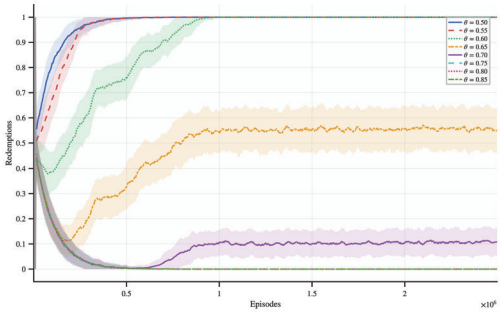
It is interesting to compare Panel 10b in Figure 10 with Panel 17c in Figure 17. The two differ in one important dimension. In Panel 17c, for some intermediate levels of fundamentals, the curves flatten and the shaded areas around them stay constant. This suggests, in line with the theoretical predictions, that the failure of algorithms to coordinate could be an equilibrium outcome in the case with fundamental uncertainty. This has also implications for fragility, as we consider next.



(a) Fully precise signal $\eta = 0$



(b) High precision signal $\eta = 0.01$



(c) Low precision signal $\eta = 0.05$

Figure 17: **Redemptions over episodes when the fund invests in equity securities in the presence of fundamental uncertainty.** The figure illustrates the evaolution of the average share of early redemptions over twenty-five rounds for different levels of precision of the signal, i.e., $\eta = \{0, 0.01, 0.05\}$. The figure is drawn for $\lambda = 0.25$. All other parameters are as specified in Section 3.1.

The role that fund illiquidity plays for fragility is illustrated in Figure 18. There are two main insights. First, fragility always increases with illiquidity, irrespective of the precision of the signal, thus confirming

[12]Section D includes plots of the frequency of episodes in which the machines do not coordinate on either redeeming at $t = 1$ or waiting until $t = 2$ for the three different levels of precision of the signal.

hypothesis H4. Second, fragility is ceteris paribus higher when fundamental uncertianty is higher (i.e., the signal is less precise). Interestingly, even when the precision is low, the level of fragility in the presence of fundamental uncertainty and no payoff uncertainty is always below that in the absence of fundamental uncertainty when the fund invests in debt securities (i.e., when payoff uncertainty is present). This can be seen by comparing levels on the y-axis in the above figure to those in Figure 12 and suggests that payoff uncertainty is more detrimental to fragility than fundamental uncertainty.

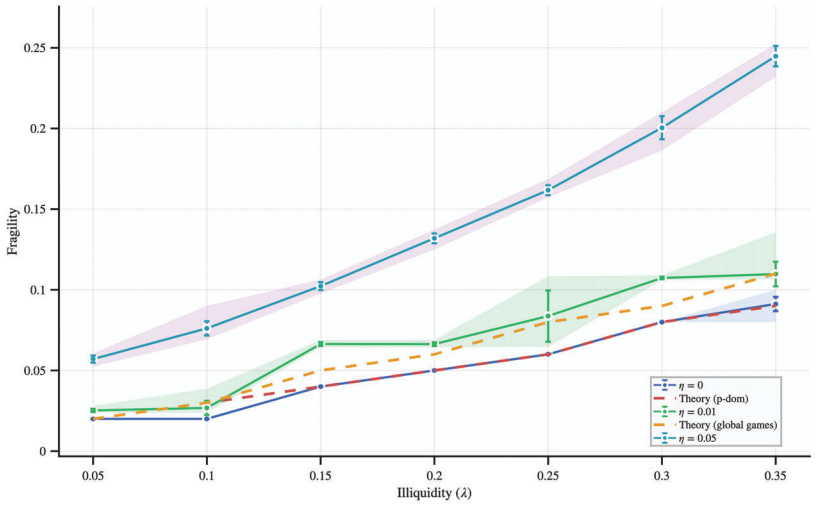


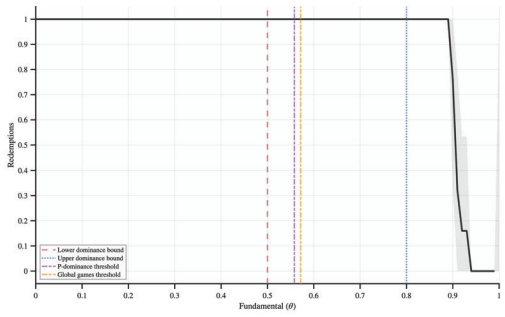
Figure 18: **Fragility and fund illiquidity in an equity fund.** The figure illustrates how fragility changes with illiquidity for different levels of signal precision, i.e., $\eta = 0$ (dark blue line), $\eta = 0.01$ (green line) and $\eta = 0.05$ (light blue line). The figure is drawn for $\lambda = 0.25$, while all other parameters are as specified in Section 3.1.

We now move on to consider a bond fund in the presence of fundamental uncertainty. In this case, the algorithms take their redemption decisions in an environment characterized by both payoff and fundamental uncertainty. The results of the simulations are illustrated in Figures 19 and 20.

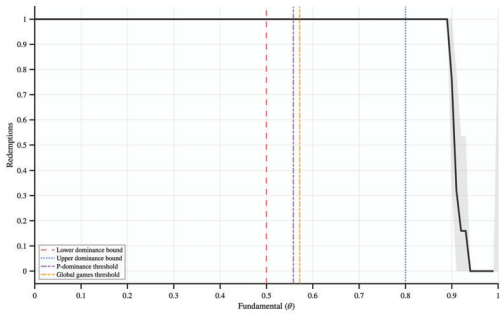
As before, the first panel in which figure, corresponding to the case $\eta = 0$ confirms the result in Section 4.1 and offers a useful benchmark to understand the role of fundamental uncertainty. Payoff uncertainty compounds the increase in fragility associated with the change in the precision of the signal. As a result, as shown in Panels 19c and 19b, the curves significantly move onto the right. The comparison with the corresponding figures when the fund invests in equity securities (Figures 16) highlights a significant increase in the average share of redemptions at $t = 1$ and so overall in fragility.[13]

We conclude the analysis of the effect of fundamental uncertainty on the behavior of Q-learning algorithms testing hypothesis H4 on the role of illiquidity for fragility. In the presence of fundamental uncertainty, the limited dependence of fragility on illiquidity, which is the result of payoff uncertainty, emerged in Figure 7 is confirmed. This can be seen from the very limited slopes of the three solid curves in Figure 21 below.

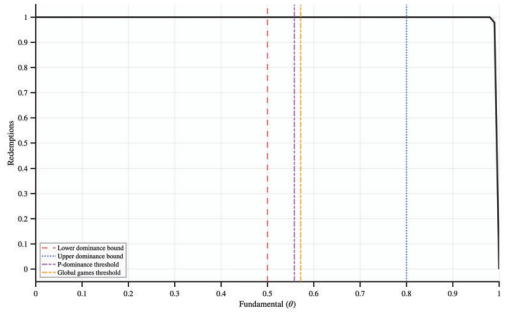
[13] In Section D, we again include plots of the frequency of episodes in which the machines do not coordinate on either redeeming at $t = 1$ or waiting until $t = 2$ for the three different levels of precision of the signal.



(a) Fully precise signal $\eta = 0$

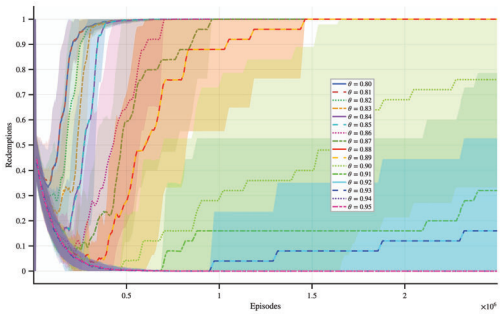


(b) High precision signal $\eta = 0.01$

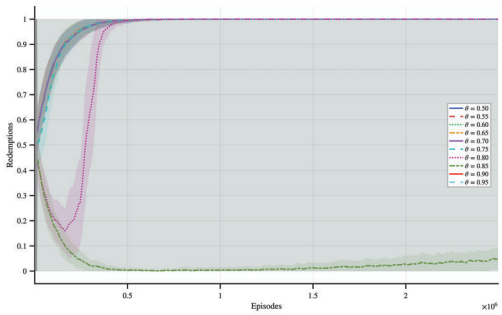


(c) Low precision signal $\eta = 0.05$

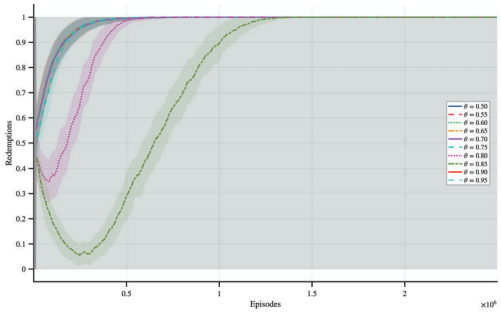
Figure 19: **Multiple Q-learning algorithms when the fund invests in debt securities in the presence of fundamental uncertainty.** The figure illustrates the algorithms' redemption decision as a function of the fundamentals θ for different levels of precision of the signal, i.e., $\eta = \{0, 0.01, 0.05\}$. The figure is drawn for $\lambda = 0.25$. All other parameters are as specified in Section 3.1.



(a) Fully precise signal $\eta = 0$



(b) High precision signal $\eta = 0.01$



(c) Low precision signal $\eta = 0.05$

Figure 20: **Redemptions over episodes when the fund invests in debt securities in the presence of fundamental uncertainty.** The figure illustrates the evaolution of the average share of early redemptions over twenty-five rounds for different levels of precision of the signal, i.e., $\eta = \{0, 0.01, 0.05\}$. The figure is drawn for $\lambda = 0.25$. All other parameters are as specified in Section 3.1.

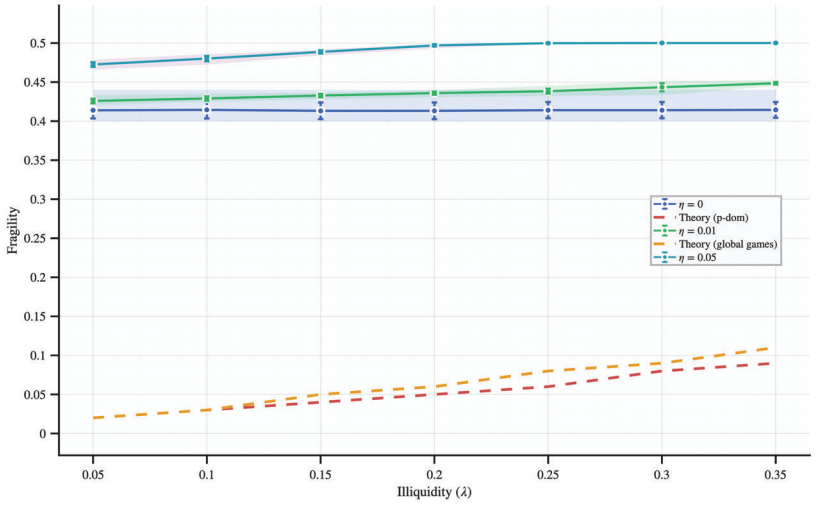


Figure 21: **Fragility and fund illiquidity in bond funds.** The figure illustrates how fragility changes with illiquidity for different levels of signal precision, i.e., $\eta = 0$ (dark blue line), $\eta = 0.01$ (green line) and $\eta = 0.05$ (light blue line). The figure is drawn for $\lambda = 0.25$, while all other parameters are as specified in Section 3.1.

5 Conclusion

This study set out to examine whether AI agents could themselves become a source of financial fragility (or a force for stability) when faced with classic coordination problems like bank runs. Our key finding is that the equilibrium selection in coordination games can depend critically on the AI agents' decision-making architectures. AI agents, if properly understood, have the potential to act as *deus ex machina*, guiding markets to efficient outcomes. However, as we have shown, they can also bring their own forms of fragility, coordinating at superhuman speed on outcomes that may not always be desirable. Our work contributes to the nascent literature on AI in finance by highlighting that the "behavioral" differences between AI systems matter. It is not just whether AI is used, but how it's designed.

As AI-driven decision-making becomes more prevalent in finance (from trading algorithms to automated fund managers and robo-advisors), understanding the emergent behavior of these agents in aggregate is crucial for preventing undesirable outcomes like bank runs, market crashes, or systemic coordination failures. Our paper demonstrates that AI agents can exhibit coordination dynamics analogous to (and sometimes more extreme than) humans. Regulators and central banks tasked with financial stability should consider incorporating AI-based agents into their stress-testing frameworks. Traditional stress tests often model worst-case scenarios by assuming human actors all panic or all stay calm, but AI agents might have different thresholds or patterns. Policymakers should invest in understanding these AI behaviors to avoid being blindsided by a new form of digital herd behavior.

Our results also highlight challenges for AI alignment when multiple AI agents interact. Much of AI alignment research focuses on aligning a single AI's objectives with human values or intentions. However, when many AI agents operate in the same environment, we face a multi-agent alignment

problem: ensuring that their interactions lead to socially desirable outcomes. In our game, each AI was "aligned" to maximize its individual reward (and loosely, to follow the game's instructions), yet collectively this sometimes led to the worst global outcome (everyone running, everyone getting less). Even perfectly rational, profit-driven AI agents can collectively generate irrational market outcomes due to strategic uncertainty, much as humans do (Lorè and Heydari, 2024). Furthermore, AI agents might reach these outcomes faster or in unison because of their speed and homogeneity.

Finally, our work has implications beyond finance. Coordination failures and multi-agent dilemmas appear in many domains (e.g., self-driving cars at an intersection, or multiple AI systems allocating resources in a supply chain). In summary, the policy implications of LLM-based agent behavior in financial coordination problems are clear: traditional financial safeguards remain vital (high liquidity, insurance, transparency), but they must be augmented with new tools for AI behavior monitoring and alignment. The literature so far provides optimism that with careful design, LLM agents could become an integral part of financial ecosystems, offering new insights and capabilities, so long as we also heed the lessons about coordination and stability.

References

Akata, E., L. Schulz, J. Coda-Forno, S. J. Oh, M. Bethge, and E. Schulz (2023). Playing repeated games with large language models. arXiv:2305.16867.

Banchio, M. and G. Mantegazza (2022). Artificial intelligence and spontaneous collusion. In Proceedings of the 2022 ACM Conference on Economics and Computation, pp. 1–2.

Banchio, M. and A. Skrzypacz (2022). Market design for ai algorithms. SSRN 3695048.

Bebchuk, L. and I. Goldstein (2011). Self-fulfilling credit market freezes. Review of Financial Studies 24(11), 3519–3555.

Calvano, E., G. Calzolari, V. Denicolo, and S. Pastorello (2020). Artificial Intelligence, Algorithmic Pricing, and Collusion. American Economic Review 110(10), 3267–3297.

Camerer, C. and T.-H. Ho (1999). Experience-weighted attraction learning in normal-form games. Econometrica 67(4), 827–874.

Carlsson, H. and E. Van Damme (1993). Global games and equilibrium selection. Econometrica: Journal of the Econometric Society, 989–1018.

Chen, Q., I. Goldstein, and W. Jiang (2010). Payoff Complementarities and Financial Fragility: Evidence from Mutual Fund Outflows. Journal of Financial Economics 97(2), 239–62.

Clark, C. L. (2010, March). Controlling Risk in a Lightning-Speed Trading Environment. Chicago Fed Letter (272). Reports that algorithmic high-frequency trading accounted for about 70% of U.S. equity trading volume in 2009.

Colliard, J.-E., T. Foucault, and S. Lovo (2025). Algorithmic pricing and liquidity in securities markets. HEC Paris Research Paper No. FIN-2022-1459.

Cong, L. W., K. Tang, J. Wang, and Y. Zhang (2023). Alphaportfolio: Direct construction through reinforcement learning and interpretable ai. SSRN 4124085.

Cont, R. and W. Xiong (2024). Dynamics of market making algorithms in dealer markets: Learning and tacit collusion. Mathematical Finance 34(2), 467–521.

Diamond, D. and P. Dybvig (1983). Bank Runs, Deposit Insurance and Liquidity. Journal of Political Economy 91(3), 401–19.

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

Erev, I. and A. E. Roth (1998). Predicting how people play games: Reinforcement learning in experimental games with unique mixed-strategy equilibria. American Economic Review 88(4), 848–881.

Goldstein, I., H. Jiang, and D. T. Ng (2017). Investor flows and fragility in corporate bond funds. Journal of Financial Economics 126(3), 592–613.

Goldstein, I. and A. Pauzner (2005). Demand Deposity Contracts and the Probability of Bank Runs. Journal of Finance 60, 1293–1327.

Heinemann, F., R. Nagel, and P. Ockenfels (2004). The theory and experiment of strategic complementarities in games. Econometrica 72(5), 1583–1599.

Hendershott, T., C. M. Jones, and A. J. Menkveld (2011). Does algorithmic trading improve liquidity? Journal of Finance 66(1), 1–33.

Horton, J. J. (2023). Large language models as simulated economic agents: What can we learn from homo silicus? NBER Working Paper 31282.

Johnson, N., F. Zhao, E. Hunsader, J. Meng, S. Ravindar, S. Carran, and B. Tivnan (2013). Abrupt jumps and ultrafast market dynamics. Journal of Economic Behavior and Organization 85(1), 44–58.

Kyle, A. S. (1985). Continuous auctions and insider trading. Econometrica: Journal of the Econometric Society, 1315–1335.

Lorè, N. and B. Heydari (2024). Strategic Behavior of Large Language Models and the Role of Game Structure versus Contextual Framing. Scientific Reports 14(1), 18490.

Morris, S. and H. S. Shin (1998). Unique equilibrium in a model of self-fulfilling currency attacks. American Economic Review, 587–597.

Morris, S. and H. S. Shin (2002). Measuring Strategic Uncertainty. Princeton University. www. princeton. edu/ hsshin/www/barcelona. pdf.

Yang, H. (2024). AI Coordination and Self-fulfilling Financial Crises.

Yang, J. (2025). Ai coordination and self-fulfilling financial crises. Working Paper, Toulouse School of Economics.

A Proof of Proposition 1

The proof is straightforward and boils down to characterize the equilibria in Pure strategies.

For extreme values of fundamentals θ, investors have a dominant action. We start with the low values of θ. Irrespective of what other investors choose, redeeming his share at $t = 1$ is optimal for an investor when even the highest payoff he can accrue at $t = 2$, which corresponds to the payoff accrued if no other investors redeem, is smaller than R_1. Formally, this is the case when $R_1 R\theta < R_1$ that is when $\theta < \underline{\theta} \equiv \frac{1}{R}$.

Symmetrically, waiting until $t = 2$ is a dominant action when even the worst payoff an investor expects to receive at the final date, which corresponds to the payoff accrued if all $A-1$ investors redeem, is larger than R_1. Formally, this is the case when $R_1 R\theta \frac{N-(A-1)}{N-(A-1)(1+\lambda)} < R_1$ that is when $\theta > \overline{\theta} = \frac{1}{R} \frac{N-(A-1)}{N-(A-1)(1+\lambda}$.

When $\theta \in (\underline{\theta}, \overline{\theta})$, both redeeming at $t = 1$ and not redeeming are equilibria as if an investors expect the others to redeem, it is otpimal to redeem as he otherwise will accrue a payoff at $t = 2$ that is smaller than R_1. Simularly, if no one else redeem at $t = 1$ is optimal for an investor not to redeem since $R_1 R\theta > R_1$ for all $theta > \underline{\theta}$. This completes the proof. □

B Proof of Proposition 2

The proof adapts the standard approach in the global game literature (Goldstein and Pauzner, 2005) and Chen et al. (2010) to the case with a discrete number of players. The arguments in their proof establish that there is a unique equilibrium in which investors reedem at $t = 1$ if and only if the signal they receive is below a common signal θ^*, which is the signal at which an investor is indifferent between reedeeming at $t = 1$ and $t = 2$ given what he believes about the signals received by other investors and, in turn, their behaviours.

We starts by characterizing investors' decisions in two extreme ranges of fundamentalss where investors have a dominant action. These two ranges corresponds to the one characterized in Proposition 1. When $\theta < \underline{\theta}$ redeeming at $t = 1$ is a dominant action and we refer to this range as the lower dominance region. When $\theta > \overline{\theta}$, redeeming at $t = 2$ is the dominant action and we refer to this range as the upper dominance region.

Consider now the intermediate region where $\theta \in (\underline{\theta}, \overline{\theta})$. Assume that investors behave according to a threshold strategy: they redeem their shares if they receive a signal below θ^* and wait until $t = 2$ otherwise. Given that the signal is uniformly distributed, for each investor, the probability of receiving a signal below θ^* is $\frac{\theta^* - \theta + \eta}{2\eta}$. Building on this, we can compute the share of investors receiving the signal below the cutoff, which is given by

$$n = \sum_{i=1}^{A} \{\theta_i < \theta^*\} \sim Binomial\left(A, \frac{\theta^* - \theta + \eta}{2\eta}\right). \tag{5}$$

A simple calculation of the expected value, gives the expression in the proposition. It follows then that

the probability that n out of A investors have received a signal below θ^* is given by:

$$f(n,A) = \binom{A}{n}\left(\frac{\theta^* - \theta + \eta}{2\eta}\right)^n \left(1 - \frac{\theta^* - \theta + \eta}{2\eta}\right)^{A-n}. \tag{6}$$

Using the above, we can then compute the probability that the depositor that receives the signal θ^* assigns to n out of $A-1$ investors redeeming at $t=1$ as follows:

$$\frac{\binom{A-1}{n}}{2\eta}\int_{\theta^*-\eta}^{\theta^*+\eta} \frac{(\theta^* - \theta + \eta)^n(\theta - \theta^* - \eta)^{A-1-n}}{(2\eta)^{A-1}} d\theta \tag{7}$$

As shown in Morris and Shin (2002), this probability is equal to $\frac{1}{1+A-1}$.

Hence, the indifference condition that characterizes the run threshold θ^* reads:

$$\sum_{n=0}^{A-1} \frac{1}{A}\frac{N - n(1+\lambda)}{N-n} R_1 R\theta = R_1, \tag{8}$$

which gives the expression (2) in the proposition. □

C Proof of Proposition 3

While the definition of risk-dominant equilibrium in a N-players game with symmetric strategies is in principle ambigous, we here characterize the strategy based on the so-called p-dominance criterion and prove that this is consistent with the Q-algorithms' behaviours.

This strategy relies on the comparison between expected payoffs under the belief that $A-1$ other players choose a given strategy with some probability p. Then, a player compares the expected payoff from choosing run with the expected payoff from choosing no run (which depends on how many others choose run). The critical probability p^* is pinned down as the one at which the expected utilities are equal. Then, playing run is the risk-dominant equilibrium as long as $p^* > 0$.

Given our assumption on fund illiquidity, i.e., $A < \frac{N}{1+\lambda}$, the payoff from redeeming at date 1, which we denote as u_1 is always R_1 irrespective of what others do. The payoff at date 2, which we denote as u_2, is, instead, given by

$$u_2 = \frac{N - n(1+\lambda)}{N-n} R_1 R\theta \tag{9}$$

Denote as p, the probability that each of the other $A-1$ players choose to redeem. The number of players (out of $A-1$) choosing to redeem (i.e., n) is $n \sim Bin(A-1,p)$. The expected payoff from choosing not to redeem is then equal to

$$u_2 = \sum_{n=0}^{A-1} \binom{A-1}{n} p^n (1-p)^{A-1-n} R_1 \frac{N - n(1+\lambda)}{N-n} R\theta. \tag{10}$$

Equating u_1 to u_2 gives the following condition

$$1 = \sum_{n=0}^{A-1} \binom{A-1}{n} p^n (1-p)^{A-1-n} \frac{N - n(1+\lambda)}{N-n} R\theta, \tag{11}$$

whose solution is p^*. As we are interested in finding a cutoff for θ, denote it as θ^{RD} at which the risk-dominant equilibrium changes, we first note that u_2 increases with θ and decreases with p so that redeeming is going to be the risk-dominant equilibrium when $\theta < \theta^{RD}$, while for $\theta > \theta_{RD}$ the risk-dominant equilibrium is not redeeming. Evaluating (11) at $p = 0.5$ allows us to find the cutoff θ^{RD}. Since the RHS in (11) is decreasing in λ and increasing in θ, the comparative statics with respect to λ follows directly. The numerical values for the various θ^{RD} presented in the figures are obtained evaluating (11) using $A = 70$, $N = 100$, $\lambda = 0.05, 0.1, 0.2$, respectively. Hence, the proposition follows. □

D Additional figures

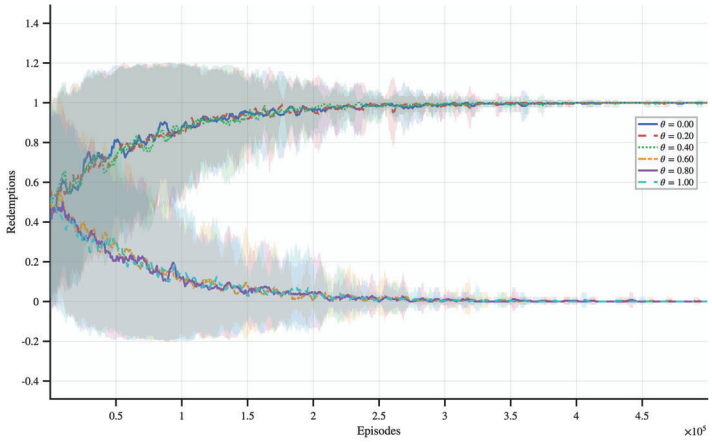


Figure 22: **Redemptions over episodes for a single Q-learning algorithm when the fund invests in equity securities.** The figure illustrates that the algorithm converges at around the 250000th episode either to the no redemption equilibrium or to the full redemption depending on different values of fundamentals θ.

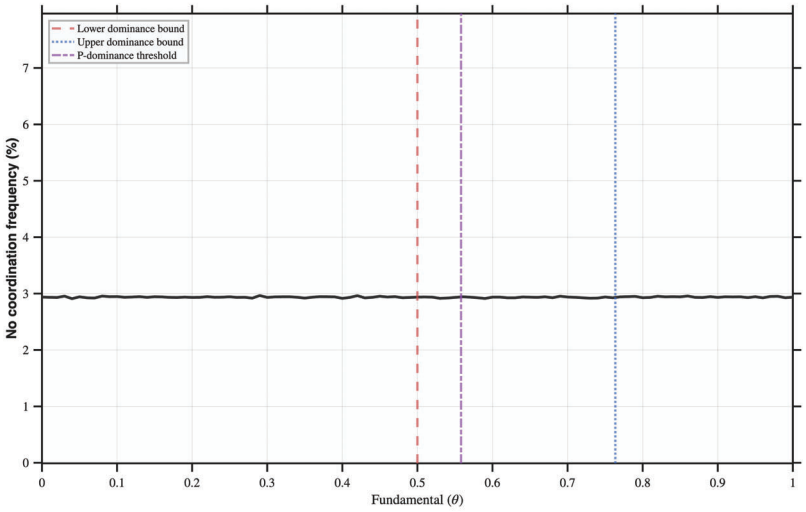


Figure 23: **Frequency of episodes in which algorithms fail to coordinate** The figure illustrates that, for all θ, the average share of episodes in which algorithms do not coordinate is about 3 per cent.

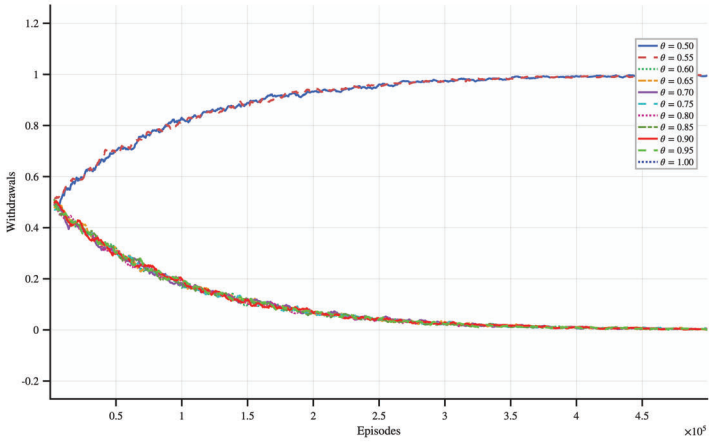
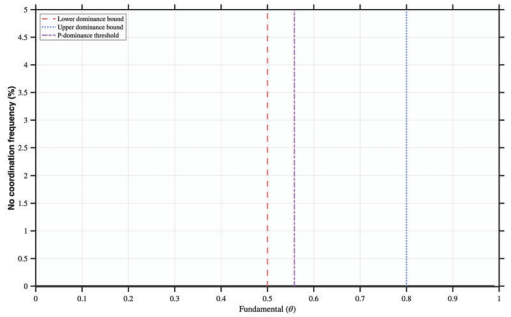
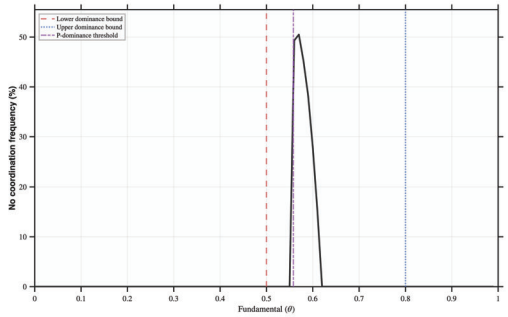


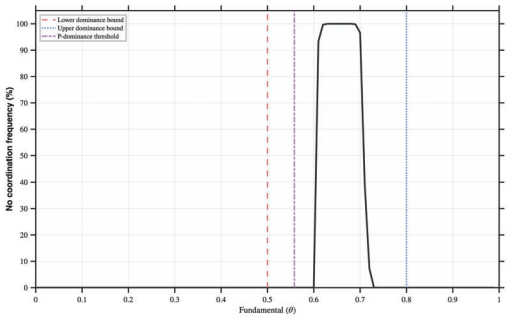
Figure 24: **Redemptions over episodes in the presence of fundamental uncertainty** The figure illustrates that the algorithms converge at around the 300000th episode either to the no redemption equilibrium or to the full redemption depending on different values of fundamentals θ.



(a) Precise signal $\eta = 0$

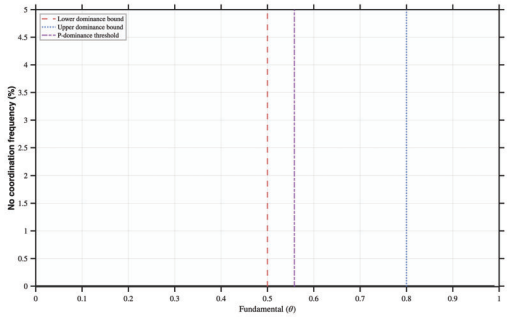


(b) High precision signal $\eta = 0.01$

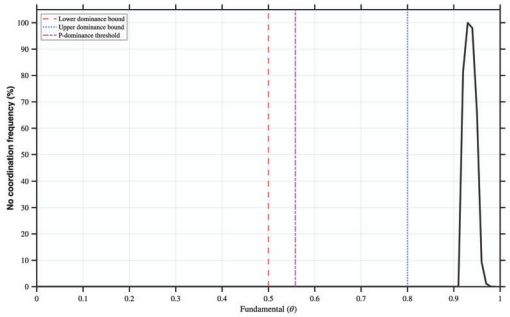


(c) LOw precision signal $\eta = 0.05$

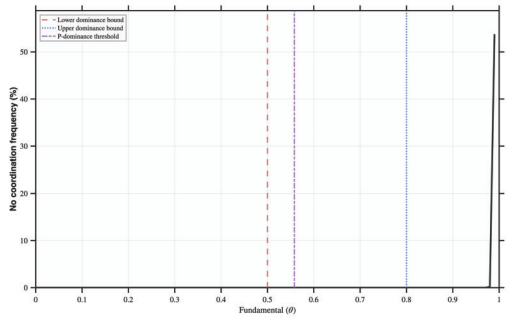
Figure 25: **Lack of coordination when the fund invests in equity securities in the presence of fundamental uncertainty.** The figure illustrates the average share of episodes (over the 25 rounds) in which algorithms do not coordinate either on redeeming at $t = 1$ or at $t = 2$ for different levels of precision of the signal, i.e., $\eta = \{0, 0.01, 0.05\}$. The figure is drawn for $\lambda = 0.25$. All other parameters are as specified in Section 3.1.



(a) Precise signal $\eta = 0$



(b) High precision signal $\eta = 0.01$



(c) LOw precision signal $\eta = 0.05$

Figure 26: **Lack of coordination when the fund invests in debt securities in the presence of fundamental uncertainty.** The figure illustrates the average share of episodes (over the 25 rounds) in which algorithms do not coordinate either on redeeming at $t = 1$ or at $t = 2$ for different levels of precision of the signal, i.e., $\eta = \{0, 0.01, 0.05\}$. The figure is drawn for $\lambda = 0.25$. All other parameters are as specified in Section 3.1.

E Fund illiquidity and payoff uncertaity at date 1

In this section, we relax the assumption concerning fund illiquidity and assume that $A > \frac{N}{1+\lambda}$. This implies that in the presence of a sufficiently large number of early redemptions, i.e., when $n \geq \overline{n} \equiv \frac{N}{1+\lambda}$, the fund liquidates its entire portfolio at date 1 and default. In this circumstance, we assume that early redeeming investors receive nothing due to the presence of full bankruptcy costs. This modification relative to the benchmark model introduces payoff uncertainty at date 1 and, thus, allows us to check whether our results concerning to the Q-algorithms' behaviour in the presence of payoff uncertainty are robust.[14] As it will become useful for the characterization of the case with fundamental uncertainty, in line with the literature, e.g., Goldstein and Pauzner (2005), we assume that when $\theta = 1$, $\lambda = 0$ and the liquidation value of the fund's investment jumps to R. This implies that when $\theta = 1$, it is a dominant action for the investors not to redeem at date 1.

To isolate the role of payoff uncertainty at date 1, we characterize the equilibrim assuming that the bank's investment returns $R\theta$ at date 2, i..e, no payoff uncertainty at the final date. We proceed as in the main text, deriving first the equilibrium in the absence of fundamental uncertainty (i.e., θ is observable) and then assuming that investors receive an imperfect private signal on θ of the form $s_i = \theta + \varepsilon_i$, with $\varepsilon_i \sim [-\eta, +\eta]$. The following proposition characterize the equilibria in the two instances.

Proposition 5. *When the fundamentals θ are observable, all investors redeem at $t = 1$ when $\theta \leq \underline{\theta} \equiv \frac{1}{R}$ and wait until $t = 2$ when $\theta = 1$. In the range $\theta \in (\underline{\theta}, 1)$, redeeming at date 1 and at date 2 are both equilibria.*

When the fundamentals θ are not observable, model has a unique symmetric Bayes-Nash equilibrium in threshold strategies that is characterized by a critical signal

$$\theta^* = \frac{\sum_{n=0}^{\overline{n}}}{\sum_{n=0}^{\overline{n}} \frac{N-n(1+\lambda)}{N-n} R}, \tag{12}$$

such that an investor will redeem their share at $t = 1$ if and only if their signal is below θ^*.

Proof. The proof follows closely that of Propostion 1 and 2, with the only difference that, in the case of fundamental uncertainty, the indifference condition giving θ^* is equal to:

$$\sum_{n=0}^{\overline{n}} \frac{1}{A} \frac{N-n(1+\lambda)}{N-n} R_1 R\theta = \sum_{n=0}^{\overline{n}} \frac{1}{A} R_1. \tag{13}$$

The p-dominance threshold in the presence of payoff uncertainty at $t = 1$ θ_{RD} solves $u_1 = u_2$ evaluated at $p = \frac{1}{2}$, where

$$u_1 = R_1 \sum_{n=0}^{\overline{n}} \binom{A-1}{n} p^n (1-p)^{A-1-n},$$

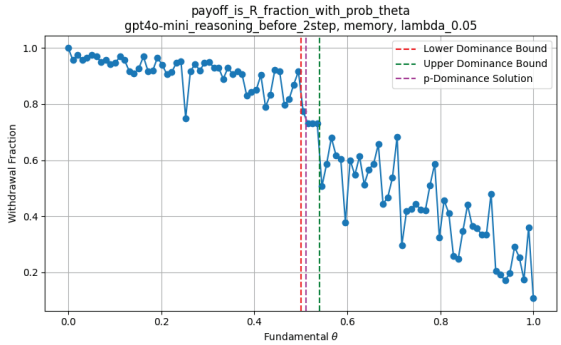
[14] Notice that the assumption of full bankrutpcy costs give rise to a payoff structure that is akin to the case where early redeeming investors are repaid according to a sequential service schedule.

and

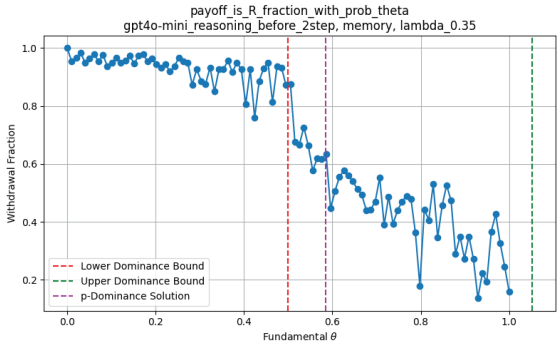
$$u_2 = R_1 \sum_{n=0}^{\bar{n}} \binom{A-1}{n} p^n (1-p)^{A-1-n} R\theta \frac{N - n(1+\lambda)}{N-n}.$$

□

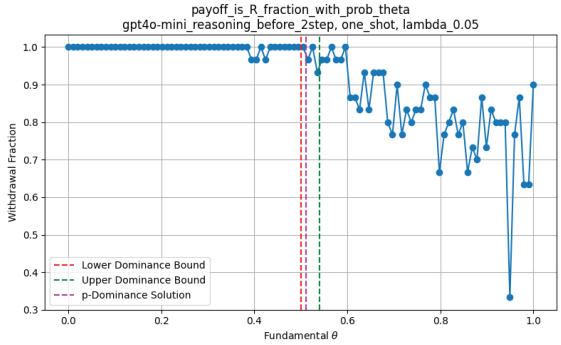
F Payoff Uncertainty – LLMs



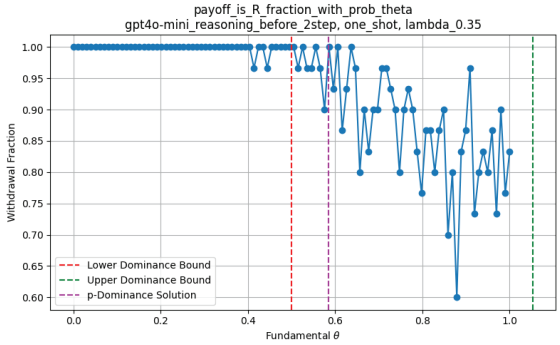
(a) 4o-mini, reasoning + memory @ $\lambda = 0.05$.



(b) 4o-mini, reasoning + memory @ $\lambda = 0.35$.

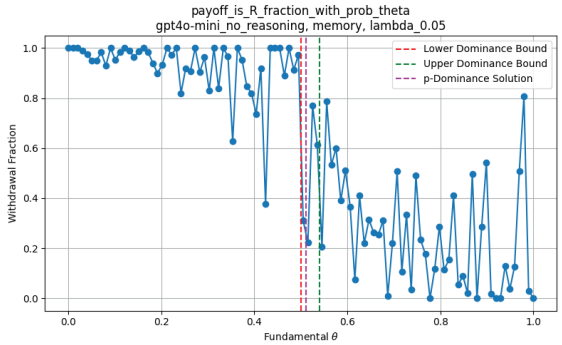


(c) 4o-mini, reasoning + one-shot @ $\lambda = 0.05$.

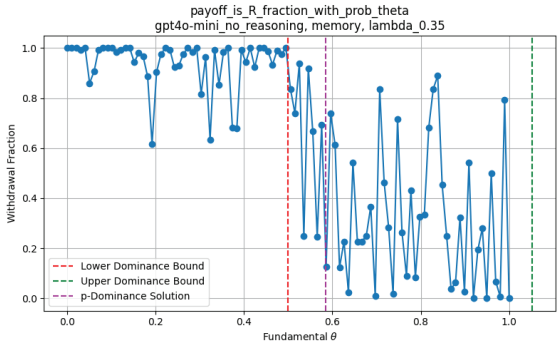


(d) 4o-mini, reasoning + one-shot @ $\lambda = 0.35$.

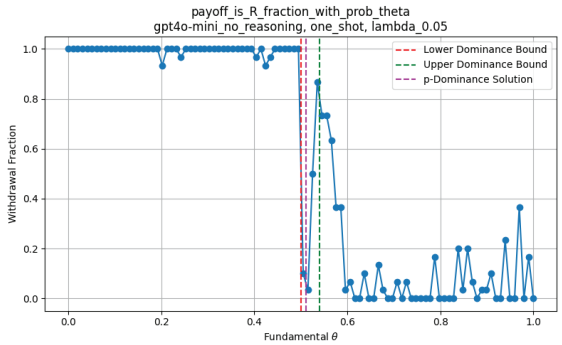
Figure 27: Withdrawal Fraction vs. Probability Parameter θ for the `4o-mini` with Reasoning
Notes: Solid blue line with markers shows the empirical share of agents choosing payoff R as θ varies. Red: lower dominance bound where always choosing R dominates. Green: upper dominance bound where always waiting dominates. Magenta: p-dominance solution—the mixed-strategy equilibrium threshold. Memory smooths the curve and shifts the drop-off point to higher θ, reflecting history-aware caution. One-shot inference produces sharp, step-like drops once the probabilistic threshold is crossed. Higher λ (right panels) penalizes late risky plays more, amplifying volatility in one-shot runs.



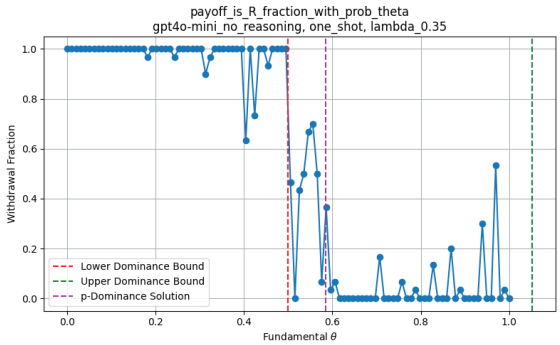
(a) 4o-mini, no reasoning + memory @ $\lambda = 0.05$.



(b) 4o-mini, no reasoning + memory @ $\lambda = 0.35$.

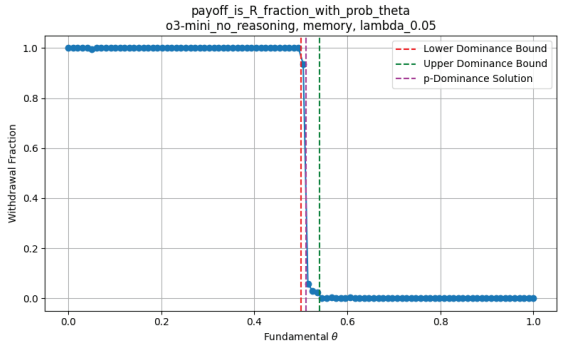


(c) 4o-mini, no reasoning + one-shot @ $\lambda = 0.05$.

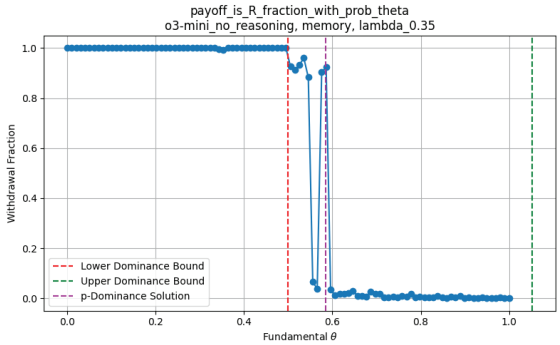


(d) 4o-mini, no reasoning + one-shot @ $\lambda = 0.35$.

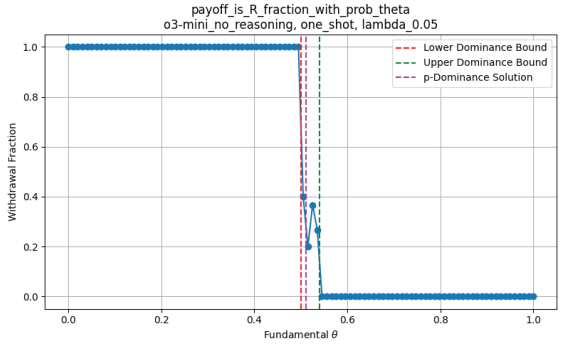
Figure 28: Withdrawal Fraction vs. Probability Parameter θ for the `4o-mini` without Reasoning
Notes: Solid blue line with markers shows the empirical share of agents choosing payoff R as θ varies. Red: lower dominance bound where always choosing R dominates. Green: upper dominance bound where always waiting dominates. Magenta: p-dominance solution—the mixed-strategy equilibrium threshold. Memory smooths the curve and shifts the drop-off point to higher θ, reflecting history-aware caution. One-shot inference yields abrupt threshold drops and greater volatility once the bound is crossed. Stronger regularization (right) amplifies differences, causing one-shot runs to break down sooner relative to the bounds.



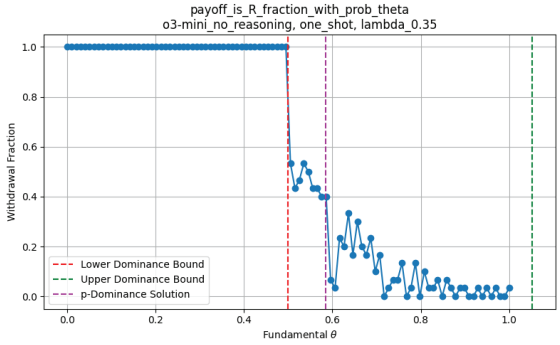
(a) o3-mini, no reasoning + memory @ $\lambda = 0.05$.



(b) o3-mini, no reasoning + memory @ $\lambda = 0.35$.



(c) o3-mini, no reasoning + one-shot @ $\lambda = 0.05$.



(d) o3-mini, no reasoning + one-shot @ $\lambda = 0.35$.

Figure 29: Withdrawal Fraction vs. Probability Parameter Θ for the `o3-mini` without Reasoning
Notes: Solid blue line with markers shows the empirical share of agents choosing payoff R as θ varies. Red: lower dominance bound where always choosing R dominates. Green: upper dominance bound where always waiting dominates. Magenta: p-dominance solution—the mixed-strategy equilibrium threshold. Memory smooths the curve and shifts the drop-off point to higher θ, reflecting history-aware caution. One-shot inference yields abrupt threshold drops and greater volatility once the bound is crossed. Higher values of lambfa amplifies differences, causing one-shot runs to break down sooner relative to the bounds.