Supplement to "Sequential Learning under Informational Ambiguity"

Jaden Yang Chen*

April 9, 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. Section S6 explores social learning under consistent ambiguity, illustrating how cascades arise from ambiguity attitudes rather than model misspecification. Section S7 investigates the case in which individuals face ambiguity about their own DGPs. Section S8 analyzes learning under heterogeneous ambiguity. Section *S*9 discusses heterogeneous preferences. Section S10 discusses an alternative updating rule: the α-maximum likelihood rule. Section S11 considers an extension in which individuals face ambiguity about the network structure. Appendix A contains omitted proofs.

JEL Classification: D81; D83; C72

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

Contents

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 exists some $\alpha > 0$ such that $F^0(x) = O(x^\alpha)$ 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 a 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 the 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 $a_i = 0$.

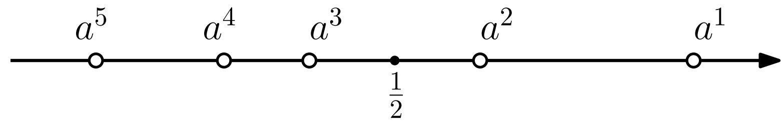

Figure 1: Linear Utility Functions

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. Furthermore, this section shows that under sufficient ambiguity: (i) at most two actions will be chosen in the limit, and (ii) these two actions must exhibit a form of symmetry. Thus, the assumption of a binary and symmetric action space is, in a sense, without loss of generality.

S3.1 Linear utility functions

Suppose that the action space is $A = \{a^1, ..., a^k\} \subset [0,1]$. First consider a simple case where the utility function is linear in a, that 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 $(0,\infty)$.[2] 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. Geometrically, it represents the set of actions closest 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.*

[2]Note that here we assume that signals are unbounded, but the analysis can be extended to bounded signals, as discussed in the next subsection.

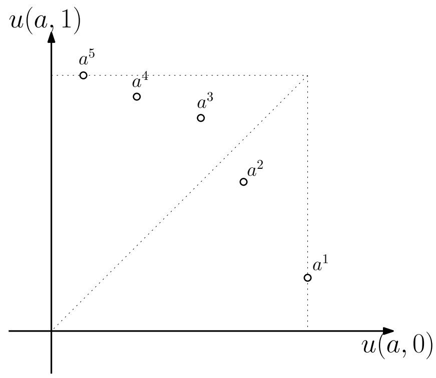

Figure 2: General Utility Functions

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. 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 two safe actions, a^2 and a^3, are equally distanced from 1/2.

Remark S1. 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.

S3.2 General utility functions

The result extends to general utility functions when ambiguous priors are also considered. Under sufficient ambiguity regarding both information and states, society will ultimately converge to at most two actions, which exhibit a form of symmetry.

From now on, I assume the following: (i) individuals form a set of priors with the prior likelihood set $L_0 = [1/R_0, R_0]$, where $R_0 > 1$ measures the degree of ambiguity about the true state; and (ii) individuals consider all possible DGPs on $[1/\gamma, \gamma]$. I impose the following regularity conditions:

Assumption S4. (No Redundancy) *For all $a, a' \in A$ with $a' \neq a$, $u(a, \theta) \neq u(a', \theta)$ in at* *least one state.*

Assumption S5. (No Strictly Dominated Action) *For all $a \in A$, there is no $a' \in A$ such that $u(a, \theta) \geq u(a', \theta)$ in both states, and the inequality is strict in at least one state.*

The set of safe actions is similarly defined as:

$$A^s = \left\{ a \in A : \min_{\theta} u(a, \theta) \geq \min_{\theta} u(a', \theta), \forall a' \in A \right\}.$$

This set also contains at most two actions, and when $|A^s| = 2$, the payoff-minimizing states must be different.

Proposition S4. *There exists $R \in \mathbb{R} \cup \{+\infty\}$ such that*

$$\lim_{t \to \infty} \mathbb{P}^* (a_t \in A^s) = 1,$$

for all $R_0 \geq R$, and we can find some $R < \infty$ when signals are bounded.

It shows that society will settle on safe actions under sufficient prior ambiguity. These actions are also symmetric, but in a weaker sense. In Figure 2, $A^s = \{a^2, a^3\}$ and they are "lower symmetric" with respect to the 45-degree line in the sense that: (i) the minimum utility is obtained at different states, i.e., they are on different sides of the 45-degree line; and (ii) the minimum utility levels are equal, i.e., $u(a^2, 1) = u(a^3, 0)$. Symmetrically, when individuals have max-max EU preference, the set of limit actions, $A^h = \{a^1, a^5\}$, are "upper symmetric" with respect to the 45-degree line, which means that the maximum utility levels are obtained at different states and must be equal.

I now characterize the **equilibrium strategy**. In what follows, I assume that A^s contains two elements (if A^s is a singleton, the equilibrium strategy becomes trivial in the limit). Suppose that $A^s = \left\{a^l, a^h\right\}$, where a^l achieves its minimum utility in state 0, and a^h achieves its minimum utility in state 1.

Proposition S5. (Equilibrium Strategy Multiple Actions) *Let $u = u\left(a^l, 0\right) = u\left(a^h, 1\right)$, $u^l = u\left(a^l, 1\right)$ and $u^h = u\left(a^h, 0\right)$. When R_0 is sufficiently large, we have*

$$\begin{matrix} a_i = a^l \\ a_i = a^h \end{matrix} \quad \text{if } \lambda_i \begin{matrix} < \\ > \end{matrix} \frac{\left(u^h - u^l\right) \underline{l}_i + \sqrt{\left(u^h - u^l\right)^2 \underline{l}_i^2 + 4\left(u^l - u\right)\left(u^h - u\right) \bar{l}_i \underline{l}_i}}{2\left(u^l - u\right) \bar{l}_i \underline{l}_i} \equiv \mathcal{X}_i,$$

and the strategy at $\lambda_i = \mathcal{X}_i$ is determined by the tie-breaking rule.

Note that if a^l and a^h are also "upper symmetric", i.e., $u^l = u^h$, the equilibrium cutoff simplifies to

$$\mathcal{X}_i = 1/\sqrt{\bar{l}_i \underline{l}_i},$$

which takes the exact same form as in the benchmark model. If they are not upper symmetric, but individuals hold sufficiently ambiguous beliefs (i.e., $\bar{l}_i$ is very large and $\underline{l}_i$ is very small), then

$$\mathcal{X}_i \approx \sqrt{\frac{u^h - u}{u^l - u}} / \sqrt{\bar{l}_i \underline{l}_i},$$

which differs from the previous characterization only by a constant. Thus, the equilibrium characterization in the binary-action case also serves as a good benchmark for the multi-action setting under sufficient ambiguity. Therefore, an information cascade still arises with probability 1 under sufficient ambiguity.

S4 Multiple States

When there are multiple states, the equilibrium becomes more difficult to characterize, but the key insights still hold.[3] 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 S6. *The set of perceived DGP, $\mathcal{G}_0$, contains all G such that*

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

[3] 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.

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.

Proposition S6. *Under Assumption S6, 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

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

From the perspective of individual 2, she will follow the first individual if

$$\min_{\pi \in \Pi_2} \sum_{\theta} \frac{\pi(\theta)}{\pi(\theta')} \times \frac{d\mathbb{G}_2(s_2|\theta)}{d\mathbb{G}_2(s_2|\theta')} > \min_{\pi \in \Pi_2} \sum_{\theta} \frac{\pi(\theta)}{\pi(\theta_1)} \times \frac{d\mathbb{G}_2(s_2|\theta)}{d\mathbb{G}_2(s_2|\theta_1)}. \tag{1}$$

Notice that

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

where the inequality comes from that $\frac{\pi(\theta_1)}{\pi(\theta)} \geq 1$ and $\frac{\pi(\theta)}{\pi(\theta')} \geq 1/R\gamma$ for all $\pi \in \Pi_2$. In addition, it can be verified that the R.H.S. of (1)$\leq \frac{K}{R} + 1$. As such 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 S6 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 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:

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

  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:

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

  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 S7. *Suppose individuals have preferences for randomization. Then, a mixed-strategy equilibrium exists, characterized as follows:*

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

where $\sigma_i^*(a_i = 1)$ represents the probability that individual i chooses action 1 *in the equilibrium. The indifference cases in* (2) *are determined by tie-breaking rules.*

The proof can be found in Appendix A.4. 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 (2) is equivalent to:

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

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 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.6. 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.

S6 Consistent Ambiguity

In the baseline model, individuals can perceive an arbitrary $\mathcal{F}_0$, so incorrect perceptions and MEU preferences can appear together, making it unclear how ambiguity attitude alone drives a cascade. This section discusses a setup that separates ambiguity attitudes from model misspecification.

S6.1 Consistent ambiguity

To isolate the effect of ambiguity attitude, this section discusses a special case in which individuals' perceptions of DGPs must be **consistent** with the objective probability—so individuals are both ambiguous and correctly specified. This section adopts the following assumption:

Assumption S7. (Consistent Ambiguity) *Suppose that $\mathbb{F}_i$ and $\mathcal{F}_0$ jointly satisfy*

$$\forall i, \theta, \lambda: \quad \mathbb{F}_i^{\theta}(\lambda) = \int F^{\theta}(\lambda) d\mu(F) \quad \textit{with } \mathcal{F}_0 = supp(\mu),$$

where μ is a second-order distribution over DGPs.

Assumption S7 can be interpreted as a situation in which each individual's DGP is drawn according to a second-order distribution μ, and individuals correctly perceive all possible DGPs. All probabilities are computed based on the ex ante signal distribution $\mathbb{F}_i$. In this case, ambiguity manifests in an objective manner—where the perceived DGP set equals the ex ante support. This allows us to isolate the role of ambiguity attitudes from that of misspecification.

S6.2 Conditions for information cascades

Although consistent ambiguity imposes additional restrictions on $\mathbb{F}_i$ and $\mathcal{F}_0$, the next proposition shows that *all* results about information cascades remain intact.

Proposition S8. (Cascade with consistent ambiguity). *Suppose that $\mathbb{F}_i$ and $\mathcal{F}_0$ jointly satisfy Assumption S7, and that either (i) $\mathcal{F}_0 = \mathcal{F}$, or (ii) $\mathbb{F}_i$ is bounded and $\mathcal{F}_0$ satisfies conditions in Theorem 2 in the paper, then an information cascade occurs $\mathbb{P}^*$-almost surely, and both correct and incorrect cascades occur with $\mathbb{P}^*$-positive probability.*

Proposition S8 corresponds to special cases of Theorems 1 and 2 in the paper. It is easy to verify that the conditions of both theorems are satisfied, so the proof is omitted. The difference is that under the conditions of Proposition S8, DGP perceptions are consistent with the objective probability, so it shows that MEU preferences alone—instead of misspecification—can produce an information cascade under sufficient ambiguity. Below is a concrete example:

Example S1. Consider the same signal structures as in Example 1 in the paper. Further suppose that every individual's signal precision $\gamma_i \overset{I.I.D.}{\sim} \mu$. Individuals know μ but not the realizations of other individuals' signal precision. Let $g_\mu(s|\theta) = \int g_\gamma(s|\theta)d\mu(\gamma)$ denote the ex ante signal distribution, where g_γ denotes the DGP with precision γ. All events are evaluated using g_μ.

- **Expected utility**: When individuals have expected-utility preferences, this corresponds to the standard SSLM with *correct* DGP specification, g_μ. In this scenario, different learning outcomes can occur, depending on the properties of μ. For example, when the support of μ is unbounded, complete learning occurs (Smith and Sørensen (2000)). When μ has a discrete and finite support, an information cascade occurs almost surely (Bikhchandani et al. (1992)).
- **MEU**: When individuals have MEU preferences, an information cascade occurs almost surely for *all* possible μ, regardless of whether it has bounded or unbounded support or whether the implied g_μ satisfies the increasing hazard ratio property or not.

In both cases, individuals are correctly specified. The key distinction lies in their preferences, underscoring the role of ambiguity attitude itself in generating cascades.

S6.3 Incorrect herding with unbounded signals

Under consistent ambiguity, a cascade with unbounded signals requires extreme ambiguity. This subsection shows that herding can still arise under moderate ambiguity, parallel to the results in the paper. For tractability, I focus on DGPs from a parametric family with power tails:

Assumption S8. (Power Tails) $\mathcal{F}_0 = \{F(\cdot, \alpha)\}_{\alpha \in \mathcal{A}}$, *where: (i)* $\mathcal{A} = [\underline{\alpha}.\overline{\alpha}] \subset \mathbb{R}_{++}$, *and (ii) we have* $F^0(x, \alpha) = O(x^\alpha)$ *as* $x \to 0$ *for all* $\alpha \in \mathcal{A}$.

I also assume that μ admits a density function over $\mathcal{A}$, and let $\mu(\alpha)$ denote the density function with some abuse of notation.[4] We have the following proposition:

Proposition S9. (Herding with consistent ambiguity). *Suppose that Assumptions S7 and S8 hold, and that* $\mu(\alpha) \leq C \times (\alpha - \underline{\alpha})^k$ *as* $\alpha \to \underline{\alpha}$ *for some* $C, k > 0$, *then herding occurs* $\mathbb{P}^*$*-almost surely, and an incorrect herd occurs with a* $\mathbb{P}^*$*-strictly positive probability.*

Proposition S9 parallels Theorem 3 in the main paper and uses a similar proof strategy. It says that if the second-order distribution μ is controlled by some power function as $\alpha \to \underline{\alpha}$, then an incorrect herd occurs with strictly positive probability. The intuition is as follows. First notice that under MEU preferences, individuals are mostly influenced by the most informative DGP (i.e., the one with power $\underline{\alpha}$), because it creates a strong cascading force that can't be fully mitigated by other DGPs. When μ is controlled by some power function, highly informative DGPs are sufficiently rare such that individuals essentially overweight the highly informative DGPs in their decisions to a sufficient large extent such that a wrong herd persists with a positive probability. Below is a concrete example:

Example S2. Consider the same setup as an example in the paper, where each individual's DGP is parameterized by m_i. Suppose that m_i is i.i.d. drawn from $M_\varepsilon = [m_0 - \varepsilon, m_0 + \varepsilon]$ according to a generalized beta distribution μ, where

$$\mu(m) = (m_0 + \varepsilon - m)^a \times (m - m_0 + \varepsilon)^b \times c$$

where $a, b > 0$, and $c > 0$ is a normalizing constant. All individuals correctly specify μ. When individuals have expected utility preferences, complete learning occurs. However, when individuals have MEU preferences, and for any $\varepsilon > 0$, complete learning collapses, and an incorrect herd occurs with a strictly positive probability.

Notice that in this example, individuals only consider ex ante correct DGPs, yet incorrect herding still emerges, even if ambiguity is arbitrarily small. This illustrates the fragility of complete learning, which can break down under consistent ambiguity.[5]

[4] The density function satisfies $\mu(\alpha) = \frac{d\mu(F(\cdot,\alpha))}{d\alpha}$.

[5] It is worth noting that the fragility of complete learning is less pronounced under consistent ambiguity (compared to that in the main paper). To achieve fragility, we require that the most informative DGP carries zero density, i.e., $\mu(\alpha) \to 0$ as $\alpha \to \underline{\alpha}$. Suppose instead that the most informative DGP is realized with a strictly positive probability; then complete learning can still occur.

S7 Ambiguity about Self-signals

The baseline model assumes that individuals face ambiguity about others' DGPs but are confident in interpreting their own signals. This section extends the model to the case where individuals face ambiguity about their own DGPs as well.

S7.1 Setup

For convenience, I work with nominal signals. Suppose that each individual receives a signal $s_i \in S = [\underline{s}, \overline{s}]$ drawn from a DGP G_i. Individuals face ambiguity about both their own DGPs and those of others, considering a common set of possible DGPs $\mathcal{G}_0$. Assume that every $G \in \mathcal{G}_0$ has full support and satisfies the monotone likelihood ratio property (MLRP):

$$\forall s \geq s' : \quad \frac{dG^1(s)}{dG^0(s)} \geq \frac{dG^1(s')}{dG^0(s')} \quad \forall G \in \mathcal{G}_0.$$

Furthermore, the maximum likelihood ratio is γ and the minimum is $1/\gamma$, so:

$$\frac{dG^1(\overline{s})}{dG^0(\overline{s})} = \gamma, \text{ and } \frac{dG^1(\underline{s})}{dG^0(\underline{s})} = 1/\gamma \quad \forall G \in \mathcal{G}_0.$$

In this setup, individuals have only a coarse understanding of the true signal distributions: (i) they know that a higher signal is more indicative of state 1; and (ii) they know the most precise signals are $\overline{s}$ and $\underline{s}$ along with their induced likelihood ratios. These are weak regularity assumptions that encompass a broad class of environments.[6]

S7.2 Equilibrium strategy

When individual i receives a signal $s_i \in S$, she faces ambiguity about the likelihood ratio it induces. Specifically, she considers the following set of possible likelihood ratios:

$$\Lambda(s_i) = \left\{ \frac{dG^1(s_i)}{dG^0(s_i)} : G \in \mathcal{G}_0 \right\}.$$

Let $\underline{\lambda}(s_i)$ and $\overline{\lambda}(s_i)$ denote the infimum and supremum likelihood ratios in the set $\Lambda(s_i)$. I also define the **average private likelihood ratio** as:

$$\lambda_i^{avg} = \sqrt{\overline{\lambda}(s_i) \times \underline{\lambda}(s_i)},$$

[6]The MLRP is primarily adopted for expositional simplicity and is nonessential. Extensions and generalizations are discussed in Section S7.4.

which is the geometric average of the highest and lowest likelihood ratios perceived by individual i. The equilibrium strategy can be characterized as follows:

Proposition S10. *In the equilibrium, we have:*

$$a_i = \begin{cases} 1 & \text{if } r_i \times \lambda_i^{avg} > 1 \\ 0 & \text{if } r_i \times \lambda_i^{avg} < 1 \end{cases}, \tag{3}$$

where r_i is the average public likelihood ratio as defined in the paper.

Proof. We have:

$$\underline{\pi}_i(1) = \frac{\underline{\lambda}_i \cdot \underline{l}_i}{\underline{\lambda}_i \cdot \underline{l}_i + 1} \text{ and } \underline{\pi}_i(0) = \frac{1}{\overline{\lambda}_i \cdot \overline{l}_i + 1}.$$

Individual i takes action 1 if $\underline{\pi}_i(1) > \underline{\pi}_i(0)$ and action 0 otherwise, which implies the proposition. □

Notice that the equilibrium strategy is **almost identical** to that in the main paper, except that the private signal λ_i is replaced with the average private likelihood ratio, λ_i^{avg}. As a result, the main results in the paper continue to apply under this extension.

S7.3 Cascades with ambiguity about self-signals

I first consider an extreme-ambiguity case in which individuals consider the set of all DGPs that satisfy the regularity conditions in Section S7.1, denoted by $\mathcal{G}$. Under this perception, we can express the average private likelihood ratio induced by each signal as:

$$\lambda_i^{avg} = \begin{cases} \gamma & \text{if } s_i = \overline{s} \\ 1 & \text{if } s_i \in (\underline{s}, \overline{s}) \\ 1/\gamma & \text{if } s_i = \underline{s} \end{cases}, \tag{4}$$

Equation 4 implies that, under $\mathcal{G}$, individuals can interpret *only* the two extreme signals, $\underline{s}$ and $\overline{s}$, in a precise way. For any intermediate signal $s_i \in (\underline{s}, \overline{s})$, individuals face maximal ambiguity—it can be flexibly interpreted as favoring either state. In such cases, we have $\underline{\lambda}(s_i) = \gamma$ and $\overline{\lambda}(s_i) = 1/\gamma$, so the average private likelihood ratio is equal to 1. Further impose a tie-breaking rule that whenever indifferent, individuals randomize between two actions with equal probability.

Proposition S11. *Suppose $\mathcal{G}_0 = \mathcal{G}$. Then, an information cascade occurs $\mathbb{P}^*$-almost surely, and both correct and incorrect cascades occur with strictly positive $\mathbb{P}^*$-probability.*

The proof is similar to that in the paper. Indeed, the dynamics of r_i are nearly identical to those in Theorem 1 of the main paper. To clarify the intuition, I now explain why the **asymmetric effect** that drives cascades in the paper still holds here:

- For individual 1: By Proposition S10 and Equation (4) (along with the tie-breaking rule), we have: (i) when $s_1 = \overline{s}$, individual 1 chooses action 1; (ii) when $s_1 = \underline{s}$, individual 1 chooses action 0; and (iii) when $s_1 \in (\underline{s}, \overline{s})$, individual 1 randomizes.
- For individual 2: Suppose $a_1 = 1$. Based on this observation, she could infer $s_1 > \underline{s}$.
  - The supremum public likelihood ratio is obtained when individual 1 has the most precise DGP. That is, $G_1 = \overline{G}$, where

    $$\text{supp}\left(\overline{G}\right) = \{\underline{s}, \overline{s}\}.$$

    In this case, the only signal that leads to action 1 is $\overline{s}$, which induces a likelihood ratio equal to γ.
  - The infimum public likelihood ratio is obtained when G_1 is atomless at $\underline{s}$. Then,

    $$\mathbb{P}^1_{G_1}(s_1 > \underline{s}) = \mathbb{P}^0_{G_1}(s_1 > \underline{s}),$$

    which implies that individual 1's action is uninformative.
- For any individual $i + 1$: Suppose that $a_1 = ... = a_i = 1$, and that an information cascade has not yet occurred. Further, suppose that $i+1$ receives signal $\underline{s}$, the strongest signal supporting state 0.
  - The worst-case scenario for deviating from the herd occurs when predecessors' DGPs are all equal to $\overline{G}$, in which case their actions perfectly reveal i strong signals—each equal to $\overline{s}$.
  - The worst-case scenario for following the herd is when all predecessors' DGPs are atomless at $\underline{s}$, in which case predecessors' actions are uninformative.

As i increases, individual $i + 1$ would need to act against an increasing number of strong signals $\overline{s}$ to deviate from the herd, whereas following the herd means acting only against her own private signal $\underline{s}$. As a result, an information cascade eventually takes place.

S7.4 Remarks

In Proposition S11, individuals face huge ambiguity regarding their own signals—they do not know how to interpret any signal except for the most extreme ones.[7] The result illustrates that even a very coarse understanding of one's own signal structure is sufficient to trigger an information cascade. Following the arguments in the main paper, it is conceivable that Proposition S11 can be extended to accommodate more general perceptions of DGPs. A natural conjecture is that a cascade will still occur whenever the set $\mathcal{G}_0$ contains some DGP G with sufficiently thick tails.

It is important to note that to make social learning nontrivial, individuals must possess at least *some* understanding of how to interpret their signals, even if that understanding is very limited. Suppose instead that individuals face even more ambiguity than in Proposition S11—specifically, they don't know how to interpret any signal and consider all DGPs on S as possible. In this case, every signal is perceived as symmetrically favorable to both states. Formally, we would have $\lambda^{avg}(s) = 1$ for all $s \in S$; that is, every signal is effectively uninformative. As a result, individuals' actions carry no information, and there is nothing to learn from observing others.

S8 Heterogeneous Ambiguity

This section explores extensions of the paper's main results to settings with heterogeneous ambiguity. In the baseline model, individuals share a common set of models, $\mathcal{F}_0$. This assumption implies two aspects of homogeneity: (i) individuals' signal structures appear homogeneously ambiguous to others, and (ii) individuals face homogeneous ambiguity about others' signal structures.[8] Below, I discuss how my results can be relaxed in these two directions.

S8.1 Individuals have heterogeneously ambiguous DGPs

Suppose instead that individuals' DGPs are heterogeneously ambiguous. There are two types, $t_i \in \{H, L\}$. If individual i has type t, other individuals believe that her DGP $F_i \in \mathcal{F}^t$. Suppose that $\mathcal{F}^L \subset \mathcal{F}^H$, so H-type individuals have more ambiguous DGPs. Here, we can think of L-type individuals as the "famous" individuals, whose signal structures are better known. For simplicity, I assume that all types are commonly known. Also, suppose

[7] Note that this result relies very little on the MLRP, so it applies to a broader class of DGPs.

[8] Formally, let $\mathcal{F}_{ij}$ denotes the set of DGPs that individual i considers possible for individual j. (i) says that for each i, we have $\mathcal{F}_{ij} = \mathcal{F}_{ij'}$ for all j, j'. (ii) says that for each j, we have $\mathcal{F}_{ij} = \mathcal{F}_{i'j}$ for all i, i'.

that the distance between the i-th and $i+1$-th t-type individuals is bounded by a fixed constant for all i, t. This assumption guarantees that no type will vanish in the limit.

Proposition S12. *When there is sufficient ambiguity for H-type individuals, e.g., when $\mathcal{F}^H$ satisfies the conditions in Theorem 2 in the paper, an information cascade occurs $\mathbb{P}^*$-almost surely.*

Proof. Following the proof of Theorem 2, we can show that if a cascade has not occurred, there exists some $\beta > 1$ such that r_i will increase or decrease by a factor of β after a H-type individual's action. Because H-type individuals have bounded distance, we can find a constant $K < \infty$ such that K identical actions can trigger a cascade. The rest of the argument follows as in Theorem 2. □

Notice that the proposition imposes no restriction on the fraction of high-ambiguity individuals, so an information cascade can emerge even when there is an ε-fraction of high-ambiguity individuals.[9] Also, the proposition imposes no restriction on $\mathcal{F}^L$. If we take $\mathcal{F}^L$ to be the true DGP, the proposition further implies that an information cascade can arise even when a small fraction of individuals have ambiguous DGPs, whereas the majority's DGPs are commonly known.

S8.2 Individuals face heterogeneous ambiguity about others

Still suppose there are two types, $t_i \in \{H, L\}$, and let $\mathcal{F}^t$ denote the set of perceived DGPs for individuals of type t, where $\mathcal{F}^L \subset \mathcal{F}^H$. Here, L-type individuals can be viewed as facing less ambiguity, since they consider a smaller set of possible DGPs about others.

Proposition S13. *If both types of individuals face substantial ambiguity, i.e., when both $\mathcal{F}^L$ and $\mathcal{F}^H$ satisfy the conditions in Theorem 2 in the paper, an information cascade occurs with strictly positive $\mathbb{P}^*$-probability.*

Proof. Let r_i^t denote the average public likelihood ratio for type t. Theorem 2 implies that when both $\mathcal{F}^L$ and $\mathcal{F}^H$ are sufficiently large, both r_i^H and r_i^L will enter the cascade set after finite number of identical actions, so an information cascade must occur with strictly positive probability. □

[9]For example, suppose that $t_i = H$ if $i \in \{1, n+1, 2n+1, 3n+1, ...\}$, and $t_i = L$ otherwise, where n is a positive integer. The fraction of H-type individuals in the whole population is $\lim_{k\to\infty} \frac{\sum_{i\le k} 1_{\{t_i = H\}}}{k} \to 1/n$, which can be an arbitrarily small number.

S9 Heterogeneous Preferences

This section explores an extension in which individuals have heterogeneous preferences. I show that the paper's main result is robust to preference heterogeneity. I consider the following two cases.

S9.1 No preference uncertainty

Suppose that individuals' preferences are commonly known. In this case, individuals interpret the implications of their predecessors' actions in the same way as in the baseline model. They may take different actions due to differences in preferences, but belief dynamics remain unchanged from those in the paper.

S9.2 Preference uncertainty

A more interesting case arises when preferences are heterogeneous and individuals do not know each other's preferences. Suppose there are two preference types $\tau_i \in T = \{A, B\}$, where

$$u^A(a, \theta) = 1_{\{a=\theta\}} \quad \text{and} \quad u^B(a, \theta) = 1_{\{a \neq \theta\}},$$

meaning that type A seeks to match the true state, whereas type B seeks to mismatch it. Types are i.i.d. with the following distribution:

$$\tau_i = \begin{cases} A & \text{w.p. } 1-p \\ B & \text{w.p. } p \end{cases},$$

where $p < 1/2$ stands for the probability of the opposing type.[10] Individuals know their own types but not those of others, though they correctly perceive the distribution of types. In this case, we can show that an information cascade still occurs almost surely.

Proposition S14. *Suppose $\mathcal{F}_0 = \mathcal{F}$, then an information cascade occurs $\mathbb{P}^*$ almost surely, and both correct and incorrect cascades occur with strictly positive $\mathbb{P}^*$-probability.*

Proof. To see this, consider the belief dynamics. Suppose an information cascade has not yet emerged. Then:

[10] The case for $p > 1/2$ is symmetric. One can also allow individuals to consider a set of possible values for p. However, we must exclude the case $p = 1/2$ (or a set symmetric around $1/2$), because in such cases, the history becomes effectively uninformative—any action can be interpreted as equally favorable for either state, and hence there is essentially nothing to learn. This is similar to the situation discussed in Section S7.4.

- The lowest public likelihood ratio $\underline{l}_{i+1}$ satisfies:

$$\begin{aligned}\underline{l}_{i+1} &= \underline{l}_i \times \inf_{F\in\mathcal{F}} \frac{\sum_{\tau_i\in T} \mathbb{P}_F^1\left(a_i=1|h_i,\tau_i\right)p\left(\tau_i\right)}{\sum_{\tau_i\in T} \mathbb{P}_F^0\left(a_i=1|h_i,\tau_i\right)p\left(\tau_i\right)} \\ &= \underline{l}_i \times \inf_{F\in\mathcal{F}} \frac{\left(1-p\right)\left(1-F^1\left(1/r_i\right)\right)+pF^1\left(1/r_i\right)}{\left(1-p\right)\left(1-F^0\left(1/r_i\right)\right)+pF^0\left(1/r_i\right)},\end{aligned}$$

  where the second equality comes from that if individual i is type A, she chooses action 1 if $\lambda_i \times r_i > 1$, and if her type is B, she chooses action 1 if $\lambda_i \times r_i < 1$. Setting $\eta = \frac{1-p}{1-2p} > 1$, it follows that:

$$\underline{l}_{i+1} = \underline{l}_i \times \inf_{F\in\mathcal{F}} \frac{\eta - F^1\left(1/r_i\right)}{\eta - F^0\left(1/r_i\right)} \geq \underline{l}_i.$$

  The inequality holds because $F^0 \geq F^1$. Since $r_i \geq 1$, the infimum can be attained at the uninformative DGP.

- The highest public likelihood ratio $\bar{l}_{i+1}$ satisfies:

$$\begin{aligned}\bar{l}_{i+1} &= \bar{l}_i \times \sup_{F\in\mathcal{F}} \frac{\left(1-p\right)\left(1-F^1\left(1/r_i\right)\right)+pF^1\left(1/r_i\right)}{\left(1-p\right)\left(1-F^0\left(1/r_i\right)\right)+pF^0\left(1/r_i\right)} \\ &\geq \bar{l}_i \times \frac{\left(1-p\right)\left(1-F_\gamma^1\left(1/r_i\right)\right)+pF_\gamma^1\left(1/r_i\right)}{\left(1-p\right)\left(1-F_\gamma^0\left(1/r_i\right)\right)+pF_\gamma^0\left(1/r_i\right)},\end{aligned}$$

  where F_γ is the "most precise" DGP that generates only signals γ and $1/\gamma$. Thus,

$$\bar{l}_{i+1} \geq \bar{l}_i \times \frac{\left(1-p\right)\frac{\gamma}{1+\gamma} + p\frac{1}{1+\gamma}}{\left(1-p\right)\frac{1}{1+\gamma} + p\frac{\gamma}{1+\gamma}} = \bar{l}_i \times \xi.$$

  Since $\gamma > 1$ and $p < 1/2$, it follows that $\xi > 1$, so the highest public likelihood ratio increases by a factor of at least ξ.

Therefore, $r_{i+1}/r_i \geq \sqrt{\xi} > 1$ when $a_i = 1$. Similarly, $r_{i+1}/r_i \leq \sqrt{1/\xi} < 1$, when $a_i = 0$. This satisfies Lemma 3 in the paper, so r_i enters the cascade set almost surely, and each cascade set is visited with a strictly positive probability. □

Remark S2. With preference heterogeneity, an information cascade occurs almost surely, but individuals do not necessarily reach action consensus. For example, suppose $r_i \in C_1 = [\gamma, \infty]$. In this case:

- type-A individuals will choose action 1 regardless of their private signals.

- type-B individuals, who want to mismatch the state, will choose action 0 regardless of private signals.

Therefore, while information stops aggregating due to cascades, actions remain heterogeneous across preference types.

S10 The α-Maximum Likelihood Rule

The occurrence of a cascade is not unique to the full Bayesian rule. This section discusses an alternative updating rule—the **α-maximum likelihood rule** (α-MLE) as in Epstein and Schneider (2007). The updating rule requires that

$$\mathcal{F}_{-i} \mid h_i = \left\{ F_{-i} : \mathbb{P}_{F_{-i}}\left(h_i|\sigma_{-i}\right) \geq \alpha \cdot \sup_{F_{-i} \in \mathcal{F}_{-i}} \mathbb{P}_{F_{-i}}\left(h_i|\sigma_{-i}\right) \right\}$$

where $\alpha \in [0,1]$, $F_{-i} = (F_1, ..., F_{i-1})$, and $\mathcal{F}_{-i} \mid h_i$ denotes the updated set of perceived DGPs after history h_i. Under this updating rule, individuals entertain only the DGPs that pass a likelihood test, where $\alpha = 1$ corresponds to the maximum likelihood updating, and $\alpha = 0$ corresponds to the full Bayesian updating.

Proposition S15. *Suppose that $\mathcal{F}_0 = \mathcal{F}$ and signals are bounded. Under α-MLE, an information cascade occurs with strictly positive $\mathbb{P}^*$-probability for all $\alpha \in [0, 1)$.*

Proof. By chain rule,

$$\mathbb{P}_{F_{-i}}\left(h_i\right) = \mathbb{P}_{F_{-i}}\left(a_1\right) \mathbb{P}_{F_{-i}}\left(a_2|a_1\right) ... \mathbb{P}_{F_{-i}}\left(a_{i-1}|a_1, a_2, ..., a_{i-2}\right).$$

Consider a history where $a_1 = a_2 = ... = a_{i-1} = 1$, and let $F^*_{-i} = \left(F^*_1, ..., F^*_{i-1}\right) \in \arg\max \mathbb{P}_{F_{-i}}\left(h_i\right)$; i.e., the DGPs that maximize the probability of history h_i.[11] With some abuse of notation, let $F_{-i} = \left(F^*_1, ..., F^*_{i-2}, F_{i-1}\right)$. By definition, $F_{-i} \in \mathcal{F}_{-i} \mid h_i$ if and only if $\mathbb{P}_{F_{-i}}\left(h_i\right) \geq \alpha \cdot \mathbb{P}_{F^*_{-i}}\left(h_i\right)$, or equivalently,

$$\mathbb{P}_{F_{-i}}\left(a_{i-1}|h_{i-1}\right) \geq \alpha \mathbb{P}_{F^*_{-i}}\left(a_{i-1}|h_{i-1}\right) = \alpha.$$

This implies:

$$\mathbb{P}_{F_{-i}}\left(a_{i-1}|h_{i-1}; \theta = 0\right) \mathbb{P}_{F_{-i}}\left(\theta = 0|h_{i-1}\right) + \mathbb{P}_{F_{-i}}\left(a_{i-1}|h_{i-1}; \theta = 1\right) \mathbb{P}_{F_{-i}}\left(\theta = 1|h_{i-1}\right) \geq \alpha. \quad (5)$$

[11] The maximum exists because (i) $\mathbb{P}_{F_1}\left(a_1\right) = \frac{1}{2}$ for all F_1 continuous at 1, and (ii) if we let $F^*_2 = ... = F^*_{i-1}$ be uninformative DGP, we have $\mathbb{P}_{F_2}\left(a_2|a_1\right) = ... = \mathbb{P}_{F_{-i}}\left(a_{i-1}|a_1, a_2, ..., a_{i-2}\right) = 1$.

When $h_{i-1} = \{1, ..., 1\}$, we have $\mathbb{P}_{F_{-i}}(\theta = 1|h_{i-1}) \geq \mathbb{P}_{F_{-i}}(\theta = 0|h_{i-1})$ for all $F_{-i} \in \mathcal{F}_{-i}$; also, since $a_{i-1} = 1$,we have $\mathbb{P}_{F_{-i}}(a_{i-1}|h_{i-1}; \theta = 1) \geq \mathbb{P}_{F_{-i}}(a_{i-1}|h_{i-1}; \theta = 0)$.[12] As a consequence,

$$\begin{aligned}&\mathbb{P}_{F_{-i}}(a_{i-1}|h_{i-1}; \theta = 0)\,\mathbb{P}_{F_{-i}}(\theta = 0|h_{i-1}) + \mathbb{P}_{F_{-i}}(a_{i-1}|h_{i-1}; \theta = 1)\,\mathbb{P}_{F_{-i}}(\theta = 1|h_{i-1})\\ \geq&\mathbb{P}_{F_{-i}}(a_{i-1}|h_{i-1}; \theta = 0)\frac{1}{2} + \mathbb{P}_{F_{-i}}(a_{i-1}|h_{i-1}; \theta = 1)\frac{1}{2},\end{aligned}$$

so inequality (5) holds if

$$\mathbb{P}_{F_{-i}}(a_{i-1}|h_{i-1}; \theta = 0)\frac{1}{2} + \mathbb{P}_{F_{-i}}(a_{i-1}|h_{i-1}; \theta = 1)\frac{1}{2} \geq \alpha. \tag{6}$$

Suppose that $i \geq 2$ and there is no information cascade yet, i.e., $r_i \in (1, \gamma)$. Consider a discrete F_i where $\operatorname{supp}(F_i) = \left\{\frac{1}{\gamma}, 1, \gamma\right\}$. Let f_i^θ be the p.m.f. of F_i^θ. Suppose that $f_i^0(\gamma) = f_i^1\left(\frac{1}{\gamma}\right) = p$, thus $f_i^0\left(\frac{1}{\gamma}\right) = f_i^1(\gamma) = p\gamma$, where $p \in \left[0, \frac{1}{\gamma+1}\right]$. Since $r_i \in (1, \gamma)$, we have

$$\begin{aligned}\mathbb{P}_{F_i}(a_i|h_i; 0) &= 1 - F^0(1/r_i) = 1 - p\gamma\\ \mathbb{P}_{F_i}(a_i|h_i; 1) &= 1 - F^1(1/r_i) = 1 - p.\end{aligned}$$

Then (6) implies $p \leq \frac{2-2\alpha}{1+\gamma}$, so the F_i with $p = \frac{2-2\alpha}{1+\gamma}$ belongs to $\mathcal{F}_{-i} \mid h_i$. When $\alpha \in [0, 1)$, we have

$$\frac{r_{i+1}}{r_i} = \frac{1 - p\gamma}{1 - p} > 1 \text{ for all } r_i \in (1, \gamma),$$

so an information cascade occurs after finite steps and hence with strictly positive probability.

□

Notice that a cascade may not occur at $\alpha = 1$, i.e., the maximum likelihood updating (MLU). This is because the MLU can lead to an "over-fitting problem": Under MLU, individuals may continue to entertain only uninformative DGPs. Since a herd can occur with probability 1 when followers have no informative signals, beliefs stop updating after the first individual. As a result, an information cascade may not arise under MLU.

S11 Ambiguity over the Network Structure

The discussion can be extended to **ambiguous networks**. This section shows that when individuals face ambiguity about other people's observation structures, and when the am-

[12] The inequalities come from the equilibrium strategy and the fact that $F^0(x) \geq F^1(x)$ in Lemma A.1 of Smith and Sørensen (2000)

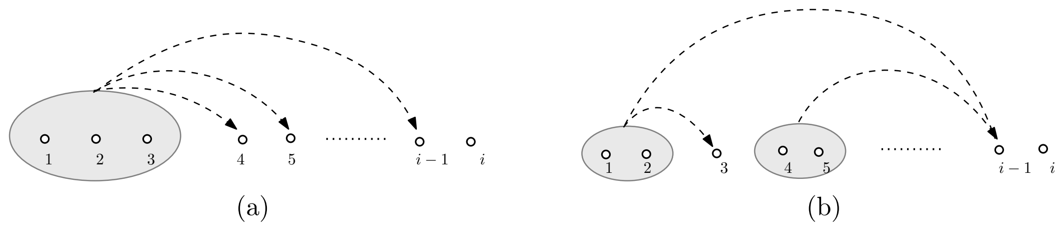

Figure 3: Ambiguous Networks

Note: The dashed curves represent the observation structure. In the first graph, individuals can observe only the actions taken by individuals in $I = \{1, 2, 3\}$. In the second graph, individuals can observe only actions taken by individuals in $I = \{1, 2\} \cup \{4, 5\}$.

biguity is sufficiently large, an information cascade occurs almost surely for all bounded signals.[13]

A network structure is denoted by $G = (G_1, G_2, ...)$, where $G_i \subset \{1, ..., i-1\}$ represents the set of individuals whose actions are observable to individual i. Individuals are located in a linear network but face ambiguity about the network structure. Let $\mathcal{G}$ represent the set of all possible network structures. Let $\mathcal{G}_0 \subset \mathcal{G}$ denote the set of network structures perceived by society. Formally, individual i believes that her predecessors' observation set can be any $(G_1, ..., G_{i-1})$ consistent with $\mathcal{G}_0$. Signals are i.i.d. according to $\mathbb{F}$, where $\mathbb{F}$ is continuous and has full support on $[1/\gamma, \gamma]$ where $\gamma \in (1, \infty)$. To emphasize the effect of network ambiguity, I assume that individuals correctly understand $\mathbb{F}$, i.e., there is no ambiguity about the DGP.

Lemma S1. *If $\mathcal{G}_0 = \mathcal{G}$, an information cascade occurs $\mathbb{P}^*$-almost surely.*

The lemma says that when individuals consider all networks as possible, an information cascade will occur almost surely. While Lemma S1 requires extreme ambiguity about the network, we only need a weaker condition:

Definition S3. A network $G = (G_1, G_2, ...)$ is *bounded by K* if there exists some $K < \infty$ such that $\max_{i,k} \{k : k \in G_i\} \leq K$.

A network is bounded if only a finite number of individuals are observable to society. The concept is illustrated in Figure 3. If individual i considers the network structure in Figure 3a, then she finds it possible that her predecessors can only observe the first three individuals. Similarly, in Figure 3b, her predecessors may only observe individuals from $\{1, 2\}$ and $\{4, 5\}$.

Proposition S16. *There exists some $K < \infty$ such that if there exists some $G \in \mathcal{G}_0$ that is bounded by K, then an information cascade occurs $\mathbb{P}^*$-almost surely.*

[13]When signals are unbounded, information cascade is a very strong concept, and ambiguity over networks may not lead to cascades on its own. However, it is conceivable that ambiguous networks can still lead to incorrect herding, so complete learning does not hold.

Proposition S16 says that if it is possible that all observations come from the first K individuals, an information cascade will occur almost surely. To explain the intuition, let's consider an extreme case where individuals consider a network G with $G_i = \emptyset$ for all i. If G is the true network, then no one observes any prior actions, so each action reflects only the individual's private signal, and actions are independent. In this case, the informativeness of each action will not diminish as the line grows, so a cascade will take place after finite actions. Following the paper's arguments, we can show that the cascade force introduced by G cannot be offset by other networks, so an information cascade always occurs as long as individuals consider G as possible.

One may wonder if cascades occur only when individuals consider small networks, i.e., when K is small. The following corollary shows that a cascade can still occur even if individuals consider arbitrarily large networks.

Corollary S2. *Suppose that there is some $G \in \mathcal{G}_0$ under which the first K actions are publicly observable, i.e.,*

$$G_i = \{1, 2, ..., K \wedge i\} \quad \forall i.$$

Then for all $K < \infty$, an information cascade occurs $\mathbb{P}^*$-almost surely.

It says that a cascade will occur almost surely as long as it is possible that only the first finite number of actions are publicly observable. Corollary S2 implies that non-cascade outcomes are not robust with respect to network ambiguity in the following sense:

Example S3. Let G^K be the network in Corollary S2, i.e., the first K individuals are observable. Suppose that individuals consider the following set of networks:

$$\mathcal{G}_n = \left\{G^K : K \geq n\right\},$$

which means that at least the first n individuals are publicly observable. Note that $\mathcal{G}_n \supset \mathcal{G}_{n+1} \supset \mathcal{G}_{n+2} \cdots$, and as $n \to \infty$, $\mathcal{G}_n$ converges to the linear network $\{G^\infty\}$. When $n = \infty$, the occurrence of an information cascade depends on the properties of $\mathbb{F}$. However, for all $n < \infty$, an information cascade occurs for any bounded $\mathbb{F}$. It provides another example in which the non-cascade results appear extreme.

A Omitted Proofs in the Supplementary Materials

A.1 Proof of Theorem S1

I first introduce the notion of local instability:

Definition S4. 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 are 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 S2. *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 S3. *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 S17. *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 $\overline{\alpha} \coloneqq \mathcal{P}(\mathbb{F})$, $\alpha_{max} \coloneqq \max_{F \in \mathcal{F}_0} \mathcal{P}(F)$ and $\alpha_{min} \coloneqq \min_{F \in \mathcal{F}_0} \mathcal{P}(F)$. Let F_{max} and F_{min} be the DGPs that attain the maximum and minimum powers, respectively.

Proof. **Proof of Proposition S17 (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,

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

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 $\mathcal{P}(F) \ge \mathcal{P}(\mathbb{F})$, that is, $\alpha_{min} \ge \overline{\alpha}$. 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 $i \ge 0$,

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

As a consequence, when r_0 is sufficiently large,

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

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 S17 (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

$$\mathbb{P}_{r_0}^{*}(H_1) = \prod_{i=1}^{\infty} \left[1 - F^0\left(\frac{1}{r_i}\right)\right] \geq \prod_{i=1}^{\infty} \left[1 - (1 + \varepsilon_1) \times \overline{C} \times \frac{1}{r_i^{\overline{\alpha}}}\right]. \tag{7}$$

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

$$r_{i+1} \geq r_i \times \left(1+\frac{\beta}{r_i^{\alpha_{min}}}\right)^{1/\alpha_{min}} = \left(r_i^{\alpha_{min}}+\beta\right)^{1/\alpha_{min}} \Rightarrow r_i \geq \left(r_0^{\alpha_{min}}+\beta\times i\right)^{1/\alpha_{min}}. \quad (8)$$

Combining (7) and (8), 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 S17 (c) & (d)

The proofs of Proposition S17 (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 $\mathbb{F}^1(x)=O\left(x^{\overline{\alpha}+1}\right)$ as $x\to 0$.[14] 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 S3, Proposition S17 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.[15]

Lemma S4. *For all $i\geq 1$, individual i only 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 a. By definition,

$$V_i(a)=\begin{cases}\frac{\lambda_i\underline{l}_i}{1+\lambda_i\underline{l}_i}a+\frac{1}{1+\lambda_i\underline{l}_i}(1-a) & a\in\left[a^h,a^H\right]\\ \frac{\lambda_i\overline{l}_i}{1+\lambda_i\overline{l}_i}a+\frac{1}{1+\lambda_i\overline{l}_i}(1-a) & a\in\left[a^L,a^l\right]\end{cases}, \tag{9}$$

[14] 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 $\mathbb{F}^1(x)=O\left(x^{\overline{\alpha}+1}\right)$ as $x\to 0$.

[15] 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.

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

Lemma S5. *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) $A^s = \left\{a^l\right\}$, (ii) $A^s = \left\{a^h\right\}$, and (iii) $A^s = \left\{a^l, a^h\right\}$. Since the logic is parallel across these cases, I focus on the case where $A^s = \left\{a^l\right\}$, i.e., $a^l + a^h > 1/2$. 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 (9), 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 $p_2 = \mathbb{F}^0\left(\lambda_2^*\right)$. Suppose $a_2 = a^l$. Then:

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

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 S5 holds. Now suppose $a_2 = a^h$. Then:

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

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 he is 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. □

A.3 Proof of Proposition S4

Individual i's minimum expected utility satisfies

$$V_i(a) = \begin{cases} \frac{\lambda_i \bar{l}_i}{1+\lambda_i \bar{l}_i} u(a,1) + \frac{1}{1+\lambda_i \bar{l}_i} u(a,0) & \text{if } u(a,0) > u(a,1) \\ \frac{\lambda_i \underline{l}_i}{1+\lambda_i \underline{l}_i} u(a,1) + \frac{1}{1+\lambda_i \underline{l}_i} u(a,0) & \text{if } u(a,1) > u(a,0) \end{cases}.$$

For individual 1, we have:

$$V_1(a) \to \min_{\theta \in \{0,1\}} u(a,\theta) \quad \text{as } R_0 \to \infty,$$

so we can find R_0 sufficiently large such that

$$a_1 \in \arg\max_{a \in A} \left[\min_\theta u(a,\theta)\right] = A^s$$

for all possible values of λ_1. When λ_1 is bounded, the threshold R_0 is also bounded. If A^s is a singleton, Proposition S4 holds trivially. Suppose that A^s contains two actions, a^l and a^h, and that the minimum utility is obtained in states 0 and 1, respectively. Then it can be verified that:

$$\begin{array}{ll} a_1 = a^l & \text{if } \lambda_1 \begin{array}{c} < \\ > \end{array} \dfrac{\left(u^h - u^l\right) \underline{l}_0 + \sqrt{\left(u^h - u^l\right)^2 \underline{l}_0^2 + 4\left(u^l - u\right)\left(u^h - u\right) \bar{l}_0 \underline{l}_0}}{2\left(u^l - u\right) \bar{l}_0 \underline{l}_0} \equiv \mathcal{X}_0, \\ a_1 = a^h & \end{array}$$

where $u = u\left(a^l, 0\right) = u\left(a^h, 1\right)$, $u^l = u\left(a^l, 1\right)$ and $u^h = u\left(a^h, 0\right)$.

Lemma S6. *When R_0 is sufficiently large,*

$$\underline{l}_i \le \underline{l}_0 \times 2\exp\left|\log\sqrt{\frac{u^h - u}{u^l - u}}\right| \quad \textit{and } \bar{l}_i \ge \bar{l}_0 \times \frac{1}{2}\exp\left(-\left|\log\sqrt{\frac{u^h - u}{u^l - u}}\right|\right).$$

Proof. Let $\rho_{hl} \equiv \sqrt{\frac{u^h - u}{u^l - u}}$. Without loss of generality, suppose that $\rho_{hl} \ge 1$.

Case 1: Suppose that $\rho_{hl} \in (1, \gamma)$. Note that $\mathcal{X}_0 \to \rho_{hl}$ as $R_0 \to \infty$, so individual 1 will choose a^h if her signal $\lambda_1 > \mathcal{X}_0 \approx \rho_{hl}$ and choose a^l otherwise. Suppose that $a_1 = a^h$. Then:

$$\bar{l}_2 = \bar{l}_1 \times \sup_F \frac{1 - F^1(\mathcal{X}_0)}{1 - F^0(\mathcal{X}_0)} = \gamma \times \bar{l}_0$$
$$\underline{l}_2 = \underline{l}_1 \times \inf_F \frac{1 - F^1(\mathcal{X}_0)}{1 - F^0(\mathcal{X}_0)} = \mathcal{X}_0 \times \underline{l}_0,$$

and for sufficiently large R_0,

$$\mathcal{X}_2 \approx \sqrt{\frac{u^h - u}{u^l - u}} \times \frac{1}{\sqrt{\bar{l}_2 \underline{l}_2}} \le \frac{\rho_{hl}}{\sqrt{\gamma \mathcal{X}_0}} \frac{1}{\sqrt{\bar{l}_2 \underline{l}_2}} \le 1.$$

Therefore, if $a_2 = a^h$, we have

$$\underline{l}_3 = \underline{l}_2 \times \inf_F \frac{1 - F^1(\mathcal{X}_2)}{1 - F^0(\mathcal{X}_2)} = \underline{l}_2 = \mathcal{X}_0 \times \underline{l}_0.$$

If $a_3 = a^l$, we have $\underline{l}_3 = \frac{1}{\gamma} \times \underline{l}_2 \le \mathcal{X}_0 \times \underline{l}_0$. Extending the argument to all $i \ge 2$, we obtain

$$\underline{l}_i \le \mathcal{X}_0 \times \underline{l}_0 < 2\exp\left|\log\sqrt{\frac{u^h - u}{u^l - u}}\right| \times \underline{l}_0$$

for sufficiently large R_0. Symmetrically, $\underline{l}_i \ge \frac{1}{2}\exp\left|\log\sqrt{\frac{u^h - u}{u^l - u}}\right| \bar{l}_0$ for sufficiently large R_0.

Case 2: Suppose that $\rho_{hl} = 1$. Then $\mathcal{X}_i = 1/\sqrt{\bar{l}_i \underline{l}_i}$, which reduces to the equilibrium strategy in the main paper. Using the same logic as Case 1, we obtain: $\underline{l}_i \le \underline{l}_0$ and $\bar{l}_i \ge \bar{l}_0$.

Case 3: Suppose that $\rho_{hl} = \gamma$. Then, we need to compare the magnitude between $\mathcal{X}_0$

and γ, when R_0 is sufficiently large. If $\mathcal{X}_0 > \gamma$ for large R_0, individual 1 always chooses a^l regardless of her signal, so an information cascade occurs. In addition, $\underline{l}_i = \underline{l}_0$ and $\bar{l}_i \geq \bar{l}_0$ for all $i \geq 1$. If $\mathcal{X}_0 < \gamma$ for large R_0, the analysis is identical to Case 1.

Case 4: Suppose that $\rho_{hl} > \gamma$. Then individual 1 always chooses a^l, and a cascade occurs with $\underline{l}_i = \underline{l}_0$ and $\bar{l}_i \geq \bar{l}_0$ for all i. □

As a result, when R_0 is sufficiently large, we can ensure that $\underline{l}_i$ remains sufficiently small and $\bar{l}_i$ sufficiently large for all i. Then individuals will eventually choose only from A^s. Furthermore, as shown in the proof, society will settle on a single action with probability one.

A.4 Proof of Proposition S7

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

A.5 Proof of Proposition S9

The proof can be decomposed into the following lemmas:

Lemma S7. *Proposition S9 holds if*

$$\sum_{t=1}^{\infty} \mathbb{F}^0 \left(1/\left(\beta t+1\right)^{1/\underline{\alpha}}\right) < \infty,$$

for some $\beta > 0$, where $\mathbb{F}^{\theta}(x) = \int_{\underline{\alpha}}^{\overline{\alpha}} F(x,\alpha)\mu(\alpha)d\alpha$.

Proof. From the discussion in the paper, we know that an incorrect herd occurs with strictly positive probability (i.e., state 1 is locally stable) if $\sum_{i=1}^{\infty} \mathbb{F}^0(1/r_i) < \infty$, where r_i represents the average public likelihood ratio during the action-1 herd. We also know that: (i) $r_i \to +\infty$ as $i \to \infty$, and (ii) for all i,

$$r_{i+1} \geq \sqrt{\frac{1-F^1\left(1/r_i,\underline{\alpha}\right)}{1-F^0\left(1/r_i,\underline{\alpha}\right)}} \times r_i = \sqrt{G_{\underline{\alpha}}\left(r_i\right)} \times r_i,$$

and (iii) $\sqrt{G_{\underline{\alpha}}(r)} \sim 1 + \frac{1}{2}F^0(1/r) \sim 1 + \frac{1}{2}C(\underline{\alpha}) \times \frac{1}{r^{\underline{\alpha}}}$ as $r \to \infty$. Let $\beta = \frac{C(\underline{\alpha})\underline{\alpha}}{3}$. Then:

$$\begin{aligned}\lim_{r\to\infty} \frac{\sqrt{G_{\underline{\alpha}}(r)}-1}{\left(1+\frac{\beta}{r^{\underline{\alpha}}}\right)^{1/\underline{\alpha}}-1} &= \lim_{r\to\infty} \frac{\sqrt{G_{\underline{\alpha}}(r)}-1}{\frac{\beta}{r^{\underline{\alpha}}}} \times \lim_{r\to\infty} \frac{\frac{\beta}{r^{\underline{\alpha}}}}{\left(1+\frac{\beta}{r^{\underline{\alpha}}}\right)^{1/\underline{\alpha}}-1} \\ &= \frac{\frac{1}{2}C(\underline{\alpha}) \times \frac{1}{r^{\underline{\alpha}}}}{\frac{C(\underline{\alpha})\underline{\alpha}}{3} \times \frac{1}{r^{\underline{\alpha}}}} \times \underline{\alpha} = \frac{3}{2} > 1.\end{aligned}$$

So, for sufficiently large I, we have $\sqrt{G_{\underline{\alpha}}(r_i)} \geq \left(1+\frac{\beta}{r_i^{\underline{\alpha}}}\right)^{1/\underline{\alpha}}$ for all $i \geq I$, which implies that $r_{I+t} \geq (\beta t+1)^{1/\underline{\alpha}}$ for all $t \geq 1$. Since $\mathbb{F}^0(x)$ is an increasing function, to show $\sum_{i=1}^{\infty} \mathbb{F}^0(1/r_i) < \infty$, it suffices to show that $\sum_{t=1}^{\infty} \mathbb{F}^0\left(1/(\beta t+1)^{1/\underline{\alpha}}\right) < \infty$. Moreover, the local stability of state 1 implies that of state 0, so we can further show that herding occurs almost surely. □

Lemma S8. $\sum_{t=1}^{\infty} \mathbb{F}^0\left(1/(\beta t+1)^{1/\underline{\alpha}}\right) < \infty$.

Proof. Under the assumption that $\mu(\alpha) \leq C \times (\alpha-\underline{\alpha})^k$ as $\alpha \to \underline{\alpha}$, there exists some $\varepsilon > 0$ such that

$$\mathbb{F}^0(x) \leq C \times \int_{\underline{\alpha}}^{\underline{\alpha}+2\varepsilon} F^0(x,\alpha)(\alpha-\underline{\alpha})^k d\alpha + \int_{\underline{\alpha}+2\varepsilon}^{\overline{\alpha}} F^0(x,\alpha)\mu(\alpha)d\alpha.$$

Since $F^0(x,\alpha) \sim C(\alpha) \times x^{\alpha}$, to establish the convergence of $\sum_{t=1}^{\infty} \mathbb{F}^0\left(1/(\beta t+1)^{1/\underline{\alpha}}\right)$, it suffices to establish the convergence of the following two infinite series:

$$S_1 = \sum_{t=1}^{\infty}\left[\int_{\underline{\alpha}}^{\underline{\alpha}+2\varepsilon} \frac{(\alpha-\underline{\alpha})^k}{(\beta t+1)^{\alpha/\underline{\alpha}}} d\alpha\right] \quad \text{and} \quad S_2 = \sum_{t=1}^{\infty}\left[\int_{\underline{\alpha}+2\varepsilon}^{\overline{\alpha}} \frac{1}{(\beta t+1)^{\alpha/\underline{\alpha}}}\mu(\alpha)d\alpha\right].$$

(i) The convergence of S_2. Let's first establish the convergence of S_2. First note that

$$\begin{aligned}
\lim_{t\to\infty} \frac{\int_{\underline{\alpha}+2\varepsilon}^{\overline{\alpha}} \frac{1}{(\beta t+1)^{\alpha/\underline{\alpha}}}\mu(\alpha)d\alpha}{\frac{1}{(\beta t+1)^{\frac{\underline{\alpha}+\varepsilon}{\underline{\alpha}}}}} &= \lim_{t\to\infty} \int_{\underline{\alpha}+2\varepsilon}^{\overline{\alpha}} (\beta t+1)^{-\frac{\alpha-(\underline{\alpha}+\varepsilon)}{\underline{\alpha}}} \mu(\alpha)d\alpha \\
&= \int_{\underline{\alpha}+2\varepsilon}^{\overline{\alpha}} \lim_{t\to\infty} (\beta t+1)^{-\frac{\alpha-(\underline{\alpha}+\varepsilon)}{\underline{\alpha}}} \mu(\alpha)d\alpha \\
&= 0.
\end{aligned}$$

where the second equality is implied by the dominated convergence theorem (since $(\beta t+1)^{-\frac{\alpha-(\underline{\alpha}+\varepsilon)}{\underline{\alpha}}} \leq 1$ for all $\alpha \in [\underline{\alpha}+2\varepsilon, \overline{\alpha}]$). Therefore, we can find some $T > \infty$ such that

$$\sum_{t\geq T}^{\infty}\left[\int_{\underline{\alpha}+2\varepsilon}^{\overline{\alpha}} \frac{1}{(\beta t+1)^{\alpha/\underline{\alpha}}}\mu(\alpha)d\alpha\right] < \sum_{t\geq T}^{\infty} \frac{1}{(\beta t+1)^{\frac{\underline{\alpha}+\varepsilon}{\underline{\alpha}}}}.$$

Since $\frac{\underline{\alpha}+\varepsilon}{\underline{\alpha}} > 1$, we know that $\sum_{t\geq T}^{\infty} \frac{1}{(\beta t+1)^{\frac{\underline{\alpha}+\varepsilon}{\underline{\alpha}}}}$ converges, which establishes the convergence of S_2.

(ii) The convergence of S_1. Let's now prove the convergence of S_1. Let $x = (\beta t+1)^{-1/\underline{\alpha}}$ and consider the integral $I(x) = \int_{\underline{\alpha}}^{\overline{\alpha}} x^{\alpha} \times (\alpha - \underline{\alpha})^k \, d\alpha$. It can be verified that

$$I(x) = \underbrace{\frac{-x^{\underline{\alpha}}}{(-\log x)^{k+1}} \Gamma\left(k+1, -\log(x) \times (\overline{\alpha} - \underline{\alpha})\right)}_{I_1(x)} + \underbrace{\frac{x^{\underline{\alpha}}}{(-\log x)^{k+1}} \Gamma(k+1)}_{I_2(x)},$$

where $\Gamma(m,n)$ denotes the upper incomplete Gamma function, i.e., $\Gamma(m,n) = \int_n^{\infty} t^{m-1} \times e^{-t} dt$. The gamma function, $\Gamma(m)$, corresponds to the special case in which $n=0$. To establish the convergence of S_1, we need to establish the convergence of

$$\mathcal{I}_1 = \sum_{t=1}^{\infty} I_1\left(\frac{1}{(\beta t+1)^{1/\underline{\alpha}}}\right) \quad \text{and} \quad \mathcal{I}_2 = \sum_{t=1}^{\infty} I_2\left(\frac{1}{(\beta t+1)^{1/\underline{\alpha}}}\right).$$

(ii.a) The convergence of $\mathcal{I}_2$ is straightforward, since

$$\mathcal{I}_2 = \underline{\alpha}^{k+1} \Gamma(k+1) \times \sum_{t=1}^{\infty} \frac{1}{(\beta t+1) \times \log^{k+1}(\beta t+1)} < \infty.$$

This employs the fact that $\sum \frac{1}{n \times \log^s n}$ converges if $s > 1$ (and diverges when $s \leq 1$).

(ii.b) Let's then investigate the convergence of $\mathcal{I}_1$. The idea is to bound the gamma function using a simpler function $-x^{\epsilon} \log(x)$. First, note that when $\epsilon > 0$ is sufficiently small, $\Gamma(k+1, -\log(x) \times (\overline{\alpha} - \underline{\alpha}))$ is an infinitesimal of higher order than $x^{\epsilon} \log(x)$.

$$\begin{aligned}
\lim_{x \to 0} \frac{\Gamma(k+1, -\log(x) \times (\overline{\alpha} - \underline{\alpha}))}{-x^{\epsilon} \log(x)} &= \lim_{x \to 0} \frac{\int_{-\log(x) \times (\overline{\alpha} - \underline{\alpha})}^{\infty} u^k \times e^{-u} du}{-x^{\epsilon} \log(x)} \\
&= \lim_{x \to 0} \frac{(-\log x)^k (\overline{\alpha} - \underline{\alpha})^{k+1} x^{\overline{\alpha} - \underline{\alpha} - 1}}{-x^{\epsilon - 1} - \epsilon x^{\epsilon - 1} \log x} \\
&= \lim_{x \to 0} \frac{(-\log x)^k (\overline{\alpha} - \underline{\alpha})^{k+1} x^{\overline{\alpha} - \underline{\alpha} - \epsilon}}{-1 - \epsilon \log x} \\
&= 0,
\end{aligned}$$

where the second equality comes from L'Hopital's rule. I define an alternative infinite series as below

$$\overline{I}_1(x) = \frac{x^{\underline{\alpha}}}{(-\log x)^{k+1}} x^{\epsilon} \log(x) \quad \text{and} \quad \overline{\mathcal{I}}_1 = \sum_{t=1}^{\infty} \overline{I}_1\left(\frac{1}{(\beta t+1)^{1/\underline{\alpha}}}\right).$$

Since $\Gamma(k+1, -\log(x) \times (\overline{\alpha} - \underline{\alpha}))$ is an infinitesimal of higher order than $x^{\epsilon} \log(x)$, we know that $\mathcal{I}_1$ converges if $\overline{\mathcal{I}}_1$ converges. Notice that

$$\overline{\mathcal{I}}_1 = \sum_{t=1}^{\infty} \frac{\underline{\alpha}^k}{(\beta t+1)^{\frac{\alpha+\epsilon}{\underline{\alpha}}} \times \log^k(\beta t+1)} < \infty,$$

where the convergence comes from the fact that $\sum \frac{1}{n^{s_1} \times \log^k n}$ converges if $s_1 > 1$. Therefore, $\mathcal{I}_1$ converges, so does S_1. □

A.6 Proof of Theorem S2

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

Lemma S9. *For all $i \geq 1$, suppose $a_1 = .. = a_i = 1$, then $\bar{l}_{i+1} = \gamma$ 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 S7 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 $\lambda_{k+1} \in \left[1/\bar{l}_{k+1}, 1/\underline{l}_{k+1}\right] = [1/\gamma, 1]$, 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 $\bar{l}_{k+1} = \bar{l}_k = \gamma$ and $\underline{l}_{k+1} = \underline{l}_k = 1$.

□

Lemma S10. *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 S9. Let $F_\emptyset$ denote the uninformative signal structure and F_γ denote the most informative signal structure, i.e., $\text{supp}(F_\emptyset) = \{1\}$ and $\text{supp}(F_\gamma) = \{1/\gamma, \gamma\}$. Therefore, we have $\bar{l}_{i+2} = \bar{l}_{i+1} \times \frac{F^1_\emptyset(1)}{F^0_\emptyset(1)} = \bar{l}_{i+1} = \gamma$ and $\underline{l}_{i+2} = \underline{l}_{i+1} \times \frac{F^1_\gamma(1)}{F^0_\gamma(1)} = \underline{l}_{i+1} \times 1/\gamma = 1/\gamma$. □

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 S11. *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 $\mathbb{P}^*(\mathcal{H}^{mix}) = 1$. ☐

A.7 Proof of Proposition S11

Proof. Suppose $r_i \in (1/\gamma, \gamma)$, i.e., an information cascade hasn't emerged. When $a_i = 1$, the updating rule of r_i satisfies:

$$r_{i+1} = \sqrt{\sup_{G \in \mathcal{G}_0} \frac{\mathbb{P}_G^1(a_i = 1|h_i)}{\mathbb{P}_G^0(a_i = 1|h_i)} \times \inf_{G \in \mathcal{G}_0} \frac{\mathbb{P}_G^1(a_i = 1|h_i)}{\mathbb{P}_G^0(a_i = 1|h_i)}} \times r_i, \tag{10}$$

where

$$\frac{\mathbb{P}_G^1(a_i = 1|h_i)}{\mathbb{P}_G^0(a_i = 1|h_i)} = \frac{\mathbb{P}_G^1(s_i = \overline{s}) + \frac{1}{2} \times \mathbb{P}_G^1(s_i \in (\underline{s}, \overline{s}))}{\mathbb{P}_G^0(s_i = \overline{s}) + \frac{1}{2} \times \mathbb{P}_G^0(s_i \in (\underline{s}, \overline{s}))} \geq 1, \tag{11}$$

where the inequality comes from the MLRP, which implies that G^1 first-order stochastically dominates G^0. The lower bound can be obtained when G is atomless at $\underline{s}$, so we have:

$$\inf_{G \in \mathcal{G}_0} \frac{\mathbb{P}_G^1(a_i = 1|h_i)}{\mathbb{P}_G^0(a_i = 1|h_i)} = 1.$$

The supermum likelihood ratio can be obtained when $\text{supp}(G) = \{\underline{s}, \overline{s}\}$, so that the only signal that leads to action 1 is $\overline{s}$, yielding a likelihood ratio of γ. Therefore, when $a_i = 1$, we have $r_{i+1} = \sqrt{\gamma} \times r_i$. Symmetrically, when $a_i = 0$, we have $r_{i+1} = 1/\sqrt{\gamma} \times r_i$. These dynamics satisfy Lemma 3 of the main paper, implying that an information cascade occurs almost surely. ☐

A.8 Proof of Proposition S16

Proof. Let $r(h_i, G_i)$ denote the threshold value for individual i when her observation structure is G_i and the history is h_i; that is, individual i chooses action 1 if $\lambda_i \cdot r(h_i, G_i) \geq 1$, and action 0 otherwise. Define

$$\overline{R}(k) = \max\left\{r(h_k, G_k) : h_k \in \{0,1\}^{k-1} \text{ and } G_k \subset \{1, ..., k-1\}\right\},$$

which denotes the highest possible threshold value for individual k. Similarly, let $\underline{R}(k)$ denote the lowest threshold value for k. Let

$$\overline{K} \equiv \sup\left\{k \in N : \gamma > \overline{R}(k+1) \geq \underline{R}(k+1) > \frac{1}{\gamma}\right\}.$$

It can be verified that $\overline{K} \geq 1$, so the definition is meaningful.[16] Define $K = \min\left\{\overline{K}, M\right\}$, where M is some finite constant. Now consider any individual $i > K$. Suppose $a_i = 1$. Then:

$$\underline{l}_{i+1} = \underline{l}_i \times \min_{G \in \mathcal{G}_0} \frac{1 - F^1\left(\frac{1}{r(h_i, G_i)}\right)}{1 - F^0\left(\frac{1}{r(h_i, G_i)}\right)} \geq \underline{l}_i.$$

Let $\hat{G}$ be a network bounded by K. By definition, all actions after individual K are not observable under $\hat{G}$, so for all $i > K$, we have

$$r\left(h_i, \hat{G}_i\right) = r\left(h_{K+1}, \hat{G}_{K+1}\right).$$

By definition, $r\left(h_{K+1}, \hat{G}_{K+1}\right) \leq \overline{R}(K+1) < \gamma$, so

$$\begin{aligned}
\bar{l}_{i+1} \geq \bar{l}_i \cdot \frac{1 - F^1\left(\frac{1}{r\left(h_i, \hat{G}_i\right)}\right)}{1 - F^0\left(\frac{1}{r\left(h_i, \hat{G}_i\right)}\right)} &= \bar{l}_i \cdot \frac{1 - F^1\left(\frac{1}{r\left(h_{K+1}, \hat{G}_{K+1}\right)}\right)}{1 - F^0\left(\frac{1}{r\left(h_{K+1}, \hat{G}_{K+1}\right)}\right)} \\
> \bar{l}_i \cdot \frac{1 - F^1\left(\frac{1}{\overline{R}(K+1)}\right)}{1 - F^0\left(\frac{1}{\overline{R}(K+1)}\right)} &\equiv \bar{l}_i \cdot \beta.
\end{aligned}$$

[16] To see this, note that when $k = 2$, we have $h_2 \in \{\{0\}, \{1\}\}$, $G_2 \in \{\emptyset, \{1\}\}$. Suppose that $h_2 = \{1\}$, i.e., individual 1 took action 1. If $G_2 = \emptyset$, we have $r(h_2, \emptyset) = 1 \in (1/\gamma, \gamma)$; if $G_2 = \{1\}$, then it becomes the standard model, where $r(h_2, \{1\}) = \frac{1-F^1(1)}{1-F^0(1)}$. Since $\text{supp}(F) = [1/\gamma, \gamma]$, we have $r(h_2, \{1\}) \in (1/\gamma, \gamma)$. The case where $h_2 = \{0\}$ is symmetric, so $r(h_2, G_2) \in (1/\gamma, \gamma)$ for all possible h_2 and G_2.

Since $\overline{R}(K+1) < \gamma$, we have $\beta > 1$. Thus,

$$r_{i+1} = \sqrt{\underline{l}_{i+1}\bar{l}_{i+1}} \geq \sqrt{\beta} \cdot r_i.$$

Similarly, when $a_i = 0$, we have $r_{i+1} \leq \sqrt{1/\beta} \cdot r_i$, so an information cascade occurs with probability 1 from Lemma 3 in the paper. □

A.9 Proof of Corollary S2

Proof. Let G denote the network structure in which only the first K individuals are observable. Let E denote the event that an information cascade does not arise before K. In other words,

$$E = \{s^\infty \in \mathcal{S}^\infty : r(h_K, G_K) \in (1/\gamma, \gamma)\}.$$

Let $E_n = \{s^\infty \in \mathcal{S}^\infty : r(h_K, G_K) \in [1/\gamma + 1/n, \gamma - 1/n]\}$, so $E = \cup_n E_n$. From the proof of Proposition S16, we know that on each E_n, for all $i > K$,

$$r_{i+1} \geq \sqrt{\frac{1 - F^1\left(\frac{1}{\gamma - 1/n}\right)}{1 - F^0\left(\frac{1}{\gamma - 1/n}\right)}} \cdot r_i \equiv \beta \times r_i,$$

when $a_i = 1$; and $r_{i+1} \leq \frac{1}{\beta} \cdot r_i$ when $a_i = 0$. Levy's 0-1 implies that on each E_n, an information cascade occurs except for null events. Note that E is countable union of E_n, so whenever E occurs, an information cascade must also occur except for null events. In addition, when E^c occurs, an information cascade occurs by definition. As a consequence, an information cascade must occur with probability 1. □

Bibliography

Arieli, Itai and Manuel Mueller-Frank (2021) "A general analysis of sequential social learning," *Mathematics of Operations Research*, 46 (4), 1235–1249. 10

Bikhchandani, Sushil, David Hirshleifer, and Ivo Welch (1992) "A theory of fads, fashion, custom, and cultural change as informational cascades," *Journal of political Economy*, 100 (5), 992–1026. 15

Epstein, Larry G and Martin Schneider (2007) "Learning under ambiguity," *Review of Economic Studies*, 74 (4), 1275–1303. 24

Herrera, Helios and Johannes Hörner (2012) "A necessary and sufficient condition for information cascades." 5, 6

Ke, Shaowei and Qi Zhang (2020) "Randomization and ambiguity aversion," *Econometrica*, 88 (3), 1159–1195. 12

Saito, Kota (2015) "Preferences for flexibility and randomization under uncertainty," *American Economic Review*, 105 (3), 1246–1271. 12

Smith, Lones and Peter Sørensen (2000) "Pathological outcomes of observational learning," *Econometrica*, 68 (2), 371–398. 15, 25

Smith, Lones, Peter Norman Sørensen, and Jianrong Tian (2021) "Informational herding, optimal experimentation, and contrarianism," *Review of Economic Studies*, 88 (5), 2527–2554. 6