Supplemental Appendix

Sequential Learning under Informational Ambiguity

Jaden Yang Chen*

September 28, 2025

Abstract

This online appendix provides the following materials: Section S1 provides a necessary and sufficient condition for complete learning with power-tail DGPs. Section S2 provides conditions that are close to necessary and sufficient for information cascades. Sections S3 and S4 explore extensions to settings with multiple actions and multiple states, respectively. Section S5 examines an extension in which individuals use mixed strategies to hedge against ambiguity. Appendix A contains omitted proofs.

JEL Classification: D81; D83; C72

*University of North Carolina at Chapel Hill, E-mail address: yangch@unc.edu

S1 Conditions for Complete Learning

This section presents a **necessary and sufficient condition** for complete learning within a class of DGPs that have power tails. For simplicity, I assume that all signals are i.i.d., and the true DGP is $\mathbb{F}$.

Definition S1. A DGP F has a *power tail* if there exist $c, \alpha > 0$ such that $F^0(x) \sim cx^\alpha$ as $x \to 0$. The exponent α is referred to as the power of F, denoted by $\mathcal{P}(F)$.

A DGP has a power tail if it can be approximated by a power function when x is close to 0. It is easy to see that a power-tail DGP is unbounded. The power provides an intuitive measure of informativeness: if F has a larger power, it means that its tails are thinner, so the DGP is less "informative" . This section focuses on the power-tail DGPs and imposes the following assumptions:

Assumption S1. *$\mathbb{F}$ has a power tail, and $\mathcal{F}_0$ contains only DGPs with power tails.*

Assumption S2. *$\mathcal{F}_0$ contains finitely many DGPs, and every DGP has a different power and is differentiable.*

Assumption S1 says that the true DGP has a power tail, and individuals only perceive DGPs with power tails. Assumption S2 is imposed for simplicity of analysis and can be relaxed. Theorem S1 provides a necessary and sufficient condition for complete learning under these two assumptions:

Theorem S1. *Under Assumptions S1 and S2, complete learning occurs* ***if and only if*** *$\mathcal{F}_0$ satisfies:*

(i) for all $F \in \mathcal{F}_0$, we have $\mathcal{P}(F) \geq \mathcal{P}(\mathbb{F})$, and

(ii) there exists some $F \in \mathcal{F}_0$ such that $\mathcal{P}(F) < \mathcal{P}(\mathbb{F}) + 1$.

Theorem S1 says that to establish complete learning, we need to impose restrictions from two directions. On one hand, all perceived DGPs cannot be too informative: their power must be greater than or equal to that of the true DGP. On the other hand, some perceived DGP must be adequately informative in the sense that its power does not exceed that of the true model by more than 1. Before explaining the intuition, let's examine what happens when the conditions in Theorem S1 are violated.

Corollary S1. *Under Assumptions S1 and S2, (i) if there exists some $F \in \mathcal{F}_0$ such that $\mathcal{P}(F) < \mathcal{P}(\mathbb{F})$, an incorrect herd occurs with strictly positive $\mathbb{P}^*$-probability; (ii) if for all $F \in \mathcal{F}_0$, $\mathcal{P}(F) \geq \mathcal{P}(\mathbb{F}) + 1$, actions do not converge $\mathbb{P}^*$-almost surely.*

First, when individuals perceive some highly informative DGP, an incorrect herd occurs with strictly positive probability. The mechanism has been explained in the paper. Second, when all DGPs considered by individuals are inadequately informative, actions will not converge. This stems from the fact that if individuals underestimate predecessors' informativeness, they are more likely to break away from a herd, so society may end up reaching no consensus. Corollary S1 implies that to achieve complete learning, we must exclude two sources of incomplete learning: incorrect herding and action nonconvergence. To prevent incorrect herding, $\mathcal{F}_0$ must not contain highly informative DGPs, which corresponds to Theorem S1 (i). To prevent action nonconvergence, $\mathcal{F}_0$ must not only contain DGPs that are too uninformative, which corresponds to Theorem S1 (ii).

S2 Conditions for Information Cascades

This section further provides two conditions that are close to necessary and sufficient for information cascades when signals are bounded. Proposition S1 provides a necessary and sufficient condition for a cascade to occur under some non-trivial prior. Proposition S2 provides a necessary and sufficient condition for the posterior monotonicity property, a concept closely related to information cascades. Both conditions use a modified version of the hazard ratio from Herrera and Hörner (2012), which I introduce below:

Definition S2. Let $h_F^\theta(x) = \frac{f^\theta(x)}{1-F^\theta(x)}$ and $H_F(x) = h_F^1(x)/h_F^0(x)$, where $H_F(x)$ is called the *hazard ratio* at x under F. For any set $\mathcal{F}_0$, define

$$H_{\mathcal{F}_0}(x) \equiv \sqrt{\sup_{F\in\mathcal{F}_0} H_F(x) \cdot \inf_{F\in\mathcal{F}_0} H_F(x)},$$

which is referred to as the **average hazard ratio** at x under $\mathcal{F}_0$.

For convenience, I impose the following assumption:

Assumption S3. *$\mathcal{F}_0$ contains finitely many DGPs. Every DGP in $\mathcal{F}_0$ is continuous and admits a full-support density function on $[1/\gamma, \gamma]$.*

The following proposition provides a necessary and sufficient condition for an information cascade to occur under some prior l_0 in the non-cascade region:

Proposition S1. *An information cascade occurs with strictly positive $\mathbb{P}^*$-probability for some prior $r_0 \in (1/\gamma, \gamma)$ **if and only if** $\mathcal{F}_0$ satisfies:*

$$H_{\mathcal{F}_0}(x) \geq \gamma \text{ or } H_{\mathcal{F}_0}(x) \leq 1/\gamma$$

for some $x \in (1/\gamma, \gamma)$.

Proof. Equivalently, we need to show that r_{i+1} enters the cascade set for some $r_i \in (1/\gamma, \gamma)$. By definition, when $a_i = 1$, we have

$$
\begin{aligned}
r_{i+1} &= \sqrt{\max_{F \in \mathcal{F}_0} \frac{1 - F^1(1/r_i)}{1 - F^0(1/r_i)} \times \min_{F \in \mathcal{F}_0} \frac{1 - F^1(1/r_i)}{1 - F^0(1/r_i)}} \times r_i \\
&= \sqrt{\max_{F \in \mathcal{F}_0} \frac{1 - F^1(1/r_i)}{1 - F^0(1/r_i)} \times \min_{F \in \mathcal{F}_0} \frac{1 - F^1(1/r_i)}{1 - F^0(1/r_i)}} \times \frac{f^0(1/r_i)}{f^1(1/r_i)} = \frac{1}{H_{\mathcal{F}_0}(1/r_i)}.
\end{aligned}
$$

When $a_i = 0$, we have

$$
\begin{aligned}
r_{i+1} &= \sqrt{\max_{F \in \mathcal{F}_0} \frac{F^1(1/r_i)}{F^0(1/r_i)} \times \min_{F \in \mathcal{F}_0} \frac{F^1(1/r_i)}{F^0(1/r_i)}} \times r_i \\
&= \sqrt{\max_{F \in \mathcal{F}_0} \frac{1 - F^0(r_i)}{1 - F^1(r_i)} \times \min_{F \in \mathcal{F}_0} \frac{1 - F^0(r_i)}{1 - F^1(r_i)}} \times \frac{f^1(r_i)}{f^0(r_i)} = H_{\mathcal{F}_0}(r_i),
\end{aligned}
$$

where the second equality employs the symmetry of signals.[1] The proposition then follows directly. □

In addition to this condition, I then provide a necessary and sufficient condition for a closely related concept—**posterior monotonicity**, which means that after any observation, the posterior is monotonically increasing in the prior. This concept is important in the cascade literature because it provides a sufficient condition for information cascades *not* to occur. Smith et al. (2021) showed that posterior monotonicity is equivalent to the log-concavity of the signal distribution. When the action space is binary, the condition is equivalent to the increasing hazard ratio (and decreasing failure ratio) in Herrera and Hörner (2012). Under ambiguity, we have a similar condition:

Proposition S2. *r_{i+1} is strictly increasing in r_i **if and only if** $H_{\mathcal{F}_0}(x)$ is a strictly increasing function in $(1/\gamma, \gamma)$.*

Proof. This follows directly from the proof of Proposition S1. □

Proposition S2 says that the **increasing average hazard ratio property** (IAHRP) is a necessary and sufficient condition for the posterior average likelihood ratio to be increasing in the prior average likelihood ratio. If the IAHRP holds, r_i is trapped in the non-cascade

[1] Without the symmetry, we need introduce another concept—the failure ratio—to characterize beliefs after $a_i = 0$.

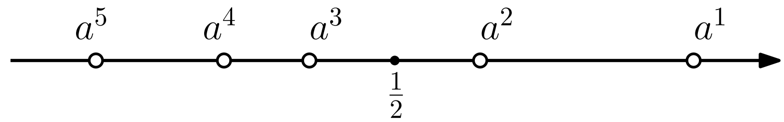

Figure 1: Linear Utility Functions

set, so an information cascade cannot occur. In other words, for an information cascade to occur, the IAHRP must be violated. This provides a necessary condition for information cascades.

S3 Multiple Actions

The paper's results extend to a multiple-action setting. Suppose that the action space is $A = \left\{a^1, ..., a^k\right\} \subset [0, 1]$. The utility function is:

$$u(a, \theta) = \begin{cases} a & \theta = 1 \\ 1 - a & \theta = 0 \end{cases}.$$

Suppose that: (i) individuals have MEU preferences and consider all DGPs as possible; and (ii) signals are i.i.d. according to $\mathbb{F}$, and $\mathbb{F}$ is continuous and has full-support on $(0, \infty)$. The set of **safe actions** is defined as:

$$A^s \equiv \{a \in A : \min\{a, 1 - a\} \geq \min\{a', 1 - a'\}, \forall a' \in A\},$$

which is the set of actions with the highest minimum payoff. It is easy to verify that A^s contains one or two actions, and when A^s contains two actions, these two actions must be symmetric with respect to 1/2. Figure 1 provides an example in which the safe actions, a^2 and a^3, are equally distanced from 1/2.

Proposition S3. *We have* $\lim_{t\to\infty} \mathbb{P}^*(a_t \in A^s) = 1$, *that is, society will only settle on* A^s *in the limit.*

This result follows from the fact that when ambiguity is sufficiently large, individuals ultimately hold highly ambiguous beliefs, which push them to choose only the safest actions to hedge against ambiguity. Furthermore, as shown in the proof, an information cascade of safe actions occurs almost surely—that is, individuals will choose the safe action(s) with probability 1 regardless of their private signals.

Remark S1. (Ambiguity attitude) A similar result also holds when individuals are ambiguity-loving. For example, under **max-max EU** preferences, society will settle on the actions with the highest maximum payoff:

$$A^h \equiv \{a \in A : \max\{a, 1-a\} \geq \max\{a', 1-a'\}, \forall a' \in A\}.$$

Geometrically, A^h consists of actions with the largest distance from $1/2$, and it also contains at most two actions. In Figure 1, we have $A^h = \{a^1, a^5\}$, so individuals will choose either a^1 or a^5 in the limit. We can see that ambiguity attitude affects which actions individuals choose in the limit—under ambiguity aversion, individuals settle on safe actions, whereas under ambiguity-loving preferences, they settle on risky actions.

S4 Multiple States

When there are multiple states, the equilibrium becomes more difficult to characterize, but the key insights still hold.[2] This section shows that in a simple case, an information cascade can still arise. Suppose that the state space $\Theta = \{0, 1, ..., K\}$, and the action space $A = \Theta$. Individuals share a flat prior, $\pi_0 = \left(\frac{1}{K+1}, ..., \frac{1}{K+1}\right)$. The utility function is

$$u(a, \theta) = \begin{cases} 1 & a = \theta \\ 0 & a \neq \theta \end{cases},$$

that is, individuals get a payoff of 1 if the action matches the true state and 0 if otherwise. Every individual has MEU preferences and updates beliefs using the full Bayesian rule. The true DGP, $\mathbb{G}_i$, satisfies:

$$\frac{d\mathbb{G}_i(s|\theta)}{d\mathbb{G}_i(s|\theta')} \in \left[\frac{1}{\gamma}, \gamma\right], \quad \forall s \in S,$$

I then consider a specific class of perceptions and show that large ambiguity can produce cascades.

Assumption S4. *The set of perceived DGP, $\mathcal{G}_0$, contains all G such that*

$$\frac{dG(s|\theta)}{dG(s|\theta')} \in \left[\frac{1}{R\gamma}, R\gamma\right], \quad \forall s \in S,$$

for some $R \geq 1$.

As R becomes larger, it reflects a higher degree of ambiguity. The following proposition shows that under sufficiently large ambiguity, an information cascade occurs almost surely.

[2] Arieli and Mueller-Frank (2021) extended the SSLM to a general state and action space. Their paper focused on correctly specified Bayesian agents, so the techniques cannot be applied here.

Proposition S4. *Under Assumption S4, there exists $R_0 < \infty$ such that an information cascade occurs $\mathbb{P}^*$-almost surely for all $R \geq R_0$.*

Proof. Suppose that $a_1 = \theta_1$. This reveals that

$$d\mathbb{G}_1 (s_1|\theta_1) / d\mathbb{G}_1 (s_1|\theta') \geq 1 \quad \forall \theta' \in \Theta. \tag{1}$$

For individual 2, she will follow the first individual if for all $\theta' \neq \theta_1$,

$$\sup_{\pi \in \Pi_2} \sum_{\theta \in \Theta} \frac{\pi(\theta)}{\pi(\theta')} \times \frac{d\mathbb{G}_2 (s_2|\theta)}{d\mathbb{G}_2 (s_2|\theta')} > \sup_{\pi \in \Pi_2} \sum_{\theta \in \Theta} \frac{\pi(\theta)}{\pi(\theta_1)} \times \frac{d\mathbb{G}_2 (s_2|\theta)}{d\mathbb{G}_2 (s_2|\theta_1)}. \tag{2}$$

Notice that

$$\begin{aligned}\text{L.H.S of (2)} &= \sup_{\pi \in \Pi_2} \left(1 + \frac{\pi(\theta_1)}{\pi(\theta')} \times \frac{d\mathbb{G}_2 (s_2|\theta_1)}{d\mathbb{G}_2 (s_2|\theta')} + \sum_{\theta \neq \theta_1, \theta'} \frac{\pi(\theta)}{\pi(\theta')} \times \frac{d\mathbb{G}_2 (s_2|\theta)}{d\mathbb{G}_2 (s_2|\theta')} \right) \\ &\geq 1 + R\gamma \times \frac{d\mathbb{G}_2 (s_2|\theta_1)}{d\mathbb{G}_2 (s_2|\theta')} + \sum_{\theta \neq \theta_1, \theta'} \frac{d\mathbb{G}_2 (s_2|\theta)}{d\mathbb{G}_2 (s_2|\theta')} \geq 1 + R + \frac{K-1}{\gamma},\end{aligned}$$

where: (i) the first inequality comes from that $\mathcal{G}_0$ consists of all DGPs with likelihood ratios between $1/R\gamma$ and $R\gamma$, so there exists some DGP $G_1 \in \mathcal{G}_0$ under which $\pi(\theta_1)/\pi(\theta') = R\gamma$ and $\pi(\theta)/\pi(\theta') = 1$,[3] and (ii) the second inequality comes from that $\frac{d\mathbb{G}_2(s_2|\theta)}{d\mathbb{G}_2(s_2|\theta')} \geq 1/\gamma$ for all θ, θ'. Furthermore, (1) implies $\pi(\theta_1) \geq \pi(\theta)$ for all $\theta \in \Theta$ and $\pi \in \Pi_2$, so

$$\text{R.H.S of (2)} \leq 1 + \sum_{\theta \neq \theta_1} \frac{d\mathbb{G}_2 (s_2|\theta)}{d\mathbb{G}_2 (s_2|\theta')} \leq 1 + K\gamma.$$

Thus, for sufficiently large R, the L.H.S. is greater than the R.H.S. for all possible s_2, so individual 2 will follow individual 1 immediately, and a cascade is triggered. □

It is worth noting that the type of ambiguity in Assumption S4 represents a very special case. An interesting direction for future research is to explore more general conditions under which a cascade occurs.

S5 Mixed-strategy Equilibrium

In this section, I explore an extension in which individuals can use mixed strategies. I show that when individuals have preferences for randomization, a mixed-strategy information

[3] Here is one example. Suppose $S = \{s_0, s_1, ..., s_K\}$ and G_1 satisfies: (i) $g_1(s_k|\theta) = c$ when $\theta \neq k$, and (ii) $g_1(s_k|\theta) = c \times R\gamma$ when $\theta = k$, where c is a normalization constant.

cascade occurs almost surely. During this cascade, individuals play the same mixed strategy regardless of their private information.

S5.1 Preferences for randomization

Before presenting the main results, I first distinguish between two different cases of mixed strategies:

- **Case 1**: Suppose that the mixing probabilities appear *outside* the minimum expected utility. Under strategy σ, individual i's utility is given by:

$$V_i(\sigma) = \sigma \min_{\pi \in \Pi_i} \mathbb{E}_\pi^i U(1) + (1-\sigma) \min_{\pi \in \Pi_i} \mathbb{E}_\pi^i U(0),$$

  where σ denotes the probability of taking action 1. In this case, individuals cannot hedge against ambiguity using mixed strategies. They will assign probability 1 to the action that maximizes their worst-case payoff, except in cases of indifference.

- **Case 2:** Suppose that the mixing probabilities appear *inside* the minimum expected utility. Then, under strategy σ, individual i's utility is:

$$V_i(\sigma) = \min_{\pi \in \Pi_i} \left[\sigma \mathbb{E}_\pi^i U(1) + (1-\sigma) \mathbb{E}_\pi^i U(0)\right].$$

  In this case, individuals exhibit **preferences for randomization** and can use mixed strategies to hedge against ambiguity.

The appropriate formulation of mixed strategies under ambiguity remains an ongoing discussion in the literature (e.g., Saito (2015) and Ke and Zhang (2020)). Notice that in the first case, individuals have no incentives to randomize, so the paper's analysis is without loss of generality. Therefore, the rest of this section focuses on the second case, assuming that individuals have preferences for randomization.

S5.2 Equilibrium strategy

I now characterize individuals' equilibrium strategy:

Proposition S5. *Suppose individuals have preferences for randomization. Then, a mixed-*

strategy equilibrium exists, characterized as follows:

$$\sigma_i^* (a_i = 1) = \begin{cases} 0 & \lambda_i \cdot \bar{l}_i < 1 \\ 1/2 & \lambda_i \cdot \underline{l_i} < 1 < \lambda_i \cdot \bar{l}_i \\ 1 & \lambda_i \cdot \underline{l_i} > 1 \end{cases}, \tag{3}$$

where $\sigma_i^* (a_i = 1)$ represents the probability that individual i chooses action 1 in the equilibrium. The indifference cases in (3) are determined by tie-breaking rules.

The proof can be found in Appendix A.3. To interpret the proposition, consider the following equivalent characterization: Let $\underline{\pi_i}$ and $\overline{\pi}_i$ denote individual i's minimum and maximum posterior beliefs about state 1. Then, equation (3) is equivalent to:

$$\sigma_i^* (a_i = 1) = \begin{cases} 0 & \overline{\pi}_i < 1/2 \\ 1/2 & \underline{\pi_i} < 1/2 < \overline{\pi}_i \\ 1 & \underline{\pi_i} > 1/2 \end{cases}.$$

Thus, individuals choose action θ with probability 1 if state θ is more likely to be the true state under **all** posteriors. Otherwise, they mix between the two actions with equal probability. In other words, individuals play a pure strategy only if all posteriors unambiguously support a state. When beliefs are sufficiently ambiguous, they mix actions to hedge against ambiguity.

S5.3 Information cascades with mixed strategy

For convenience, we impose a tie-breaking rule such that whenever individuals are indifferent, they randomize over actions. Based on the equilibrium strategy, we define the following **cascade sets**:

$$C_0 = \left\{ \left(\underline{l}_i, \bar{l}_i \right) : 0 \leq \underline{l}_i \leq \bar{l}_i < 1/\gamma \right\} \text{ and } C_1 = \left\{ \left(\underline{l}_i, \bar{l}_i \right) : \gamma < \underline{l}_i \leq \bar{l}_i \right\},$$

which represent the sets of public beliefs—characterized by $\bar{l}_i$ and $\underline{l}_i$—such that individuals will choose only action 0 or only action 1, respectively. Similarly, we define:

$$C_{1/2} = \left\{ \left(\underline{l}_i, \bar{l}_i \right) : \underline{l}_i \leq \frac{1}{\gamma}, \bar{l}_i \geq \gamma \right\},$$

which represents the set of public beliefs under which individuals randomize between the two actions. This is referred to as the **cascade set of the mixed strategy**. Once public beliefs enter this set, we say that a mixed-strategy information cascade occurs. We now state the following result:

Theorem S2. *Suppose that $\mathcal{F}_0$ consists of all DGPs with support in $[1/\gamma, \gamma]$. Then, a mixed-strategy information cascade occurs $\mathbb{P}^*$-almost surely.*

During a mixed-strategy information cascade, individuals randomize between the two actions regardless of their private signals. Since the mixing probability is 1/2, Theorem S2 implies that, in the limit, the fraction of individuals choosing each action is 1/2. In this case, even though actions continue to oscillate indefinitely, information ceases to aggregate after a finite number of periods.

The proof can be found in Appendix A.4. The intuition behind this result is as follows: In a social learning environment, individuals inevitably observe both actions. Ambiguity-averse individuals interpret actions inconsistent with state θ as negative signals about that state. As such signals accumulate, committing to a pure strategy becomes increasingly unattractive. Ultimately, society settles on a mixed strategy as a way to hedge against ambiguity.

A Omitted Proofs in the Supplementary Materials

A.1 Proof of Theorem S1

I first introduce the notion of local instability:

Definition S3. State 0 (or state 1) is ***locally unstable*** if there is some $r \in \mathbb{R}_{++}$ (or $R \in \mathbb{R}_{++}$) such that $\mathbb{P}^*_{r_0}(r_i > r \text{ for some } i) = 1$ (or $\mathbb{P}^*_{r_0}(r_i < R \text{ for some } i) = 1$) for all prior sets Π_0 with r_0 sufficiently small (or sufficiently large).

In other words, state θ is locally unstable if posteriors escape from a small neighborhood around δ_θ almost surely, where beliefs are described by the average likelihood ratio. The notion of local stability is defined in the appendix to the main paper, which says that beliefs remain in the neighborhood with strictly positive probability, and is omitted here. We begin with two lemmas:

Lemma S1. *Complete learning occurs if and only if $r_i \to 0$ with probability 1.*

Proof. First, complete learning requires that a herd of action 0 occurs eventually, which implies $r_i \to 0$ with probability 1 by Lemma 6 in the paper. Second, if $r_i \to 0$ with probability 1, a herd of action 0 occurs almost surely, also by Lemma 6, which implies complete learning. □

Lemma S2. *Complete learning occurs if 0 is locally stable and state 1 is locally unstable.*

Proof. Since state 1 is locally unstable, beliefs will enter $\{r_i < R\}$ infinitely many often. Whenever $r_i < R$, we can find a finite K such that K consecutive actions of 0 drive $r_i < r$. Since state 0 is locally stable, once $r_i < r$, we have $r_i \to 0$ with positive probability. Therefore, the probability of $r_i \to 0$ is greater than some positive constant across all histories, and complete learning occurs from Levy's 0-1 Law. □

Now, we characterize local stability under the assumptions of the theorem.

Proposition S6. *Under Assumptions S1 and S2, we have:*

(a) if for all $F \in \mathcal{F}_0$, $\mathcal{P}(F) \geq \mathcal{P}(\mathbb{F})$, state 1 is locally unstable;

(b) if there exists some $F \in \mathcal{F}_0$ such that $\mathcal{P}(F) < \mathcal{P}(\mathbb{F})$, state 1 is locally stable;

(c) if for all $F \in \mathcal{F}_0$, $\mathcal{P}(F) \geq \mathcal{P}(\mathbb{F}) + 1$, state 0 is locally unstable;

(d) if there exists some $F \in \mathcal{F}_0$ such that $\mathcal{P}(F) < \mathcal{P}(\mathbb{F}) + 1$, state 0 is locally stable.

Let $\overline{\alpha} := \mathcal{P}(\mathbb{F})$, $\alpha_{max} := \max_{F\in\mathcal{F}_0} \mathcal{P}(F)$ and $\alpha_{min} := \min_{F\in\mathcal{F}_0} \mathcal{P}(F)$. Let F_{max} and F_{min} be the DGPs that attain the maximum and minimum powers, respectively.

Proof. **Proof of Proposition S6 (a):** Given r_0, the probability of a herd of action 1 is:

$$\lim_{i\to\infty} \mathbb{P}^*_{r_0}(a_1 = a_2 = ...a_i = 1) = \prod_{i=1}^{\infty} \mathbb{P}^*_{r_0}(a_i = 1|h_i) = \prod_{i=1}^{\infty}\left[1 - \mathbb{F}^0\left(\frac{1}{r_i}\right)\right],$$

where r_i is the average likelihood ratio after $h_i = (1, 1, ..., 1)$. The probability is zero if and only if $\sum \mathbb{F}^0\left(\frac{1}{r_i}\right) = \infty$, or equivalently, $\sum \frac{1}{r_i^{\overline{\alpha}}} = \infty$. The sequence $\{r_i\}$ evolves according to:

$$r_{i+1} = r_i \times \sqrt{\max_{F\in\mathcal{F}_0} \frac{1 - F^1(1/r_i)}{1 - F^0(1/r_i)} \times \min_{F\in\mathcal{F}_0} \frac{1 - F^1(1/r_i)}{1 - F^0(1/r_i)}}.$$

When r_0 is sufficiently large, $\frac{1-F^1(1/r_i)}{1-F^0(1/r_i)} \sim 1 + F^0(1/r_i)$ for all i, so its maximum is obtained at F_{min} and it minimum is obtained at F_{max}. Therefore, when r_0 is sufficiently large,

$$r_{i+1} = r_i \times \sqrt{\frac{1 - F^1_{min}(1/r_i)}{1 - F^0_{min}(1/r_i)} \times \frac{1 - F^1_{max}(1/r_i)}{1 - F^0_{max}(1/r_i)}} \leq r_i \times \frac{1 - F^1_{min}(1/r_i)}{1 - F^0_{min}(1/r_i)}.$$

By the definition of F_{min}, we have $\frac{1-F^1_{min}(1/r_i)}{1-F^0_{min}(1/r_i)} \sim 1 + C_{min} \times \frac{1}{r_i^{\alpha_{min}}}$, for some constant $C_{min} > 0$. Suppose that for all $F \in \mathcal{F}_0$, we have $\mathcal{P}(F) \geq \mathcal{P}(\mathbb{F})$, that is, $\alpha_{min} \geq \overline{\alpha}$. Then,

$$\begin{aligned}
\lim_{r\to\infty} \frac{\frac{1-F^1_{min}(1/r)}{1-F^0_{min}(1/r)} - 1}{\left(1 + \frac{2\overline{\alpha}C_{min}}{r^{\overline{\alpha}}}\right)^{1/\overline{\alpha}} - 1} &= \lim_{r\to\infty} \frac{\frac{1-F^1_{min}(1/r)}{1-F^0_{min}(1/r)} - 1}{\frac{2\overline{\alpha}C_{min}}{r^{\overline{\alpha}}}} \times \frac{\frac{2\overline{\alpha}C_{min}}{r^{\overline{\alpha}}}}{\left(1 + \frac{2\overline{\alpha}C_{min}}{r^{\overline{\alpha}}}\right)^{1/\overline{\alpha}} - 1} \\
&= \lim_{r\to\infty} \frac{C_{min} \times \frac{1}{r^{\alpha_{min}}}}{\frac{2\overline{\alpha}C_{min}}{r^{\overline{\alpha}}}} \times \overline{\alpha} \\
&= \frac{1}{2} \times \lim_{r\to\infty} \frac{1}{r^{\alpha_{min}-\overline{\alpha}}} = \begin{cases} 0 & \alpha_{min} > \overline{\alpha} \\ \frac{1}{2} & \alpha_{min} = \overline{\alpha} \end{cases} < 1,
\end{aligned}$$

so $\frac{1-F^1_{min}(1/r_i)}{1-F^0_{min}(1/r_i)} < \left(1 + \frac{2\overline{\alpha}C_{min}}{r_i^{\overline{\alpha}}}\right)^{1/\overline{\alpha}}$. Therefore, for all $i \geq 0$,

$$\begin{aligned}
r_{i+1} &< \left(1 + \frac{2\overline{\alpha}C_{min}}{r_i^{\overline{\alpha}}}\right)^{1/\overline{\alpha}} \times r_i = \left(r_i^{\overline{\alpha}} + 2\overline{\alpha}C_{min}\right)^{1/\overline{\alpha}} \\
r_{i+2} &< \left(r_{i+1}^{\overline{\alpha}} + 2\overline{\alpha}C_{min}\right)^{1/\overline{\alpha}} < \left(r_i^{\overline{\alpha}} + 2\overline{\alpha}C_{min} \times 2\right)^{1/\overline{\alpha}} \\
&\dots \\
r_{i+t} &< \left(r_i^{\overline{\alpha}} + 2\overline{\alpha}C_{min} \times t\right)^{1/\overline{\alpha}}.
\end{aligned}$$

As a consequence, when r_0 is sufficiently large,

$$\sum_{i=1}^{\infty} \frac{1}{r_i^{\overline{\alpha}}} > \sum_{i=1}^{\infty} \frac{1}{r_0^{\overline{\alpha}} + 2\overline{\alpha} C_{min} \times i} = \infty,$$

so a herd of action 1 occurs with probability 0. This property holds for all $r_0 \in \mathbb{R}_{++}$, so state 1 is unstable.

Proof of Proposition S6 (b)

To show that state 1 is locally stable, we need to show that the probability of an action-1 herd is greater than some $\varepsilon > 0$ when r_0 is large. Recall that

$$\mathbb{P}_{r_0}^{*}(H_1) = \lim_{i \to \infty} \mathbb{P}_{r_0}^{*}(a_1 = a_2 = ...a_i = 1) = \prod_{i=1}^{\infty} \left[1 - \mathbb{F}^0\left(\frac{1}{r_i}\right)\right].$$

In order to establish local stability, we need to find a *uniform* lower bound for the probability on the right-hand side for all large r_0. Suppose that $\mathbb{F}^0(x) \sim \overline{C} \times x^{\overline{\alpha}}$ for some constant $\overline{C} > 0$. On one hand, we can find a sufficiently large R such that whenever $r_0 \geq R$, we have $\frac{\mathbb{F}^0(1/r_i)}{\overline{C} \times (1/r_i)^{\overline{\alpha}}} \in [1 - \varepsilon_1, 1 + \varepsilon_1]$ for some $\varepsilon_1 > 0$, so

$$\mathbb{P}_{r_0}^{*}(H_1) = \prod_{i=1}^{\infty} \left[1 - F^0\left(\frac{1}{r_i}\right)\right] \geq \prod_{i=1}^{\infty} \left[1 - (1 + \varepsilon_1) \times \overline{C} \times \frac{1}{r_i^{\overline{\alpha}}}\right]. \tag{4}$$

We also want R to be sufficiently large such that the infinite product on the right-hand side is strictly positive. On the other hand, recall that

$$r_{i+1} = r_i \times \sqrt{\frac{1 - F_{min}^1(1/r_i)}{1 - F_{min}^0(1/r_i)} \times \frac{1 - F_{max}^1(1/r_i)}{1 - F_{max}^0(1/r_i)}}.$$

Define $\beta = (1-\varepsilon)\frac{C_{min}\times\alpha_{min}}{2}$ for some small $\varepsilon > 0$, then we have

$$\begin{aligned}
\lim_{r\to\infty}\frac{\sqrt{\frac{1-F^1_{min}(1/r_i)}{1-F^0_{min}(1/r_i)}\times\frac{1-F^1_{max}(1/r_i)}{1-F^0_{max}(1/r_i)}}-1}{\left(1+\frac{\beta}{r^{\alpha_{min}}}\right)^{1/\alpha_{min}}-1} &= \lim_{r\to\infty}\frac{\sqrt{\frac{1-F^1_{min}(1/r_i)}{1-F^0_{min}(1/r_i)}\times\frac{1-F^1_{max}(1/r_i)}{1-F^0_{max}(1/r_i)}}-1}{\sqrt{\frac{1-F^1_{min}(1/r_i)}{1-F^0_{min}(1/r_i)}}-1}\times\lim_{r\to\infty}\frac{\sqrt{\frac{1-F^1_{min}(1/r_i)}{1-F^0_{min}(1/r_i)}}-1}{\left(1+\frac{\beta}{r^{\alpha_{min}}}\right)^{1/\alpha_{min}}-1} \\
&= 1\times\lim_{r\to\infty}\frac{\sqrt{\frac{1-F^1_{min}(1/r_i)}{1-F^0_{min}(1/r_i)}}-1}{\left(1+\frac{\beta}{r^{\alpha_{min}}}\right)^{1/\alpha_{min}}-1} \\
&= \lim_{r\to\infty}\frac{\sqrt{\frac{1-F^1_{min}(1/r_i)}{1-F^0_{min}(1/r_i)}}-1}{\frac{\beta}{r^{\alpha_{min}}}}\times\lim_{r\to\infty}\frac{\frac{\beta}{r^{\alpha_{min}}}}{\left(1+\frac{\beta}{r^{\alpha_{min}}}\right)^{1/\alpha_{min}}-1} \\
&= \frac{C_{min}\times\alpha_{min}}{2\beta} = \frac{1}{1-\varepsilon} > 1.
\end{aligned}$$

When R sufficiently large, we have

$$r_{i+1} \geq r_i \times \left(1+\frac{\beta}{r_i^{\alpha_{min}}}\right)^{1/\alpha_{min}} = \left(r_i^{\alpha_{min}}+\beta\right)^{1/\alpha_{min}} \Rightarrow r_i \geq \left(r_0^{\alpha_{min}}+\beta\times i\right)^{1/\alpha_{min}}. \tag{5}$$

Combining (4) and (5), we obtain:

$$\begin{aligned}
\mathbb{P}^*_{r_0}(H_1) &\geq \prod_{i=1}^{\infty}\left[1-(1+\varepsilon_1)\times\overline{C}\times\frac{1}{r_i^{\overline{\alpha}}}\right] \\
&\geq \prod_{i=1}^{\infty}\left[1-(1+\varepsilon_1)\times\overline{C}\times\frac{1}{\left(r_0^{\alpha_{min}}+\beta\times i\right)^{\overline{\alpha}/\alpha_{min}}}\right] \\
&\geq \prod_{i=1}^{\infty}\left[1-(1+\varepsilon_1)\times\overline{C}\times\frac{1}{\left(R^{\alpha_{min}}+\beta\times i\right)^{\overline{\alpha}/\alpha_{min}}}\right]
\end{aligned}$$

for all $r_0 \geq R$. Again, R is chosen to be sufficiently large such that each term is strictly positive. Suppose that there exists some $F \in \mathcal{F}_0$ such that $\mathcal{P}(F) < \mathcal{P}(\mathbb{F})$. This fact implies that $\alpha_{min} < \overline{\alpha}$, so

$$\sum\frac{1}{\left(R^{\alpha_{min}}+\beta\times i\right)^{\overline{\alpha}/\alpha_{min}}} < \infty,$$

which further implies that

$$\mathbb{P}^*_{r_0}(H_1) \geq \prod_{i=1}^{\infty}\left[1-(1+\varepsilon_1)\times\overline{C}\times\frac{1}{\left(R^{\alpha_{min}}+\beta\times i\right)^{\overline{\alpha}/\alpha_{min}}}\right] =: \delta > 0,$$

for all $r_0 \geq R$. In other words, the probability of an action-1 herd is greater than $\delta > 0$, which proves that state 1 is locally stable.

Proof of Proposition S6 (c) & (d)

The proofs of Proposition S6 (c) and (d) are almost identical to those of (a) and (b). The only difference is that the cutoff value becomes $\mathcal{P}(\mathbb{F})+1$. To see why this new cutoff arises, note that the probability of an action-0 herd is

$$\mathbb{P}^*_{r_0}(H_0)=\lim_{i\to\infty}\mathbb{P}^*_{r_0}(a_1=a_2=...a_i=0)=\prod_{i=1}^{\infty}\mathbb{F}^0\left(\frac{1}{r_i}\right)=\prod_{i=1}^{\infty}\left[1-\mathbb{F}^1(r_i)\right],$$

where r_i denotes the average likelihood ratio after $h_i=(0,...,0)$. An action-0 herd occurs with strictly positive probability if and only if $\sum\mathbb{F}^1(r_i)<\infty$. During a herd of action 0, we have $r_i\to 0$; besides, it can be verified that $\mathbb{F}^1(x)\sim C_1x^{\overline{\alpha}+1}$ as $x\to 0$ for some $C_1>0$.[4] As a consequence, an action-0 herd occurs with a strictly positive probability if and only if $\sum r_i^{\overline{\alpha}+1}<\infty$. The remainder of the proofs follows exactly the same logic as in parts (a) and (b). □

From Lemma S2, Proposition S6 implies Theorem S1, so the theorem is proved.

A.2 Proof of Proposition S3

Without loss of generality, I index all actions in descending order, i.e., $a^1>a^2>...>a^k$. The proof focuses on the case in which $a^k<1/2<a^1$, since the case in which all actions belong to one side of $1/2$ is a simple extension of this benchmark. Define the following four actions:

$$a^L=a^k,a^H=a^1,a^l=\max\{a\in A:a\leq 1/2\}\text{, and }a^h=\min\{a\in A:a>1/2\}.$$

Also, suppose that these four actions are different.[5]

Lemma S3. *For all $i\geq 1$, individual i will a.s. choose from $A^*=\left\{a^L,a^H,a^l,a^h\right\}$.*

Proof. Let $V_i(a)$ denote the minimum expected utility of individual i from choosing action

[4]Recall that $\mathbb{F}^0(x)\sim\overline{C}\times x^{\overline{\alpha}}$ as $x\to 0$, so

$$\lim_{x\to 0}\frac{\mathbb{F}^1(x)}{x^{\overline{\alpha}+1}}=\lim_{x\to 0}\frac{\mathbb{F}^1(x)}{(\overline{\alpha}+1)x^{\overline{\alpha}}}=\frac{1}{\overline{\alpha}+1}\lim_{x\to 0}\frac{f^0(x)}{x^{\overline{\alpha}-1}}=\frac{\overline{\alpha}}{\overline{\alpha}+1}\lim_{x\to 0}\frac{\mathbb{F}^0(x)}{x^{\overline{\alpha}}}=\frac{\overline{\alpha}}{\overline{\alpha}+1}\overline{C},$$

hence $\mathbb{F}^1(x)\sim\frac{\overline{\alpha}}{\overline{\alpha}+1}\overline{C}\times x^{\overline{\alpha}+1}$ as $x\to 0$.

[5]It is possible that some actions may coincide. For example, if there is only one action below 1/2, then $a^l=a^L$. The analysis can be easily extended to incorporate such cases.

a. By definition,

$$V_i(a) = \begin{cases} \frac{\lambda_i \underline{l}_i}{1+\lambda_i \underline{l}_i} a + \frac{1}{1+\lambda_i \underline{l}_i}(1-a) & a > 1/2 \\ \frac{\lambda_i \bar{l}_i}{1+\lambda_i \bar{l}_i} a + \frac{1}{1+\lambda_i \bar{l}_i}(1-a) & a \leq 1/2 \end{cases}, \tag{6}$$

which is a piecewise linear function, so the optimal a can be only obtained at one of the endpoints, i.e., in A^*. □

Lemma S4. *All actions in $A^* \backslash A^s$ will be chosen with probability zero in the limit.*

Proof. First, it is easy to verify that the first person will only choose a^L or a^H, and $a_1 = \begin{cases} a^L & \text{if } \lambda_1 < 1 \\ a^H & \text{if } \lambda_1 > 1 \end{cases}$. Without loss of generality, I assume that $a_1 = a^H$. There are three cases to consider: (i) $A^s = \{a^l\}$, (ii) $A^s = \{a^h\}$, and (iii) $A^s = \{a^l, a^h\}$. Since the logic is parallel across these cases, I focus on the case where $A^s = \{a^l\}$, i.e., $a^l + a^h > 1$.[6] Because $a_1 = a^H$, we have $\bar{l}_2 = \infty$ and $\underline{l}_2 = 1$. Substituting $\bar{l}_2$ and $\underline{l}_2$ into (6), individual 2's optimal choice is:

$$a_2 = \begin{cases} a^H & \lambda_2 > 1 \\ a^h & \lambda_2 \in (\lambda_2^*, 1) \\ a^l & \lambda_2 < \lambda_2^* \end{cases},$$

where λ_2^* is the signal such that individual 2 is indifferent between a^h and a^l, so it satisfies

$$a^l = \frac{\lambda_2^*}{1+\lambda_2^*} a^h + \frac{1}{1+\lambda_2^*}\left(1-a^h\right).$$

Since $a^l < 1/2$, it follows that $\lambda_2^* < 1$. Let p_i denote the probability that individual i chooses a^l. Then $p_2 = \mathbb{F}^0(\lambda_2^*)$. Suppose $a_2 = a^l$. Then:

$$\bar{l}_3 = \bar{l}_2 \times \sup_F \frac{F^1(\lambda_2^*)}{F^0(\lambda_2^*)} = \infty \times \lambda_2^* = \infty \quad \text{and} \quad \underline{l}_3 = \underline{l}_2 \times \inf_F \frac{F^1(\lambda_2^*)}{F^0(\lambda_2^*)} = 0.$$

Substituting them into the utility functions yields:

$$V_3\left(a^L\right) = a^L, V_3\left(a^l\right) = a^l, V_3\left(a^h\right) = 1 - a^h, \text{ and } V_3\left(a^H\right) = 1 - a^H.$$

Therefore, individual 3 will choose action a^l regardless of her private signal, i.e., $p_3 = 1$, and an information cascade on a^l begins. Therefore, Lemma S4 holds. Now suppose $a_2 = a^h$.

[6] For the other two cases, we can follow similar arguments and show that posteriors will become extremely ambiguous, i.e., $\bar{l}_i = \infty$ and $\underline{l}_i = 0$, after finitely many individuals.

Then:

$$\bar{l}_3 = \infty \quad \text{and} \quad \underline{l}_3 = \underline{l}_2 \times \inf_F \frac{F^1(1) - F^1(\lambda_2^*)}{F^0(1) - F^0(\lambda_2^*)} \leq \underline{l}_2 = 1.$$

From the perspective of individual 3, her optimal choice is

$$a_3 = \begin{cases} a^H & \lambda_3 > 1/\underline{l}_3 \\ a^h & \lambda_3 \in (\lambda_3^*, 1/\underline{l}_3) \\ a^l & \lambda_3 < \lambda_3^* \end{cases},$$

where λ_3^* solves:

$$a^l = \frac{\lambda_3^* \underline{l}_3}{1 + \lambda_3^* \underline{l}_3} a^h + \frac{1}{1 + \lambda_3^* \underline{l}_3} \left(1 - a^h\right).$$

Thus, $\lambda_3^* = \lambda_2^*/\underline{l}_3 \geq \lambda_2^*$. The probability of individual 3 choosing a^l is $p_3 = \mathbb{F}^0(\lambda_3^*) \geq p_2$. Suppose that $a_2 = a^H$, then we still have $\bar{l}_3 = \infty$ and $\underline{l}_3 = 1$, so individual 3 will act as if she were individual 2, and hence $p_3 = p_2$. To summarize, we have $p_3 \geq p_2$ regardless of individual 2's action. Analogously, we have $p_i \geq p_2$ for all $i \geq 2$. Levy's 0-1 Law implies that a^l will almost surely be taken by some individual i. Once it is taken, $\underline{l}_{i+1}$ becomes 0, and an information cascade of action a^l is triggered. Hence, in the limit, only actions in A^s will be chosen; furthermore, there is an information cascade. □

A.3 Proof of Proposition S5

Proof. Suppose that individuals favor randomization, i.e., when mixing different strategies, the mixing probability appears inside the minimum. That is,

$$V_i(\sigma) = \inf_{\pi \in \Pi_i\left(I_i, \sigma_{-i}^*\right)} \left[\sum_{a \in A} \sigma(a) \mathbb{E}_\pi U(a, \theta)\right].$$

Let $\sigma = \sigma(a = 1)$ and $\pi = \pi(1)$, then we have

$$\begin{aligned} \sum_{a \in A} \sigma(a) \mathbb{E}_\pi U(a, \theta) &= \sigma\pi + (1 - \sigma)(1 - \pi) \\ &= (2\sigma - 1)\pi + (1 - \sigma). \end{aligned}$$

So, the utility becomes:

$$V_i(\sigma) = \begin{cases} (2\sigma - 1)\underline{\pi_i} + (1 - \sigma) = \left(2\underline{\pi_i} - 1\right)\sigma + 1 - \underline{\pi_i} & \sigma \in [1/2, 1] \\ (2\sigma - 1)\overline{\pi}_i + (1 - \sigma) = (2\overline{\pi}_i - 1)\sigma + 1 - \overline{\pi}_i & \sigma \in [0, 1/2] \end{cases}.$$

Now consider the following cases:

1. Suppose $\underline{\pi_i} \leq \overline{\pi}_i \leq 1/2$. Then,

$$V_i(\sigma) \text{ is decreasing over } \sigma \in [0,1],$$

so the optimal strategy is $\sigma_i^* = 0$, i.e., individual i chooses action 0 with probability 1.

2. Suppose $\underline{\pi_i} \leq 1/2 \leq \overline{\pi}_i$. Then,

$$V_i(\sigma) \text{ is increasing on } [0,1/2] \text{ and decreasing on } [1/2,1],$$

so the optimal strategy is $\sigma_i^* = 1/2$, i.e., individual i randomizes evenly between the two actions.

3. Suppose $1/2 \leq \underline{\pi_i} \leq \overline{\pi}_i$. Then,

$$V_i(\sigma) \text{ is increasing over } \sigma \in [0,1],$$

so the optimal strategy is $\sigma_i^* = 1$, i.e., individual i chooses action 1 with probability 1.

Finally, noting that $\overline{\pi}_i = \frac{\lambda_i \overline{l}_i}{1+\lambda_i \overline{l}_i}$ and $\underline{\pi_i} = \frac{\lambda_i \underline{l_i}}{1+\lambda_i \underline{l_i}}$, the proposition follows directly. □

A.4 Proof of Theorem S2

Without loss of generality, suppose that $a_1 = 1$. Then, we have $\overline{l}_2 = \gamma$ and $\underline{l}_2 = 1$. I first state the following claim:

Lemma S5. *For all $i \geq 1$, suppose $a_1 = .. = a_i = 1$, then $\overline{l}_{i+1} = \gamma$ and $\underline{l}_{i+1} = 1$.*

Proof. The lemma clearly holds for $i = 1$. Suppose it also holds for $i = k$, and that $a_{k+1} = 1$. From Proposition S5 and the tie-breaking rule, $a_{k+1} = 1$ occurs in one of the following two cases:

- If $\lambda_{k+1} > 1/\underline{l}_{k+1} = 1$, then individual $k+1$ always takes action 1;
- If $\lambda_{k+1} \in \left[1/\overline{l}_{k+1}, 1/\underline{l}_{k+1}\right] = [1/\gamma, 1]$, then individual $k+1$ takes action 1 with a probability of 1/2.

Thus, for all signals $\lambda \in [1/\gamma, \gamma]$, individual $k+1$ takes action 1 with positive probability, and this probability is the same in both states. As a result, the public belief remains unchanged, implying that $\overline{l}_{k+1} = \overline{l}_k = \gamma$ and $\underline{l}_{k+1} = \underline{l}_k = 1$.

□

Lemma S6. *For all $i \geq 1$, suppose $a_1 = .. = a_i = 1$ and $a_{i+1} = 0$. Then, $\bar{l}_{i+2} = \gamma$ and $\underline{l}_{i+2} = 1/\gamma$.*

Proof. Note that $a_{i+1} = 0$ occurs only when

$$\lambda_{i+1} \in \left[1/\bar{l}_{i+1}, 1/\underline{l}_{i+1}\right] = [1/\gamma, 1],$$

where the equality comes from Lemma S5. Let $F_\emptyset$ denote the uninformative signal structure and F_γ denote the most informative signal structure, i.e., $\operatorname{supp}(F_\emptyset) = \{1\}$ and $\operatorname{supp}(F_\gamma) = \{1/\gamma, \gamma\}$. Therefore, we have $\bar{l}_{i+2} = \bar{l}_{i+1} \times \frac{F_\emptyset^1(1)}{F_\emptyset^0(1)} = \bar{l}_{i+1} = \gamma$ and $\underline{l}_{i+2} = \underline{l}_{i+1} \times \frac{F_\gamma^1(1)}{F_\gamma^0(1)} = \underline{l}_{i+1} \times 1/\gamma = 1/\gamma$. □

Now define $\mathcal{H}_{i+1}^{mix} = \{a_1 = ... = a_i \neq a_{i+1}\}$, which represents the event that a herd of action 1 is disrupted by individual $i+1$. The previous two lemmas establish that once $\mathcal{H}_{i+1}^{mix}$ occurs, a mixed-strategy information cascade occurs.

Lemma S7. *Let $\mathcal{H}^{mix} = \cup_i \mathcal{H}_{i+1}^{mix}$, then $\mathbb{P}^*(\mathcal{H}^{mix}) = 1$.*

Proof. Suppose that $a_1 = 1$. Then,

$$\mathbb{P}^*(a_{i+1} = 0 | a_1 = ... = a_i = 1) = \mathbb{P}^*(a_{i+1} = 0 | \lambda_{i+1} \in [1/\gamma, 1]) = \frac{\mathbb{F}^0(1)}{2}.$$

Given $a_1 = 1$, we have

$$\mathbb{P}^*(a_1 = ... = a_{i+1} | a_1 = 1) = \left[1 - \frac{\mathbb{F}^0(1)}{2}\right]^i \to 0 \text{ as } i \to \infty.$$

Therefore, $\mathbb{P}^*(\mathcal{H}^{mix} | a_1 = 1) = 1$. The case for $a_1 = 0$ is identical, so $\mathbb{P}^*(\mathcal{H}^{mix}) = 1$. □

Bibliography

Arieli, Itai and Manuel Mueller-Frank (2021) "A general analysis of sequential social learning," *Mathematics of Operations Research*, 46 (4), 1235–1249. 6

Herrera, Helios and Johannes Hörner (2012) "A necessary and sufficient condition for information cascades." 3, 4

Ke, Shaowei and Qi Zhang (2020) "Randomization and ambiguity aversion," *Econometrica*, 88 (3), 1159–1195. 8

Saito, Kota (2015) "Preferences for flexibility and randomization under uncertainty," *American Economic Review*, 105 (3), 1246–1271. 8

Smith, Lones, Peter Norman Sørensen, and Jianrong Tian (2021) "Informational herding, optimal experimentation, and contrarianism," *Review of Economic Studies*, 88 (5), 2527–2554. 4