WORKING PAPERS

N° 1680

October 2025

"Endogenous Quality in Social Learning"

Georgy Lukyanov, Konstantin Shamruk and Ekaterina Logina

Toulouse School of Economics

Endogenous Quality in Social Learning*

Georgy Lukyanov† Konstantin Shamruk‡ Ekaterina Logina§

Abstract

We study a dynamic reputation model with a fixed posted price where only purchases are public. A long-lived seller chooses costly quality; each buyer observes the purchase history and a private signal. Under a Markov selection, beliefs split into two cascades—where actions are unresponsive and investment is zero—and an interior region where the seller invests. The policy is inverse-U in reputation and produces two patterns: Early Resolution (rapid absorption at the optimistic cascade) and Double Hump (two investment episodes). Higher signal precision at fixed prices enlarges cascades and can reduce investment. We compare welfare and analyze two design levers: flexible pricing, which can keep actions informative and remove cascades for patient sellers, and public outcome disclosure, which makes purchases more informative and expands investment.

Keywords: Reputation; Social learning; Informational cascades; Product quality; Dynamic games.

JEL Classification Numbers: D82; D83; C73; L15.

1 Introduction

Many fixed-price service markets reveal only *actions* to outside observers. Walk-in restaurants and cafés make prices and queues publicly visible, yet bystanders do not observe realized satisfaction; they see who buys and who passes by. Similar observability arises for ticketed events with posted prices and visible attendance and, in digital

*This paper previously circulated under the title "Reputation and Social Learning." We thank Kirill Bukin, Miaomiao Dong, Olivier Gossner, Christian Hellwig, Natalia Kovaleva, Yukio Koriyama, Harry Pei, and participants at the 30th Stony Brook International Conference on Game Theory, the 2019 Econometric Society European Meeting (ESEM) in Manchester, and internal seminars at ICEF and NES for helpful comments. All remaining errors are our own.

†Toulouse School of Economics, 1 Esp. de l'Université, 31080 Toulouse Cedex, `georgy.lukyanov@tse-fr.eu`.

‡Toulouse School of Economics, 1 Esp. de l'Université, 31080 Toulouse Cedex, `c.shamruk@gmail.com`.

§Higher School of Economics, Shabolovka 26 Moscow, Russia, 119049, `loginka1995@gmail.com`.

markets, for subscription services where outsiders see uptake but not post-purchase outcomes. In such environments, buyers use observed purchase/non-purchase decisions as social information, and sellers can influence the informativeness of those actions by choosing costly quality that shifts the distribution of private experiences. This paper studies the dynamic incentives that follow.[1]

Fixed posted prices are pervasive in IO settings—menus, platform subscription fees, and ticketed events—and onlookers often observe only actions (queues, attendance, subscription counts) rather than realized outcomes. We study how a seller's dynamic quality choice interacts with this action-based social learning environment. Relative to canonical IO reputation models with outcome observability or price adjustment, the information constraint here is sharper: the seller's quality primarily matters by sustaining the informativeness of actions. This generates distinct comparative statics and clear design levers that align with platform practice (pricing bands; outcome disclosure).

We model a long-lived seller who, in each period, chooses a costly quality action before meeting a short-lived buyer. The buyer observes the history of purchases and a private signal about match quality and then decides whether to buy at a fixed posted price.[2] The key institutional feature is that outsiders observe only *actions* (purchase/non-purchase), not realized outcomes; hence the public state evolves via the informativeness of actions themselves.[3] The seller trades off current revenue and cost against the dynamic value of pushing the public belief in a favorable direction by keeping actions informative.

Three findings organize the analysis. First, public beliefs partition into two *cascade* regions—pessimistic (no one buys) and optimistic (everyone buys)—and an interior *experimentation* region in which actions respond to private information. The seller optimally does not invest in the cascades and invests only in the interior (Theorem 2 and Corollary 2). The underlying force is a single marginal-value difference across the two next-period belief updates that the buyer's action can induce; this difference pins down when quality investment pays (Proposition 1).

Second, within the experimentation region the equilibrium quality policy is *inverse-U* in reputation: investment is highest at intermediate beliefs and falls to zero as beliefs approach either cascade (Proposition 2, with sufficient curvature conditions in Appendix B). This shape yields two dynamic patterns. In an *Early-Resolution* regime, beliefs quickly drift to the optimistic cascade and investment winds down; in a *Double-Hump* regime, investment appears in two episodes separated by a pause strictly inside the experimentation region (Theorem 3 and Proposition 5). When the per-period social value of information is concave in "log-odds", Early Resolution dominates in welfare because the interior

[1]For classic social-learning foundations and recent operations work on signaling via queue disclosure, see Bikhchandani et al. (1992); Banerjee (1992); Guo et al. (2023); Guo (2022).

[2]Short-run price rigidity is common due to menu costs, posted menus, platform pricing rules, or regulation. Allowing price choice changes information design rather than basic incentives; Section 6 shows the seller can keep actions informative by selecting prices from an implementability set.

[3]We abstract from ratings and textual reviews to highlight the informational role of actions. In many markets credible cross-platform reviews are sparse or delayed, and on-site observers cannot condition on them even if they exist. Endogenizing public outcome disclosure is analyzed in Section 6.2.

pause in Double Hump slows learning without improving decisions (Proposition 7).

Third, comparative statics reverse a common intuition from exogenous-state social learning. When private signals become more precise, buyers are more prone to herd at a fixed price, shrinking the experimentation region. As a result, the seller's reputational return to quality can *fall* even though individual information improves (Proposition 3). The model thus explains why "better recommendations" or stronger priors in the background may cool investment incentives when outsiders can only see actions.

Methodologically, we select a Markov equilibrium using a fixed tie-break at buyer indifference and a vanishing public tremble at the cascade boundaries. This resolves the well-known boundary indeterminacies, restores a clean Markov structure in finite and infinite horizons, and delivers existence and continuity of the equilibrium value function (Theorem 1). The value-iteration algorithm we use for figures proceeds on a "log-odds" grid and is summarized in Appendix C.

We also study two extensions that connect directly to institutional design. Allowing the seller to choose the (still posted) price creates, at each belief, an implementability set of prices that keep actions informative.[4] Patient sellers optimally remain in that set to avoid cascades and preserve learning; pointwise, they select the top of the set to maximize static revenue while maintaining informativeness (Propositions 8 and 9). Separately, making post-purchase outcomes publicly observable strengthens each observed purchase as a learning event and expands the investment region; equilibrium investment increases with outcome precision (Proposition 10 and Corollary 3). Both changes accelerate information aggregation and raise total surplus under the benchmark primitives.

Related Literature

Our paper sits at the intersection of three strands: (i) social learning with action-only observability and informational cascades; (ii) dynamic reputation with a long-lived seller under imperfect public information; and (iii) models where informativeness is endogenous—through costly information acquisition or through choices that shape the information content of observed actions. We contribute by endogenizing the "state" (quality) inside an action-observed social-learning environment, characterizing a Markov equilibrium with two cascade regions separated by an experimentation region, and showing an inverse-U relationship between reputation and investment, together with a comparative-statics reversal for private-signal precision.

The canonical foundations are Banerjee (1992) and Bikhchandani et al. (1992), where agents move sequentially, observe predecessors' *actions* (but not their private signals), and may rationally disregard private information once the action history is sufficiently informative. Early work clarified the conditions under which cascades arise and how quickly beliefs converge: Lee (1993) analyzes convergence, Vives (1993) studies learn-

[4] The intuition is close in spirit to "reputation traps" (e.g., Levine (2021)), but the environment differs crucially: here actions—not outcomes—drive belief updates, and price is exogenous in the benchmark. This shift changes both the partition of beliefs and the geometry of incentives.

ing speed, and Smith and Sørensen (2000) document "pathological" outcomes in observational learning that limit information aggregation even with rich signals. Laboratory evidence of cascades is provided by Anderson and Holt (1997) and their classroom formulation in Anderson and Holt (1996). Broader perspectives and applications are surveyed in Bikhchandani et al. (1998) and synthesized in Chamley (2004).

Several extensions examine the channels by which information flows. Ellison and Fudenberg (1995) study word-of-mouth transmission, while Cao and Hirshleifer (2000) incorporate conversational exchange alongside observational learning, and Bala and Goyal (2001) analyze networked observation and the tension between conformism and diversity. Moscarini et al. (1998) consider learning when the underlying state changes over time. Relative to this tradition, our environment retains the action-only observability of Banerjee (1992) and Bikhchandani et al. (1992) but departs by *endogenizing* the state: the seller chooses current quality, which alters how informative actions are for future buyers. This generates a middle region in which the seller invests precisely to keep actions informative, flanked by down-/up-cascade regions where actions become unresponsive and investment shuts down—an inverse-U pattern that does not appear when the state is exogenous.

Classic reputation models—Kreps and Wilson (1982) and Milgrom and Roberts (1982)—show how incomplete information about a long-lived player can discipline behavior when opponents observe payoffs or informative public signals; Mailath and Samuelson (2006) provide a comprehensive treatment. In our setting, however, only *actions* are publicly observed; outcomes remain private, and the seller's current choice determines future informativeness. This places us closer to work in which reputation evolves through noisy or incomplete records. Liu (2011) study optimal information acquisition and reputation dynamics; Liu and Skrzypacz (2014) analyze "limited records" and show how thin public histories can produce reputation bubbles—a force akin to our action-only observability. Dilmé (2019) examines reputation building when adjusting behavior is costly; we share the dynamic investment perspective but focus on how quality *endogenously* shapes the informativeness of future actions and thereby creates cascade/experimentation regions. Most directly, we connect to the "reputation traps" mechanism in Levine (2021): self-fulfilling pessimism can sustain low reputation.

In our model, a similar force appears at low beliefs because investment itself changes whether actions will be informative; we propose an equilibrium selection that preserves a tractable Markov structure while acknowledging this multiplicity.

A complementary literature studies how agents choose the information content that others learn. Burguet and Vives (2000) and Ali (2018) investigate social learning when information acquisition is costly; agents' responsiveness depends on endogenous attention. Bohren (2016) shows how misspecified beliefs can generate herding that differs from Bayesian predictions. Our paper maintains Bayesian rationality and fixed private-signal technology for buyers; instead, the *seller's* quality choice endogenizes the informativeness of future actions. This mapping from investment to public informativeness interacts with standard strategic-complements structure (Vives, 1990), delivering monotone best responses, cutoff rules, and the clean partition of the belief space that underpins our inverse-U result.

Four features of our contribution differentiate it within the above literatures. First, we endogenize the state in an action-only social-learning environment and show that the seller's investment is *inverse-U in reputation*: zero inside both cascades and positive in the experimentation region. Second, we provide an explicit dynamic program with a belief-based Bayes map and show how static cutoffs and dynamic incentives jointly pin down the cascade thresholds. Third, we identify a comparative-statics reversal: increasing private-signal precision can *reduce* the responsiveness of actions to information in equilibrium by shrinking the experimentation region where investment pays. This mechanism is distinct from behavioral departures (e.g., Bohren, 2016) or costly attention on the *buyer* side (e.g., Ali, 2018; Burguet and Vives, 2000). Fourth, we articulate an equilibrium selection that preserves Markov perfection in the face of self-fulfilling multiplicity at low beliefs (cf. Levine, 2021), avoiding the history dependence that arises from naive mixing at cascade boundaries in finite horizons.

We provide welfare comparisons, a value-iteration recipe in "log-odds", and two design extensions: price choice yields an implementability set that keeps actions informative (eliminating cascades for sufficiently patient sellers), and public outcome observability strengthens each purchase as a learning event and expands the investment region.

Our contribution is complementary to classic IO accounts where reputation is built through outcomes and/or price premia—e.g., Klein and Leffler, 1981, Shapiro, 1983, and the modern repeated-games literature (Hörner, 2002; Bar-Isaac and Tadelis, 2008; Board and Meyer-ter Vehn, 2013). We instead keep prices fixed in the benchmark and make only actions observable; reputation capital is accumulated by keeping behavior informative rather than by extracting premia. This difference yields the inverse-U investment profile and a comparative-statics reversal in signal precision at fixed prices, while our price-choice extension shows how moving within an implementability set recovers a no-cascade policy that resembles the classic IO logic. The outcome-disclosure extension maps to platform design choices that strengthen the informativeness of observed actions.

The remainder of the paper proceeds as follows. Section 2 introduces the environment, timing, and selection. Section 3 characterizes the buyer's decision regions in the static game. Section 4 formulates the dynamic program and proves existence under our selection. Section 5 develops the inverse-U policy, the Early-Resolution versus Double-Hump taxonomy, and the welfare comparison. Section 6 studies price choice and public outcome observability. All proofs appear in Appendix A; curvature conditions, numerics, the finite-horizon formulation are collected in Appendices B–D.

2 Model

2.1 Environment and primitives

Time is discrete, $t = 1, 2, \ldots$. A long-lived seller chooses the period-t *quality* $\theta_t \in \{0, 1\}$, where $\theta_t = 1$ ("high") costs $c > 0$ to the seller and $\theta_t = 0$ ("low") is costless. In each period a short-lived buyer arrives, observes public information described below, receives a private signal $s_t \in \{H, L\}$ about the current quality, and then chooses an action $a_t \in \{0, 1\}$, where $a_t = 1$ denotes purchase and $a_t = 0$ denotes no purchase. The good is sold at a fixed price $p > 0$.

The buyer's (flow) payoff is

$$u(a_t, \theta_t) = \begin{cases} v - p & \text{if } a_t = 1 \text{ and } \theta_t = 1, \\ -p & \text{if } a_t = 1 \text{ and } \theta_t = 0, \\ 0 & \text{if } a_t = 0, \end{cases}$$

with $v > p$ so that a buyer strictly prefers to purchase when she believes the good is high with sufficiently high probability. The seller's (flow) payoff is $p \cdot a_t - c\,\theta_t$. The seller discounts at factor $\delta \in (0, 1)$.

Signals are conditionally i.i.d. across buyers with precision $q \in (1/2, 1)$:

$$\Pr(s_t = H \mid \theta_t = 1) = q, \qquad \Pr(s_t = H \mid \theta_t = 0) = 1 - q.$$

Define the *signal likelihood ratio* $z \equiv \frac{q}{1-q} > 1$. We also use the *purchase indifference threshold* for a buyer with posterior probability π that $\theta_t = 1$:

$$\pi \geq \pi^\star \iff \text{buy}, \qquad \pi^\star \equiv \frac{p}{v}.$$

Equivalently, in odds form with $K \equiv \frac{\pi^\star}{1-\pi^\star} = \frac{p}{v-p}$, the buy rule is "posterior odds $\geq K$."

2.2 Information, observables, and public belief

Only *actions* are publicly observed; consumption outcomes are private and leave no public trace. Let h_t be the public history of actions up to $t - 1$. The public state is the *reputation* (belief) $\lambda_t \in [0, 1]$, a sufficient statistic for h_t that players use to forecast period-t quality under the seller's equilibrium strategy. We take an initial belief $\lambda_1 \in (0, 1)$ as given.

Timing and information within period t.

1. The public belief λ_t is observed by the seller and buyer t.
2. The seller chooses $\theta_t \in \{0, 1\}$ (privately). She pays c if $\theta_t = 1$.
3. Buyer t observes a private signal $s_t \in \{H, L\}$ about *current* quality and then chooses $a_t \in \{0, 1\}$.

4. The action a_t becomes public; payoffs realize; the public belief updates to $\lambda_{t+1} = \Phi(\lambda_t \mid a_t)$ via Bayes' rule (defined below).

Thus, outsiders only see $\{a_\tau\}_{\tau \le t}$ (or, equivalently, λ_{t+1}), never signals s_τ nor realized qualities θ_τ.

2.3 Strategies and equilibrium

A (pure) *Markov strategy* for the seller is a measurable function $\theta : [0,1] \to \{0,1\}$; we allow mixed strategies $\theta(\lambda) \in [0,1]$ interpreted as the probability of choosing high quality at belief λ. A (pure) *Markov strategy* for buyer t is a measurable function $a : [0,1] \times \{H, L\} \to \{0,1\}$ mapping (λ_t, s_t) to an action; again, mixing is allowed.

Definition 1. A Markov Perfect Equilibrium (MPE) consists of seller and buyer strategy profiles $(\theta(\cdot), a(\cdot,\cdot))$ and a belief-update rule $\Phi(\cdot \mid \cdot)$ such that: (i) given Φ and a, the seller's θ maximizes her discounted expected payoff at every λ; (ii) given Φ and θ, the buyer's a is myopically optimal at every (λ, s); (iii) Φ is obtained from Bayes' rule applied to actions and the equilibrium strategies whenever possible, and by a fixed tie-breaking convention otherwise.

At buyer indifference (posterior exactly $\pi^\star$) we fix a deterministic tie-break.

In addition, we adopt a vanishing public "tremble" in the buyers' signal distribution as a selection device: formally, we analyze equilibria of ε-perturbed environments and let $\varepsilon \downarrow 0$.[5] This selection guarantees a *Markov* structure even near cascade boundaries and rules out pathologies in finite-horizon backward induction where naive mixing can make optimal actions depend on lagged beliefs.

2.4 Static decision rules and cascade thresholds

Let $r \equiv \lambda/(1-\lambda)$ denote prior odds that $\theta_t = 1$.[6] After observing $s \in \{H, L\}$, posterior odds are

$$r(H) = r \cdot z, \qquad r(L) = \frac{r}{z}.$$

The buyer's optimal action is: buy after s iff $r(s) \ge K$ (equivalently, posterior probability $\ge \pi^\star$). Hence:

Up-cascade (buy regardless of s):	$r \ge Kz$,
Down-cascade (never buy):	$r \le K/z$,
Experimentation (signal-sensitive):	$K/z < r < Kz$,

[5]The fixed tie-break removes knife-edge cycling at indifference; the public tremble puts strictly positive probability on both actions in cascades. Taking the limit as tremble $\to 0$ restores the benchmark while preserving Markov structure.

[6]"log-odds" linearize Bayesian updates and simplify drift calculations. Several of our curvature arguments are stated in "log-odds" because concavity is preserved under translation by $\pm \log z$.

which, in probabilities, correspond to thresholds

$$\underline{\lambda} = \frac{K/z}{1+K/z}, \qquad \overline{\lambda} = \frac{Kz}{1+Kz}.$$

In the experimentation region the equilibrium buyer rule is "buy after H, do not buy after L." We use this partition repeatedly below and in Section 4.

2.5 Action-based belief updating

Given a seller strategy $\theta(\cdot)$ and the buyer's best response, the public belief updates from λ to $\lambda' = \Phi(\lambda \mid a)$ after observing the public action $a \in \{0,1\}$. Let

$$\psi_1(\lambda) \equiv \Pr(a = 1 \mid \theta = 1, \lambda), \qquad \psi_0(\lambda) \equiv \Pr(a = 1 \mid \theta = 0, \lambda),$$

which are determined by the region of λ:

$$(\psi_1, \psi_0) = \begin{cases} (1,1) & \text{if } \lambda \geq \overline{\lambda} \quad \text{(up-cascade)}, \\ (q,\, 1-q) & \text{if } \underline{\lambda} < \lambda < \overline{\lambda} \quad \text{(experimentation)}, \\ (0,0) & \text{if } \lambda \leq \underline{\lambda} \quad \text{(down-cascade)}. \end{cases}$$

When the seller mixes at λ with probability $\theta(\lambda)$, the *prior* odds of high quality are $r(\lambda) \equiv \theta(\lambda)/(1-\theta(\lambda))$. The *posterior* odds after observing a are

$$\mathcal{R}(\lambda \mid a) = \begin{cases} \dfrac{\theta(\lambda)}{1-\theta(\lambda)} \cdot \dfrac{\psi_1(\lambda)}{\psi_0(\lambda)}, & a = 1, \\ \dfrac{\theta(\lambda)}{1-\theta(\lambda)} \cdot \dfrac{1-\psi_1(\lambda)}{1-\psi_0(\lambda)}, & a = 0, \end{cases}$$

and the *next* public belief is $\lambda' = \Phi(\lambda \mid a) \equiv \frac{\mathcal{R}(\lambda|a)}{1+\mathcal{R}(\lambda|a)}$.

Remark. In the experimentation region, where the buyer's action coincides with her signal, $\psi_1(\lambda) = q$ and $\psi_0(\lambda) = 1-q$, so observing a purchase multiplies odds by z and observing no purchase divides odds by z. In the cascade regions, actions are uninformative and beliefs remain locally constant (up to the vanishing-tremble selection).

2.6 Seller's objective and dynamic program

Let $V(\lambda)$ be the seller's continuation value from public belief λ. Given the buyer's static rule and the Bayes map Φ, the seller's Bellman equation is

$$V(\lambda) = \max_{\theta \in [0,1]} \Big\{ p \cdot \gamma(\lambda, \theta) \; - \; c\,\theta \; + \; \delta \cdot \mathbb{E}\big[\, V\big(\Phi(\lambda \mid A)\big) \,\big|\, \lambda, \theta \big] \Big\},$$

where $\gamma(\lambda,\theta) \equiv \theta\,\psi_1(\lambda) + (1-\theta)\,\psi_0(\lambda)$ is the purchase probability at belief λ when quality is chosen according to θ, and the expectation is over the realized action $A \in \{0,1\}$ induced by the buyer's rule. Section 4 analyzes this problem and delivers the cascade–experimentation–cascade partition and the inverse-U investment result.

3 Static Game

Before turning to dynamics, we characterize a single period. The static analysis pins down the buyer's decision regions, the action likelihoods under each quality, and a myopic benchmark for the seller. These objects are the building blocks of the dynamic program: they determine when actions are informative and how beliefs update after a purchase or a non-purchase.

Fix a public belief $\lambda \in (0,1)$ about current quality. The seller chooses $\theta \in \{0,1\}$ (high quality $\theta = 1$ costs $c > 0$), a buyer receives a private signal $s \in \{H, L\}$ with precision $q \in (1/2, 1)$, and then chooses $a \in \{0,1\}$ (purchase if $a = 1$). The price is $p > 0$, the buyer's value if $\theta = 1$ is $v > p$, and outcomes are not publicly observed.

Throughout, write the signal likelihood ratio $z \equiv \frac{q}{1-q} > 1$, the buyer's indifference probability $\pi^\star \equiv \frac{p}{v}$, and the corresponding odds threshold

$$K \equiv \frac{\pi^\star}{1-\pi^\star} = \frac{p}{v-p}.$$

Let $r \equiv \frac{\lambda}{1-\lambda}$ denote public odds.

The first step is to characterize how a buyer maps her private signal and the public belief into a purchase decision at the posted price. This delivers the familiar three-region partition that underlies all subsequent dynamics.

Lemma 1. *Given prior odds r and signal $s \in \{H, L\}$, posterior odds are $r(H) = rz$ and $r(L) = r/z$. The buyer's optimal action is:*

$$a(s) = \begin{cases} 1 & \textit{iff } r(s) \geq K, \\ 0 & \textit{iff } r(s) < K. \end{cases}$$

Equivalently, let $\underline{r} \equiv K/z$ and $\overline{r} \equiv Kz$. Then:

$$\begin{aligned} &\textit{(Up-cascade)} && r \geq \overline{r} \Rightarrow a(H) = a(L) = 1, \\ &\textit{(Down-cascade)} && r \leq \underline{r} \Rightarrow a(H) = a(L) = 0, \\ &\textit{(Experimentation)} && \underline{r} < r < \overline{r} \Rightarrow a(H) = 1,\ a(L) = 0. \end{aligned}$$

The posted price defines a posterior cutoff. When public odds are above the upper threshold, even a weak signal suffices and everyone buys; below the lower threshold, even a strong signal is not enough and nobody buys. Only between these thresholds do actions load on private information, and only there can the seller's quality choice affect informativeness.

Definition 2. Let $\underline{\lambda} \equiv \frac{\underline{r}}{1+\underline{r}} = \frac{K/z}{1+K/z}$ and $\overline{\lambda} \equiv \frac{\overline{r}}{1+\overline{r}} = \frac{Kz}{1+Kz}$. Then the three regions in Lemma 1 correspond to $\lambda \leq \underline{\lambda}$ (down-cascade), $\underline{\lambda} < \lambda < \overline{\lambda}$ (experimentation), and $\lambda \geq \overline{\lambda}$ (up-cascade).

Given the decision regions, it is useful to summarize the purchase probabilities under each quality. This reduces the dynamic problem to two primitives: the informativeness of actions and the static gain from a sale.

Lemma 2. *Let* $\psi_1(\lambda) \equiv \Pr(a = 1 \mid \theta = 1, \lambda)$ *and* $\psi_0(\lambda) \equiv \Pr(a = 1 \mid \theta = 0, \lambda)$*. Then*

$$(\psi_1(\lambda), \psi_0(\lambda)) = \begin{cases} (1,1) & \text{if } \lambda \geq \overline{\lambda} \quad (\text{up-cascade}), \\ (q, 1-q) & \text{if } \underline{\lambda} < \lambda < \overline{\lambda} \quad (\text{experimentation}), \\ (0,0) & \text{if } \lambda \leq \underline{\lambda} \quad (\text{down-cascade}). \end{cases}$$

In the experimentation region, the purchase decision satisfies MLRP with respect to θ *and has likelihood ratio* $\frac{\psi_1}{\psi_0} = \frac{q}{1-q} = z$ *for* $a = 1$ *and* $\frac{1-\psi_1}{1-\psi_0} = \frac{1-q}{q} = 1/z$ *for* $a = 0$*.*

Inside the experimentation region, a purchase is a disguised signal realization: under high quality it occurs with probability q, under low quality with $1 - q$. At the cascade boundaries, actions cease to be informative because the buyer's decision stops reacting to the signal.

With action likelihoods in hand, the belief update becomes a simple multiplicative map in odds. We record it here for later use in the dynamic program.

Corollary 1. *In the experimentation region, observing* $a = 1$ *multiplies public odds by* z *and observing* $a = 0$ *divides them by* z*:*

$$r' = \begin{cases} rz & \text{if } a = 1, \\ r/z & \text{if } a = 0. \end{cases} \qquad \lambda' = \frac{r'}{1 + r'}.$$

In the cascade regions, actions are locally uninformative: $(\psi_1, \psi_0) = (1,1)$ *or* $(0,0)$ *and* λ' *equals* λ *up to the vanishing-tremble selection used in the dynamic analysis.*

A purchase moves "log-odds" up by $\log z$ and a non-purchase moves them down by $\log z$.[7] This symmetry is what lets us analyze drift and hitting times in the dynamic section.

As a benchmark, we compute the seller's one-period best response, holding fixed the informativeness of actions. This isolates the static force against which dynamic incentives must overcome.

Lemma 3. *Let* $\eta \equiv \frac{c}{p(2q-1)}$*. If* $\lambda \notin (\underline{\lambda}, \overline{\lambda})$*,* $\psi_1 = \psi_0$ *and the myopic gain from high quality is zero, so* $\theta^{myop}(\lambda) = 0$*. If* $\underline{\lambda} < \lambda < \overline{\lambda}$ *(experimentation), the myopic gain from raising* θ *is* $p\big(\psi_1 - \psi_0\big) = p(2q - 1)$*; hence*

$$\theta^{myop}(\lambda) = \begin{cases} 1 & \text{if } \eta < 1, \\ 0 & \text{if } \eta > 1, \end{cases} \qquad \text{and any mix if } \eta = 1.$$

[7] In cascades the update is frozen under our selection; with a small tremble, beliefs drift only at $O(\varepsilon)$, and the drift vanishes as $\varepsilon \to 0$.

Because actions are uninformative in cascades, static investment there is wasteful. Inside the experimentation region, the static gain rises in q and the posted price, but the static calculus ignores how today's informativeness changes tomorrow's demand.

Lemma 3 provides an upper benchmark for dynamic investment: the dynamic policy never invests outside the experimentation region and cannot profitably exceed the myopic incentive unless continuation values strictly increase with θ via the Bayes map.

4 Dynamic Infinite-Horizon Problem

We now embed the static building blocks in an infinite-horizon problem. The key state is the public belief, which moves only when actions are informative. The seller trades off current revenue and cost against the dynamic value of steering beliefs by keeping actions informative. We adopt a transparent selection at the cascade boundaries to rule out knife-edge multiplicities and to obtain a clean Markov equilibrium.

This section formulates the seller's dynamic program, establishes existence of a Markov Perfect Equilibrium (MPE) under our selection, and characterizes the equilibrium policy. We maintain the timing and observables from Section 2, the static best responses and thresholds from Section 3, and the tie-breaking convention at buyer indifference together with a vanishing public tremble in the buyers' signal distribution (letting $\varepsilon \downarrow 0$).

4.1 Bellman equation and Bayes-step notation

Let $V(\lambda)$ denote the seller's continuation value starting from public belief $\lambda \in [0,1]$. Write $z \equiv q/(1-q) > 1$, $K \equiv p/(v-p)$, and the static thresholds

$$\underline{\lambda} = \frac{K/z}{1+K/z}, \qquad \overline{\lambda} = \frac{Kz}{1+Kz},$$

as in Definition 2. In odds $r = \lambda/(1-\lambda)$, define the *Bayes steps*

$$r^+(\lambda) \equiv r\,z, \qquad r^-(\lambda) \equiv r/z, \qquad \lambda^+(\lambda) \equiv \frac{r^+(\lambda)}{1+r^+(\lambda)}, \qquad \lambda^-(\lambda) \equiv \frac{r^-(\lambda)}{1+r^-(\lambda)}.$$

In the experimentation region $(\underline{\lambda}, \overline{\lambda})$, a purchase multiplies odds by z and a non-purchase divides odds by z. In both cascade regions, actions are locally uninformative and, under the ε-selection, $\lambda^{\pm}(\lambda) = \lambda$ in the $\varepsilon \downarrow 0$ limit.

Let $\psi_1(\lambda) = \Pr(a=1 \mid \theta=1, \lambda)$ and $\psi_0(\lambda) = \Pr(a=1 \mid \theta=0, \lambda)$ as in Lemma 2. The probability of purchase given mixing $\theta \in [0,1]$ is

$$\gamma(\lambda,\theta) \;\equiv\; \theta\,\psi_1(\lambda) + (1-\theta)\,\psi_0(\lambda).$$

The seller's Bellman equation is

$$V(\lambda) \;=\; \max_{\theta\in[0,1]} \Big\{ p\,\gamma(\lambda,\theta) \;-\; c\,\theta \;+\; \delta\,\mathbb{E}\big[V(\lambda') \mid \lambda,\theta\big] \Big\}, \tag{1}$$

where (using the buyer's static rule) the law of λ' is

$$\lambda' = \begin{cases} \lambda^+(\lambda) & \text{w.p. } \gamma(\lambda, \theta), \\ \lambda^-(\lambda) & \text{w.p. } 1 - \gamma(\lambda, \theta). \end{cases}$$

4.2 Existence and selection

Boundary indifference can generate spurious history dependence. We therefore impose a fixed tie-break at buyer indifference and a vanishing public tremble at cascades. This yields a well-behaved Bellman operator and restores a clean Markov structure.

To rule out boundary cycling and preserve a Markov structure, we adopt a fixed tie-break at buyer indifference together with a vanishing public tremble.

Theorem 1. *Fix the buyer tie-breaking rule at indifference and consider the ε-perturbed environments in which the public action likelihoods are smoothed by a vanishing tremble. For each $\varepsilon > 0$, the Bellman operator induced by* (1) *is a contraction on* $(\mathcal{B}([0,1]), \|\cdot\|_\infty)$ *with modulus δ, admits a unique fixed point V_ε, and yields a Markov optimal policy $\theta^\star_\varepsilon(\cdot)$. Along any sequence $\varepsilon_n \downarrow 0$, there exists a subsequence for which $V_{\varepsilon_n} \to V$ uniformly and $\theta^\star_{\varepsilon_n}(\lambda) \to \theta^\star(\lambda)$ for almost every λ, where $(V, \theta^\star)$ constitute a Markov Perfect Equilibrium of the original environment.*

Proof. See Appendix A. □

The tremble keeps transition kernels strictly inside the simplex and the fixed tie-break removes cycling at the boundaries. Passing to the limit delivers the desired equilibrium without forcing outcomes that rely on arbitrary history-dependent rules.

Two general properties of the value function simplify later arguments: monotonicity in the public belief and continuity on the compact state space.

Lemma 4. *V is weakly increasing in λ. Moreover, V is continuous on* $(0, 1)$ *and right- (left-) continuous at $\underline{\lambda}$ ($\overline{\lambda}$).*

Proof. See Appendix A. □

Higher reputation both raises the chance of a sale and improves the expected continuation, so the value is monotone. Continuity follows from discounting and the continuity of the transition map.

4.3 A sufficient-statistic for investment incentives

Because (1) is *linear* in θ, the optimal policy is bang-bang (mixing only at indifference). The following decomposition is central.

To understand investment, we decompose the gain from raising quality into a static term and a continuation term that depends on the value gap between the two next-period beliefs.[8]

[8]This decomposition separates a static sale effect from the dynamic value of information—the value spread between the two next-period beliefs induced by a purchase vs. a non-purchase.

The equilibrium investment decision is pinned down by a single marginal-incentive term that decomposes into a static sale effect and a continuation-value difference.

Proposition 1. *Fix λ. The one-shot gain from raising quality from $\theta = 0$ to $\theta = 1$ equals*

$$\Delta(\lambda) \;=\; p\big(\psi_1(\lambda) - \psi_0(\lambda)\big) \;+\; \delta\Big(\mathbb{E}[V(\lambda') \mid \lambda, \theta{=}1] - \mathbb{E}[V(\lambda') \mid \lambda, \theta{=}0]\Big) \;-\; c.$$

Hence the optimal policy is

$$\theta^\star(\lambda) = \begin{cases} 1 & \text{if } \Delta(\lambda) > 0, \\ 0 & \text{if } \Delta(\lambda) < 0, \\ \text{any } \theta \in [0,1] & \text{if } \Delta(\lambda) = 0. \end{cases}$$

In the experimentation region $(\underline{\lambda}, \overline{\lambda})$, this simplifies to

$$\Delta(\lambda) \;=\; (2q-1)\left[\, p \;+\; \delta\left(V(\lambda^+) - V(\lambda^-)\right)\right] \;-\; c, \tag{2}$$

whereas in either cascade region $\psi_1 = \psi_0$ and $\lambda^+ = \lambda^- = \lambda$, so $\Delta(\lambda) = -c < 0$.

Proof. See Appendix A. □

Investment pays when the sale probability differential is large enough and when the belief gap is valuable enough. The latter is strongest where actions are most informative, which anticipates the inverse-U policy.[9]

4.4 Partition and inverse-U investment

Combining the marginal incentive with the buyer's decision rule yields a simple partition of the belief space into cascades and an interior experimentation region.

Theorem 2. *In any MPE, $\theta^\star(\lambda) = 0$ for $\lambda \in [0, \underline{\lambda}] \cup [\overline{\lambda}, 1]$. Moreover, $\theta^\star(\lambda) > 0$ for all λ in the nonempty set*

$$\mathcal{I} \;\equiv\; \Big\{\lambda \in (\underline{\lambda}, \overline{\lambda}) \;:\; (2q-1)\,[\, p + \delta\,(V(\lambda^+) - V(\lambda^-))\,] \;>\; c\Big\}.$$

Proof. See Appendix A. □

The seller never invests in cascades because actions there carry no information. All investment is therefore concentrated in the interior, where actions move beliefs.

The next figure illustrates the partition and the implied drift in a calibrated example. Investment is strictly positive only in the interior and the policy is inverse-U shaped. Where $\theta^*(\lambda)! >!1/2$, "log-odds" drift upward; outside, drift is zero by definition of cascades (Figure 1).

[9]Single-peakedness of the finite-difference gradient $V(\lambda^+) - V(\lambda^-)$ is the key. When that gradient is bimodal (e.g., for highly patient sellers with intermediate costs), the policy can exhibit two investment windows; see Section 5.

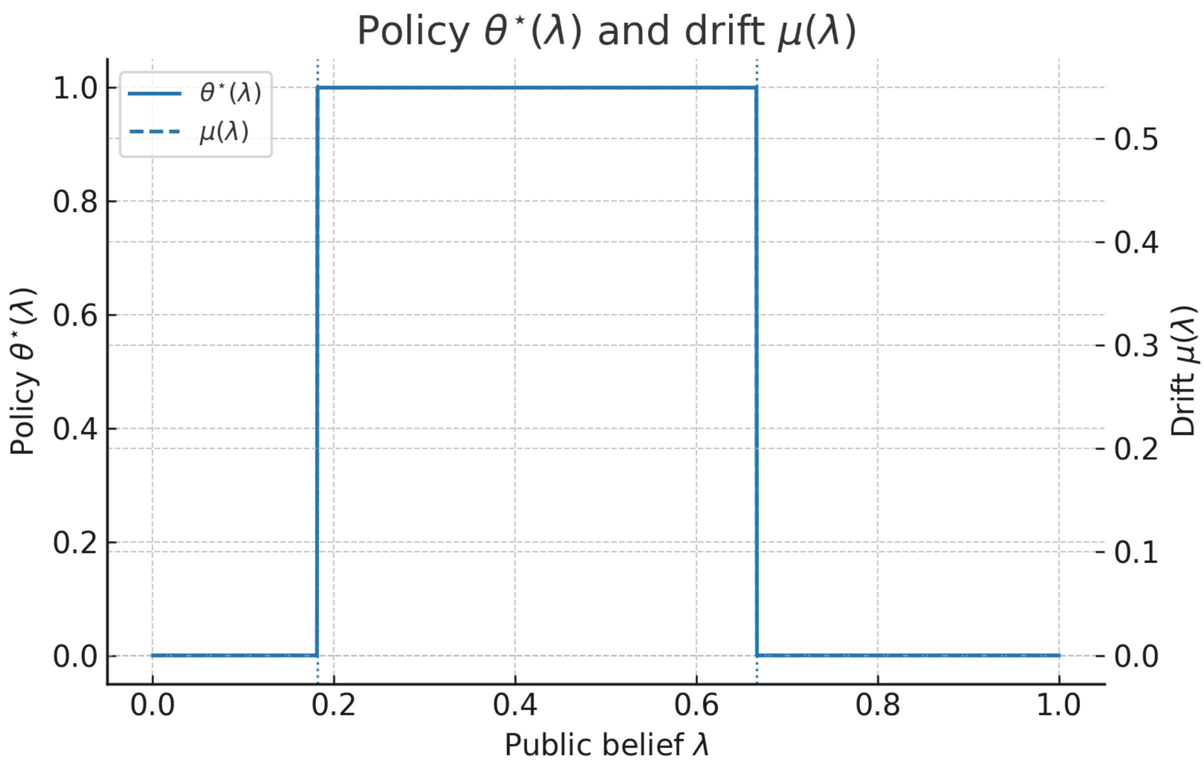

Figure 1: **Policy and drift.** The equilibrium policy $\theta^\star(\lambda)$ is zero in cascades $[0, \underline{\lambda}] \cup [\overline{\lambda}, 1]$ and positive only on $(\underline{\lambda}, \overline{\lambda})$, yielding an inverse-U. The implied drift in "log-odds", $\mu(\lambda) = (2q-1)\big(2\theta^\star(\lambda) - 1\big)\log z$, is positive where $\theta^\star(\lambda) > \frac{1}{2}$, zero at the cut, and negative otherwise. Parameters: $v = 1$, $p = 0.40$, $q = 0.75$ ($z = 3$), $\delta = 0.92$, $c = 0.22$.

The shape of the continuation term controls the shape of the policy. Under mild curvature conditions, the value difference across the two updates is unimodal, which forces a single connected investment window.

Proposition 2. *Suppose V is twice continuously differentiable in odds on $(\underline{\lambda}, \overline{\lambda})$ and V is strictly concave in $\log r$.*[10] *Then the map*

$$\lambda \;\mapsto\; \Delta(\lambda) = (2q-1)\Big[p + \delta\big(V(\lambda^+) - V(\lambda^-)\big)\Big] - c$$

is single-peaked on $(\underline{\lambda}, \overline{\lambda})$. In particular, there exist at most two cutoffs

$$\underline{\lambda} < \lambda_L \le \lambda_H < \overline{\lambda}$$

such that $\theta^\star(\lambda) = 0$ for $\lambda \le \lambda_L$ and $\lambda \ge \lambda_H$, and $\theta^\star(\lambda) = 1$ (or mixes at the boundary) for $\lambda \in (\lambda_L, \lambda_H)$. Hence the equilibrium investment policy is inverse-U in reputation: zero at the extremes and positive in the middle.

[10]This holds, for example, under value iteration limits when $\lambda^\pm(\lambda)$ induce a mean-preserving spread in $\log r$ and the stage payoff is affine in λ. See Appendix B for sufficient conditions.

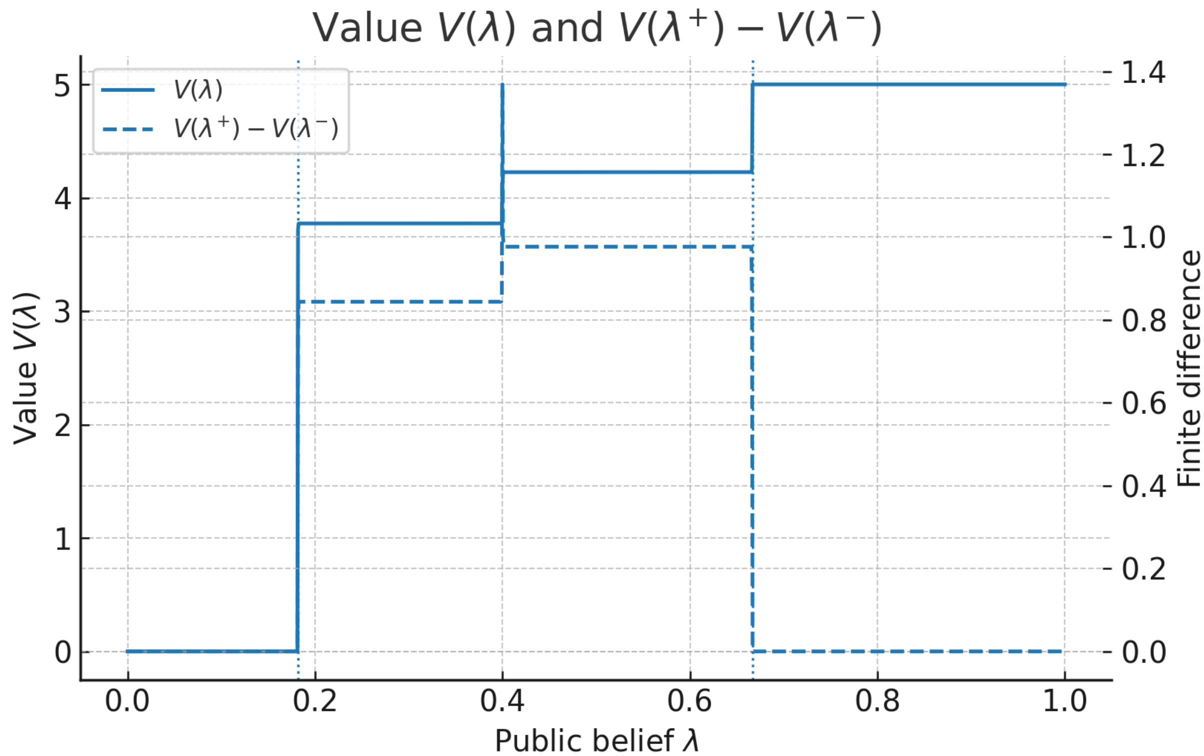

Figure 2: **Value and finite difference.** The value function $V(\lambda)$ (solid) is increasing and concave in "log-odds". The discrete gradient $V(\lambda^+) - V(\lambda^-)$ (dashed) is unimodal, implying that the marginal incentive $\Delta(\lambda) = (2q-1)[p + \delta(V(\lambda^+) - V(\lambda^-))] - c$ is single-peaked and the investment region is connected. Same parameters as Figure 1.

Proof. See Appendix A. □

Figure 2 illustrates the value function and its discrete gradient, showing the single-peaked marginal incentive that underlies the inverse-U policy. As beliefs approach either cascade, an incremental action has little effect on future beliefs. The marginal return to quality is therefore low near the edges and highest in the middle.

Figure 2 shows $V(\lambda)$ and the discrete gradient $V(\lambda^+)!-!V(\lambda^-)$ that enters the marginal incentive $\Delta(\lambda)$. The gradient is unimodal over $(\underline{\lambda}, \overline{\lambda})$, which delivers a connected investment region.

An immediate consequence of the partition is that the seller does not invest in cascades.

Corollary 2. *For all parameters,* $\theta^\star(\lambda) = 0$ *on* $[0, \underline{\lambda}] \cup [\overline{\lambda}, 1]$.

Once actions stop reacting to signals, quality no longer changes what future observers learn, so any cost is wasted.

4.5 Comparative statics and patterns

The framework delivers a sharp prediction about private-signal precision: it raises herding pressure at fixed prices and can therefore depress investment.

Proposition 3. *(i) The experimentation interval widens with signal precision:* $\partial\underline{\lambda}/\partial z < 0$ *and* $\partial\overline{\lambda}/\partial z > 0$.
(ii) Define $\Delta_V(\lambda) \equiv V(\lambda^+) - V(\lambda^-)$. *If* $\Delta_V(\lambda)$ *is decreasing in z on a subinterval of* $(\underline{\lambda}, \overline{\lambda})$, *then the equilibrium policy is (weakly) decreasing in z on that subinterval. In particular, for parameters with p small and* δ *moderate so that* $\Delta(\lambda)$ *in* (2) *is dominated by the continuation term, increasing z can reduce* $\theta^\star(\lambda)$*—a reversal relative to exogenous-state social learning.*

Proof. See Appendix A. □

More precise signals improve individual inferences but make the buyer's action more decisive; the interior region shrinks as cascades set in sooner, lowering the value of keeping actions informative.[11]

It is helpful to bound investment from above with a simple cost condition. This provides a quick test for parameterizations with no interior investment.

Proposition 4. *Let* $\eta \equiv c/(p(2q-1))$ *as in Lemma 3. For all* λ, $\theta^\star(\lambda) = 0$ *outside* $(\underline{\lambda}, \overline{\lambda})$, *and inside it*

$$\theta^\star(\lambda) = 0 \quad \textit{whenever} \quad \eta \geq 1 + \delta\,\frac{\Delta_V(\lambda)}{p}.$$

In particular, if $\eta \geq 1$ *and* δ *is small, the dynamic policy coincides with the myopic benchmark and the experimentation region is investment-free.*

Proof. See Appendix A. □

If the cost exceeds a myopic term plus a scaled continuation term, the marginal incentive cannot turn positive, so the inverse-U collapses to zero everywhere.

In finite horizons, naive mixing at the up-cascade boundary can make the buyer's optimal action depend on *lagged* public beliefs (since the current belief may embed the seller's previous mixing probability), breaking Markovianity. Our selection (fixed tie-break plus vanishing trembles) pins down the mapping from beliefs to actions at the boundary and restores a measurable Markov structure in the infinite-horizon limit (see Appendix D for a worked example).

The next section (§5) applies the characterization above to describe two equilibrium patterns—*Early Resolution* and *Double Hump*—and maps out the parameter regions that support each.

[11] The result holds at a fixed posted price. With flexible pricing, the seller can keep actions informative by adjusting price within the implementability set (Section 6.1), reversing the herding effect.

5 Equilibrium Patterns: Early Resolution vs. Double Hump

Having characterized optimal investment and belief movements, we classify the qualitative shapes that equilibria can take. The central object is the marginal value of investment, whose shape determines whether the seller invests in a single connected window (Early Resolution) or in two separated windows (Double Hump). We link these shapes to drift of beliefs, sample-path behavior, and welfare.

We now classify Markov equilibria by the geometry of the investment set and the induced drift of public beliefs inside the experimentation region $(\underline{\lambda}, \overline{\lambda})$.

5.1 Preliminaries: "log-odds" dynamics and drift

Let $r = \lambda/(1-\lambda)$ and $\ell \equiv \log r$. In the experimentation region, observing $a = 1$ moves "log-odds" by $+\log z$ and $a = 0$ by $-\log z$. Given a Markov policy $\theta(\lambda) \in [0,1]$, the purchase probability is $\gamma(\lambda, \theta) = (1-q) + \theta(2q-1)$, so the *one-step expected change* in "log-odds" equals

$$\mu(\lambda) \;\equiv\; \mathbb{E}[\ell' - \ell \mid \lambda] = \big(2\gamma(\lambda, \theta(\lambda)) - 1\big) \log z = (2q-1)\,(2\theta(\lambda) - 1)\, \log z. \tag{3}$$

We translate the policy into expected movement of beliefs. The sign of drift is pinned down by whether the policy exceeds one half.

Lemma 5. *Inside* $(\underline{\lambda}, \overline{\lambda})$, $\mu(\lambda) > 0$ *iff* $\theta(\lambda) > \frac{1}{2}$, $\mu(\lambda) = 0$ *iff* $\theta(\lambda) = \frac{1}{2}$, *and* $\mu(\lambda) < 0$ *iff* $\theta(\lambda) < \frac{1}{2}$.

When the seller invests with probability above one half, purchases are more likely and beliefs tend to rise; the reverse holds below one half.

By Theorem 2, $\theta^\star(\lambda) = 0$ outside $(\underline{\lambda}, \overline{\lambda})$, and beliefs then remain (locally) constant. Hence all movement occurs while λ stays in $(\underline{\lambda}, \overline{\lambda})$.

5.2 Investment set and pattern definitions

Recall $\Delta(\lambda)$ from Proposition 1 and define the (open) set of beliefs where investment is strictly profitable,

$$\mathcal{I} \;\equiv\; \{\lambda \in (\underline{\lambda}, \overline{\lambda}) :\; \Delta(\lambda) > 0\}.$$

Continuity of Δ implies $\mathcal{I}$ is a finite union of open intervals.

Definition 3. An MPE is of *Early-Resolution* type if $\mathcal{I}$ is a single interval $(\lambda_L, \lambda_H) \subset (\underline{\lambda}, \overline{\lambda})$ and $\theta^\star(\lambda) \geq \frac{1}{2}$ for all $\lambda \in (\lambda_L, \lambda_H)$. In this case $\mu(\lambda) \geq 0$ on (λ_L, λ_H) and the belief process almost surely hits $\overline{\lambda}$ in finite expected time when started in (λ_L, λ_H).[12]

[12]Symmetrically, one can define a downward Early-Resolution variant if $\theta^\star \leq \frac{1}{2}$ on the active interval, leading to absorption at $\underline{\lambda}$. The "upward" version is empirically more relevant for a seller who benefits from demand.

Definition 4. An MPE is of *Double-Hump* type if $\mathcal{I}$ is the union of two disjoint nonempty intervals $(\lambda_1, \lambda_2) \cup (\lambda_3, \lambda_4)$ with $\underline{\lambda} < \lambda_1 < \lambda_2 < \lambda_3 < \lambda_4 < \overline{\lambda}$, and

$$\theta^\star(\lambda) > \tfrac{1}{2} \text{ on } (\lambda_1, \lambda_2), \qquad \theta^\star(\lambda) < \tfrac{1}{2} \text{ on } (\lambda_2, \lambda_3), \qquad \theta^\star(\lambda) > \tfrac{1}{2} \text{ on } (\lambda_3, \lambda_4).$$

Thus the drift is upward in the lower hump, downward in the middle gap, and upward again in the upper hump, allowing beliefs to *re-enter* an investment region without ever leaving $(\underline{\lambda}, \overline{\lambda})$.

5.3 When do these patterns arise?

The shape of $\mathcal{I}$ is governed by $\Delta(\lambda) = (2q-1)\left[p + \delta\big(V(\lambda^+) - V(\lambda^-)\big)\right] - c$. Two forces operate: a *static* term $p(2q-1)$ and a *continuation* term $(2q-1)\delta\left[V(\lambda^+) - V(\lambda^-)\right]$. The static term is constant across λ, whereas the continuation term is shaped by the curvature of V in "log-odds".

If the marginal incentive is single-peaked and the policy exceeds one half on the investment set, beliefs drift upward and hit the optimistic boundary quickly.

Theorem 3. *Suppose V is strictly concave in $\log r$ on $(\underline{\lambda}, \overline{\lambda})$ and*

$$p(2q-1) - c + \delta(2q-1)\left(V(\lambda^+) - V(\lambda^-)\right)$$

is single-peaked in λ with a unique maximizer $\hat{\lambda}$ and strictly positive on a connected interval (λ_L, λ_H).

If, in addition, $\theta^\star(\lambda) \geq \frac{1}{2}$ on (λ_L, λ_H) (equivalently, $\Delta(\lambda) \geq c - (2q-1)p$ exceeds the myopic half-investment threshold), then the unique MPE is Early-Resolution in the sense of Definition 3. From any initial $\lambda_0 \in (\lambda_L, \lambda_H)$ the belief process hits $\overline{\lambda}$ in finite expected time and remains there thereafter.

Proof. See Appendix A. □

The "log-odds" process behaves like a biased random walk with an absorbing upper boundary.[13] Once absorbed, actions are uninformative and investment shuts down.

When the continuation term is bimodal, the investment set can split into two windows, creating a region in the interior where the seller temporarily stops investing.

Proposition 5. *Suppose $V(\cdot)$ is sufficiently nonconcave in $\log r$ so that the map $\lambda \mapsto V(\lambda^+) - V(\lambda^-)$ is* bimodal. *For costs c in an intermediate range and discount factors δ large enough that the continuation term dominates the static term, $\Delta(\lambda)$ crosses zero at four points*

$$\underline{\lambda} < \lambda_1 < \lambda_2 < \lambda_3 < \lambda_4 < \overline{\lambda},$$

with $\Delta > 0$ on (λ_1, λ_2) and (λ_3, λ_4). If, in addition, the myopic term is small enough that $\theta^\star(\lambda) < \frac{1}{2}$ on (λ_2, λ_3) (e.g., because c is high relative to $p(2q-1)$), there exists a Double-Hump MPE as in Definition 4.

[13] In "log-odds" the process is a biased random walk with absorbing boundaries. Standard hitting-time bounds imply finite expected time to the optimistic boundary when the drift is nonnegative on the investment set.

Proof. See Appendix A. □

Beliefs can move up in the first window, drift back in the gap, and then re-enter a second window. This produces two investment episodes on the way to absorption.[14]

In both patterns, once a cascade boundary is crossed, actions become uninformative and beliefs stop moving under our selection; re-entry into $(\underline{\lambda}, \overline{\lambda})$ is therefore impossible *from a cascade*. In the Double-Hump case, re-entry occurs entirely *within* the experimentation region: upward drift in the lower hump moves beliefs into the middle gap, where downward drift returns them toward the lower hump or moves them on toward the upper hump.

5.4 Implications for paths and welfare

We summarize how these policies translate into realized paths of beliefs.

Proposition 6. *In Early-Resolution MPE, for any initial $\lambda_0 \in (\underline{\lambda}, \overline{\lambda})$ the stopping time $\tau \equiv \inf\{t : \lambda_t \notin (\underline{\lambda}, \overline{\lambda})\}$ satisfies $\mathbb{E}[\tau] < \infty$, and $\lambda_t = \overline{\lambda}$ for all $t \geq \tau$. In Double-Hump MPE, on any sample path starting in (λ_1, λ_4) there are two disjoint time intervals on which $\theta^\star(\lambda_t) > 0$ separated by a no-investment spell when $\lambda_t \in [\lambda_2, \lambda_3]$.*

Proof. See Appendix A. □

Figure 3 shows short sample paths under the policy: in Early Resolution paths move monotonically to the optimistic boundary, whereas Double Hump exhibits two visible investment episodes separated by a pause. Early Resolution yields short interior spells and rapid absorption; Double Hump yields alternation between investment and no-investment before absorption.

Simulated paths in Figure 3 demonstrate Early Resolution: starting anywhere in the experimentation region, beliefs reach the up-cascade boundary in finite expected time and remain there. A Double-Hump parameterization (not shown) would exhibit two investment episodes separated by a no-investment spell within the experimentation region.

Finally, we compare welfare across patterns when the per-period social value of information is concave in "log-odds".

Proposition 7. *Expected buyer surplus and total surplus are (weakly) higher in Early-Resolution than in Double-Hump whenever the per-period social value of information is concave in $\log r$ and the middle gap (λ_2, λ_3) is nontrivial. The gap arises because the no-investment spell reduces the speed of information aggregation precisely when beliefs are most influential.*

Proof. See Appendix A. □

[14] Empirically, such "two-wave" behavior can arise when early adopters trigger a reputational push, followed by a lull as beliefs hover near an interior threshold, and a second push once fresh information revives informativeness.

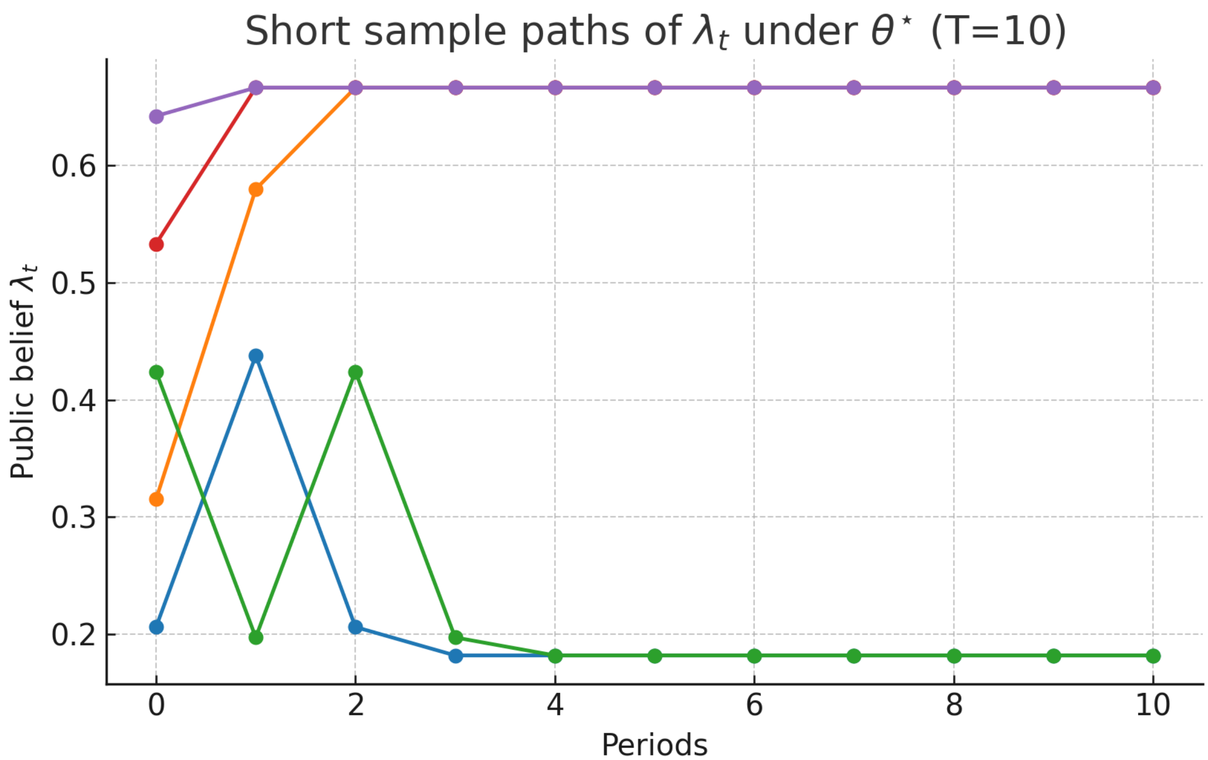

Figure 3: **Short sample paths of λ_t under $\theta^\star$.** From diverse initial beliefs inside $(\underline{\lambda}, \overline{\lambda})$, beliefs move in visible steps over $T{=}10$ periods and approach the up-cascade boundary $\overline{\lambda}$ (Early-Resolution pattern; Proposition 6). Same parameters as Figures 1–2.

Because Double Hump spends time in an interior no-investment spell, it slows information aggregation exactly where it is most valuable, producing a welfare loss relative to Early Resolution.[15]

6 Extensions: Price Choice and Outcome Observability

The benchmark imposes a fixed posted price and keeps outcomes private. This section shows how relaxing either assumption changes the information design of the environment. Allowing the seller to choose price creates an implementability set that can keep actions informative at every belief; revealing outcomes strengthens the information conveyed by each purchase. Both interventions alter investment incentives by changing the return to keeping actions informative.

[15]We evaluate welfare as buyer surplus net of seller costs (and revenue transfers cancel). Alternative surplus measures that weight speed of learning similarly deliver the same ranking when per-period value is concave in "log-odds".

We consider two modifications of the benchmark: (i) the seller may choose the period price p_t before the buyer moves; (ii) consumption outcomes become publicly observable (in addition to actions). In both cases we keep the timing of Section 2 and the Markov restriction.

6.1 Flexible price

Suppose the seller chooses $p_t \in (0, v)$ at the beginning of period t. Given p_t, the buyer's static purchase cutoff is

$$\pi^\star(p_t) = \frac{p_t}{v}, \qquad K(p_t) = \frac{\pi^\star(p_t)}{1 - \pi^\star(p_t)} = \frac{p_t}{v - p_t}.$$

For any public odds $r = \lambda/(1-\lambda)$, the experimentation (signal-sensitive) region at price p is the interval $K(p)/z < r < K(p)\,z$ (Lemma 1). The next lemma shows the seller can *always* select a price that induces experimentation at the current belief.

With flexible posted prices, the seller can select a price that keeps the buyer at the information-sensitive margin. We characterize the set of such prices at each belief.

Lemma 6. *Fix $\lambda \in (0,1)$ with odds $r = \lambda/(1-\lambda)$ and $z > 1$. The set of prices that induce experimentation is the nonempty open interval*[16]

$$\mathcal{P}(\lambda) \;=\; \Big(\, p_L(\lambda),\, p_H(\lambda) \,\Big), \qquad p_L(\lambda) = v\frac{r}{r^{+}z}, \qquad p_H(\lambda) = v\frac{rz}{1+rz}.$$

For any $p \in \mathcal{P}(\lambda)$, the buyer purchases after H and not after L, so the public belief updates by the $\pm \log z$ rule as in Corollary 1.

Proof. See Appendix A. □

Prices map to posterior odds thresholds. Keeping the threshold between the two signal-induced odds pins the buyer to an action that reacts to the signal, preserving informativeness.

With flexible prices the seller can (i) keep actions informative at *every* belief by choosing $p_t \in \mathcal{P}(\lambda_t)$, thereby precluding informational cascades under our selection, or (ii) deliberately move to pooling (buy-all or buy-none) by setting p below/above $\mathcal{P}(\lambda)$. Importantly, when $p \in \mathcal{P}(\lambda)$, the period purchase probability in the experimentation region is

$$\gamma(\lambda, \theta) = (1-q) + \theta(2q-1),$$

which is *independent of p*; price affects only the static revenue $p \cdot \gamma$, not belief transitions (provided p stays in $\mathcal{P}(\lambda)$).

[16] Within $\mathcal{P}(\lambda)$ demand in the experimentation region does not depend on the exact price, so pointwise the seller prefers the highest implementable price. Stepping outside $\mathcal{P}(\lambda)$ induces buy-all or buy-none pooling and shuts down learning.

A simple policy chooses a price from the implementability set and then applies the benchmark investment rule. This eliminates cascades while preserving the belief-update geometry.

Proposition 8. *Consider the dynamic problem in which the seller chooses (p_t, θ_t) each period. The policy that (i) sets any $p_t \in \mathcal{P}(\lambda_t)$ and (ii) sets θ_t by the benchmark rule (Proposition 1) with p replaced by p_t induces no cascades and preserves the Bayes $\pm \log z$ transitions at all beliefs. Under this policy, the marginal incentive is*

$$\Delta(\lambda; p) \;=\; (2q-1)\Big[p + \delta\big(V(\lambda^+) - V(\lambda^-)\big)\Big] - c,$$

and the value function solves the same Bellman equation (1) *with p replaced by a chosen $p(\lambda) \in \mathcal{P}(\lambda)$.*

Proof. See Appendix A. □

Once price holds the buyer at the margin, actions remain informative everywhere and belief movements follow the same $\pm \log z$ steps as in the benchmark interior.

Inside experimentation, demand $\gamma(\lambda, \theta)$ does not depend on p. Thus, conditional on keeping actions informative, the seller prefers the *largest* price in $\mathcal{P}(\lambda)$ pointwise:

$$p^\star(\lambda) \in \arg\max_{p \in \mathcal{P}(\lambda)} p \;=\; p_H(\lambda).$$

Choosing $p > p_H(\lambda)$ triggers buy-none pooling (no information); choosing $p < p_L(\lambda)$ triggers buy-all pooling.

Avoiding cascades is not always optimal when the seller is impatient. Flexible price introduces a trade-off between static revenue extraction and maintaining informativeness; patience pushes the solution toward the latter.

Proposition 9. *Let $p_H(\lambda)$ be as above. There exists $\bar{\delta} \in (0,1)$ such that for all $\delta \geq \bar{\delta}$ the seller's optimal policy keeps $p_t \in \mathcal{P}(\lambda_t)$ (hence avoids cascades) for all λ_t, while for δ small enough the seller optimally chooses a pooling price (buy-all near high beliefs or buy-none near low beliefs) on a nontrivial subset of beliefs.*

Proof. See Appendix A. □

Pooling prices give a one-time static gain but halt learning. When the seller values the future enough, the loss from shutting down belief movement dominates.

Flexible pricing gives the seller a second instrument to manage informativeness. With sufficiently patient sellers, the dynamic value of information dominates, and the seller sets $p_t \in \mathcal{P}(\lambda_t)$, eliminating cascades. The inverse-U shape of $\theta^\star(\lambda)$ *within* experimentation can persist (via the continuation term), but the zero-investment corners associated with cascades disappear in equilibrium when p is used to keep actions informative.

6.2 Publicly observed outcomes

Now suppose that if a purchase occurs ($a_t = 1$), a public outcome $y_t \in \{G, B\}$ is observed with precision $\rho \in (1/2, 1]$:[17]

$$\Pr(y_t = G \mid \theta_t = 1) = \rho, \qquad \Pr(y_t = G \mid \theta_t = 0) = 1 - \rho.$$

Actions remain public when $a_t = 0$, but no outcome is realized. The public update thus uses *both* the action and (when available) the outcome.

Let $z \equiv q/(1-q)$ and $w \equiv \rho/(1-\rho)$. In the experimentation region, the action equals the private signal, so if $a_t = 1$ then $s_t = H$ and we subsequently observe $y_t \in \{G, B\}$. The public odds update is

$$r' = \begin{cases} r \cdot z \cdot w & \text{if } a_t = 1 \text{ and } y_t = G, \\ r \cdot z/w & \text{if } a_t = 1 \text{ and } y_t = B, \\ r/z & \text{if } a_t = 0, \end{cases} \qquad \text{with } r = \frac{\lambda}{1-\lambda}.$$

If $\rho = 1$ (perfectly revealing outcome), purchases generate $r' \in \{0, \infty\}$ (absorbing beliefs) under our selection; if no purchase occurs, beliefs move by $1/z$ as before.

If post-purchase outcomes become public, the action still filters the private signal, but each observed purchase is now followed by an additional signal. We describe the resulting belief jumps.

Lemma 7. *For fixed price p, the buyer's static cutoffs and the cascade thresholds $(\underline{\lambda}, \overline{\lambda})$ are unchanged by public outcomes. Inside experimentation, the Bayes steps become $(+\log z \pm \log w)$ after a purchase and $-\log z$ after a non-purchase. Hence, for $\rho > 1/2$ the expected* magnitude *of belief movements weakly increases relative to the benchmark.*

Proof. See Appendix A. □

Outcomes multiply the action-based likelihood ratio. Purchases now move beliefs by a larger amount on average, while non-purchases move beliefs as before.

Let $V^{\text{out}}(\lambda)$ denote the continuation value with public outcomes. In the experimentation region, the purchase probability is still $\gamma(\theta) = (1-q) + \theta(2q-1)$; conditional on purchase, $y = G$ occurs with probability ρ under $\theta = 1$ and $1 - \rho$ under $\theta = 0$. A computation analogous to Proposition 1 yields:

We update the marginal-incentive decomposition to account for outcome realizations after a purchase. The result reveals how outcome precision shifts investment.

Proposition 10. *In the experimentation region, the one-step gain from raising quality is*

$$\Delta^{\text{out}}(\lambda) \;=\; (2q-1)\,p \;+\; \delta\Big\{(2q-1)\,V^{\text{out}}(\lambda^{-}) \;+\; (2q-1)(2\rho-1)\,\big(V^{\text{out}}(\lambda^{+G}) - V^{\text{out}}(\lambda^{+B})\big)\Big\} \;-\; c,$$

[17] Higher outcome precision strengthens learning only when purchases occur; non-purchases remain as in the benchmark. Thus the effect works by increasing the informational content of purchase events, not by changing the threshold for buying.

where λ^- is the no-purchase update, and $\lambda^{+G}, \lambda^{+B}$ are the post-purchase updates after $y \in \{G, B\}$. In particular, $\partial\Delta^{\text{out}}/\partial\rho \geq 0$ whenever $V^{\text{out}}(\lambda^{+G}) \geq V^{\text{out}}(\lambda^{+B})$.

Proof. See Appendix A. □

Outcomes raise the value spread between good and bad post-purchase states. Higher outcome precision therefore strengthens the continuation motive for investment.

The preceding expression delivers a clean monotonicity: higher outcome precision expands the investment region and accelerates learning.

Corollary 3. *For $\rho > 1/2$, the experimentation region's investment set under public outcomes weakly contains the benchmark's set; moreover, $\Delta^{\text{out}}(\lambda)$ is (weakly) increasing in ρ pointwise. If $\rho = 1$ and there is any vanishing purchase tremble (as in our selection), down-cascades are less stable in the sense that the expected hitting time to the upper boundary from any interior starting point weakly decreases.*

Proof. See Appendix A. □

Perfect outcomes make each purchase almost fully revealing, shortening the time beliefs spend in the interior and increasing investment incentives along the way.

6.3 Discussion

Public outcomes make each *observed* purchase disproportionately informative, strengthening dynamic incentives to invest in quality by raising the value difference between good and bad post-purchase states. The inverse-U shape can persist (via the concavity arguments in Appendix B), but the active window widens and shifts up with ρ. If prices are flexible *and* outcomes are public, the seller can keep actions informative at all beliefs and accelerate learning through larger post-purchase jumps; in this joint extension, cascades can be avoided in equilibrium for sufficiently patient sellers (Proposition 9), and investment is strictly higher throughout the interior for larger ρ.

Both extensions—flexible price (used to stay inside $\mathcal{P}(\lambda)$) and public outcomes—speed up information aggregation. Flexible price trades off revenue extraction against informativeness, whereas public outcomes relax the information constraint directly. For δ large, both interventions unambiguously raise total surplus by reducing time spent in uninformative regions.

7 Conclusion

This paper analyzes reputation formation when only actions are public and current quality is chosen by a long-lived seller. The central message is simple: reputational investment is strongest precisely where actions remain informative. In equilibrium, public beliefs fall into two cascade regions—one pessimistic, one optimistic—where actions no longer react to private information and the seller has no reason to invest. Between them lies an experimentation region where investment is positive because it keeps actions informative and shifts future demand. The resulting policy is inverse-U in reputation: zero at the extremes, positive in the middle.

Two dynamic patterns organize the behavior we obtain. In the first, beliefs move monotonically toward the optimistic cascade and investment winds down quickly (Early Resolution). In the second, investment appears in two distinct episodes separated by a pause within the interior region (Double Hump). Which pattern arises depends on primitives that govern the trade-off between current revenue, the cost of quality, and the value of accelerating information aggregation.

A practical virtue of the analysis is that equilibrium incentives can be expressed in terms of a single continuation-value difference across the two immediate belief updates generated by a purchase versus a non-purchase. This decomposition both explains why the policy is inverse-U and clarifies the comparative statics. In particular, improving the precision of private signals does not necessarily make observed actions more responsive. When precision pushes the system more quickly toward a cascade, the marginal reputational return to quality falls and investment can decline despite better information in the background.

Two extensions illustrate how institutional design reshapes these forces. Allowing the seller to choose price lets her keep actions informative at any belief by setting a price that places buyers at the information-sensitive margin. Patient sellers then prefer to avoid cascades altogether, because the dynamic value of information dominates the short-run gains from pooling prices. Making consumption outcomes public strengthens each observed purchase as a learning event and expands the region where investment pays. Either change speeds up learning; together, they can eliminate cascades in equilibrium for sufficiently patient sellers.

The framework yields concrete empirical and experimental implications in settings where outsiders see who subscribes or adopts but not their satisfaction. Investment effort should be highest at intermediate reputation and lowest at the extremes; increases in background signal precision can reduce action responsiveness and investment by hastening herding; flexible pricing used to keep actions informative should make cascades rare, while fixed prices make them common; credible public outcome disclosure should widen the investment window and accelerate convergence.

There are several natural directions for further work. On the market side, introducing a second long-lived seller would allow investment races and strategic substitution in informativeness; even basic existence and selection questions become interesting in that competition. Allowing a fraction of repeat buyers would add memory to the public state and soften the cascade boundaries. On the information side, platforms often choose recommendation intensity or targeting policies that effectively set signal precision; analyzing that choice as a mechanism-design problem could connect platform tools to welfare through the size of the experimentation region. Price dynamics deserve their own treatment, including temporary sales, commitment to price paths, and the interaction between pricing and disclosure. Finally, a continuous-time formulation provides a transparent benchmark with an HJB characterization; it would be useful to map more precisely when the diffusion limit preserves the inverse-U shape and when it overturns it.

The framework yields testable predictions in settings where outsiders observe uptake but not outcomes. Investment effort should be highest at intermediate reputation and lowest at the extremes. Exogenous increases in private-signal precision (for example, improved off-platform recommendations or tighter targeting) can reduce action responsiveness and investment by hastening herding. When prices are flexible and used to maintain informativeness, cascades should be rare, whereas fixed prices make them common. Credible public outcome disclosure (e.g., post-purchase ratings visible to outsiders) widens the investment window and speeds convergence.

The broader takeaway is that reputational incentives are governed as much by the *informativeness of actions* as by the quality choice itself. When institutions or platform choices make actions uninformative—by fixing prices at pooling levels or by suppressing outcome disclosure—investment stalls and beliefs calcify. When actions remain informative, the seller invests, learning proceeds, and the market moves toward the efficient side of the state space.

A Proofs of Main Results

Proofs from Section 3

Proof of Lemma 1. Let prior odds be $r = \lambda/(1-\lambda)$ and likelihood ratio $z = q/(1-q) > 1$. By Bayes, posterior odds after H and L are $r(H) = rz$ and $r(L) = r/z$. A buyer who purchases when her posterior probability π that $\theta = 1$ exceeds $\pi^\star = p/v$ is indifferent at odds $K = \pi^\star/(1-\pi^\star) = p/(v-p)$. Thus $a(H) = 1$ iff $rz \geq K$ and $a(L) = 1$ iff $r/z \geq K$. The three regions follow: if $r \geq Kz$ both inequalities hold (buy regardless of s); if $r \leq K/z$ neither holds (never buy); otherwise $a(H) = 1$ and $a(L) = 0$. □

Proof of Lemma 2. Fix λ. In the up-cascade, the buyer buys regardless of s, so $(\psi_1, \psi_0) = (1, 1)$. In the down-cascade, the buyer never buys, so $(\psi_1, \psi_0) = (0, 0)$. In the experimentation region, the buyer's action equals her signal: under $\theta = 1$ the signal is H with prob. q and under $\theta = 0$ with prob. $1-q$, hence $(\psi_1, \psi_0) = (q, 1-q)$. For $a = 1$, the likelihood ratio equals $\psi_1/\psi_0 = q/(1-q) = z$, and for $a = 0$ it equals $(1-\psi_1)/(1-\psi_0) = (1-q)/q = 1/z$, establishing MLRP. □

Proof of Corollary 1. In the experimentation region we have $\psi_1 = q$ and $\psi_0 = 1-q$. By Bayes' rule on odds, $r' = r \cdot z$ after $a = 1$ and $r' = r/z$ after $a = 0$, yielding the claimed update for $\lambda' = r'/(1+r')$. In cascades, $\psi_1 = \psi_0 \in \{0, 1\}$, so actions are uninformative and the public belief is unchanged (up to the vanishing-tremble selection). □

Proof of Lemma 3. The seller's one-period gain from $\theta = 1$ over $\theta = 0$ equals $p(\psi_1 - \psi_0) - c$. This is zero in cascades since $\psi_1 = \psi_0$, so the myopic best response is $\theta = 0$ there. In the experimentation region, $\psi_1 - \psi_0 = (q - (1-q)) = 2q - 1$, giving gain $p(2q-1) - c$. Hence invest myopically iff $c < p(2q-1)$, as stated. □

Proofs from Section 4

Proof of Theorem 1. For $\varepsilon > 0$, smooth the action likelihoods at cascade boundaries by setting $(\psi_1, \psi_0) = (1-\varepsilon, 1-\varepsilon)$ in the up-cascade and $(\varepsilon, \varepsilon)$ in the down-cascade. The Bellman operator

$$(T_\varepsilon V)(\lambda) = \max_{\theta \in [0,1]} \{\, p\,\gamma(\lambda, \theta) - c\theta + \delta\,\mathbb{E}[V(\lambda') \mid \lambda, \theta]\,\}$$

is a contraction on $(\mathcal{B}([0,1]), \|\cdot\|_\infty)$ with modulus δ by Blackwell's sufficient conditions. Thus T_ε has a unique fixed point V_ε, and there exists a measurable maximizer $\theta^\star_\varepsilon(\cdot)$ (Berge's maximum theorem and measurable selection). Any sequence $\varepsilon_n \downarrow 0$ admits a uniformly convergent subsequence $V_{\varepsilon_n} \to V$ by boundedness and equicontinuity on a compact state space; along the same subsequence, $\theta^\star_{\varepsilon_n}(\lambda)$ converges a.e. to $\theta^\star(\lambda) \in [0,1]$. As $\varepsilon \to 0$, the smoothed transition kernels converge pointwise to the equilibrium transition induced by the buyer's static rule with a fixed tie-break at indifference, so $(V, \theta^\star)$ is a Markov PBE of the original problem. □

Proof of Lemma 4. Fix $\lambda_1 \le \lambda_2$. For any control sequence, the purchase probability and the transition under λ_2 first-order stochastically dominate those under λ_1 (given the buyer's monotone static rule), and the period payoff and continuation are increasing in both the purchase indicator and next-period belief. Taking sup over controls preserves the inequality, hence $V(\lambda_1) \le V(\lambda_2)$. Continuity follows from the ε–approximation and bounded convergence. □

Proof of Proposition 1. Linearity of the Bellman equation in θ gives

$$\Delta(\lambda) = p(\psi_1 - \psi_0) + \delta\Big(\mathbb{E}[V(\lambda') \mid \theta = 1] - \mathbb{E}[V(\lambda') \mid \theta = 0]\Big) - c.$$

In cascades, $\psi_1 = \psi_0$ and $\lambda' = \lambda$, so $\Delta(\lambda) = -c$. In the experimentation region, $\psi_1 = q$ and $\psi_0 = 1 - q$, and

$$\begin{aligned}\mathbb{E}[V(\lambda') \mid \theta = 1] - \mathbb{E}[V(\lambda') \mid \theta = 0] &= (q - (1-q))\big(V(\lambda^+) - V(\lambda^-)\big)\\ &= (2q-1)\big(V(\lambda^+) - V(\lambda^-)\big),\end{aligned}$$

yielding the expression in the main text. □

Proof of Theorem 2. The zero-investment claim on $[0, \underline{\lambda}] \cup [\overline{\lambda}, 1]$ follows from $\Delta(\lambda) = -c$ by Proposition 1. On $(\underline{\lambda}, \overline{\lambda})$, investment is strictly profitable exactly when the marginal incentive exceeds zero, defining the nonempty set $\mathcal{I}$. □

Proof of Proposition 2. See Appendix B. In brief: concavity of V in "log-odds" implies unimodality of the finite-difference gradient $V(\lambda^+) - V(\lambda^-)$, and hence single-peakedness of the marginal incentive. □

Proof of Corollary 2. Immediate from $\Delta(\lambda) = -c$ in cascades and Proposition 1. □

Proof of Proposition 3. Part (i): $\underline{\lambda} = \frac{K/z}{1+K/z}$ and $\overline{\lambda} = \frac{Kz}{1+Kz}$ with $K > 0$ give $\partial\underline{\lambda}/\partial z < 0$ and $\partial\overline{\lambda}/\partial z > 0$. Part (ii): Differentiate the marginal incentive with respect to z inside the experimentation region, noting that z pushes $\lambda^{\pm}$ outward (toward cascades). If the induced decrease in $V(\lambda^+) - V(\lambda^-)$ dominates the direct static effect, then the marginal incentive falls and the policy weakly decreases. □

Proof of Proposition 4. From the marginal-incentive expression, $\Delta(\lambda) \leq (2q-1)\big[p + \delta\,\Delta_V(\lambda)\big] - c$, where $\Delta_V(\lambda) = V(\lambda^+) - V(\lambda^-)$. If $c \geq p(2q-1)\big[\,1 + \delta\,\Delta_V(\lambda)/p\,\big]$, then $\Delta(\lambda) \leq 0$, and the optimal policy is $\theta^\star(\lambda) = 0$ at that λ. The cascade claim uses Proposition 1. □

Proofs from Section 5

Proof of Lemma 5. Inside $(\underline{\lambda}, \overline{\lambda})$, $\Pr(a = 1 \mid \theta) = \gamma(\theta) = (1-q) + \theta(2q-1)$. The "log-odds" step is $\pm \log z$ with probabilities $\gamma(\theta)$ and $1 - \gamma(\theta)$, so $\mu(\lambda) = \mathbb{E}[\ell' - \ell] = (2\gamma(\theta) - 1)\log z = (2q-1)(2\theta-1)\log z$. The sign follows immediately. □

Proof of Theorem 3. Under the hypotheses, $\mathcal{I} = (\lambda_L, \lambda_H)$ is connected and $\theta^\star(\lambda) \geq 1/2$ there. Then the "log-odds" random walk has nonnegative drift on (λ_L, λ_H). Standard arguments for biased random walks with absorbing boundaries imply the hitting time of $\overline{\lambda}$ has finite expectation when started in (λ_L, λ_H). After absorption, actions are uninformative by definition of the up-cascade, so the process remains there. □

Proof of Proposition 5. Assume $\lambda \mapsto V(\lambda^+) - V(\lambda^-)$ is bimodal and δ large with c in an intermediate range. Then the marginal incentive crosses zero at four points, with positive values on the two outer intervals and negative in between. If in addition the static component is small enough that $\theta^\star(\lambda) < 1/2$ on the middle gap, drift is downward there and upward on both outer investment windows, yielding the stated pattern. □

Proof of Proposition 6. Early-Resolution: from any interior starting point the process has nonnegative drift and an absorbing upper boundary; expected hitting time is finite. Double-Hump: the drift alternates sign across the inactive middle interval; with positive probability a sample path enters the first investment window, then the no-investment interval, and then the second window before absorption, establishing two investment episodes. □

Proof of Proposition 7. Let social surplus per period be concave in "log-odds". In Early-Resolution, expected time spent in the interior is shorter and belief moves are biased toward the optimistic cascade. In Double-Hump, a no-investment spell in the interior delays learning without improving decision accuracy, lowering the expected integral of the concave surplus along the path. Formalization uses optional stopping for the "log-odds" submartingale/supermartingale. □

Proofs from Section 6

Proof of Lemma 6. Experimentation requires $r/z < K < rz$. Since $p = \frac{vK}{1+K}$ is strictly increasing in K, the corresponding price interval is $\left(\frac{v(r/z)}{1+r/z}, \frac{v(rz)}{1+rz}\right)$, which simplifies to $(p_L(\lambda), p_H(\lambda))$ with $p_L(\lambda) = vr/(r^+z)$ and $p_H(\lambda) = vrz/(1+rz)$. Nonemptiness uses $z > 1$. □

Proof of Proposition 8. Choosing any $p \in \mathcal{P}(\lambda)$ induces experimentation by Lemma 6, hence $\psi_1 = q$ and $\psi_0 = 1-q$ and the $\pm \log z$ Bayes steps are preserved at all beliefs. The Bellman equation remains the same with p replaced by the chosen $p(\lambda)$, and cascades are precluded under this policy since actions never pool unless $p \notin \mathcal{P}(\lambda)$. □

Proof of Proposition 9. Relative to $p^\star(\lambda) = p_H(\lambda)$ (which preserves experimentation), moving p to a pooling level changes the static flow by at most a bounded amount while eliminating belief movement, which reduces the continuation value by $\delta\,[V(\lambda^+) - V(\lambda^-)] > 0$ on compact interior subsets. For δ sufficiently large, this continuation loss dominates the static gain, so the optimal policy keeps $p \in \mathcal{P}(\lambda)$ at all beliefs. For small δ, the opposite holds on a nontrivial subset, and pooling prices are optimal there. □

Proof of Lemma 7. The buyer's static cutoff depends only on p and her private signal, which arrives before any public outcome. Hence the thresholds $\underline{\lambda}, \overline{\lambda}$ are unchanged. In the experimentation region, a purchase implies $s = H$; observing $y \in \{G, B\}$ multiplies the odds by w or $1/w$ respectively, giving the three update cases stated. □

Proof of Proposition 10. In the experimentation region, $\Pr(a = 1 \mid \theta = 1) - \Pr(a = 1 \mid \theta = 0) = 2q - 1$. Conditional on $a = 1$, we have $\Pr(y = G \mid \theta = 1) - \Pr(y = G \mid \theta = 0) = 2\rho - 1$. The difference in expected continuation values between $\theta = 1$ and $\theta = 0$ therefore equals

$$(2q-1)\,V^{\text{out}}(\lambda^-) \;+\; (2q-1)(2\rho-1)\left(V^{\text{out}}(\lambda^{+G}) - V^{\text{out}}(\lambda^{+B})\right),$$

and adding the static gain $p(2q-1)$ and subtracting c yields the formula. Monotonicity in ρ is immediate from $V^{\text{out}}(\lambda^{+G}) \geq V^{\text{out}}(\lambda^{+B})$. □

Proof of Corollary 3. Because $V^{\text{out}}(\lambda^{+G}) - V^{\text{out}}(\lambda^{+B}) \geq 0$, $\partial\Delta^{\text{out}}/\partial\rho \geq 0$. Hence the set where investment is profitable weakly expands with ρ. For $\rho = 1$, purchase outcomes are perfectly revealing and the magnitude of post-purchase jumps increases, reducing the expected time to reach the optimistic boundary from any interior starting point. □

B Curvature, Concavity, and the Inverse-U Policy

Let $\ell \equiv \log r = \log \frac{\lambda}{1-\lambda}$ denote "log-odds" and $\Delta\ell \equiv \log z > 0$. In the experimentation region $(\underline{\lambda}, \overline{\lambda})$, a purchase moves ℓ to $\ell + \Delta\ell$ and a non-purchase to $\ell - \Delta\ell$. For any fixed $\theta \in [0,1]$, recall $\gamma(\theta) = (1-q) + \theta(2q-1) \in (0,1)$ and note that *inside* $(\underline{\lambda}, \overline{\lambda})$ the transition weights $\gamma(\theta)$ and $1-\gamma(\theta)$ are *independent of* λ.

Lemma 8. *If $W(\ell)$ is concave on $(\underline{\ell}, \overline{\ell})$, then for any fixed $\theta \in [0,1]$ the map*

$$(T_\theta W)(\ell) \;\equiv\; p\,\gamma(\theta) - c\,\theta \;+\; \delta\big(\gamma(\theta)W(\ell+\Delta\ell) + (1-\gamma(\theta))W(\ell-\Delta\ell)\big)$$

is concave on $(\underline{\ell}, \overline{\ell})$.

Proof. $T_\theta W$ is an affine combination (with constant weights) of two translations of W plus an affine term; concavity is preserved under translations, convex combinations, and addition of affine functions. □

Lemma 9. *If $\{f_\alpha(\ell)\}_{\alpha\in\mathcal{A}}$ are concave, then $f(\ell) \equiv \sup_\alpha f_\alpha(\ell)$ is concave.*

Theorem 4. *Consider the Bellman operator on the experimentation region,*

$$(TV)(\ell) \equiv \sup_{\theta\in[0,1]} \big\{p\,\gamma(\theta) - c\theta + \delta\big(\gamma(\theta)V(\ell+\Delta\ell) + (1-\gamma(\theta))V(\ell-\Delta\ell)\big)\big\}.$$

If $V_0(\ell)$ is concave, then $V_n \equiv T^n V_0$ is concave for all n, hence the fixed point V is concave. Moreover, the one-sided limits connect continuously to the cascade boundaries, where $V(\underline{\lambda}) = 0$ and $V(\overline{\lambda}) = \frac{p}{1-\delta}$.

Proof. Concavity preservation follows by Lemma 8 (for each fixed θ) and Lemma 9 (sup over θ). Contraction implies convergence; boundary values follow from no-investment, uninformative actions, and the seller choosing $\theta = 0$ at cascades (down: zero sales forever; up: sale each period at zero cost). □

Lemma 10. *If $V(\ell)$ is concave, then the finite-difference gradient*

$$D(\ell) \;\equiv\; V(\ell+\Delta\ell) - V(\ell-\Delta\ell)$$

is single-peaked on $(\underline{\ell}, \overline{\ell})$ (unimodal).

Proof. For concave V, the discrete slope $V(\cdot+\Delta) - V(\cdot)$ is nonincreasing. Then $D(\ell) = \big[V(\ell+\Delta) - V(\ell)\big] + \big[V(\ell) - V(\ell-\Delta)\big]$ is the sum of a nonincreasing and a nondecreasing function, hence unimodal. □

Theorem 5 (Proof of Proposition 2)**.** *Inside $(\underline{\lambda}, \overline{\lambda})$,*

$$\Delta(\lambda) \;=\; (2q-1)\Big[p + \delta\big(V(\lambda^+) - V(\lambda^-)\big)\Big] - c \;=\; (2q-1)\big[p + \delta\, D(\ell)\big] - c.$$

By Lemma 10, $D(\ell)$ is single-peaked. Adding the constant p and multiplying by $(2q-1)\delta > 0$ preserves single-peakedness, so $\Delta(\lambda)$ is single-peaked. Hence the investment set $\{\lambda : \Delta(\lambda) > 0\}$ is a (possibly empty) interval, and $\theta^\star$ is inverse-U.

C Numerical Implementation and Replication

State grid. Work in "log-odds" $\ell \in [\underline{\ell}, \overline{\ell}]$ with $\underline{\ell} = \log \frac{\underline{\lambda}}{1-\underline{\lambda}}$ and $\overline{\ell} = \log \frac{\overline{\lambda}}{1-\overline{\lambda}}$. Let $\Delta\ell \equiv \log z$. Choose a grid aligned to the step: $\ell_k = \underline{\ell} + k\,\Delta\ell$, $k = 0, \dots, K$ with $\ell_K \approx \overline{\ell}$. (If $\overline{\ell}$ is not an integer multiple of $\Delta\ell$ from $\underline{\ell}$, cap at the largest $\ell_K \le \overline{\ell}$ and treat $\overline{\ell}$ as an absorbing state handled separately.)

Boundary conditions. Set $V(\underline{\lambda}) = 0$ and $V(\overline{\lambda}) = \frac{p}{1-\delta}$ (seller chooses $\theta = 0$; actions are uninformative but buyers purchase in the up-cascade).

Value iteration. Initialize $V^{(0)}(\ell_k)$ (e.g., zeros). Iterate for $n = 0, 1, 2, \dots$:

$$V^{(n+1)}(\ell_k) = \max_{\theta \in \{0,1\}} \Big\{ p\,\gamma(\theta) - c\theta + \delta\big(\gamma(\theta)\,V^{(n)}(\ell_{k+1}) + (1-\gamma(\theta))\,V^{(n)}(\ell_{k-1})\big) \Big\},$$

where $\gamma(1) = q$, $\gamma(0) = 1-q$. At $k = 0$ and $k = K$, use the boundary values above. Converge when $\|V^{(n+1)} - V^{(n)}\|_\infty < \text{tol}$.

Policy recovery and plots. Compute $\Delta(\ell_k) = (2q-1)\big[p + \delta\,(V(\ell_{k+1}) - V(\ell_{k-1}))\big] - c$ and set $\theta^\star(\ell_k) = 1\{\Delta(\ell_k) > 0\}$ (mix only at $\Delta = 0$). Plot $\theta^\star(\lambda)$, $V(\lambda)$, and $V(\lambda^+) - V(\lambda^-)$ by mapping $\ell \leftrightarrow \lambda$.

Notes. (i) No interpolation is needed if the grid is aligned; otherwise use linear interpolation for $V(\ell \pm \Delta\ell)$. (ii) Contraction guarantees convergence. (iii) For figures in Section 5, choose parameters to obtain single- vs. double-hump patterns (e.g., small p, large δ, and intermediate c for a double hump).

D Finite Horizon: Pathology and Selection

D.1 Backward induction and boundary mixing

Consider a T-period version. Let λ_t denote the public belief at period t. Backward induction determines $(\theta_t(\cdot), a_t(\cdot,\cdot))$ from $t = T$ to 1. At the up-cascade boundary $\overline{\lambda}$, buyers are indifferent ($r = Kz$) between buying after L and not buying after H. If the seller mixes in period $t-1$ at λ_{t-1} in a way that puts λ_t exactly at $\overline{\lambda}$, then buyer t's optimal action can depend on *the mixing probability used at* $t-1$, hence on λ_{t-1}, breaking Markovianity.

Example. With $T = 2$, suppose $\lambda_2 = \overline{\lambda}$ induced by mixing at $t = 1$. Buyer 2 is indifferent at $\overline{\lambda}$. If buyer 2 follows any history-dependent tie-break (e.g., buy when the seller mixed with probability above 1/2), then buyer 1's optimal period-1 choice depends on λ_1 *and* the intended tie-break, and buyer 2's action depends on λ_1 even conditional on λ_2, violating Markov perfection.

D.2 Selection that restores Markov structure

We adopt the same buyer tie-break and vanishing-tremble selection as in the infinite-horizon section; the next result shows that this delivers a unique Markov equilibrium in finite horizon and matches the infinite-horizon selection in the limits $\varepsilon \downarrow 0$ and $T \to \infty$.

Proposition 11. *Fix the buyer tie-breaking rule at indifference and perturb the period-t action likelihoods by a public tremble $\varepsilon > 0$ (so that at cascades $\psi_1 = \psi_0 = 1 - \varepsilon$ or ε). For each T and $\varepsilon > 0$, backward induction yields a Markov equilibrium $(\theta^\star_{t,\varepsilon}(\lambda), a^\star_{t,\varepsilon}(\lambda, s))$. Along any sequence $\varepsilon_n \downarrow 0$, the limit strategies are Markov and coincide with the infinite-horizon selection when $T \to \infty$.*

Proof. Fix a finite horizon $T \in \mathbb{N}$, discount factor $\delta \in (0,1)$, and a buyer tie-break at indifference (e.g., $a = 0$). For $\varepsilon > 0$ introduce a *public tremble* that makes actions strictly mixed in cascades: in the down-cascade set $(\psi_1^\varepsilon, \psi_0^\varepsilon) = (\varepsilon, \varepsilon)$ and in the up-cascade set $(\psi_1^\varepsilon, \psi_0^\varepsilon) = (1-\varepsilon, 1-\varepsilon)$; in the experimentation region keep $(\psi_1^\varepsilon, \psi_0^\varepsilon) = (q, 1-q)$. Let $z \equiv q/(1-q) > 1$. For a given public belief λ with odds $r = \lambda/(1-\lambda)$, the action-induced next belief is

$$\lambda' \;=\; \begin{cases} \lambda^+ \equiv \dfrac{rz}{1+rz}, & \text{if } a = 1, \\ \lambda^- \equiv \dfrac{r/z}{1+r/z}, & \text{if } a = 0, \end{cases}$$

with the understanding that at the cascade thresholds the buyer's action is determined by the tie-break and tremble.[18] Given a seller quality choice $\theta \in [0,1]$, the purchase probability under the ε-selection is

$$\gamma_\varepsilon(\lambda, \theta) \;=\; \psi_0^\varepsilon(\lambda) \;+\; \theta \left(\psi_1^\varepsilon(\lambda) - \psi_0^\varepsilon(\lambda)\right),$$

which equals $(1-q) + \theta(2q-1)$ in the experimentation region, $(1-\varepsilon)$ in the up-cascade, and ε in the down-cascade. The one-period flow payoff is $p\,\gamma_\varepsilon(\lambda,\theta) - c\,\theta$.

Step 1. Define the finite-horizon dynamic program backward:

$$V_{T+1}^\varepsilon(\lambda) \equiv 0, \qquad V_t^\varepsilon(\lambda) \;=\; \max_{\theta\in[0,1]} \left\{ p\,\gamma_\varepsilon(\lambda,\theta) - c\,\theta \;+\; \delta\,\mathbb{E}\left[V_{t+1}^\varepsilon(\lambda') \mid \lambda, \theta\right] \right\}, \quad t = T, \ldots, 1.$$

For fixed t and λ, the objective is *affine* in θ. Hence optimal $\theta_t^\varepsilon(\lambda)$ is bang–bang:

$$\theta_t^\varepsilon(\lambda) = 1\{\Delta_t^\varepsilon(\lambda) > 0\} \quad \text{with} \quad \Delta_t^\varepsilon(\lambda) = \left(\psi_1^\varepsilon - \psi_0^\varepsilon\right)\left[p + \delta\left(V_{t+1}^\varepsilon(\lambda^+) - V_{t+1}^\varepsilon(\lambda^-)\right)\right] - c.$$

By Berge's maximum theorem and the continuity of γ_ε and the transition map, there exists a measurable maximizer. We *fix the seller's tie-break* at indifference by setting $\theta_t^\varepsilon(\lambda) = 0$ whenever $\Delta_t^\varepsilon(\lambda) = 0$; this pins down a unique equilibrium policy profile $\{\theta_t^\varepsilon\}_{t=1}^T$. Given

[18] Thus, for any $\varepsilon > 0$, both $a = 0$ and $a = 1$ have strictly positive probability at any λ, which yields a continuous transition kernel in what follows.

the fixed buyer tie-break and the public tremble, the buyer's best response is single-valued at every state and the belief process $\{\lambda_t\}_{t=1}^{T+1}$ is a time-inhomogeneous Markov chain with Borel transition kernel (since $\lambda^{\pm}(\cdot)$ are continuous and γ_ε is continuous). Thus, for every T and $\varepsilon > 0$, there is a *unique* Markov Perfect Bayesian Equilibrium (value V_t^ε, policy θ_t^ε).

Step 2. We prove by backward induction that each V_t^ε is continuous on $[0, 1]$ and that $V_t^\varepsilon \to V_t$ uniformly as $\varepsilon \downarrow 0$, where V_t is the value under the knife-edge (no-tremble) selection with the same buyer tie-break. The base case $V_{T+1}^\varepsilon \equiv 0$ is trivial. Suppose $V_{t+1}^\varepsilon \to V_{t+1}$ uniformly and each V_{t+1}^ε is continuous. Then the map

$$\lambda \mapsto p\,\gamma_\varepsilon(\lambda, \theta) - c\,\theta + \delta\Big(\gamma_\varepsilon(\lambda, \theta)\, V_{t+1}^\varepsilon(\lambda^+) + \big(1 - \gamma_\varepsilon(\lambda, \theta)\big)\, V_{t+1}^\varepsilon(\lambda^-)\Big)$$

is jointly continuous in (λ, θ) for each ε, and converges uniformly (in (λ, θ)) to the corresponding no-tremble objective as $\varepsilon \downarrow 0$ because $\psi_i^\varepsilon(\lambda) \to \psi_i(\lambda)$ pointwise and the state space is compact. By Berge's maximum theorem, $V_t^\varepsilon \to V_t$ uniformly and V_t^ε is continuous. The policy selectors $\theta_t^\varepsilon(\cdot)$ converge pointwise except possibly on the closed indifference set $\{\Delta_t = 0\}$; our seller tie-break $\theta_t = 0$ identifies a unique limit policy θ_t.

Step 3. For each fixed $\varepsilon \geq 0$, the finite-horizon values satisfy the monotonicity $V_1^{\varepsilon,T} \uparrow V^\varepsilon$ as $T \to \infty$ (where V^ε solves the infinite-horizon Bellman equation with the same selection), because adding periods weakly increases attainable discounted payoffs under discounting. The corresponding policies admit convergent subsequences by compactness; the limit is Markov and attains the infinite-horizon value. Combining this with Step 2 (letting $\varepsilon \downarrow 0$) shows that the law of motion for beliefs and the value/policy under the finite-horizon ε-selection converge to those under the infinite-horizon knife-edge selection used in the main text: belief transitions are the $\pm \log z$ updates in the experimentation region and (by construction) the process is absorbed in cascades.

Putting Steps 1–3 together proves: (i) for every T and $\varepsilon > 0$ there is a unique Markov PBE under the tie-breaks; (ii) as $\varepsilon \downarrow 0$, $(V_t^\varepsilon, \theta_t^\varepsilon) \to (V_t, \theta_t)$ uniformly/pointwise; and (iii) as $T \to \infty$ the finite-horizon objects converge to the infinite-horizon selection used in the main analysis. □

References

Ali, S. N. (2018). Herding with costly information. *Journal of Economic Theory*, 175:713 – 729.

Anderson, L. R. and Holt, C. A. (1996). Classroom games: Information cascades. *Journal of Economic Perspectives*, 10(4):187–193.

Anderson, L. R. and Holt, C. A. (1997). Information cascades in the laboratory. *The American Economic Review*, 87(5):847–862.

Bala, V. and Goyal, S. (2001). Conformism and diversity under social learning. *Economic Theory*, 17(1):101–120.

Banerjee, A. V. (1992). A simple model of herd behavior. *The Quarterly Journal of Economics*, 107(3):797–817.

Bar-Isaac, H. and Tadelis, S. (2008). *Seller Reputation*. Foundations and trends in microeconomics. Now Publishers.

Bikhchandani, S., Hirshleifer, D., and Welch, I. (1992). A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of Political Economy*, 100(5):992–1026.

Bikhchandani, S., Hirshleifer, D., and Welch, I. (1998). Learning from the behavior of others: Conformity, fads, and informational cascades. *Journal of Economic Perspectives*, 12(3):151–170.

Board, S. and Meyer-ter Vehn, M. (2013). Reputation for quality. *Econometrica*, 81(6):2381–2462.

Bohren, J. A. (2016). Informational herding with model misspecification. *Journal of Economic Theory*, 163:222 – 247.

Burguet, R. and Vives, X. (2000). Social learning and costly information acquisition. *Economic Theory*, 15(1):185–205.

Cao, H. H. and Hirshleifer, D. (2000). Conversation , observational learning , and informational cascades.

Chamley, C. P. (2004). *Rational Herds: Economic Models of Social Learning*. Cambridge University Press.

Dilmé, F. (2019). Reputation building through costly adjustment. *Journal of Economic Theory*, 181:586 – 626.

Ellison, G. and Fudenberg, D. (1995). Word-of-mouth communication and social learning*. *The Quarterly Journal of Economics*, 110(1):93–125.

Guo, P. (2022). Optimal queue length information disclosure when service quality is uncertain. *Production and Operations Management*, 31(5):1912–1927.

Guo, P., Haviv, M., Luo, Z., and Wang, Y. (2023). Signaling service quality through queue disclosure. *Manufacturing & Service Operations Management*, 25(2):543–562.

Hörner, J. (2002). Reputation and competition. *American Economic Review*, 92(3):644–663.

Klein, B. and Leffler, K. B. (1981). The role of market forces in assuring contractual performance. *Journal of Political Economy*, 89(4):615–641.

Kreps, D. M. and Wilson, R. (1982). Reputation and imperfect information. *Journal of Economic Theory*, 27(2):253 – 279.

Lee, I. H. (1993). On the convergence of informational cascades. *Journal of Economic Theory*, 61(2):395 – 411.

Levine, D. K. (2021). The reputation trap. *Econometrica*, 89(6):2659–2678.

Liu, Q. (2011). Information acquisition and reputation dynamics. *The Review of Economic Studies*, 78(4):1400–1425.

Liu, Q. and Skrzypacz, A. (2014). Limited records and reputation bubbles. *Journal of Economic Theory*, 151:2 – 29.

Mailath, G. and Samuelson, L. (2006). *Repeated Games and Reputations: Long-Run Relationships*. Oxford University Press.

Milgrom, P. and Roberts, J. (1982). Predation, reputation, and entry deterrence. *Journal of Economic Theory*, 27(2):280 – 312.

Moscarini, G., Ottaviani, M., and Smith, L. (1998). Social learning in a changing world. *Economic Theory*, 11(3):657–665.

Shapiro, C. (1983). Premiums for high quality products as returns to reputations*. *The Quarterly Journal of Economics*, 98(4):659–679.

Smith, L. and Sørensen, P. (2000). Pathological outcomes of observational learning. *Econometrica*, 68(2):371–398.

Vives, X. (1990). Nash equilibrium with strategic complementarities. *Journal of Mathematical Economics*, 19(3):305 – 321.

Vives, X. (1993). How fast do rational agents learn? *The Review of Economic Studies*, 60(2):329–347.