Supplement to "Sequential Learning under Informational Ambiguity"

Jaden Yang Chen

November 3, 2022

Abstract

This online appendix provides the following materials. Section S1 provides a necessary and sufficient condition for complete learning within power-tail DGPs. Section S2 provides two conditions that are close to necessary and sufficient for information cascades. Section S3 considers a situation in which individuals face ambiguity and are correctly specified with respect to the true measure. Sections S4 and S5 discuss multiple actions and multiple states. Section S7 discusses alternative updating rules. Section S8 discusses an extension in which individuals face ambiguity about the network structure.

JEL Classification: D81; D83; C72

Contents

S1 Conditions for Complete Learning

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

Definition S1. A DGP $F \in \mathcal{F}$ has a *power tail* if there exists some $\alpha > 0$ such that $F^0(x) = O(x^\alpha)$ as $x \to 0$. The power of F, denoted by $\mathcal{P}(F)$, is defined to be α.

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 models and imposes the following assumptions.

Assumption S1. *$\overline{F}$ has a power tail, and $\mathcal{F}_0$ only contains 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 in 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 $\mathcal{P}(F) \geq \mathcal{P}\left(\overline{F}\right)$, and

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

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 higher than the power of the true DGP. On the other hand, some perceived DGP has to be adequately informative in the sense that its power does not exceed that of the true model by 1. Before explaining the intuition, let's see what will happen if 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 $\mathcal{P}(F) < \mathcal{P}\left(\overline{F}\right)$, an incorrect herd occurs with a $\mathbb{P}^*$-strictly positive probability; (ii) if for all $F \in \mathcal{F}_0$, $\mathcal{P}(F) \geq \mathcal{P}\left(\overline{F}\right) + 1$, 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 models considered by individuals are inadequately informative, actions will not converge. This comes from the fact that if individuals underestimate predecessors' informativeness, they are more likely to break a herd, so the 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 non-convergence. To prevent incorrect herding, $\mathcal{F}_0$ must not contain highly informative DGPs, which correspond to Theorem S1 (i). To prevent action non-convergence, $\mathcal{F}_0$ 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 which 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, which is a highly relevant concept for information cascade. Both conditions employ a modified version of the hazard ratio in Herrera and Hörner (2012), which I introduce below.

Definition S2. Denote by $h_F^\theta(x) \equiv \frac{f^\theta(x)}{1-F^\theta(x)}$ and by $H_F(x) \equiv h_F^1(x)/h_F^0(x)$, where $H_F(x)$ is referred to as the *hazard ratio* at x under F. For a model set $\mathcal{F}_0$, denote by

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

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

The average hazard ratio is defined using density functions, and it requires taking the maximum and minimum over the model set. For convenience, I impose the following assumption.

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

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 a $\mathbb{P}^*$-strictly positive probability for some prior $l_0 \in (1/\gamma, \gamma)$ **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 $a_i = 1$

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

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. ☐

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

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 w.r.t. 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 as follows.

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. It 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 w.r.t. to the prior average likelihood ratio. If the IAHRP holds, r_i is trapped in the non-cascade set, so an information cascade will not occur. In other words, for an information cascade to occur, the IAHRP must be violated, which provides a necessary condition for information cascades.

S3 Ambiguous and Correct Model Perceptions

This section considers a special case in which individuals are **ambiguous and correctly specified** in the following sense. Suppose that that $\overline{F}_i$s are i.i.d. distributed according to a second-order distribution, $h \in \Delta(\mathcal{F})$. Individuals only perceive objectively possible models, i.e., $\mathcal{F}_0 = \text{supp}(h)$. In addition, every event of interest is calculated using h, so I refer to the measure induced by h as the true measure. In this case, individuals are correctly specified in that their model perceptions are correct with respect to the true measure. For convenience, I assume that the implied signal distribution, $H^\theta(x) = \mathbb{E}_h F^\theta(x)$, is continuous and has support $[1/\gamma, \gamma]$, where $\gamma > 1$. For subsetions S3.1 and S3.2, individuals have max-min EU preferences.

S3.1 Bounded Signals: Information Cascades

Under ambiguity and correct specification, we have the almost sure occurrence of a cascade as follows.

Corollary S2. *If supp$(h) = \mathcal{F}$, an information cascade occurs h-almost surely.*

Corollary S3. *When signals are bounded, an information cascade occurs h-almost surely as long as there exists some model $F \in$ supp(h) that satisfies conditions in Theorem 2 in the paper.*

Corollaries S2 and S3 show that when there is sufficient ambiguity, or when supp(h) is very large, an information cascade occurs almost surely irrespective of the details of H^θ. We further notice that the conditions for cascades are exactly the same as in the benchmark model, so we can also construct examples where non-cascades represent knife-edge situations.[2]

[2] The main reason is that an information cascade is a finite-time event, so it is affected very little by the extra consistency between the perceived models and the true measure. In the proof, we simply need to replace the ex post distribution with the ex ante one, and everything else remains unchanged.

S3.2 Unbounded Signals: Incorrect Herding

When signals are unbounded, Corollary S2 shows that complete learning breaks down under extreme ambiguity. This subsection further establishes that complete learning can be fragile with respect to small ambiguity. Below, I focus on a situation in which h is defined on a parametric family, $\{F(x,\alpha)\}$, where each model has a **power tail**. Formally,

$$\text{supp}(h) = \{F(x,\alpha), \alpha \in [\underline{\alpha}.\overline{\alpha}] \subset \mathbb{R}_{++}\},$$

and each $F(x,\alpha) \in \text{supp}(h)$ satisfies

$$F^0(x,\alpha) \sim C(\alpha) \times x^{\alpha} \quad \text{as } x \to 0,$$

where $C(\alpha) > c$ for all $\alpha \in [\underline{\alpha}.\overline{\alpha}]$. To simplify notations, I use $h(\alpha)$ to denote the density of $F(x,\alpha)$.

Proposition S3. *Suppose that*

$$h(\alpha) < C \times (\alpha - \underline{\alpha})^k, \quad \text{as } \alpha \to \underline{\alpha}$$

for some $C, k > 0$, then an incorrect herd occurs with a h-strictly positive probability.

Proposition S3 shows that if h decreases to 0 at a sub-polynomial speed as $\alpha \to \underline{\alpha}$, an incorrect herd will occur with a strictly positive probability. The intuition is as follows. Recall that the most "precise" DGP is the one with the thickest tail, that is, the DGP with power $\underline{\alpha}$. If h decreases very fast near $\underline{\alpha}$, it means that the precise DGPs are very rare, so the frequency of precise signals is insufficient to break all incorrect herds, which leads to incomplete learning. Below is an example in which complete learning is not robust in a sense.

Example S1. As in the main paper, I consider an example in which $S = [0,1]$, and

$$g_m^0(s) = (m+1)(1-s)^m \quad \text{and} \quad g_m^1(s) = (m+1)s^m.$$

Each individual's DGP is parametrized by m_i, which is i.i.d. drawn from $M_\varepsilon = [m_0 - \varepsilon, m_0 + \varepsilon]$ according to some distribution h. Suppose that h takes the following form

$$\forall m \in M_\varepsilon : \qquad h(m) = \begin{cases} (m_0 + \varepsilon - m)^a \times (m - m_0 + \varepsilon)^b \times c & \text{when } \varepsilon > 0 \\ \delta_{m_0} & \text{when } \varepsilon = 0 \end{cases},$$

where $a, b > 0$, $c > 0$ is a normalizing constant, and δ_x denote the degenerate distribution. All individuals have max-min EU preference and correctly perceive the feasible model set, M_ε. When $\varepsilon = 0$, the support collapses to a singleton $M = \{m_0\}$, and complete learning obtains almost surely (h). In contrast, for an arbitrarily small $\varepsilon > 0$, complete learning breaks down, and an incorrect herd occurs with a strictly positive probability (h).

S3.3 Smooth Ambiguity Preference

The max-min preference may seem extreme because the worst-case utility may be obtained by a DGP whose actual probability is 0.[3] This subsection discusses the **smooth ambiguity** preference (Klibanoff et al., 2005), in which the utility is not determined by any single DGP.

[3] E.g., in the extreme ambiguity, the utility-minimizing model for breaking a herd is obtained at the perfectly informative DGP, whose ex ante probability is zero. Also, in Proposition S3, the lowest power, $\underline{\alpha}$, has zero probability and density.

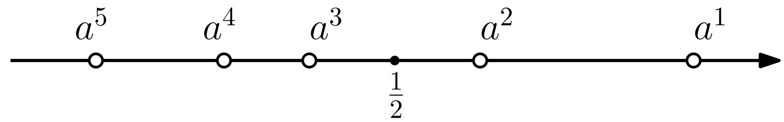

Figure 1: Linear Utility Functions

When signals are bounded, Section 7.2 of Chen (2019) shows that an information cascade occurs with a probability greater than $1-\varepsilon$ for all $\varepsilon>0$ when individuals are sufficiently *ambiguity sensitive*. When signals are **unbounded**, a similar result still holds. Below is an example.

Example S2. (Unbounded signals) Let's consider the setup in Example 1 of the main paper. That is, each individual i receives a signal $s_i \in \{H, L\}$ and has DGP $g_i(s|\theta)$ with

$$\frac{g_i^1(H)}{g_i^1(L)} = \frac{g_i^0(L)}{g_i^0(H)} = \gamma_i \in (1, \infty),$$

where $\gamma_i \overset{I.I.D.}{\sim} f$, and f is continuous and has full-support. Individuals have the following preference

$$V_i(a) = \left[\int_{\Gamma^{i-1}} \left[\mathbb{E}_{\gamma_1,\dots,\gamma_{i-1}} u(a,\theta)\right]^{1-\sigma} df(\gamma_1, \dots, \gamma_{i-1})\right]^{\frac{1}{1-\sigma}}.$$

Further suppose that $f(\gamma) \geq C \times \gamma^{-\alpha}$ for some $C, \alpha > 0$ as $\gamma \to \infty$, i.e., f is controlled by a power function near the upper tail. When $\sigma = 0$, it corresponds to the correct Bayesian case, in which complete learning occurs almost surely (f). When $\sigma \geq \overline{\sigma}$ for some finite $\overline{\sigma}$—i.e., individuals are sufficiently ambiguity-sensitive—an **information cascade** occurs almost surely, and an incorrect cascade occurs with a strictly positive probability (f). The proof can be found in the Appendix.

In Example S2, not only does an incorrect herd occur, but an information cascade also occurs almost surely. Recall that in the max-min preference, an information cascade occurs because individuals use the perfectly informative DGP to evaluate the contrarian decision. This represents an extreme case and may lead to the conjecture that a cascade occurs because individuals essentially perceive a DGP whose probability is actually 0. In contrast, Example S2 shows that a cascade still occurs when every DGP has some influence on the decision. It shows that the cascade result for the unbounded signals is less extreme than it appears in the max-min case.

Remark S1. Example S2 assumes a binary signal space, but it actually nests very general cases, because there is a second-order distribution f.[4] Also, Example S2 focuses on a specific case $\phi(x) = x^{1-\sigma}$, but it is conceivable that the result can be extended to all ϕs that can be approximated by a power function near 0.

S4 Multiple Actions

The paper's results can be extended to the multiple-action space. Furthermore, this section finds that under sufficient ambiguity, (i) at most two actions will be chosen in the limit, and (ii) these two actions must be symmetric in some sense. Therefore, the binary and symmetric action space is in some sense W.L.O.G..

[4]The compound distribution $\int f \times g$ can cover all unbounded distributions that are controlled by a power function and have full support.

S4.1 Linear Utility Function

Suppose that the action space is $A = \left\{a^1, ..., a^k\right\} \subset [0,1]$. 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 max-min EU preference and consider all DGPs on $(0,\infty)$ as possible; (ii) signals are i.i.d. according to $\overline{F}$, and $\overline{F}$ is continuous and has full-support on $(0,\infty)$.[5] 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 stands for the set of actions with the smallest distance to $1/2$.

Proposition S4. $\lim_{t\to\infty} \mathbb{P}^*(a_t \in A^s) = 1$, *that is, the society will only settle on* A^s *in the end.*

The result comes from the fact that when the ambiguity is adequately large, individuals will end up holding highly ambiguous beliefs, which push them to only choose 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** w.r.t. $1/2$. Figure 1, provides an example in which there are two safest actions, a^2 and a^3, and they are equally distanced from $1/2$.

Remark S2. Note that similar result also holds when individuals are ambiguity-loving. For example, when individuals have **max-max EU** preference, the society will settle on the actions with the highest maximum payoff, A^h, where

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

Geometrically, A^h means the actions with largest distance from $1/2$, and it also contains at most two actions. In Figure 1, it represents actions a^1 and a^5. It shows that when there are multiple actions, ambiguity attitude can affect which actions will be chosen in the limit.

S4.2 General Utility Functions

The result can be extended to general utility functions if we also allow for ambiguous priors. It turns out that under sufficient ambiguity w.r.t. both information and states, the society will settle on at most two actions in the end, and these two actions must be symmetric in some sense.

From now on, I assume that: (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 ambiguity about the true state; (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.*

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

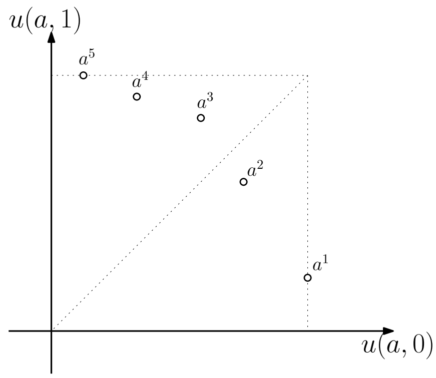

Figure 2: General Utility Functions

The set of safest actions can be similarly defined as follows,

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

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

Proposition S5. *There exists $R \in \mathbb{R} \cup \{+\infty\}$ such that*

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

for all $R_0 \geq R$,, where $R < \infty$ when signals are bounded.

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

I then discuss the **equilibrium strategy** in the multi-action case. In the following, I assume that A^s contains two elements (if A^s is a singleton, the equilibrium strategy becomes trivial in the limit). More specifically, $A^s = \left\{a^l, a^h\right\}$, where a^l achieves its minimum utility at state 0, and a^h achieves its minimum utility at state 1.

Proposition S6. (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 $u^h = u\left(a^l, 0\right)$. 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.

We first notice that if a^l and a^h are also "upper symmetric", i.e., $u^l = u^h$, the equilibrium cutoff becomes

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

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., when $\bar{l}_i$ is very large and $\underline{l}_i$ is very small, we have

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

which only differs from the previous characterization by a constant. As can be seen, the equilibrium characterization in the binary-action case also serves as a good benchmark for multi-action situation when there is sufficient ambiguity. Therefore, an information cascade also arises with probability 1 under sufficient ambiguity.

S5 Multiple States

When there are multiple states, the equilibrium becomes more difficult to characterize, but the key insights still hold.[6] This section shows that in a simple case how 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 a payoff of 0 if otherwise. Every individual has max-min EU preference and updates beliefs according to the full Bayesian rule. The true DGP, $\overline{g_i}$, satisfies that

$$\frac{\overline{g_i}(s|\theta)}{\overline{g_i}(s|\theta')} \in \left[\frac{1}{\gamma}, \gamma\right], \quad \forall \theta \neq \theta' \text{ and } \forall s \in S,$$

In the following, I simply refer to $\left[\frac{1}{\gamma}, \gamma\right]$ as the support of the DGP. In the below, I consider a specific class of perceptions and show that large ambiguity can produce cascades.

Assumption S6. *Individuals consider all DGPs with support in* $\left[\frac{1}{R\gamma}, R\gamma\right]$, *where* $R \geq 1$.

When $R = 1$, individuals correctly specify the support of the signals as in the paper. When R gets larger, it corresponds to a higher degree of ambiguity. The following proposition shows that under sufficient large ambiguity, an information cascade occurs almost surely.

Proposition S7. *There exists* $R_0 < \infty$ *such that an information cascade occurs* $\mathbb{P}^*$*-almost surely for all* $R \geq R_0$.

Proof. Suppose that $a_1 = \theta_1$. Then we have

$$\overline{g_1}(s_1|\theta_1) / \overline{g_1}(s_1|\theta') \geq 1 \quad \forall \theta' \in \Theta.$$

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

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

$$\min_{\pi\in\Pi_2}\sum_{\theta}\frac{\pi(\theta)}{\pi(\theta')}\times\frac{\underline{g_2}(s_2|\theta)}{\underline{g_2}(s_2|\theta')} > \min_{\pi\in\Pi_2}\sum_{\theta}\frac{\pi(\theta)}{\pi(\theta_1)}\times\frac{\overline{g_2}(s_2|\theta)}{\underline{g_2}(s_2|\theta_1)}. \tag{1}$$

Notice that

$$\text{L.H.S of (1)} = \min_{\pi\in\Pi_2}\left(\frac{\pi(\theta_1)}{\pi(\theta')}\times\frac{\underline{g_2}(s_2|\theta_1)}{\overline{g_2}(s_2|\theta')}+\sum_{\theta\neq\theta_1,\theta'}\frac{\pi(\theta)}{\pi(\theta')}\times\frac{\underline{g_2}(s_2|\theta)}{\overline{g_2}(s_2|\theta')}+1\right)\geq\frac{\underline{g_2}(s_2|\theta_1)}{\overline{g_2}(s_2|\theta')}+\frac{K-1}{R\gamma^2}+1.$$

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. □

S6 Heterogeneous Ambiguity

This section discusses how to extend the paper's main results to heterogeneous ambiguity. Recall that in the paper, individuals share a common set of models $\mathcal{F}_0$. This assumption implies two aspects of homogeneity: (i) individuals' signal structures seem homogeneously ambiguous to others and (ii) individuals are homogeneously ambiguous about others' signal structures. Below, I discuss how my results can be relaxed in these two directions.

S6.1 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 perceive 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; all types are commonly known. I also assume 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 S8. *When there is sufficient ambiguity for high-ambiguity individuals, e.g., when $\mathcal{F}^H$ satisfies the conditions in Theorem 2 (Chen, 2019), an information cascade occurs $\mathbb{P}^*$-almost surely.*

Proof. The proof of Theorem 2 shows that if $\mathcal{F}^H$ satisfies the conditions in Theorem 2, 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 if a cascade has not occurred. Because H-type individuals have bounded distance, we can find a constant $K<\infty$ such that K identical actions can trigger a cascade. Following a similar argument, we can establish the almost sure occurrence of a cascade. □

Notice that the proposition imposes no restriction on the fraction of high-ambiguity individuals, so an information cascade can emerge even when there are an ε-fraction of high-ambiguity individuals.[7] Also, the proposition imposes no restriction on $\mathcal{F}^L$. If we take $\mathcal{F}^L$ to be the true model, 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.

[7] 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 $\lim_{k\to\infty}\frac{\sum_{i\leq k}1_{\{t_i=H\}}}{k}\to 1/n$, which can be an arbitrarily small number.

S6.2 Individuals are heterogeneously ambiguous about others

Still suppose that there are two types, $t_i \in \{H, L\}$, and that individuals with type t hold a belief set $\mathcal{F}^t$ about other DGPs, where $\mathcal{F}^L \subset \mathcal{F}^H$. Here, we can think of the L-type individuals as more knowledgeable in the sense that they entertain a smaller set of possible DGPs about others.

Proposition S9. *If both types of individuals are sufficiently ambiguous, i.e., when both $\mathcal{F}^L$ and $\mathcal{F}^H$ satisfy the conditions in Theorem 2 (Chen, 2019), an information cascade occurs with a $\mathbb{P}^*$-strictly positive probability.*

Proof. Let r_i^t denote the average public likelihood ratio after history h_i if individual i were of 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 r_i must also enter the cascade set. Therefore, an information cascade occurs with a strictly positive probability. □

The proposition shows that qualitative result in the paper still holds—an information cascade can occur under sufficient, but not necessarily homogeneous, ambiguity.

S7 The α-Maximum Likelihood Rule

The occurrence of a cascade is not unique to the full Bayesian rule. In the main text, I showed that under the smooth ambiguity model, an information cascade can occur under *any* updating rule of the second-order belief that generates a full-support posterior. This section focuses on the max-min model and investigates 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 $\alpha \in [0,1]$, $F_{-i} \equiv (F_1, ..., F_{i-1})$, and $\mathcal{F}_{-i} \mid h_i$ denotes the set of predecessors' models after history h_i. Under this updating rule, individuals only entertain the models that pass some likelihood test, where $\alpha = 1$ corresponds to the maximum likelihood updating, and $\alpha = 0$ corresponds to the full Bayesian updating.

Proposition S10. *Suppose that $\mathcal{F}_0 = \mathcal{F}$ and signals are bounded. Under α-MLE, an information cascade occurs with strictly positive 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 an action profile $a_1 = a_2 = ... = a_{i-1} = 1$, and denote by $F_{-i}^* = \left(F_1^*, ..., F_{i-1}^*\right) \in \arg\max \mathbb{P}_{F_{-i}}\left(h_i\right)$; i.e., the DGPs that maximize the probability of history h_i.[8] With some abuse of notation, define $F_{-i} \equiv \left(F_1^*, ..., F_{i-2}^*, F_{i-1}\right)$. 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,

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

[8] 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 $\mathbb{P}_{F_2}\left(a_2|a_1\right) = ... = \mathbb{P}_{F_{-i}}\left(a_{i-1}|a_1, a_2, ..., a_{i-2}\right) = 1$.

which implies that

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

When $h_{i-1}=\{1,...,1\}$, we have $\mathbb{P}_{F_{-i}}\left(\theta=1|h_{i-1}\right)\geq\mathbb{P}_{F_{-i}}\left(\theta=0|h_{i-1}\right)$ for all $F_{-i}\in\mathcal{F}_{-i}$; also, since $a_{i-1}=1$,we have $\mathbb{P}_{F_{-i}}\left(a_{i-1}|h_{i-1};\theta=1\right)\geq\mathbb{P}_{F_{-i}}\left(a_{i-1}|h_{i-1};\theta=0\right)$.[9] As a consequence,

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

so inequality (2) holds if

$$\mathbb{P}_{F_{-i}}\left(a_{i-1}|h_{i-1};\theta=0\right)\frac{1}{2}+\mathbb{P}_{F_{-i}}\left(a_{i-1}|h_{i-1};\theta=1\right)\frac{1}{2}\geq\alpha \tag{3}$$

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

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

Then (3) 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 $\alpha=1$, the maximum likelihood updating (MLU). This is because the MLU can lead to an "over-fitting problem". Under the MLU, individuals can just keep uninformative DGPs, because a herd occurs with probability 1 when all followers have no information. As a consequence, beliefs stop updating after the first person during a herd, so an information cascade usually does not occur.

S8 Ambiguity over the Network Structure

The discussion can be extended to **ambiguous networks**. This section shows that when individuals are ambiguous about other people's observation structures, and when the ambiguity is sufficiently large, an information cascade occurs almost surely for all bounded signals.[10]

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 are ambiguous about the network structure. Let $\mathcal{G}$ represents the set of all possible network structures. Let

[9] 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)

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

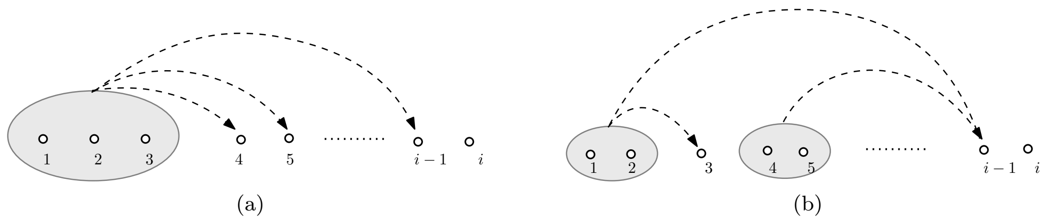

Figure 3: Ambiguous Networks

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

$\mathcal{G}_0 \subset \mathcal{G}$ denote the set of network structures perceived by the 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 $\overline{F}$, where $\overline{F}$ is continuous and has full support on $[1/\gamma, \gamma]$ with $\gamma \in (1, \infty)$. To emphasize the effect of network ambiguity, I assume that individuals correctly understand $\overline{F}$, i.e., there is no ambiguity about 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. Lemma S1 requires extreme ambiguity about the network, but we actually 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 finite number of individuals are observable to the 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 from $\{1, 2\}$ and $\{4, 5\}$.

Proposition S11. *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.*

Proposition S11 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, every individual observes no previous action, so all actions perfectly reflect private signals, and hence 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 can not 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., K is small. The following corollary shows that a cascade can still occur even if individuals consider an arbitrarily large network.

Corollary S4. *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 the first finite number of actions are publicly observable. Corollary S4 implies that non-cascade is not robust w.r.t. network ambiguity in the following sense.

Example S3. Let G^K be the network in Corollary S4, that is, 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. Notice that $\mathcal{G}_n \supset \mathcal{G}_{n+1} \supset \mathcal{G}_{n+2} \cdots$, and as $n \to \infty$, $\mathcal{G}_n$ is approaching the linear network $\{G^\infty\}$. When $n = \infty$, the occurrence of an information cascade depends on the properties of $\overline{F}$. However, for all $n < \infty$, an information cascade occurs for all possible bounded $\overline{F}$s. It provides another example in which the non-cascade result is extreme in some sense.

Bibliography

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

Chen, Jaden Yang (2019) "Sequential learning under informational ambiguity," *Available at SSRN 3480231*. 7, 11, 12, 16

Epstein, Larry G and Martin Schneider (2007) "Learning under ambiguity," *Review of Economic Studies*, 74 (4), 1275–1303. 12

Herrera, Helios and Johannes Hörner (2012) "A necessary and sufficient condition for information cascades." 4, 5

Klibanoff, Peter, Massimo Marinacci, and Sujoy Mukerji (2005) "A smooth model of decision making under ambiguity," *Econometrica*, 73 (6), 1849–1892. 6

Smith, Lones and Peter Sørensen (2000) "Pathological outcomes of observational learning," *Econometrica*, 68 (2), 371–398. 13

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

A Appendix: Omitted Proofs

A.1 Proof of Theorem S1

I first introduce the notion of local unstable as below.

Definition S4. State 0 (or state 1) is ***locally unstable*** if there is some $r \in \mathbb{R}_{++}$ 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 Π_0 with $r_0 < r$ (or $r_0 > r$).

In other words, state θ are locally unstable if posteriors will escape from a small neighborhood around δ_θ almost surely, where beliefs are described by the average likelihood ratio. We have the following lemmas.

Lemma S2. *Complete learning occurs if and only if $r_i \to 0$ with probability 1.*

Proof. First, during complete learning, there must be a herd of action 0 after some point, so $r_i \to 0$ with probability 1. Second, if $r_i \to 0$ with probability 1, a herd of action 0 will eventually occur from Lemma 4 in the paper. □

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 action 0s lead to $r_i < r$. Following the argument in Lemma 4 in Chen (2019), the probability of $r_i \to 0$ is greater than some positive constant so for all history h_i, so complete learning occurs by the Levy's 0-1 Law. □

We then have the following proposition.

Proposition S12. *Under Assumptions S1 and S2, we have:*

(a) if for all $F \in \mathcal{F}_0$, $\mathcal{P}(F) \geq \mathcal{P}\left(\overline{F}\right)$, state 1 is locally unstable;
(b) if there exists some $F \in \mathcal{F}_0$ such that $\mathcal{P}(F) < \mathcal{P}\left(\overline{F}\right)$, state 1 is locally stable;
(c) if for all $F \in \mathcal{F}_0$, $\mathcal{P}(F) \geq \mathcal{P}\left(\overline{F}\right) + 1$, state 0 is locally unstable;
(d) if there exists some $F \in \mathcal{F}_0$ such that $\mathcal{P}(F) < \mathcal{P}\left(\overline{F}\right) + 1$, state 0 is locally stable.

For simplicity, I denote by $\overline{\alpha} \coloneqq \mathcal{P}\left(\overline{F}\right)$, $\alpha_{max} \coloneqq \max_{F\in\mathcal{F}_0} \mathcal{P}(F)$ and $\alpha_{min} \coloneqq \min_{F\in\mathcal{F}_0} \mathcal{P}(F)$. The DGPs with the maximum and minimum power are denoted by F_{max} and F_{min}.

Proof. **Proof of Proposition S12 (a):** Given r_0, the probability of a herd of action 1 is

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

where r_i represents the average likelihood ratio after $h_i = (1, 1, ..., 1)$. The probability is equal to 0 if and only if $\sum \overline{F}^0\left(\frac{1}{r_i}\right) = \infty$, or equivalently, $\sum \frac{1}{r_i^{\overline{\alpha}}} = \infty$. Note that $\{r_i\}$ is determined by the following dynamics

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

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

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

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

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

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

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

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, which implies that state 1 is unstable.

Proof of Proposition S12 (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

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

In order to establish local stability, we need to find a *uniform* lower bound of the probability on the RHS for all large r_0s.

Suppose that $\overline{F}^0(x) \sim \overline{C} \times x^{\overline{\alpha}}$ 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 $\frac{\overline{F}^0(1/r_i)}{\overline{C}\times(1/r_i)^{\overline{\alpha}}} \in [1-\varepsilon_1, 1+\varepsilon_1]$ for some $\varepsilon_1 > 0$, so

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

Here, we also want R to be sufficiently large such that the infinite product on the RHS is strictly positive.

On the other hand, recall that

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

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

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

When R sufficiently large, we have

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

Combining (4) and (5), we obtain

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

for all $r_0 \geq R$. Again, R is chosen to be sufficiently large such that each term is strictly positive. Suppose that there exists some $F \in \mathcal{F}_0$ such that $\mathcal{P}(F) < \mathcal{P}\left(\overline{F}\right)$, which 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 S12 (c) & (d)

The proofs of Proposition S12 (c) and (d) are almost identical to the proofs of (a) and (b). The only difference is that the cutoff value becomes $\mathcal{P}\left(\overline{F}\right)+1$. To see why we have a different cutoff value, note that the probability of an action-0 herd is

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

where r_i denotes the average likelihood ratio after $h_i=(0,...,0)$. An action-0 herd occurs with a strictly positive probability if and only if $\sum\overline{F}^{1}\left(r_i\right)<\infty$. During a herd of action 0, we have $r_i\to 0$; besides, it can be verified that $\overline{F}^{1}\left(x\right)=O\left(x^{\overline{\alpha}+1}\right)$ as $x\to 0$. [11]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 rest of the proofs are exactly symmetric to those of (a) and (b). □

From Lemma S3, Proposition S12 implies Theorem S1, so the theorem is proved.

A.2 Proof of Proposition S3

Let $\{r_1,r_2,...\}$ be the of average likelihood ratios when all individuals take action 1. Let $H^{\theta}\left(x\right)$ be the ex ante distribution of signals, that is, $H^{\theta}\left(x\right)=\int_{\mathcal{A}}F^{\theta}\left(x,\alpha\right)h\left(\alpha\right)d\alpha$. The proof can be decomposed into the following lemmas.

Lemma S4. *Proposition S3 holds if* $\sum_{t=1}^{\infty}H^{0}\left(1/\left(t+1\right)^{1/\underline{\alpha}}\right)<\infty$.

Proof. Based on the previous discussion, an incorrect herd occurs with a strictly positive probability if $\sum_{i=1}^{\infty}H^{0}\left(1/r_i\right)<\infty$. We also know that: (i) $r_i\to+\infty$ as $i\to\infty$, (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) $\sqrt{G_{\underline{\alpha}}\left(r\right)}\sim 1+\frac{1}{2}F^{0}\left(1/r\right)\sim 1+\frac{1}{2}C\left(\underline{\alpha}\right)\times\frac{1}{r^{\underline{\alpha}}}$ as $r\to\infty$. Let $\beta=\frac{C(\underline{\alpha})\underline{\alpha}}{3}$, and we have

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

So, for sufficiently large I, we have $\sqrt{G_{\underline{\alpha}}\left(r_i\right)}\geq\left(1+\frac{\beta}{r_i^{\underline{\alpha}}}\right)^{1/\underline{\alpha}}$ for all $i\geq I$, which implies that $r_{I+t}\geq\left(\beta t+1\right)^{1/\underline{\alpha}}$ for all $t\geq 1$. Because $H^{0}(x)$ is an increasing function, so to show $\sum_{i=1}^{\infty}H^{0}\left(1/r_i\right)<\infty$, it suffices to show that $\sum_{t=1}^{\infty}H^{0}\left(1/\left(\beta t+1\right)^{1/\underline{\alpha}}\right)<\infty$. □

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

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

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

Lemma S5. $\sum_{t=1}^{\infty} H^0\left(1/\left(\beta t+1\right)^{1/\underline{\alpha}}\right) < \infty$.

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

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

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

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

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

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

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

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

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

(ii) The convergence of S_1. Let's then show 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 where $n = 0$. To show the convergence of S_1, we need to show 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

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

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 $-x^{\epsilon}\log\left(x\right)$. First, note that when $\epsilon > 0$ is sufficiently small, $\Gamma\left(k+1, -\log\left(x\right) \times \left(\overline{\alpha} - \underline{\alpha}\right)\right)$ is an infinitesimal of higher order than $x^{\epsilon}\log\left(x\right)$.

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

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

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

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

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

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. □

Combining these two lemmas, the proposition is proved.

A.3 Proof of the Claim in Example S2

Proof. Suppose that $a_1 = 1$, and that individual 2 received an opposite signal, and her signal precision is γ_2. Her utility of each action is

$$V_2\left(0\right) = \left[\int_1^{\infty} \left[\frac{\gamma_2}{\gamma_1 + \gamma_2}\right]^{1-\sigma} f\left(\gamma_1\right) d\gamma_1\right]^{\frac{1}{1-\sigma}}, \quad V_2\left(1\right) = \left[\int_1^{\infty} \left[\frac{\gamma_1}{\gamma_1 + \gamma_2}\right]^{1-\sigma} f\left(\gamma_1\right) d\gamma_1\right]^{\frac{1}{1-\sigma}}.$$

By assumption, there exists $R<\infty$ such that $f(\gamma)\geq \frac{C}{2}\gamma^{-\alpha}$ for all $\gamma\geq R$. For all $\sigma>\alpha+1$, we have

$$\begin{aligned} V_2(0) &= \left[\int_1^R \left[\frac{\gamma_2}{\gamma_1+\gamma_2}\right]^{1-\sigma} f(\gamma_1)\,d\gamma_1 + \int_R^\infty \left[\frac{\gamma_2}{\gamma_1+\gamma_2}\right]^{1-\sigma} f(\gamma_1)\,d\gamma_1\right]^{\frac{1}{1-\sigma}} \\ &\leq \left[\int_1^R \left[\frac{\gamma_2}{\gamma_1+\gamma_2}\right]^{1-\sigma} f(\gamma_1)\,d\gamma_1 + \int_R^\infty \left[\frac{\gamma_2}{\gamma_1+\gamma_2}\right]^{1-\sigma} \frac{C}{2}\gamma_1^{-\alpha}d\gamma_1\right]^{\frac{1}{1-\sigma}} \\ &= \left[M + \frac{C\gamma_2^{1-\sigma}}{2}\times\int_R^\infty \frac{(\gamma_1+\gamma_2)^{\sigma-1}}{\gamma_1^{\alpha}}d\gamma_1\right]^{\frac{1}{1-\sigma}}, \end{aligned}$$

where $M\equiv\int_1^R\left[\frac{\gamma_2}{\gamma_1+\gamma_2}\right]^{1-\sigma} f(\gamma_1)\,d\gamma_1<\infty$. Since $\sigma>\alpha+1$, we have $\frac{(\gamma_2\gamma_1+1)^{\sigma-1}}{\gamma_1^{\alpha}}\to\infty$ as $\gamma_1\to\infty$, so $\int_R^\infty\frac{(\gamma_1+\gamma_2)^{\sigma-1}}{\gamma_1^{\alpha}}d\gamma_1=\infty$, and hence $V_2(0)=0$. Also,

$$V_2(1) = \left[\int_1^\infty \left[\frac{\gamma_1}{\gamma_1+\gamma_2}\right]^{1-\sigma} f(\gamma_1)\,d\gamma_1\right]^{\frac{1}{1-\sigma}} \geq \frac{1}{1+\gamma_2} > 0,$$

so $V_2(1)>V_2(0)$ for all signals and for all $\gamma_2\in(1,\infty)$, which implies that individual 2 will follow regardless of her private signal—i.e., an information cascade occurs. The argument is symmetric for $a_1=0$, so a cascade occurs almost surely, and an incorrect cascade occurs with a strictly positive probability. □

A.4 Proof of Proposition S4

Without loss of generality, I index all actions in descending order, i.e., $a^1>a^2>...>a^k$. The proof focuses on the situation in which $a^k<1/2<a^1$ because the case in which all actions belong to one side of 1/2 is a simple extension of this benchmark. I 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.[12]

Lemma S6. *For all $i\geq 1$, individual i only will a.s. choose from $A^*=\{a^L,a^H,a^l,a^h\}$.*

Proof. Let $V_i(a)$ denote the minimum expected utility of individual i if she chose action a. By definition,

$$V_i(a)=\begin{cases}\frac{\lambda_i\underline{l}_i}{1+\lambda_i\underline{l}_i}a+\frac{1}{1+\lambda_i\underline{l}_i}(1-a) & a\in\left[a^h,a^H\right]\\ \frac{\lambda_i\bar{l}_i}{1+\lambda_i\bar{l}_i}a+\frac{1}{1+\lambda_i\bar{l}_i}(1-a) & a\in\left[a^L,a^l\right]\end{cases}. \tag{6}$$

Notice that $V_i(a)$ is a piecewise linear function, so the optimal a can only be obtained at the cutoff points, A^*. □

Lemma S7. *All actions in $A^*\backslash A^s$ will be chosen with probability 0 in the limit.*

Proof. First, it is easy to verify that the first person will only choose a^L or a^H, and $a_1=\begin{cases}a^L & \text{if } \lambda_1<1\\ a^H & \text{if } \lambda_1>1\end{cases}$.
I assume that $a_1=a^H$ WLOG. There are three possible cases: (i) $A^s=\{a^l\}$, (ii) $A^s=\{a^h\}$, and (iii)

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

$A^s = \{a^l, a^h\}$. The analysis for them is parallel, so the discussion focuses on the case $A^s = \{a^l\}$, 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 (6), individual 2's optimal choice is

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

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

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

Note that $a^l < 1/2$, so we must have $\lambda_2^* < 1$. Let p_i denote the probability of individual i choosing a^l, so $p_2 = \overline{F}^0(\lambda_2^*)$. Suppose that $a_2 = a^l$, then

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

Substituting them into the utility functions, we get

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

so individual 3 will choose action a^l regardless of his private signal, and hence $p_3 = 1$, and an information cascade on a^l starts. Therefore, Lemma S7 holds. Suppose that $a_2 = a^h$, then

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

From the perspective of individual 3, his optimal choice is

$$a_2 = \begin{cases} a^H & \lambda_2 > 1/\underline{l}_3 \\ a^h & \lambda_2 \in (\lambda_3^*, 1/\underline{l}_3) \\ a^l & \lambda_2 < \lambda_3^* \end{cases},$$

where λ_3^* satisfies

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

so $\lambda_3^* = \lambda_2^*/\underline{l}_3 \ge \lambda_2^*$. The probability of individual 3 choosing a^l is $p_3 = \overline{F}^0(\lambda_3^*) \ge p_2$. 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 $p_3 \ge p_2$ regardless of individual 2's action. Analogously, we have $p_i \ge p_2$ for all $i \ge 2$. 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, so individuals only choose from A^s. □

A.5 Proof of Proposition S5

We can write down individual i's minimum EU utility function,

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

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 λ_1s. When λ_1 is bounded, the threshold R_0 is also bounded. When A^s is a singleton, Proposition S5 is trivially true. Suppose that A^s contains two actions, a^l and a^h, and that the minimum utility is obtained in state 0 and 1 respectively. It can be verified that

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

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

Lemma S8. *When R_0 is sufficiently large,*

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

Proof. Denote by $\rho_{hl} \equiv \sqrt{\frac{u^h - u}{u^l - u}}$. W.L.O.G., suppose that $\rho_{hl} \ge 1$. First, 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 (except for the tie-case). Suppose that $a_1 = a^h$, then we have

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

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 $\underline{l}_3 = \frac{1}{\gamma} \times \underline{l}_2 \leq \mathcal{X}_0 \times \underline{l}_0$. Extending the argument to all $i \geq 2$, so we have

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

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

Second, suppose that $\rho_{hl} = 1$. Then $\mathcal{X}_i = 1/\sqrt{\bar{l}_i \underline{l}_i}$, which degenerates to the equilibrium strategy in the paper. Following the same argument as in Case 1, we also have: $\underline{l}_i \leq \underline{l}_0$ and $\bar{l}_i \geq \bar{l}_0$.

Third, suppose that $\rho_{hl} = \gamma$. Then, we have 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 will choose a^l regardless of her signal, so an information cascade occurs, and $\underline{l}_i = \underline{l}_0$ 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.

Fourth, suppose that $\rho_{hl} > \gamma$. Individual 1 will choose a^l regardless of her private signal. An information cascade occurs on a^l, so $\underline{l}_i = \underline{l}_0$ and $\bar{l}_i \geq \bar{l}_0$ for all $i \geq 1$. □

As a consequence, when R_0 is sufficiently large, we can ensure that $\underline{l}_i$ is sufficiently small and $\bar{l}_i$ is sufficiently large for all i. Then, individuals will only choose actions from A^s in the end. Furthermore, as shown in the proof, the society will settle on one action with probability 1.

A.6 Proof of Proposition S11

Proof. Denote $r(h_i, G_i)$ as the threshold value of individual i when her observation structure is G_i and the history is h_i, i.e., individual i will choose 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 threshold value for individual k. Define $\underline{R}(k)$ to be 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.[13] Define $K \equiv \min \left\{ \overline{K}, M \right\}$, where M is a finite constant. For individual $i > K$, suppose that $a_i = 1$, 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 an arbitrary network bounded by K. By definition, all actions after individual K are not observable under $\hat{G}$, so for all $i > K$, we have

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

[13] To see that, when $k = 2$, $h_2 \in \{\{0\}, \{1\}\}$, $G_2 \in \{\emptyset, \{1\}\}$. 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.

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

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

Since $\overline{R}\left(K+1\right)<\gamma$, we have $\beta>1$, so

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

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

A.7 Proof of Corollary S4

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,

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

Denote by $E_n\equiv\left\{s^\infty\in\mathcal{S}^\infty:r\left(h_K,G_K\right)\in\left[1/\gamma+1/n,\gamma-1/n\right]\right\}$, so $E=\cup_nE_n$. From the proof of Proposition S11, we know that on 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 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. □