Sequential Learning under Informational Ambiguity*

Jaden Yang Chen†

First version: November 2019 This version: October 2023

Abstract

This paper studies a sequential social learning problem in which individuals face ambiguity regarding other people's signal structures. It finds that ambiguity has a significant influence on social learning and offers new insights into the mechanism driving herding behavior. In contrast to previous findings that identified various learning outcomes based on fine details of the learning environments, such as the statistical properties of the signal structures, this paper establishes information cascades as the only robust outcome under ambiguity. Specifically, it demonstrates that in the presence of sufficient ambiguity, an information cascade will occur almost surely, regardless of the statistical properties of the signal structures or other specific details of the learning environments. Furthermore, this paper highlights that some standard results that feature the absence of a cascade can become fragile in the face of ambiguity. In some cases, even a slight degree of ambiguity can trigger a cascade when signals are bounded and can lead to incorrect learning when signals are unbounded.

JEL Classification: D81, D83, C72
Keywords: Social learning, ambiguity, information cascades, herding

*The paper is based on my Ph.D. thesis at Cornell University. It was also circulated as "Information Cascades and Ambiguity". I am indebted to my committee members, David Easley (chairperson), Larry Blume and Tommaso Denti, for their encouragement and guidance. I thank Gary Biglaiser, Simon Board, Larry Epstein, Drew Fudenberg, Yoram Halevy, Fei Li, Qingmin Liu, Erik Madsen, Suraj Malladi, Peter Norman, Seth Sanders, Todd Sarver, Peter Sorensen, Omer Tamuz, Jianrong Tian, Alex Wolitzky and many conference participants for discussion and comments. Any errors are mine.

†Department of Economics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27517, USA; E-mail address: yangch@unc.edu

1 Introduction

Herding behavior is an important feature of real life. One of the most influential models to explain herding behavior is the sequential social learning model (SSLM) introduced by Banerjee (1992) and Bikhchandani et al. (1992). Under this framework, a sequence of individuals take actions to match the actual state. Individuals receive i.i.d. private signals with a commonly known data-generating process (DGP), which generates a finite number of signals. The authors showed that an ***information cascade*** will arise with probability 1 and an incorrect cascade occurs with a positive probability. In the end, all individuals will choose to ignore their own signals and follow one action, even if that action is suboptimal.

The theory of information cascades was later challenged by the following two findings by Smith and Sørensen (2000). First, when signals are unbounded, the society will eventually settle on the correct action, which implies that an incorrect herd cannot occur. Second, even if signals are bounded, an information cascade may or may not occur depending on the statistical properties of the actual DGPs. More precisely, when the DGPs satisfy the increasing hazard ratio property (IHRP) or the log-concavity property, an information cascade will not take place (Herrera and Hörner, 2012; Smith et al., 2021). Because many common distributions satisfy these properties, the SSLM cannot generate information cascades or incorrect herding in many interesting situations. Recent works on misspecified learning further suggest that if individuals consider an incorrect DGP, the learning outcome would also depend on the model perception, and it is possible that actions will oscillate forever (Bohren, 2016; Bohren and Hauser, 2021).[1] In summary, the literature concludes that social learning outcomes depend on the fine details of the true model and model perceptions. This raises the question of which learning outcome remains robust when alternative specifications of these details are considered. Furthermore, the literature predominantly focuses on cases where individuals have expected utility preferences, leaving the question of how society learns with other preferences largely unexplored.[2]

This paper studies social learning with ambiguity-averse individuals. Unlike most previous works in which individuals are certain about the society's signal structure, this paper adopts the assumption that individuals are ***ambiguous*** about predecessors' DGPs by considering a set of DGPs as possibilities. It describes a situation in which individuals face model uncertainty and cannot determine a specific DGP that explains the information of interest or form a prior distribution over DGPs. The distinction between risk and ambiguity was made by Knight (1921), and it plays an important role in understanding decision making under uncertainty.[3] In the specific context of social learning, ambiguity can emerge naturally because, in reality, individuals' information quality can vary significantly; some individuals are "experts" with better information, whereas others are "laymen" with little information who mainly follow others. Furthermore,

[1] Their setups do not nest the standard SSLM, but the insight still holds in the standard model.

[2] Bikhchandani et al. (2021) and Golub and Sadler (2017) provide excellent surveys on SSLM.

[3] See Hansen et al. (2014); Marinacci (2015); and Hansen and Marinacci (2016) for surveys on ambiguity and model uncertainty.

individuals often have limited observations from others; for example, in SSLM, individuals only observe one action from each predecessor, making it challenging to distinguish between "experts" and "laymen" or determine the distribution of these two types, resulting in ambiguity. In the benchmark model, I also assume that individuals have the ***max-min expected utility*** (MEU) preferences as in Wald (1950) and Gilboa and Schmeidler (1989), so they are ambiguity-averse and choose an action to maximize the expected utility in the worst-case scenario. The motivation for MEU preferences is twofold. First, it relaxes the independence axiom of the expected utility and is compatible with some experimental evidence, e.g., the Ellsberg paradox (Ellsberg, 1961). Second, it describes the situations where individuals have the robustness concern in their decision making (Hansen and Sargent, 2001). In the presence of model uncertainty, individuals are worried about model misspecification and may seek to make robust decisions with respect to all models they consider possible.

In the paper, I provide detailed discussion of social learning under ambiguity and characterize learning outcomes. In contrast to previous findings that various learning outcomes can emerge depending on the modeling details, this paper finds that under sufficient ambiguity, i.e., when individuals consider adequately many models, information cascades will emerge as the only outcome in the sequential social learning environment. Therefore, it establishes information cascades as a robust result in an ambiguous environment whose occurrence relies little on various details of the learning environment—e.g., (i) the fine properties of the true DGPs, and (ii) whether a particular model is perceived by individuals. This finding may appear counter-intuitive since, from the previous literature, different DGPs can lead to different learning outcomes. It seems unclear what will happen if individuals consider multiple DGPs simultaneously, and a natural conjecture is that the learning outcomes should also exhibit a variety of forms depending on the fine details of the perceived model set. However, the paper notes that the DGPs that encourage an information cascade and discourage it are actually *asymmetric*. When the environment features sufficient ambiguity, the cascading force always dominates, establishing information cascades as the only robust result.

To illustrate the asymmetry, consider an example where customers decide which restaurants to go to and face ambiguity about other customers' informativeness. Suppose that all customers went to restaurant A, and the next customer received a signal supporting restaurant B. If the next customer went to restaurant B, the worst-case scenario is that all predecessors are experts, their signals are all very precise, and their actions reveal strong information supporting restaurant A. On the other hand, if the customer went to restaurant A, the worst-case scenario is that all predecessors are laymen, their signals are uninformative, and their actions do not reveal any information about each restaurant. As ambiguity increases, the customer would be more concerned about breaking away from the herd, because she could not rule out the possibility of acting against some highly precise signals. In contrast, the concern about following the herd is more controllable, because the customer would only act against her private signal, the precision of which is certain to her. This asymmetry in the worst-case scenarios pushes the customer to follow and creates a cascade force.

Below is a more concrete example.

Example 1. The state space $\Theta = \{0, 1\}$. The true state is unknown. Individuals share a common flat prior π_0. Every individual i takes action $a_i \in \{0, 1\}$. The utility is 1 if the action matches the state and 0 otherwise. Each individual i receives a signal $s_i \in \{H, L\}$ and has DGP $g_i(s|\theta)$ with

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

where $g_i(s|\theta)$ denotes the conditional probability of signal s in state θ, and γ_i describes individual i's signal precision. Individuals only know their own signal precision but are ambiguous about others' precision and they believe that every $\gamma_i \in \Gamma$ is possible. Suppose that the first individual's (his) action is $a_1 = 1$. Denote by $V_2(a)$ the minimum expected utility of the second individual (she) if she takes action a. We have

$$V_2(1) = \begin{cases} \gamma_2/(\gamma_2+1) & s_2 = H \\ 1/(\gamma_2+1) & s_2 = L \end{cases} \text{ and } V_2(0) = 0.$$

To see that, the worst-case scenario for $a_2 = 1$ is that individual 1 only received uninformative signals, so individual 2's utility is $\frac{\gamma_2}{\gamma_2+1}$ if her signal is H and $\frac{1}{\gamma_2+1}$ if her signal is L. On the other hand, the worst-case scenario for $a_2 = 0$ is that individual 1 received the perfectly revealing signal, so individual 2's minimum utility is 0.[4] Since $V_2(1) > V_2(0)$, individual 2 will always follow individual 1's action regardless of her private signal, so an information cascade occurs immediately.

The paper extends the insight from this example to a general framework, in which (i) signals can come from a wide class of distributions, and (ii) individuals can perceive an arbitrary model set, which can encompass correct specification, misspecification, and various forms of ambiguity as special cases. The paper characterizes social learning outcomes with this general information structure and provides conditions under which an information cascade occurs. The paper starts with the cases where signals are bounded. Theorem 2 provides sufficient conditions for an information cascade to occur almost surely for bounded signals. The theorem states that whenever the perceived model set contains an adequately informative DGP, i.e., a DGP that assigns sufficiently high weights to precise signals, an information cascade will occur almost surely. In other words, as long as individuals can't exclude the possibility of some informative DGP, a cascade will always emerge, regardless of what other DGPs individuals may also consider as possible, and regardless of the statistical properties that the true DGPs may possess. The intuition is similar to that in Example 1, where the perception of a highly informative DGP would encourage individuals to follow a herd, creating an asymmetrically high cascade force, which cannot be offset by the perception of any other DGPs. Perhaps surprisingly, Theorem 2 even implies that non-cascade results represent knife-edge cases in some interesting situations. With the introduction of a slight degree of ambiguity, an information cascade will occur almost surely even if there is no cascade in

[4] To be more precise, it is actually the infimum utility because γ_i cannot be 1 or infinity.

the case without ambiguity. The paper then discusses the cases when signals are unbounded. With sufficient ambiguity, an information cascade also occurs. To have a cascade with unbounded signals, individuals are required to consider arbitrarily informative DGPs, which represents an extreme situation. This paper then focuses on a weaker but qualitatively similar concept—herding, i.e., individuals end up taking the same action, which can be incorrect, but not necessarily ignoring their private signals. Theorem 3 provides sufficient conditions for herding to emerge, and the conditions are parallel to Theorem 2, which require individuals to consider an adequately informative DGP. Similarly, Theorem 3 also implies that complete learning result in Smith and Sørensen (2000) is not robust—in some situations, an incorrect herd can emerge when individuals consider DGPs that are arbitrarily close to the true DGP. The main paper focuses on MEU preferences, but the insights are then extended to general ambiguity preferences, as discussed later in the paper.

Technically, this paper differs from common papers in SSLM in two ways. First, under ambiguity, individuals consider a set of DGPs, which leads to a set of posteriors, so we cannot keep track of the posterior likelihood ratio as in the literature. Second, under ambiguity, the posterior under each perceived DGP may no longer be a martingale, so we cannot apply the martingale convergence theorem common in the literature. To solve the first challenge, this paper employs a simple statistic—*the average likelihood ratio*, i.e., the geometric mean of the maximum and minimum of posterior likelihood ratios generated by all perceived DGPs—and demonstrates that it serves as a sufficient statistic for social learning under MEU preferences. To solve the second challenge, this paper focuses on analyzing the dynamics of the average likelihood ratio. When signals are bounded, it turns out that the dynamics are relatively simple. When signals are unbounded, the analysis involves analyzing the local stability of the average likelihood ratio by estimating the probability of each type of herding. It turns out that whether the average likelihood ratio can settle on the correct/incorrect state is equivalent to whether a correct/herd can happen with a positive probability, which is further equivalent to the convergence of an infinite series, whose increasing rate relies on the tail properties of perceived DGPs, as will be explained later in the paper.

The paper proceeds as follows. Sections 2 and 3 lay out the model setup and characterize the equilibrium. Sections 4 to 7 provide conditions for an information cascade to occur. Section 8 discusses other ambiguity preferences. Section 10 reviews related literature. Proofs are collected in the Appendix. Other topics and extensions are presented in the Supplementary Material.

2 The Model

States and Actions. There are two possible states of world, $\Theta = \{0, 1\}$. Without loss of generality, the true state $\theta^* = 0$. A countably infinite set of individuals $N = \{1, 2, ...\}$ act sequentially. Each individual makes a choice $a \in A = \{0, 1\}$ and can observe the choices taken by all predecessors. Individuals get a payoff of 1 when their actions match the actual state and a payoff of 0 otherwise.

Information structures. Individuals do not know the true state and share a common prior π_0

which is flat.[5] Each individual i will receive a signal $s_i \in \mathcal{S} \subset \mathbb{R}$. Signals are independently, but not necessarily identically, distributed according to $\left\{\overline{G}_1^\theta, \overline{G}_2^\theta, ...\right\}$, where $\overline{G}_i^\theta : \mathcal{S} \to [0,1]$ denotes the cumulative distribution function of s_i when the actual state is θ. No signal perfectly reveals the state, so the probability measures induced by $\overline{G}_i^0$ and $\overline{G}_i^1$ are mutually absolutely continuous. Following the convention, I introduce the normalized signal, λ_i, where $\lambda_i(s) = \frac{d\overline{G}_i^1(s)}{d\overline{G}_i^0(s)}$ denotes the likelihood ratio induced by signal s. The distribution of the likelihood ratio λ_i is denoted by $\overline{F}_i^\theta$, which must satisfy $\lambda = \frac{d\overline{F}_i^1(\lambda)}{d\overline{F}_i^0(\lambda)}$ almost everywhere. The rest of the paper focuses on the normalized signal, λ, and the normalized DGP, $\overline{F}_i^\theta$. All normalized DGPs have a common support, $\Lambda = \left[\frac{1}{\gamma}, \gamma\right]$, where $\gamma > 1$. Signals are ***bounded*** if $\gamma < \infty$ and signals are ***unbounded*** if $\gamma = \infty$. For notional convenience, I assume: (i) all signals are continuous, that is, $\overline{F}_i^\theta$ is continuous for all i and θ, and (ii) signals are symmetric $\overline{F}_i^1(\lambda) = 1 - \overline{F}_i^0(1/\lambda)$ for all i and λ.[6] Let $\mathbb{F}$ denote the set of feasible DGPs with a typical element being $F = (F^0, F^1)$. Let Λ^∞ denote the set of all signal paths with a typical element being $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, ...)$ and is associated with σ-algebra $\sigma(\Lambda^\infty)$. Let $\mathbb{P}^*$ denote the probability measure on $(\Lambda^\infty, \sigma(\Lambda^\infty))$ induced by the true signal distribution $\left\{\overline{F}_1^0, \overline{F}_2^0, ...\right\}$. The paper refers to $\mathbb{P}^*$ as the ***true probability***. Without explicit mention, every event is evaluated according to the true probability.

Ambiguous Information and Beliefs. Individuals know their own DGPs and that all signals are independently distributed, but they may be ***ambiguous*** about others' DGPs by considering a set of DGPs. Specifically, individuals share a common set of models, $\mathcal{F}_0 \subset \mathbb{F}$, and believe that every other individual's DGP belongs to $\mathcal{F}_0$ but do not know which is the true DGP. The ambiguity assumption describes a situation in which individuals lack sufficient knowledge to pin down the society's signal structure. It is believed to emerge naturally in social learning, in which individuals only observe limited number of actions from other individuals, so the information is often insufficient to determine a specific DGP. The paper's setup allows for general $\mathcal{F}_0$, which can nest following special cases.

Example 2. *Correct Specification.* Suppose that signals are i.i.d. with $\overline{F}_i = F$ for all i, and that $\mathcal{F}_0 = \{F\}$. Then, individuals **correctly specify** the society's true signal distribution, which corresponds to the standard social learning models, e.g., Banerjee (1992) and Bikhchandani et al. (1992).

Example 3. *Misspecification.* Suppose that signals are i.i.d. with $\overline{F}_i = F$ for all i, and that $\mathcal{F}_0 = \left\{\hat{F}\right\}$ for some $\hat{F} \neq F$. Then, individuals **misspecify** the society's true signal distribution, which corresponds to misspecified social learning, e.g., Bohren and Hauser (2021) and Arieli et al. (2023).

[5] The paper's analysis extends to any full-support prior or a set of priors bounded away from extreme beliefs.

[6] With continuity, we avoid separate notations that deal with discontinuity points; with symmetry, we only need to characterize one size of the distribution. The paper's discussion can extend to general cases because: (i) the discontinuous case can be nested as a limit of the continuous case, and (ii) the asymmetric case can be dealt with by imposing similar conditions on the other side of the distribution.

Example 4. *Correct & incorrect ambiguity.* Suppose that $\mathcal{F}_0$ is non-singleton, so individuals are ambiguous about the true DGPs. Individuals are **correctly ambiguous** if

$$\mathcal{F}_0 \supset \left\{ \overline{F}_1, \overline{F}_2, ...\right\},$$

that is, the model set contains every individual's true DGP. Individuals are **incorrectly ambiguous** if the model set doesn't contain the true DGPs. The correct-ambiguity case is sometimes called *model uncertainty*, and the incorrect-ambiguity case is called *model misspecification*, e.g., Hansen and Marinacci (2016).

Example 5. *Entropy-based ambiguity.* Suppose that $\mathcal{F}_0$ takes the following form

$$\mathcal{F}_0 = \left\{ F \in \mathbb{F} : \int \log\left(\frac{G^\theta(\lambda)}{F^\theta(\lambda)}\right) dG^\theta(\lambda) \leq r \right\},$$

where $G \in \mathbb{F}$ and $r > 0$. It corresponds to the **entropy-based ambiguity** commonly seen in the robust control literature, e.g., Hansen and Sargent (2001), where individuals consider all signal distributions whose relative entropy w.r.t. the benchmark distribution G is less than r. Here, r measures the degree of ambiguity. When $r = 0$, $\mathcal{F}_0 = \{G\}$, so there is no ambiguity as in Example 2 and 3. As r increases, $\mathcal{F}_0$ features a higher degree of ambiguity.

Example 6. *Consistent ambiguity.* Suppose that $\mathcal{F}_0$ is non-singleton and satisfies

$$\forall i, \theta, \lambda : \quad \overline{F}_i^\theta(\lambda) = \int F^\theta(\lambda) d\mu(F) \quad \text{with } \mathcal{F}_0 = \text{supp}(\mu), \tag{1}$$

for some probability measure $\mu \in \Delta(\mathbb{F})$.[7] That is, the true DGP can be obtained by mixing DGPs in $\mathcal{F}_0$. Then, we say that the model perception features **consistent ambiguity**. To see understand it, suppose that there is an ex ante stage in which every individual's DGP is i.i.d. according to a second-order distribution μ, and that individuals only consider DGPs that are ex ante possible, so $\mathcal{F}_0 = \text{supp}(\mu)$. Besides, all events are evaluated according to the ex ante signal distribution, so the true signal distribution is taken to be $\overline{F}_i^\theta = \int F^\theta(\lambda) d\mu(F)$. In this case, individuals' model perceptions align with the true probability in the sense that they correctly specify all possible DGPs at the ex ante stage. Consistent ambiguity is useful in applications, as it allows individuals to maintain ambiguity and correctness simultaneously.[8]

Belief-updating Rule. Due to the informational ambiguity, individuals will form ambiguous beliefs in social learning. Denote by $h_i = (a_1, ..., a_{i-1})$ the history observed by individual i and by $I_i = \{\lambda_i, h_i\}$ the information available to individual i—that is, her private signal λ_i and history h_i.

[7] To save on notation, whenever $\text{supp}(\mu)$ appears, it is implicitly assumed that it is well-defined with respect to some common topology. In most examples, I simply consider DGPs that can be parameterized by real numbers and use the standard definition of the support on the real line.

[8] With some abuse of language, here the correctness means that perceptions are consistent with the true probability, which is different from the correct ambiguity in Example 4.

Let $\mathcal{I}_i$ be the set of all possible information available to i, and denote by $\sigma_i : \mathcal{I}_i \to A$ the (pure) strategy of individual i. Given strategy profile $\sigma_{-i} = (\sigma_1, ..., \sigma_{i-1})$, DGP profile $F_{-i} = (F_1, ..., F_{i-1})$ and conditional on state θ, the observed history $h_i = (a_1, ..., a_{i-1})$ is a stochastic process with a probability measure $\mathbb{P}_{F_{-i}}(.|\theta; \sigma_{-i})$. Given history h_i and strategy profile σ_{-i}, denote by $\Pi(h_i, \sigma_{-i})$ the set of beliefs generated by DGPs in $\mathcal{F}_0$, which I refer to as a ***public belief set***. That is,

$$\Pi(h_i, \sigma_{-i}) = \left\{\pi \in \Delta(\Theta) : \pi(\theta) = \mathbb{P}_{F_{-i}}(\theta|h_i; \sigma_{-i}), F_{-i} \in \mathcal{F}_0^{i-1}\right\},$$

where $\mathbb{P}_{F_{-i}}(\theta|h_i; \sigma_{-i})$ is the conditional probability on θ derived from $\mathbb{P}_{F_{-i}}(.|\theta; \sigma_{-i})$, and $\mathcal{F}_0^{i-1}$ is $i-1$ copies of $\mathcal{F}_0$. The public belief set consists of conditional probabilities generated by all possible $F_{-i} \in \mathcal{F}_0^{i-1}$ for which the conditional probabilities are well defined. Based on the public beliefs and private signal λ_i, individual i will form a belief set, $\Pi_i(I_i, \sigma_{-i})$, which I refer to as a ***private belief set***. Assuming that individuals use the full Bayesian rule (axiomatized by Pires (2002)) to update beliefs,

$$\Pi_i(I_i, \sigma_{-i}) = \left\{\pi \in \Delta(\Theta) : \pi = BU\left(\pi'; \lambda_i\right), \pi' \in \Pi(h_i, \sigma_{-i})\right\},$$

where $BU(\pi'; \lambda_i)$ denotes the Bayesian update of belief π' based on signal λ_i. In other words, individuals update the public belief set prior-by-prior using Bayes' rule. The full Bayesian rule is commonly adopted in applications, but two major criticisms of it are (i) the set of models remains unchanged after learning new information, and (ii) it can lead to dynamic inconsistency. These criticisms are of less concern in this paper, because (i) individuals observe one action from every other individual, so there is often very limited information to be learned about others' DGPs, and (ii) individuals make a once-in-a-lifetime decision, so dynamic inconsistency is not relevant here. For these reasons, the paper's main result also holds for alternative updating rules.[9]

Equilibrium Concept. Assume that individuals have ***max-min expected utility*** (MEU) preferences, as in Gilboa and Schmeidler (1989). The equilibrium is defined as follows.

Definition 1. A strategy profile $\sigma^* = (\sigma_i^*)_{i \in N}$ constitutes an equilibrium if for all $i \in N$ and all information sets $I_i \in \mathcal{I}_i$, we have

$$\sigma_i^*(I_i) \in \arg\max_{a \in \{0,1\}} \inf_{\pi \in \Pi_i\left(I_i, \sigma_{-i}^*\right)} \mathbb{E}_\pi U(a, \theta), \tag{2}$$

where $U(a, \theta)$ is the utility function that equals 1 if $a = \theta$ and 0 if $a \neq \theta$.

Where no confusion would exist, I omit the equilibrium strategy notation σ^* and denote $\Pi(h_i)$ and $\Pi_i(I_i)$ as the equilibrium public belief set and posterior set. Individuals follow some tie-breaking rule when indifferent, so Definition 1 provides a unique pure-strategy equilibrium. The

[9] For example, the α-maximum likelihood rules as discussed in Section 6 of the Supplementary Material.

choice of tie-breaking rule is not essential to the result, so I do not specify it in the paper.[10] By focusing on pure strategies, the paper implicitly assumes that agents cannot be better off by playing mixed strategies. Note that this assumption is not contradictory to ambiguity-aversion, which says that individuals have incentives to engage in ex post randomization instead of ex ante randomization, as in the mixed-strategy case.[11]

3 Equilibrium Strategies and Learning Concepts

This section first characterizes individuals' equilibrium strategies under ambiguity and then defines some learning concepts that will be used later.

3.1 Characterizations of Equilibrium Strategies

When individuals are ambiguous, it seems difficult to characterize learning dynamics because individuals now form a set of posteriors instead of a single posterior. Fortunately, the max-min model enables us to extend the concept of likelihood ratio and represent the posterior set using the ***average likelihood ratio*** of beliefs featured in it. This property leads to a simple equilibrium characterization, which enhances the tractability.

Definition 2. (Average Public Likelihood Ratio) Denote by $L(h_i) = \left\{\frac{\pi(1)}{\pi(0)} : \pi \in \Pi(h_i)\right\}$, the set of public likelihood ratios. Let $\underline{l_i} = \inf L(h_i)$ and $\overline{l_i} = \sup L(h_i)$, and denote by $r_i = \sqrt{\overline{l_i} \cdot \underline{l_i}}$, called the *average public likelihood ratio*, based on history h_i.

The average public likelihood ratio r_i is the geometric average of the highest and lowest likelihood ratios in the public belief set. It reflects how likely the public thinks state 1 is (relative to state 0) on average. Proposition 1 characterizes individuals' equilibrium strategies by employing average public likelihood ratios.

Proposition 1. (Characterizations of Equilibrium Strategies) *In the equilibrium, for any individual, $i \in N$, and information set, $I_i \in \mathcal{I}_i$, we have*

$$\sigma_i^*(I_i) = \begin{cases} 1 & \text{if } \lambda_i \cdot r_i > 1 \\ 0 & \text{if } \lambda_i \cdot r_i < 1 \end{cases},$$

and the strategy at $\lambda_i \cdot r_i = 1$ is determined by the tie-breaking rule.

[10] When signals are continuous, the tie case happens with zero probability.

[11] Although it still remains a question of whether individuals can benefit from ex ante randomization, the literature seems to suggest that indifference to ex ante randomization seems a reasonable assumption; e.g., in Seo (2009) and Klibanoff et al. (2005), individuals have expected utility preferences over second-order acts; Saito (2012) suggests that individuals have no incentive to engage in ex ante randomization under some axiom; Eichberger et al. (2016) show that dynamic consistency implies that individuals are indifferent to ex ante randomization.

Proof. Denote by $\underline{\pi}_i(\theta) = \inf\{\pi(\theta) : \pi \in \Pi_i(I_i)\}$, then $a_i = 1$ if $\underline{\pi}_i(1) > \underline{\pi}_i(0)$. Note that

$$\underline{\pi}_i(1) = \frac{\lambda_i \underline{l}_i}{1+\lambda_i \underline{l}_i} \qquad \underline{\pi}_i(0) = \frac{1}{1+\lambda_i \overline{l}_i}.$$

By solving $\underline{\pi}_i(1) > \underline{\pi}_i(0)$, we have $\lambda_i > \frac{1}{\sqrt{\overline{l}_i \cdot \underline{l}_i}} = \frac{1}{r_i}$. The other case follows symmetrically. □

The average likelihood ratio is an extension of the likelihood ratio in the standard model. It acts as a sufficient statistic for the public history in cases in which there are multiple beliefs. Proposition 1 shows that individuals' equilibrium strategies can be decomposed into two parts. The private information component is the private signal, λ_i, whereas public information is captured by the average public likelihood ratio, r_i. When the product, $\lambda_i \cdot r_i$, is greater than 1, reflecting that state 1 is more likely, individuals will choose action 1 and vice versa. For simplicity, "average public likelihood ratio" is sometimes referred to as "public belief" when there is no confusion.

3.2 Information Cascades and Some Learning Concepts

Definition 3. On a signal path $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, ...)$, an ***information cascade*** occurs if there exists some $I < \infty$ and $a \in A$ such that for all $i \geq I$, we have $\mathbb{P}^*(a_i = a | h_i) = 1$.

An information cascade occurs if, after some point, individuals will only choose one action regardless of their private signals. During a cascade, information stops aggregating and the society may settle on an incorrect action, albeit with infinitely many signals. Using Proposition 1, information cascades can be described using the average likelihood ratio.

Lemma 1. *Denoting by $C_0 = \left[0, \frac{1}{\gamma}\right]$ and $C_1 = [\gamma, \infty]$, an information cascade of action a occurs when there exists some $I < \infty$ such that $r_i \in C_a$ for all $i \geq I$.*

In the literature, C_a is referred to as the ***cascade set*** of action a. Whenever $r_i \in C_a$, the public belief that favors action a becomes so strong such that individuals will choose action a regardless of their private signals; that is, an information cascade takes place. In classical models of finite signals, an information cascade will almost surely occur, e.g., Banerjee (1992) and Bikhchandani et al. (1992). However, with more general signal structures, the occurrence of a cascade is not always guaranteed. The following outcomes are also possible.

Definition 4. On a signal path $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, ...)$, we say that (i) ***herding*** occurs if there exists some $I < \infty$ and $a \in A$ such that $a_i = a$ for all $i \geq I$; (ii) ***action non-convergence*** occurs if a_i fails to converge. We say that ***complete learning*** occurs if the correct-action herding occurs $\mathbb{P}^*$-almost surely;

The concept of herding is often confused with information cascade, since both imply the conformity of actions. The difference is that individuals during a herd would have acted differently if they received different signals, but individuals during an information cascade will choose the same action and ignore their private signals, so cascades are more stable than herding.[12] When signals are continuous and satisfy IHRP, the society achieves herding but not information cascades. In addition to the absence of a cascade, Smith and Sørensen (2000) show that complete learning occurs when signals are unbounded, so the SSLM does not explain the persistence of incorrect herds in this case. Furthermore, if individuals misspecify the actual DGPs, action non-convergence may emerge, in which case the SLLM does not necessarily lead to a consensus of actions.

4 Benchmark Case: Cascades under Extreme Ambiguity

This section considers an extreme case of ambiguity and shows that an information cascade occurs in this case. Denote by $\mathcal{F}$ the set of all DGPs on $[1/\gamma, \gamma]$, where γ denotes the highest signal. We have the following theorem.

Theorem 1. *When $\mathcal{F}_0 = \mathcal{F}$, an information cascade occurs $\mathbb{P}^*$-almost surely.*

The condition $\mathcal{F}_0 = \mathcal{F}$ describes a situation in which individuals only understand the support of signals and consider all DGPs on this support to be possible. Theorem 1 shows that in this benchmark case, an information cascade occurs almost surely. This finding is different from the standard findings in the following respects. First, in the standard literature, the occurrence of a cascade relies on specific properties of true DGPs. However, Theorem 1 does not impose any restrictions, so a cascade can occur under all DGPs on $[1/\gamma, \gamma]$. Second, in the misspecified learning literature, the learning outcome depends on the model specification; however, Theorem 1 shows that when individuals consider multiple model specifications simultaneously, information cascades will emerge as the only outcome. Therefore, Theorem 1 shows that under high ambiguity, a cascade almost surely arises without regard to many details that would matter in the standard case.

The intuition

When signals are unbounded $\gamma = \infty$, an information cascade occurs immediately after the first individual, as in Example 1. Below, I focus on the bounded-signal case $\gamma < \infty$. Suppose that the first i individuals took action 1 and that individual $i+1$ received a signal $\frac{1}{\gamma}$, the strongest signal for state 0. Suppose that an information cascade did not occur when the first i individuals made decisions. Let's consider the decision problem of individual $i+1$. As she has MEU preference, her decision is determined by the worst scenarios:

- If she follows the herd and takes action 1, the worst case is that the predecessors' DGPs are uninformative. In this case, $\lambda_1 = ... = \lambda_i = 1$. By following the herd, she would act against her private signal, $\frac{1}{\gamma}$.

[12]The concept of information cascade was proposed by Bikhchandani et al. (1992) and tested in a lab experiment by Anderson and Holt (1997). The difference between information cascade and herding was distinguished by Smith and Sørensen (2000) and tested by Çelen and Kariv (2004) in an experimental environment.

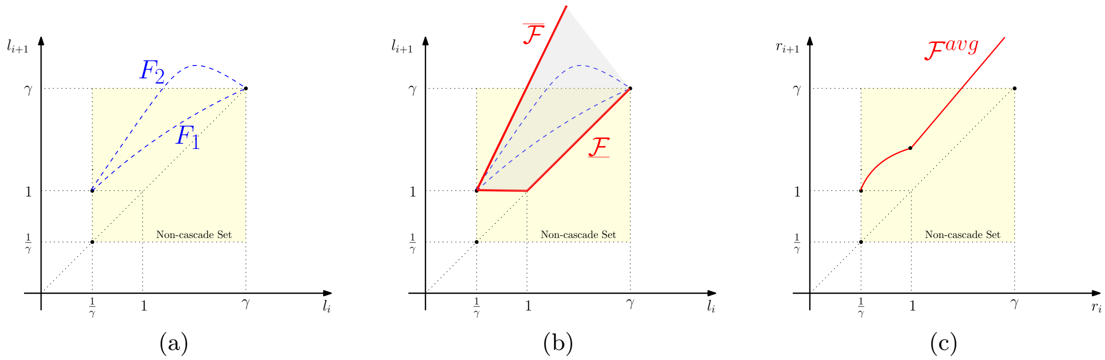


Figure 1: Information Cascades under Ambiguity

Note: The horizontal axis represents the prior likelihood ratio (between states 1 and 0), and the vertical axis represents the posterior likelihood ratio after observing an action 1 (the dynamics after an action 0 are symmetric). The yellow area represents the non-cascade region. Figure 1a depicts the likelihood curves under F_1 and F_2. Figure 1b depicts the set of likelihood curves under all DGPs in $\mathcal{F}$ (marked by the gray shaded area). Figure 1c depicts the average likelihood curve under $\mathcal{F}$.

- If she breaks the herd and takes action 0, the worst case is that every predecessor's DGP has the most precise DGP—i.e., the DGP that only generates signals γ and $1/\gamma$. In this case, the predecessors' actions reveal that their signals must be γ.[13] Hence, by taking action 0, individual $i+1$ would act against i signal γ.

As can be seen, the forces that encourage a cascade and discourage it are **asymmetric**. As i increases, individual $i+1$ would act against increasingly more signals γ in the worst case if she broke the herd; however, she would only act against one signal—her private signal—in the worst case if she followed the herd. When i is sufficiently large, individual $i+1$ would find it optimal to follow, which creates an information cascade.

Graphic illustration. Figure 1a illustrates that the occurrence of a cascade depends on the DGP if there is no ambiguity. In the figure, F_1 satisfies the IHRP, but F_2 does not. As can be seen, posteriors under F_1 are trapped in the non-cascade set, so a cascade never occurs; however, posteriors under F_2 can extend into the cascade set, so a cascade can emerge.

Figure 1b illustrates that under ambiguity, the forces that encourage a cascade and discourage it are asymmetric. When beliefs are in the non-cascade set, the upper envelope of all likelihood curves under $\mathcal{F}$ (marked by $\overline{\mathcal{F}}$) has a slope of γ, which means that observing an action 1 can at most increase the likelihood of state 1 (relative to state 0) by a factor of γ. However, the lower envelope (marked by $\underline{\mathcal{F}}$) is always bounded from below by the 45-degree line, which means that observing an action 1 cannot decrease the likelihood of state 1.[14] The asymmetry of these two curves corresponds

[13] If any of them received a signal $1/\gamma$, that individual would have taken action 0 given the presumption that a cascade did not occur.

[14] The upper envelope is obtained at the DGP that only generates the most precise signals, γ and $1/\gamma$. The lower envelope has a kink at 1. When $l_i > 1$, the lower envelope is obtained at the uninformative DGP. When $l_i < 1$, individual i's prior favors state 0, so she must have received some minimum information to take action 1. In this case, the lower envelope is obtained at the DGP that only generates signals $1/l_i$ and l_i.

to the asymmetry between cascade and non-cascade. The worst-case scenario for breaking a herd happens when beliefs are updated according to the upper envelope, in which case predecessors have the most precise DGP and individuals would act against a sequence of signal γ. In contrast, the worst-case scenario for following a herd happens when beliefs are updated according to the lower envelope, in which case the minimum expected utility is bounded from below by the expected utility when predecessors have uninformative DGPs. Figure 1c depicts the average likelihood curve (marked by $\mathcal{F}^{avg}$), which is obtained by averaging the two envelope curves. As can be seen, the average likelihood curve extends to the cascade set, so an information cascade can occur. To fully prove Theorem 1, it remains to show that the probability of a cascade is 1, which is delegated to the Appendix.

5 Information Cascades with Bounded Signals

The previous section focuses on the extreme case in which individuals consider all DGPs on the actual support to be possible. This section provides less restrictive conditions on the model setup under which an information cascade occurs. Throughout this section, I focus on the case with bounded signals, i.e., $\gamma < \infty$. We have the following result.

Theorem 2. (Cascade with bounded signals). *Suppose that there exists some $F \in \mathcal{F}_0$ such that one of the following conditions holds:*

(1) F is discrete at γ;

(2) F is continuously differentiable on $(\gamma - \varepsilon, \gamma)$ for some $\varepsilon > 0$ with $f^1(\gamma) > \frac{2}{\gamma - 1}$,

where $f^1(\gamma) = \lim_{x \nearrow \gamma} \frac{dF^1}{dx}(x)$. Then, when signals are bounded, an information cascade occurs $\mathbb{P}^*$-almost surely.

The two conditions say that individuals consider a DGP that assigns adequately large weights to high-precision signals, i.e., the tail is adequately heavy. With some abuse of language, I refer to DGPs that satisfy similar heavy-tail conditions as highly informative. Theorem 2 therefore says that if individuals find it possible that other individuals may have a highly informative DGP, an information cascade will occur almost surely.

The conditions in Theorem 2 are not very restrictive: First, it only requires that $\mathcal{F}_0$ contain one such F but does not impose other restrictions on $\mathcal{F}_0$; second, it only imposes restrictions on the F's tail but does not impose any restrictions in the middle. The intuition behind the first point is similar to that in the benchmark case: If a highly informative DGP is considered possible, it will create a strong cascade force—which, due to the asymmetry, cannot be mitigated by any other model perception, therefore an information cascade always occurs regardless of the detailed structure of the model set.

The intuition behind the second point can be illustrated using Figure 2, which depicts the likelihood curves when $r_i \in (\gamma - \varepsilon, \gamma) \equiv C_\varepsilon^1$. By definition, the upper envelope of likelihood curves

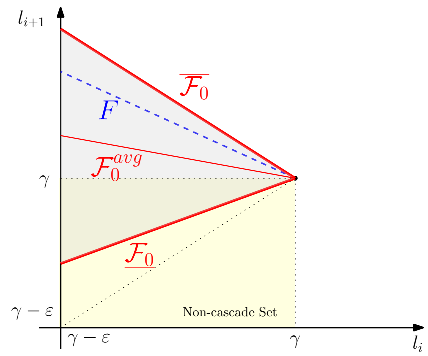


Figure 2: Sufficient Conditions for Information Cascades

Note: Figure 2 depicts the posterior likelihood curves in a small neighborhood near γ. The gray shaded area represents all likelihood curves generated by models in $\mathcal{F}_0$, and the yellow area represents the non-cascade set.

under $\mathcal{F}_0$ (marked by $\overline{\mathcal{F}_0}$) must lie above the likelihood curve under F (marked by F) for all $F \in \mathcal{F}_0$. From the previous discussion, the lower envelope curve (marked by $\underline{\mathcal{F}_0}$) is always bounded from below by the 45-degree line.[15] As a consequence, when the likelihood curve under F is adequately high, which can be guaranteed by the "highly informative" conditions in Theorem 2, the averaged likelihood curve (marked by $\mathcal{F}_0^{avg}$) will enter the cascade set. This implies that a cascade will be triggered by an action 1 whenever beliefs enters C_ε^1. It can be verified that whenever a cascade doesn't occur, there is a positive probability for beliefs to enter C_ε^1 and trigger a cascade, which implies that an information cascade occurs with probability 1. Therefore, it suffices to focus on the region C_ε^1 and restrict the tail property of F in order to produce a cascade, which explains the second point.

Remark 1. Although Theorem 2 requires the perception of a specific F, we can't simply interpret F as the "effective model" perceived by the MEU individuals. In fact, learning under $\mathcal{F}_0$ is often not observationally equivalent to learning under any $F \in \mathcal{F}_0$.[16]

Another common misinterpretation is that to have a cascade, individuals must be misspecified in the sense that they consider specific incorrect DGPs. However, as will be discussed in Section 7, an information cascade can also occur when individuals are only ambiguity-averse but do not entertain any incorrect models. Hence, information cascades are primarily driven by ambiguity attitude and are robust in the face of model misspecification, rather than being caused by it.

Theorem 2 also implies that standard results featuring the absence of an information cascade may even represent *knife-edge* cases. Below is an example.

Example 7. (ε-perturbation set) Suppose that the model set is as follows

$$\mathcal{F}_0 = (1-\varepsilon)\,G + \varepsilon\mathcal{F} \equiv \left\{F : F = (1-\varepsilon)\,G + \varepsilon F', \text{where } F' \in \mathcal{F}\right\},$$

[15] All these curves intersect at γ because when the public likelihood ratio becomes γ, the society will take action 1 for sure, so an additional action 1 brings no information, and hence γ becomes fixed point.

[16] Recall that the role of F is to provide a lower bound for the upper envelope, but the exact learning dynamics depend on both upper and lower envelopes, which often depends on the detailed structure of $\mathcal{F}_0$.

where $\mathcal{F}$ represents the set of all DGPs on $[1/\gamma, \gamma]$, and $G \in \mathcal{F}$ and $\varepsilon \in (0, 1)$. When $\varepsilon = 0$, we have $\mathcal{F}_0 = \{G\}$, which corresponds to Bayesian social learning, so we can have a variety of learning outcomes—e.g., cascade, herding and action oscillations—depending on the properties of G the true DGP. In sharp contrast, whenever $\varepsilon > 0$, an information cascade occurs almost surely for *all* possible G and true DGPs. It shows that any non-cascade result is not robust with respect to arbitrarily small perturbations.

Necessary Condition for Cascades

Notice that the conditions in Theorem 2 are sufficient but not necessary. A simple necessary condition for cascades is that $\mathcal{F}_0$ must contain a DGP whose likelihood curve enters the cascade set, that is, $\mathcal{F}_0$ must contain a DGP that *violates* the IHRP; e.g., F_2 in Figure 1a. This is because if all DGPs satisfy the IHRP, then the learning dynamics under each possible DGP are trapped in the non-cascade set, so the average likelihood curve is also trapped in the set, which implies that a cascade will not occur.

This condition, however, is *not sufficient*. In Figure 3a, the likelihood curve under $\hat{F}$ enters the cascade set; however, when individuals consider $\mathcal{F}_0 = \left\{\hat{F}, F_\emptyset\right\}$, where $F_\emptyset$ denotes the uninformative DGP, an information cascade will not occur, as shown in Figure 3b. Intuitively, $\hat{F}$ is inadequately informative in Theorem 2's sense, so when it is perceived with a highly uninformative DGP, the non-cascade force dominates and a cascade will not occur.

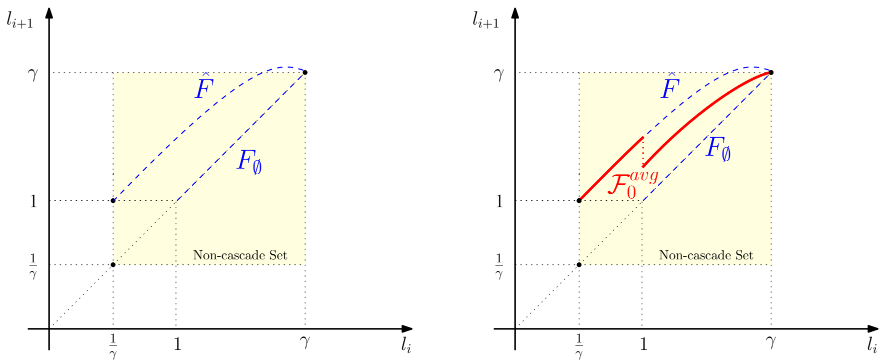
Remark 2. It is not known if there exists a simple necessary and sufficient condition for information cascades even in the standard case. The closest condition in the literature is perhaps the IHRP or log-concavity condition which is necessary and sufficient for the posterior monotonicity property, i.e., the posterior likelihood ratio is increasing in the prior likelihood ratio.[17] In Section 2 of the Supplementary Material, I provide a similar necessary and sufficient condition for the *average posterior monotonicity property*: The average posterior likelihood ratio is monotonic in the prior likelihood ratio **iff** the average hazard ratio of every DGP in $\mathcal{F}_0$ is strictly increasing. This condition ensures that a cascade will not occur under ambiguity. In other words, to have a cascade, the average posterior likelihood ratio must be non-monotonic.

6 Incorrect Herding with Unbounded Signals

This section extends Theorem 1 to unbounded signals. Note that information cascade is a restrictive concept for unbounded signals, so it is difficult to occur under small ambiguity as in the last section.[18] This section finds that we can still establish results parallel to the bounded-signal case

[17] The posterior monotonicity is a highly relevant concept because it implies that a cascade will not occur.

[18] When signals are unbounded, a cascade requires individuals to ignore arbitrarily strong signals, so the model set must also be "unbounded" in the sense that it contains arbitrarily informative DGP. It is also worth noting that although a cascade can't occur with "bounded" model set, it can occur with preferences less extreme than MEU as shown in Section 8.1.



(a) (b)

Figure 3: Necessary Conditions for Cascades

by focusing on a weaker but qualitatively similar concept—herding. I show that under moderate conditions, herding occurs almost surely, and an incorrect herd occurs with a strictly positive probability. In some interesting situations, the complete learning result featured by Smith and Sørensen (2000) no longer holds as long as there is a grain of ambiguity.

6.1 Sufficient Conditions for Herding

The following theorem provides a sufficient condition for herding.

Theorem 3. (Herding with unbounded signals). *Suppose that for all i, $\overline{F}_i^0(x) \leq ax^\alpha$ with $a, \alpha > 0$ as $x \to 0$. If there exists some $F \in \mathcal{F}_0$ such that $x^p = o\left(F^0(x)\right)$ as $x \to 0$ for some $p \in (0, \alpha)$, herding occurs $\mathbb{P}^*$-almost surely, and incorrect herding occurs with a $\mathbb{P}^*$-strictly positive probability.*

Theorem 3 is a parallel statement of Theorem 2 when signals are unbounded. The restriction $\overline{F}_i^0(x) \leq ax^\alpha$ says that the true DGP is bounded by some power function. This condition is relatively weak and can cover many interesting DGPs.[19] The condition $x^p = o\left(F^0(x)\right)$ means that the tail of $F^0(x)$ is sufficiently fat, so it is parallel to the conditions in Theorem 2. Thus, Theorem 3 says that when individuals consider a highly informative DGP, herding occurs almost surely and the herding can be incorrect. The intuition is similar to that behind Theorem 2—whenever individuals perceive a highly informative DGP, it creates a strong herding force that cannot be offset by any other model, so an incorrect herding can emerge. Also, Theorem 3 only requires $\mathcal{F}_0$ to contain a specific model without imposing restrictions on other structures, so it is easy to hold in many interesting cases as implied by the following corollary.

Corollary 1. *Suppose that signals are i.i.d. with $\overline{F}^0(x) = O(x^\alpha)$ with $\alpha > 0$ as $x \to 0$. If there exists some $F^0 \in \mathcal{F}_0$ such that $F^0 = O(x^{\alpha-\varepsilon})$ with $\varepsilon \in (0, \alpha)$ as $x \to 0$, then herding occurs $\mathbb{P}^*$-almost surely, and incorrect herding occurs with a $\mathbb{P}^*$-strictly positive probability.*

[19] For example, normal distributions, power law distributions and many others. One class of DGPs that violates the condition is $\overline{F}_i^0(x) \sim 1/|\log(x)|$ as $x \to 0$ (Rosenberg and Vieille, 2019).

Corollary 1 says that if the true DGP has a "power tail," arbitrarily small ambiguity in the power of $\overline{F}^0$ is sufficient to trigger an incorrect herding.[20] This suggests that within this class of models, complete learning is not robust in a sense. Below is a concrete example.

Example 8. For better exposition, this example focuses on the actual signal s_i (instead of normalized signals). Consider the signal space $\mathcal{S} = (0,1)$; signals are i.i.d., and the DGP takes the form of $g_m = \left(g_m^0, g_m^1\right)$, where

$$g_m^0(s) = (m+1)(1-s)^m \text{ and } g_m^1(s) = (m+1)s^m \qquad \text{for } s \in (0,1).$$

The true DGP is $g_{m_0}^\theta$ where $m_0 > 0$. It is easy to see that signals are unbounded—i.e., $g_{m_0}^0(s)/g_{m_0}^1(s)$ is unbounded, so complete learning will occur if individuals precisely perceive the true DGP. Suppose that individuals are ambiguous and perceive a set $M_\varepsilon = [m_0 - \varepsilon, m_0 + \varepsilon] \subset \mathbb{R}_+$. Corollary 1 implies that for *all* $\varepsilon > 0$, complete learning no longer holds, and the society will settle on an incorrect action with a strictly positive probability.

Sketch of the Proof

For simplicity, I sketch the proof for Corollary 1, and the proof for Theorem 3 is similar. The main technical difficulty is that under ambiguity, posteriors are no longer martingales, so we cannot apply the martingale techniques as in Smith and Sørensen (2000). This paper's approach is to evaluate the probability of both types of herding and check whether the probability of each herding is positive. First note that the probability of an incorrect herd starting from the first individual is

$$\lim_{i\to\infty} \mathbb{P}^*(a_1 = \ldots = a_i = 1) = \prod_{i=1}^{\infty}\left(1 - \overline{F}^0(1/r_i)\right),$$

where r_i denotes the average public likelihood after $h_i = (1, \ldots, 1)$. The R.H.S. is positive if and only if $\sum_i \overline{F}^0(1/r_i) < \infty$, which is further equivalent to $\sum_i \left(\frac{1}{r_i}\right)^\alpha < \infty$, where the equivalence comes from the facts that: (i) $r_i \to +\infty$ during an action-1 herd,[21] and (ii) $\overline{F}^0(x)$ can be approximated by x^α when x is close to 0. Next, we can show that the growing speed of r_i is bounded from below by a power function. Formally, if there exists some $F \in \mathcal{F}_0$ such that $F^0(x) = O(x^\beta)$ as $x \to 0$. Then we must have

$$r_i \geq (i + C)^{1/\beta} \text{ for some } C > 0, \tag{3}$$

when i is large. Note that if $\mathcal{F}_0$ contains a DGP with a smaller power (hence more informative), the R.H.S. of (3) is larger, so r_i can grow faster. Corollary 1 assumes that there is some DGP in

[20] Formally, if $F^0(x) \sim x^\alpha$ as $x \to 0$, I say that F has a *power tail*, and the *power* of F is α.

[21] The intuition is similar to the standard case: During an action-1 herd, and when signals are unbounded, the society believes state 1 is increasing more likely, and the average likelihood ratio goes to infinity.

$\mathcal{F}_0$ with power $\beta = \alpha - \varepsilon$, so (3) implies that

$$\sum_i \left(\frac{1}{r_i}\right)^\alpha \leq \sum_i \frac{1}{(i+C)^{\alpha/\beta}} = \sum_i \frac{1}{(i+C)^{\frac{\alpha}{\alpha-\varepsilon}}} < \infty, \tag{4}$$

which establishes that an incorrect herd occurs with a strictly positive probability.[22] Similarly, we can show that a correct herd also happens with a strictly positive probability. The intuition is that signals supporting the true state are more likely to realize, so the society is more likely to form a correct herd than an incorrect herd. Once we show that both herds happen with a strictly positive probability, we can employ standard 0-1 arguments to show that herding occurs almost surely.

Remark 3. *Why complete learning is not robust.* From (3) and (4), we can get an idea of why complete learning is not robust for power-tail DGPs. It turns out that with power tails, whether complete learning occurs is related to the convergence of some p-series with $p = \alpha/\beta$, where α is the power of the true DGP, and β is the minimum power of perceived DGPs.[23] The fragility of complete learning comes from the property that the p-series changes from divergence to convergence at $p = 1$. The case in Smith and Sørensen (2000) corresponds to $p = 1$, i.e., $\beta \equiv \alpha$, so the infinite series on the R.H.S. of (4) becomes

$$\sum_i \frac{1}{(i+C)^{\alpha/\beta}} = \sum_i \frac{1}{i+C} = \infty, \tag{5}$$

which is consistent with the divergence of $\sum_i \left(\frac{1}{r_i}\right)^\alpha$, which implies the absence of an incorrect herd. However, when individuals face ambiguity and perceive some DGP with power $\beta < \alpha$ regardless of how close it is to α, we have $p = \beta/\alpha > 1$, and the infinite series becomes convergent, so an incorrect herd can emerge.

Conditions for complete learning

Previous discussion shows the fragility of complete learning under ambiguity. It is natural to ask when complete learning occurs. Section 1 of the Supplementary Material provides a **necessary and sufficient** condition for complete learning when DGPs have power tails. Specifically, I show that when the true DGP and all perceived DGPs have power tails and satisfy some regularity conditions, complete learning occurs **iff**: (i) all DGPs' power is weakly greater than the true power (i.e., the power of the true DGP), and (ii) some DGP has a power strictly less than the true power plus 1. That is, when we have

$$\min \mathcal{P}(\mathcal{F}_0) \in [\overline{\mathcal{P}}, \overline{\mathcal{P}} + 1),$$

[22] To be more precise, the summation is taken for large i, i.e., $\sum_{i \geq I}$ for some finite I. But this doesn't change the convergence, so it is omitted for brevity.

[23] A p-series is $S(p) = \sum_n^\infty \frac{1}{n^p}$. The series is convergent if $p > 1$ and divergent when $p \leq 1$.

where $\overline{\mathcal{P}}$ is the power of the true DGP, and $\min \mathcal{P}(\mathcal{F}_0)$ denotes the minimum power of perceived DGPs.[24] The intuition is as follows. To achieve complete learning, we need to exclude two sources of incomplete learning—incorrect herding and action non-convergence. First, to exclude incorrect herding, $\mathcal{F}_0$ cannot contain any highly informative DGP, i.e., the power of all DGPs must be greater than some lower bound. This comes from the fact that the cascading force from a highly informative DGP can't be offset by any other DGP due to the aforementioned asymmetric effect. Second, to exclude action non-convergence, $\mathcal{F}_0$ must contain some DGP that is adequately informative, i.e., $\mathcal{F}_0$ must contain some DGP whose power is less than some upper bound. This is because if all DGPs under consideration are of little precision, individuals tend to think others' actions are not very informative, so they will constantly break a herd, which prevents actions from converging. We can further show that the upper bound and lower bound correspond to $\overline{\mathcal{P}}$ and $\overline{\mathcal{P}}+1$ respectively.

7 Cascades and Herding under Consistent Ambiguity

Previous sections allow individuals to perceive an arbitrary $\mathcal{F}_0$, so the perception of incorrect models and MEU preferences can appear together, making it not obvious how ambiguity itself can drive a cascade. To observe the direct effect of ambiguity attitude, this section discusses a special type of ambiguity in which model perceptions must be **consistent** with the true probability, so individuals are both ambiguous and correctly specified. This section adopts the following assumption.

Assumption 1. (Consistent Ambiguity) *Suppose that $\overline{F}_i$ and $\mathcal{F}_0$ jointly satisfy*

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

where $\mu \in \Delta(\mathbb{F})$ is a second-order distribution over DGPs.

As discussed in Example 6, Assumption 1 can describe the situation where everyone's DGP is distributed according to some second-order distribution μ, and individuals correctly perceive all possible DGPs. Besides, all probabilities are calculated according to the ex ante probability, i.e., the true DGP $\overline{F}_i$ is the ex ante signal distribution. In this case, ambiguity manifests itself in an objective manner—where the set of perceived DGPs is equal to the set of ex ante possible DGPs. Therefore, we can separate ambiguity attitudes from the effect of incorrect model perceptions.

7.1 Conditions for Information Cascades

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

[24] A similar condition also appears in Arieli et al. (2023). They study sequential social learning with misspecified model perception, which corresponds to the special case where $\mathcal{F}_0$ is a singleton set.

Proposition 2. (Cascade with consistent ambiguity). *Suppose that $\overline{F}_i$ and $\mathcal{F}_0$ jointly satisfy Assumption 1, and that either (i) $\mathcal{F}_0 = \mathcal{F}$, or (ii) $\overline{F}_i$ is bounded and $\mathcal{F}_0$ satisfies conditions in Theorem 2, then an information cascade occurs $\mathbb{P}^*$-almost surely.*

Proposition 2 corresponds to special cases of Theorems 1 and 2: 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 2, model perceptions are consistent with the true probability, so it shows that MEU preference itself, instead of incorrect model perceptions, can produce an information cascade under high ambiguity. Below is a concrete example.

Example 1'. Consider the same signal structures as in Example 1. 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, it corresponds to the standard SSLM with *correct* model 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 and do not consider any incorrect models. The key distinction lies in their preferences, highlighting the role of ambiguity attitude itself in triggering a cascade.

Remark 4. Combining all previous results, we can gain a comprehensive understanding of how an information cascade occurs under ambiguity. Proposition 2 demonstrates that individuals with MEU preferences, who consider only correct models, are prone to information cascades.

Theorems 1 and 2 further solidify the occurrence of cascades under general ambiguity. They not only establish that a cascade occurs under correct model perceptions but also confirm that it takes place under arbitrary model perceptions that exhibit sufficient ambiguity, which further demonstrates the robustness of the cascade result.

7.2 Incorrect Herding with Unbounded Signals

Under consistent ambiguity, a cascade also requires extreme ambiguity for unbounded signals. I then show that herding can emerge under moderate ambiguity parallel to Section 6. For tractability, this subsection focuses on the case where DGPs belong to a parametric family with power tails.

Assumption 2. (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 on $\mathcal{A}$, and I use $\mu(\alpha)$ to denote the density function with some abuse of notation.[25] We have the following proposition.

Proposition 3. (Herding with consistent ambiguity). *Suppose that Assumptions 1 and 2 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 3 is parallel to Theorem 3 and employs a similar proof technique. It says that if the second-order distribution μ is controlled by some power function as $\alpha \to \underline{\alpha}$, then an incorrect herd happens with a 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 models. 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 8'. Consider the same setup as in Example 8, 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.

Different from Example 8, here individuals only consider ex ante correct DGPs, yet incorrect herding still emerges, even if ambiguity is arbitrarily small. This further establishes the fragility of complete learning, which can occur even with consistent ambiguity.[26]

8 Other Ambiguity Preferences

The key results of this paper can be extended to a wider class of ambiguity preferences. Below are two important examples: the smooth ambiguity preference and the α-max-min EU preference.

[25] The density function $\mu(\alpha) \equiv \frac{d\mu(F(\cdot,\alpha))}{d\alpha}$.

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

8.1 Smooth Ambiguity Model

The max-min model makes a restrictive assumption whereby decisions only depend on the worst cases. To relax this assumption, I consider an extension in which individuals have the ***smooth ambiguity preferences***, as axiomatized by Klibanoff et al. (2005). Suppose that the model set is $\mathcal{F}_0 = \{F(\cdot, \alpha)\}_{\alpha \in \mathcal{A}}$, where $\mathcal{A} = [\underline{\alpha}, \overline{\alpha}] \subset \overline{\mathbb{R}}$, and that preferences satisfy

$$V_i(a) = \phi^{-1}\left(\int_{\mathcal{A}^{i-1}} \phi\left[\mathbb{E}_{\alpha_1,\ldots,\alpha_{i-1}}\left(U(a,\theta)|I_i\right)\right] d\mu(\alpha_1, \ldots, \alpha_{i-1})\right), \tag{6}$$

where (i) $\mathbb{E}_{\alpha_1,\ldots,\alpha_{i-1}}(\cdot)$ denotes the expectation when the first $i-1$ individuals' DGPs have parameters $\alpha_1, \ldots, \alpha_{i-1}$, and (ii) $\mu(\alpha_1, \ldots, \alpha_{i-1})$ stands for the second-order belief on DGPs, and suppose that μ features i.i.d. distribution and has full support on $\mathcal{A}^{i-1}$, (iii) ϕ denotes the second-order utility function and is strictly increasing, concave, and twice continuously differentiable. To quickly grasp the idea, I first present an illustrative example and then state the formal results later.

Example 1". (Cascades under smooth ambiguity preferences). Consider the same signal structures as in Example 1. Signals are i.i.d. according to $\overline{g}(s|\theta)$, which is the society's true DGP. Suppose that individuals have constant relative ambiguity aversion (CRAA) preferences

$$V_i(a) = \left[\int \left[\mathbb{E}_{\gamma_1,\ldots,\gamma_{i-1}} U(a,\theta)\right]^{1-\rho} d\mu(\gamma_1)\ldots d\mu(\gamma_{i-1})\right]^{\frac{1}{1-\rho}},$$

where ρ is the coefficient of relative ambiguity aversion, and μ satisfies some regularity conditions that will be specified later. Denote by $g_\mu(s|\theta) = \int g_\gamma(s|\theta) d\mu(\gamma)$, which is referred to as the society's perceived DGP. We have the following results.

- *Expected utility.* When $\rho = 0$, individuals have expected-utility preferences. If $g_\mu = \overline{g}$, it corresponds to correct social learning; if $g_\mu = \overline{g}$, it corresponds to misspecified social learning. As discussed earlier, we have different learning outcomes, such as cascades, complete learning, or action oscillations, depending on the properties of g_μ and $\overline{g}$.

- *Ambiguity aversion.* As $\rho \to +\infty$, we can show that the probability that an information cascade occurs approaches 1 regardless of what the learning outcome is in the expected-utility case. That is, information cascade gradually emerges as the *only* outcome as individuals become increasingly ambiguity averse.

Information Cascades with Smooth Ambiguity Preferences

I then present the general conditions for information cascades under smooth ambiguity preferences. Let $\rho_\phi(u) = -\frac{\phi''(u)}{\phi'(u)}$ denote the coefficient of ambiguity aversion and let $\underline{\rho}_\phi$ be the minimum of $\rho_\phi(u)$ over $[0,1]$. We have the following proposition.

Proposition 4. *Suppose that signals are bounded, and that $\mathcal{F}_0$ satisfies the conditions of Theorem 2. Then for all $\varepsilon > 0$, there exists $\rho_0 < \infty$ such that an information cascade occurs with a $\mathbb{P}^*$-probability greater than $1 - \varepsilon$ for all ϕ with $\underline{\rho}_\phi > \rho_0$.*

Proposition 4 shows an information cascade can occur with an arbitrarily large probability when individuals are sufficiently ambiguity-averse. Therefore, the result under max-min model can serve as a benchmark for high ambiguity aversion. The intuition comes from two facts: (i) the smooth model approaches the max-min model as $\rho_\phi \to \infty$, and (ii) an information cascade occurs in finite time. We can ensure the belief dynamics under the smooth model and the max-min model arbitrarily close up to any finite time by making ϕ_ϕ sufficiently large. As a consequence, the probability of a cascade can also be arbitrarily close to 1. Proposition 4 requires bounded signals. I then show that when signals are unbounded, an information cascade can still occur under smooth ambiguity preferences.

Assumption 3. (Adequate Ambiguity) *Let $\chi(\alpha) = \frac{F^0(1,\alpha)}{F^1(1,\alpha)}$. As $\alpha \to \overline{\alpha}$, we have: (i) $\chi(\alpha) \to +\infty$, and (ii) $\mu(\alpha) \geq C \cdot \chi^{-k}(\alpha)$ for some $C, k > 0$.*

Note that $\chi(\alpha) = +\infty$ means that $F(x, \alpha)$ is perfectly informative, so Assumption 3 (i) says that perceived DGPs can be arbitrarily informative. Assumption 3 (ii) imposes a lower bound on the right tail of second-order belief, which means that individuals believe that highly informative DGPs are realized with adequately high probability.

Proposition 5. *Suppose that Assumption 3 holds, and that ϕ is CRAA with coefficient ρ.*[27] *There exists $\rho_0 < +\infty$ such that an information cascade occurs $\mathbb{P}^*$-almost surely if $\rho > \rho_0$.*

Proposition 5 imposes no restriction on the true DGP, so an information cascade occurs for all DGPs—regardless of whether signals are bounded or unbounded—if there is adequate ambiguity in Assumption 3's sense, and if individuals are sufficiently ambiguity-averse in the CRAA sense. Interestingly, Proposition 5 shows that an information cascade with unbounded signals is less extreme than it appears. Recall that with MEU preferences, a cascade occurs with unbounded signals because individuals use the perfectly informative DGP to evaluate the worst case, which represents a very extreme case (e.g., Theorem 1). However, with smooth ambiguity preferences, a cascade can occur in less extreme cases in which the perfectly informative DGP carries zero weight.

8.2 α-MEU Model

The occurrence of a cascade can even go beyond ambiguity aversion. This subsection considers another extension in which individuals have ***α-maxmin expected utility*** ($\alpha-$MEU) preferences (Hurwicz, 1951; Ghirardato et al., 2004). With this class of preferences, individual i's utility is

$$V_i(a) = \alpha \cdot \inf_{\pi \in \Pi_i} \mathbb{E}_\pi U(a, \theta) + (1 - \alpha) \cdot \sup_{\pi \in \Pi_i} \mathbb{E}_\pi U(a, \theta),$$

[27] If ϕ is constant relative ambiguity aversion (CRAA), it satisfies $\phi(x) = \frac{x^{1-\rho}}{1-\rho}$ if $\rho > 0$ and $\rho \neq 1$, and $\phi(x) = \ln(x)$ if $\rho = 1$.

where $\alpha \in [0,1]$. Here α represents the degree of an individual's pessimism, where $\alpha = 1$ corresponds to the MEU model, and $\alpha = 0$ corresponds to the max-max EU model. We have the following proposition.

Proposition 6. *All previous results under MEU preferences hold for $\alpha-$MEU preferences.*

Proof. It can be verified that the decision rule under α-MEU preferences is the same as that under MEU preferences, i.e., individuals choose action 1 if $\lambda_i \cdot r_i > 1$ and action 0 if otherwise, so all action dynamics are identical. □

Proposition 6 implies that an information cascade can also occur when individuals are ambiguity-loving. Suppose that individuals have max-max preferences (i.e., $\alpha = 0$), then under high ambiguity, we still have the asymmetric forces between herding and contrarian: Every action in a herd can be interpreted as highly informative, so the best-case utility of herding can be very high; in contrast, the best-case utility of breaking a herd cannot exceed the case where previous actions contain no information, which can also lead to a cascade. To accommodate ambiguity-loving, a more general statement should be that an information cascade emerges when: (i) there is sufficient ambiguity—i.e., there are sufficiently many models, and (ii) individuals are sufficiently ambiguity-sensitive—i.e., their decisions are adequately influenced by the best or the worst outcomes.

Remark 5. It is worth noting that the equivalence in dynamics between MEU and α-max-min relies crucially on the binary action space. Suppose that we have general action space, then ambiguity attitudes can affect which actions will be taken in the end. For example, if individuals are ambiguity-averse, the society may settle on safe actions, while if individuals are ambiguity-loving, the society will select riskier actions.[28]

9 Discussion and Extensions

In this section, I summarize how social learning under ambiguous model perceptions differs from that under alternative model setups and also discuss some extensions.

Model uncertainty vs model certainty. Previous literature predominantly focuses on social learning with model certainty, including both learning under correct models and learning under model misspecification. Differently, this paper introduces *model uncertainty*, which describe the situations where individuals can't pin down a specific DGP. In the paper, I allow individuals to perceive an arbitrary model set, so this paper's framework can encompass model certainty as special cases. Additionally, the paper addresses the question of which social learning outcomes remain robust under alternative model specifications. As discussed earlier, with model certainty, the social learning outcome can depend intricately on the statistical properties of the true DGPs and their perceptions. This paper instead establishes information cascades as the robust social

[28] An earlier version of this paper also shows that if there is a safe action with a constant payoff (say 1/2), and if individuals have MEU preferences, the society will form an information cascade of this safe action almost surely. However, a cascade of the safe action never appears with ambiguity-loving individuals.

learning outcome when individuals have robustness concerns as in Wald (1950) and Hansen and Sargent (2001) by considering a broad range of models and maximizing the utility in the worst-case scenario.

Remark 6. One may wonder how results will change if we describe model uncertainty in a Bayesian way in which individuals form a prior over models. It turns out that the social learning outcome will depend on the properties of the prior, just as it depends on the model specification in the case of model certainty. [29] Therefore, it is not straightforward to discuss which result is robust as all of them are prior-dependent.

MEU vs expected utility. This paper also underscores the importance of studying *non-expected utility* preferences in the context of social learning by showing that social learning outcomes with MEU preferences can differ significantly from those with expected-utility preferences. To emphasize the impact of MEU preferences, it is beneficial to disentangle the effects of incorrect model perception from the effects of non-expected utility preferences. To achieve this goal, Section 7 delves into a special case of the benchmark model in which individuals face consistent ambiguity, allowing them to maintain ambiguity and correct specifications simultaneously. We see that under consistent ambiguity, despite the various outcomes that can arise with expected utility, an information cascade will happen almost surely with MEU preferences, which shows that MEU preferences itself (in contrast to the perception of incorrect models) can drive a cascade. To gain a better understanding of the role of ambiguity attitudes in triggering cascades, the paper further discusses smooth ambiguity preferences, wherein preferences can transition continuously from expected utility to MEU. From this discussion, it is observed that the probability of an information cascade approaches 1 as the degree of ambiguity aversion approaches infinity, providing a clearer picture of the role of ambiguity attitude in social learning. To summarize the previous discussion, the paper highlights that incorporating non-expected utility can be a promising direction for future research in social learning.

Extensions. I also discuss the following extensions in the Supplementary Materials. First, this paper focuses on the standard setup in which the state and action space are binary, but similar insights can still apply in cases with multiple states and actions.[30] Second, the paper assumes that all individuals share a common model set, and I discuss an extension in which individuals hold heterogeneous model sets. For example, some individuals may have less ambiguity than others by considering a smaller model set. Third, the benchmark model assumes that individuals update beliefs model-by-model, and I consider an extension in which individuals follow the α-maximum likelihood rule as in Epstein and Schneider (2007). Lastly, the paper assumes that individuals

[29] For example, if signals are unbounded and the prior assigns a strictly positive probability to the true DGPs, complete learning will occur due to a similar argument in Kalai and Lehrer (1993), but for other priors, complete learning may not occur; similarly, when signals are bounded, the occurrence of a cascade also depends on properties of the prior, e.g., how much weight the prior attaches to the cascading DGP.

[30] Arieli and Mueller-Frank (2021) extended the standard SSLM to allow for general state and action space. Differently, they focus on correct Bayesian agents, so similar techniques cannot be applied here, which prevents this paper from achieving similar generality.

are certain about the network structure. I present an extension in which individuals also face ambiguity about the network structure, meaning they do not know what their predecessors can actually observe.

10 Related Literature

This paper contributes to the growing literature on learning under ambiguity. Most works in this thread of literature focus on individual learning. Marinacci (2002) and Marinacci and Massari (2019) study in an individual learning problem in terms of whether ambiguity will fade away asymptotically. Epstein and Schneider (2007) introduce an α-maximum likelihood learning rule and investigate a dynamic portfolio choice problem. Battigalli et al. (2019) study a learning problem in which data are endogenously generated from an experimentation process. Fryer Jr et al. (2019) and Chen (2022) study learning problems where individuals are biased in interpreting ambiguous information. This paper complements the literature by investigating a social learning problem, where informational ambiguity seems to occur very naturally. A relevant paper is by Ford et al. (2013), who study a sequential trading model in which traders face ambiguity and have neo-additive capacity expected utility (CEU) preferences (Chateauneuf et al., 2007). They show that ambiguity can produce both herding and contrarian. This arises from the property that the CEU is bounded away from 0 and 1, but the ask-bid prices can fully adjust to 0 and 1, so the discrepancy provides room for herding and contrarian. In contrast, this paper's result comes from a different mechanism that employs the asymmetry between cascade and non-cascade under ambiguous information; also, the ambiguity in the paper mainly produces herding but not contrarian. In addition to the aforementioned applications, learning under ambiguity is also examined in recent works in decision theory—e.g., Cheng (2022); Kovach (2021); and Tang (2022);—and experimental economics; e.g., De Filippis et al. (2022) and Epstein et al. (2019).

This paper is also related to the literature on social learning with misspecified models. Bohren (2016) and Bohren and Hauser (2021) examine a sequential social learning problem in which individuals misspecify the true model. Bohren (2016) finds that different model specifications can lead to different learning outcomes—e.g., complete learning, incomplete learning, and cyclical actions. Bohren and Hauser (2021) incorporate these results in a more general framework. They find that complete learning is robust with respect to small misspecifications, which stands in contrast to this paper's finding that complete learning may be non-robust. The difference is driven by their assumption that the society has a positive fraction of "autarkic agents" who only act according to their private signals. (Frick et al., 2020a,b) also find that complete learning is not robust but in different settings. Specifically, Frick et al. (2020a) consider a social learning problem in which the state space is continuous and individuals with different preferences randomly meet with each other. Frick et al. (2020b) propose a local martingale-based approach and show the fragility of sequential social learning in an environment in which signals are bounded and individuals have heterogeneous risk preferences. Arieli et al. (2023) investigate the efficiency of sequential social learning with misspecified model perceptions. Compared with previous papers on misspecified social learning, this paper assumes that individuals face model uncertainty and entertain a set of models; besides, this paper studies social learning with non-expected utility preferences.

This paper also connects to the literature on social learning with non-Bayesian agents. The literature shows that incorrect learning can emerge if individuals follow some naive learning rules—for example, when they do not fully account for predecessors' inferences (Eyster and Rabin, 2010), when they follow a coarse inference rule (Guarino and Jehiel, 2013), or when they follow some average rule to aggregate the opinions from others (DeMarzo et al., 2003; Molavi et al., 2018; Dasaratha and He, 2020). In this paper, individuals are not naive, and they understand how others make inferences. The paper's deviation from the Bayesian paradigm is mainly created by ambiguity and ambiguity preferences.

11 Conclusion

This paper investigates a sequential social learning problem in which individuals face ambiguity regarding other people's DGPs. In contrast to previous findings where various learning outcomes can emerge depending on modeling details, this paper establishes information cascades as the only robust outcome under ambiguity. Specifically, under sufficient ambiguity, an information cascade almost surely occurs without regard to many details of the learning environment, such as the statistical properties of the actual signal-generating processes or the presence of a particular type of model specification. Interestingly, this paper further demonstrates that some results featuring non-cascades are fragile when subjected to small perturbations in ambiguity. The paper focuses on the sequential social learning model, but it would also be intriguing to investigate how individuals learn in other environments, such as general networks, repeated interactions, and situations with heterogeneous preferences, etc.

Bibliography

Anderson, Lisa R and Charles A Holt (1997) "Information cascades in the laboratory," *American Economic Review*, 847–862. 11

Arieli, Itai, Yakov Babichenko, Stephan Müller, Farzad Pourbabaee, and Omer Tamuz (2023) "The Hazards and Benefits of Condescension in Social Learning," *arXiv preprint arXiv:2301.11237*. 6, 19, 26

Arieli, Itai and Manuel Mueller-Frank (2021) "A general analysis of sequential social learning," *Mathematics of Operations Research*, 46 (4), 1235–1249. 25

Banerjee, Abhijit V (1992) "A simple model of herd behavior," *Quarterly Journal of Economics*, 107 (3), 797–817. 2, 6, 10

Battigalli, Pierpaolo, Alejandro Francetich, Giacomo Lanzani, and Massimo Marinacci (2019) "Learning and self-confirming long-run biases," *Journal of Economic Theory*, 183, 740–785. 26

Bikhchandani, Sushil, David Hirshleifer, Omer Tamuz, and Ivo Welch (2021) "Information cascades and social learning,"Technical report, National Bureau of Economic Research. 2

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. 2, 6, 10, 11, 20

Bohren, J Aislinn (2016) "Informational herding with model misspecification," *Journal of Economic Theory*, 163, 222–247. 2, 26

Bohren, J Aislinn and Daniel N Hauser (2021) "Learning with heterogeneous misspecified models: Characterization and robustness," *Econometrica*, 89 (6), 3025–3077. 2, 6, 26

Çelen, Boğaçhan and Shachar Kariv (2004) "Distinguishing informational cascades from herd behavior in the laboratory," *American Economic Review*, 94 (3), 484–498. 11

Chateauneuf, Alain, Jürgen Eichberger, and Simon Grant (2007) "Choice under uncertainty with the best and worst in mind: Neo-additive capacities," *Journal of Economic Theory*, 137 (1), 538–567. 26

Chen, Jaden Yang (2022) "Biased learning under ambiguous information," *Journal of Economic Theory*, 105492. 26

Cheng, Xiaoyu (2022) "Relative maximum likelihood updating of ambiguous beliefs," *Journal of Mathematical Economics*, 99, 102587. 26

Dasaratha, Krishna and Kevin He (2020) "Network structure and naive sequential learning," *Theoretical Economics*, 15 (2), 415–444. 27

De Filippis, Roberta, Antonio Guarino, Philippe Jehiel, and Toru Kitagawa (2022) "Non-Bayesian updating in a social learning experiment," *Journal of Economic Theory*, 199, 105188. 26

DeMarzo, Peter M, Dimitri Vayanos, and Jeffrey Zwiebel (2003) "Persuasion bias, social influence, and unidimensional opinions," *Quarterly Journal of Economics*, 118 (3), 909–968. 27

Eichberger, Jürgen, Simon Grant, and David Kelsey (2016) "Randomization and dynamic consistency," *Economic Theory*, 62 (3), 547–566. 9

Ellsberg, Daniel (1961) "Risk, ambiguity, and the Savage axioms," *Quarterly Journal of Economics*, 643–669. 3

Epstein, Larry G, Yoram Halevy et al. (2019) *Hard-to-interpret signals*: University of Toronto, Department of Economics. 26

Epstein, Larry G and Martin Schneider (2007) "Learning under ambiguity," *Review of Economic Studies*, 74 (4), 1275–1303. 25, 26

Eyster, Erik and Matthew Rabin (2010) "Naive herding in rich-information settings," *American Economic Journal: Microeconomics*, 2 (4), 221–43. 27

Ford, JL, David Kelsey, and Wei Pang (2013) "Information and ambiguity: herd and contrarian behaviour in financial markets," *Theory and Decision*, 75 (1), 1–15. 26

Frick, Mira, Ryota Iijima, and Yuhta Ishii (2020a) "Misinterpreting others and the fragility of social learning," *Econometrica*, 88 (6), 2281–2328. 26

——— (2020b) "Stability and robustness in misspecified learning models." 26

Fryer Jr, Roland G, Philipp Harms, and Matthew O Jackson (2019) "Updating beliefs when evidence is open to interpretation: Implications for bias and polarization," *Journal of the European Economic Association*, 17 (5), 1470–1501. 26

Ghirardato, Paolo, Fabio Maccheroni, and Massimo Marinacci (2004) "Differentiating ambiguity and ambiguity attitude," *Journal of Economic Theory*, 118 (2), 133–173. 23

Gilboa, Itzhak and David Schmeidler (1989) "Maxmin Expected Utility with Non-Unique Prior," *Journal of Mathematical Economics*, 18, pp–141. 3, 8

Golub, Benjamin and Evan Sadler (2017) "Learning in social networks," *Available at SSRN 2919146*. 2

Guarino, Antonio and Philippe Jehiel (2013) "Social learning with coarse inference," *American Economic Journal: Microeconomics*, 5 (1), 147–74. 27

Hansen, Lars Peter et al. (2014) "Uncertainty outside and inside economic models," *Journal of Political Economy*, 122 (5), 945–87. 2

Hansen, Lars Peter and Massimo Marinacci (2016) "Ambiguity aversion and model misspecification: An economic perspective," *Statistical Science*, 31 (4), 511–515. 2, 7

Hansen, Lars Peter and Thomas J Sargent (2001) "Robust control and model uncertainty," *American Economic Review*, 91 (2), 60–66. 3, 7, 25

Herrera, Helios and Johannes Hörner (2012) "A necessary and sufficient condition for information cascades." 2

Hurwicz, Leonid (1951) "Some specification problems and applications to econometric models," *Econometrica*, 19 (3), 343–344. 23

Kalai, Ehud and Ehud Lehrer (1993) "Rational learning leads to Nash equilibrium," *Econometrica*, 1019–1045. 25

Klibanoff, Peter, Massimo Marinacci, and Sujoy Mukerji (2005) "A smooth model of decision making under ambiguity," *Econometrica*, 73 (6), 1849–1892. 9, 22, 39

Knight, Frank Hyneman (1921) *Risk, uncertainty and profit*, 31: Houghton Mifflin. 2

Kovach, Matthew (2021) "Ambiguity and partial Bayesian updating," *arXiv preprint arXiv:2102.11429*. 26

Marinacci, Massimo (2002) "Learning from ambiguous urns," *Statistical Papers*, 43 (1), 143. 26

——— (2015) "Model uncertainty," *Journal of the European Economic Association*, 13 (6), 1022–1100. 2

Marinacci, Massimo and Filippo Massari (2019) "Learning from ambiguous and misspecified models," *Journal of Mathematical Economics*, 84, 144–149. 26

Molavi, Pooya, Alireza Tahbaz-Salehi, and Ali Jadbabaie (2018) "A theory of non-Bayesian social learning," *Econometrica*, 86 (2), 445–490. 27

Pires, Cesaltina (2002) "A rule for updating ambiguous beliefs," *Theory and Decision*, 53 (2), 137–152. 8

Rosenberg, Dinah and Nicolas Vieille (2019) "On the efficiency of social learning," *Econometrica*, 87 (6), 2141–2168. 16, 35

Saito, Kota (2012) "Subjective timing of randomization and ambiguity,"Technical report, mimeo. 9

Seo, Kyoungwon (2009) "Ambiguity and second-order belief," *Econometrica*, 77 (5), 1575–1605. 9

Smith, Lones and Peter Sørensen (2000) "Pathological outcomes of observational learning," *Econometrica*, 68 (2), 371–398. 2, 5, 11, 16, 17, 18, 20, 31

Smith, Lones, Peter Norman Sørensen, and Jianrong Tian (2021) "Informational herding, optimal experimentation, and contrarianism," *Review of Economic Studies*, 88 (5), 2527–2554. 2

Tang, Rui (2022) "A Theory of Contraction Updating," *Available at SSRN*. 26

Wald, Abraham (1950) "Statistical decision functions.." 3, 25

A Proofs

A.1 Proof of Theorem 1

I first present some useful results below.

Lemma 2. *For all normalized DGP, F, we have*

(1) $F^0(r) > F^1(r)$ except when both are equal to 0 or 1;

(2) $\frac{F^0(r)}{F^1(r)} \geq \frac{1}{r}$ and $\frac{1-F^1\left(\frac{1}{r}\right)}{1-F^0\left(\frac{1}{r}\right)} \geq \frac{1}{r}$ for $r \in (0,\infty)$ (strictly when $F^1(r) > 0$ and $F^0\left(\frac{1}{r}\right) < 1$);

(3) $\frac{F^0(r)}{F^1(r)}$ and $\frac{1-F^1\left(\frac{1}{r}\right)}{1-F^0\left(\frac{1}{r}\right)}$ are weakly decreasing (strictly on supp(F)).

Proof. See Lemma A.1 in Smith and Sørensen (2000). □

Lemma 3. *Suppose that $\gamma < \infty$, and that for all $r_i \in \left(\frac{1}{\gamma},\gamma\right)$, there exists some $\beta > 1$ such that*

$$\frac{r_{i+1}}{r_i}\begin{cases}\geq \beta & \text{if } a_i = 1\\ \leq 1/\beta & \text{if } a_i = 0\end{cases},$$

then an information cascade occurs $\mathbb{P}^*$-almost surely.

Proof. Suppose that for all $r_i \in \left(\frac{1}{\gamma},\gamma\right)$, the ratio $\frac{r_i}{r_{i+1}}$ is bounded away from 1. Then, there exists some $K < \infty$ such that K consecutive action θ will bring r_i into the cascade set C_θ and trigger an information cascade of action θ. Specifically, when $r_i \geq 1$, K consecutive signals $\lambda_i, \lambda_{i+1}, ..., \lambda_{i+K-1} > 1$ lead to $a_i = a_{i+1} = ... = a_{i+K-1} = 1$ and lead to a cascade of action 1. Also notice that

$$\frac{\mathbb{P}^*(\lambda_i > 1)}{1-\mathbb{P}^*(\lambda_i > 1)} = \frac{1-\overline{F}^0(1)}{\overline{F}^0(1)} = \frac{\overline{F}^1(1)}{\overline{F}^0(1)} = \frac{\int_{1/\gamma}^1 x d\overline{F}^0(x)}{\int_{1/\gamma}^1 d\overline{F}^0(x)} \geq \frac{1}{\gamma} \Rightarrow \mathbb{P}^*(\lambda_i > 1) \geq \frac{1}{1+\gamma}, \tag{7}$$

where the second equality comes from the symmetry of $\overline{F}$, and the third equality comes from the definition of normalized signal $x = \frac{d\overline{F}^1(x)}{d\overline{F}^0(x)}$. As a result, we have

$$\mathbb{P}^*(\text{Cascade}|r_i \geq 1) \geq \mathbb{P}^*(\lambda_i, \lambda_{i+1}, ..., \lambda_{i+K-1} > 1|r_i \geq 1) \geq \left(\frac{1}{1+\gamma}\right)^K > 0, \tag{8}$$

and similarly,

$$\mathbb{P}^*(\text{Cascade}|r_i < 1) \geq \left(\frac{\gamma}{1+\gamma}\right)^K > 0. \tag{9}$$

Levy's 0-1 Law shows that as $i \to \infty$, we have

$$\mathbb{P}^*(\text{Cascade}|h_i) \to \mathbb{P}^*(\text{Cascade}|h_\infty) = 1_{\text{Cascade}} \in \{0,1\} \quad \mathbb{P}^*\text{-almost surely.}$$

(8) and (9) imply that $\mathbb{P}^*(\text{Cascade}|h_i) > \left(\frac{1}{1+\gamma}\right)^K > 0$ for all i, so we must have $1_{\text{Cascade}} = 1$ $\mathbb{P}^*$-almost surely—i.e., a cascade almost surely happens. □

Proof of Theorem 1

Now we are ready to prove Theorem 1. We first show that the dynamics of the average likelihood ratio satisfy the following condition.

Lemma 4. *When $\mathcal{F}_0 = \mathcal{F}$, for all $r_i \in \left(\frac{1}{\gamma}, \gamma\right)$, we have*

$$\frac{r_{i+1}}{r_i}\begin{cases}\geq \sqrt{\gamma} & \text{if } a_i = 1\\ \leq \frac{1}{\sqrt{\gamma}} & \text{if } a_i = 0\end{cases}.$$

Proof. If $a_i = 1$, we have

$$r_{i+1} = \sqrt{\sup_{F_i \in \mathcal{F}_0} \frac{1 - F_i^1\left(\frac{1}{r_i}\right)}{1 - F_i^0\left(\frac{1}{r_i}\right)} \times \inf_{F_i \in \mathcal{F}_0} \frac{1 - F_i^1\left(\frac{1}{r_i}\right)}{1 - F_i^0\left(\frac{1}{r_i}\right)}} \times r_i.$$

Let F_γ be the DGP such that $\text{supp}(F_\gamma) = \left\{\gamma, \frac{1}{\gamma}\right\}$, i.e., the "most informative" DGP that only generates signals with the highest and the lowest likelihood ratios. For all $r_i \in \left(\frac{1}{\gamma}, \gamma\right)$, we have

$$\sup_{F_i \in \mathcal{F}_0} \frac{1 - F_i^1\left(\frac{1}{r_i}\right)}{1 - F_i^0\left(\frac{1}{r_i}\right)} \geq \frac{1 - F_\gamma^1\left(\frac{1}{r_i}\right)}{1 - F_\gamma^0\left(\frac{1}{r_i}\right)} = \gamma. \tag{10}$$

From Lemma 2 (1), we know that for all $r_i \in \left(\frac{1}{\gamma}, \gamma\right)$,

$$\inf_{F_i \in \mathcal{F}_0} \frac{1 - F_i^1\left(\frac{1}{r_i}\right)}{1 - F_i^0\left(\frac{1}{r_i}\right)} \geq 1. \tag{11}$$

Combining (10) and (11), we obtain $r_{i+1} \geq \sqrt{\gamma} \times r_i$ for all $r_i \in \left(\frac{1}{\gamma}, \gamma\right)$, when $a_i = 1$. The discussion of $a_i = 0$ is symmetric. □

It can be seen that Lemma 4 then implies Theorem 1 immediately. This is because if signals are unbounded, i.e., $\gamma = \infty$, Lemma 4 implies that $r_1 = \begin{cases}\infty & \text{if } a_1 = 1\\ 0 & \text{if } a_1 = 0\end{cases}$, so a cascade occurs immediately after the first action. Suppose signals are bounded, then Lemma 4 satisfies the condition in Lemma 3, so an information cascade occurs almost surely.

A.2 Proof of Theorem 2

Proof. **Proof of Theorem 2 (1):** Suppose that there is $F \in \mathcal{F}_0$, which is discrete at γ. Denote $p = F^0\left(\frac{1}{\gamma}\right) > 0$, which is the probability that F^0 puts on $\frac{1}{\gamma}$. Suppose that $a_i = 1$, for $r_i \in \left(\frac{1}{\gamma}, \gamma\right)$, we have:

$$\bar{l}_{i+1} = \bar{l}_i \times \sup_{F_i \in \mathcal{F}_0} \frac{1 - F_i^1\left(\frac{1}{r_i}\right)}{1 - F_i^0\left(\frac{1}{r_i}\right)} \geq \bar{l}_i \times \frac{1 - F^1\left(\frac{1}{r_i}\right)}{1 - F^0\left(\frac{1}{r_i}\right)} \geq \bar{l}_i \cdot \left[\lim_{r \to \gamma} \frac{1 - F^1\left(\frac{1}{r}\right)}{1 - F^0\left(\frac{1}{r}\right)}\right] = \bar{l}_i \cdot \frac{1 - \frac{1}{\gamma} \cdot p}{1 - p}, \quad (12)$$

where the second inequality comes from Property (3) in Lemma 2, and the last equality comes from the discreteness of signals. Also, we have $\underline{l}_{i+1} \geq \underline{l}_i$, so

$$r_{i+1} \geq \sqrt{\frac{1 - \frac{1}{\gamma} \cdot p}{1 - p}} r_i \equiv \beta \times r_i.$$

Symmetrically, when $a_i = 0$, we have $r_{i+1} \leq \frac{1}{\beta} \times r_i$. From Lemma 3, an information cascade occurs $\mathbb{P}^*$-almost surely.

Proof of Theorem 2 (2): Suppose that there exists some $F \in \mathcal{F}_0$ such that F^1 is continuously differentiable on $(\gamma - \varepsilon, \gamma)$ with $f^1(\gamma) > \frac{2}{\gamma - 1}$. When F is discrete at γ, an information cascade occurs almost surely, as implied by condition (1). I thus focus on the case in which F is continuous at γ. Suppose that $a_i = 1$; we have

$$r_{i+1} = r_i \cdot \sqrt{\sup_{F_i \in \mathcal{F}_0} \frac{1 - F_i^1\left(\frac{1}{r_i}\right)}{1 - F_i^0\left(\frac{1}{r_i}\right)} \cdot \inf_{F_i \in \mathcal{F}_0} \frac{1 - F_i^1\left(\frac{1}{r_i}\right)}{1 - F_i^0\left(\frac{1}{r_i}\right)}} \geq r_i \cdot \sqrt{\frac{1 - F^1\left(\frac{1}{r_i}\right)}{1 - F^0\left(\frac{1}{r_i}\right)}} \equiv I(r_i).$$

Let $I'(\gamma) \equiv \lim_{\delta \to 0} I'(\gamma - \delta)$ and f^0 denote the density functioon of F^0; then we have

$$I'(\gamma) = \gamma \cdot \left[\frac{1}{\gamma} + \frac{1}{2}\left(f^0(\gamma) - f^1(\gamma)\right)\right] = 1 - \left(\frac{\gamma - 1}{2}\right) f^1(\gamma) < 0,$$

where the second equality comes from $f^0(\gamma) = \frac{1}{\gamma} f^1(\gamma)$. Because F^1 is continuously differentiable on $(\gamma - \varepsilon, \gamma)$, there exists some $\varepsilon_0 > 0$ such that for all $r \in [\gamma - \varepsilon_0, \gamma)$, $I'(r) < 0$. Since $I(\gamma) = \gamma$, we have $I(r) \geq \gamma$ for all $r \in [\gamma - \varepsilon_0, \gamma]$. For all $r_i \in \left(\frac{1}{\gamma - \varepsilon_0}, \gamma - \varepsilon_0\right)$, if $a_i = 1$, we have

$$\frac{r_{i+1}}{r_i} \geq \sqrt{\frac{1 - F^1\left(\frac{1}{r_i}\right)}{1 - F^0\left(\frac{1}{r_i}\right)}} \geq \sqrt{\frac{1 - F^1\left(\frac{1}{\gamma - \varepsilon_0}\right)}{1 - F^0\left(\frac{1}{\gamma - \varepsilon_0}\right)}} > 1,$$

so there exists a $K < \infty$ such that after K action 1, we have $r_i \geq \gamma - \varepsilon_0$. Also note that if $r_i \in [\gamma - \varepsilon_0, \gamma]$ and $a_i = 1$, we have $r_{i+1} \geq I(r_i) \geq \gamma$, so $K + 1$ consecutive action 1 will trigger a cascade of action 1. Similarly, $K+1$ consecutive action 0 will trigger a cascade of action 0. Applying the proof of Lemma 3 again, we can show that r_i will enter the cascade set almost surely. □

A.3 Proof of Theorem 3

A.3.1 Local Stability under Ambiguity

Definition 5. State 0 (or state 1) is *locally stable* if there exists some $r \in \mathbb{R}_{++}$ (or $R \in \mathbb{R}_{++}$) and $\varepsilon > 0$ such that $\mathbb{P}^*_{r_0}(r_i \to 0) > \varepsilon$ (or $\mathbb{P}^*_{r_0}(r_i \to \infty) > \varepsilon$) for all prior set Π_0 with $r_0 < r$ (or $r_0 > R$).

Here $\mathbb{P}^*_{r_0}$ denotes the true probability measure conditional on the average prior likelihood ratio's being r_0. Roughly, state θ is locally stable if posteriors will converge to δ_θ with a strictly positive probability when priors are close to δ_θ. We have the following results.

Lemma 5. *Suppose that $\mathcal{F}_0$ contains a DGP with unbounded signals. Then, a herd of action 0 (or 1) occurs if and only if $r_i \to 0$ (or $r_i \to \infty$).*

Proof. Due to the symmetry, I only prove the result for the herd of action 1. First, suppose that $r_i \to \infty$; then we must have a herd of action 1, because if an action 0 is taken by an individual i, then

$$r_{i+1} = r_i \times \sqrt{\sup_{F_i \in \mathcal{F}_0} \frac{F_i^1(1/r_i)}{F_i^0(1/r_i)} \times \inf_{F_i \in \mathcal{F}_0} \frac{F_i^1(1/r_i)}{F_i^0(1/r_i)}} \leq r_i \times \sqrt{\frac{1}{r_i} \times \frac{1}{r_i}} = 1,$$

which contradicts $r_i \to \infty$. Second, suppose that a herd of action 1 occurs, then

$$r_{i+1} = r_i \times \sqrt{\sup_{F_i \in \mathcal{F}_0} \frac{1 - F_i^1(1/r_i)}{1 - F_i^0(1/r_i)} \times \inf_{F_i \in \mathcal{F}_0} \frac{1 - F_i^1(1/r_i)}{1 - F_i^0(1/r_i)}} \geq r_i,$$

so $\{r_i\}$ is an increasing sequence and has a limit in $\mathbb{R} \cup \{+\infty\}$. If r_i does not diverge to infinity, it must converge to some $R < \infty$. Let F be the unbounded DGP that $\mathcal{F}_0$ contains; then

$$r_{i+1} \geq r_i \times \sqrt{\frac{1 - F^1(1/r_i)}{1 - F^0(1/r_i)}}. \tag{13}$$

Taking the limit on both sides of (13), we obtain $R \geq \sqrt{\frac{1-F^1(1/R)}{1-F^0(1/R)}} \times R$, so $\sqrt{\frac{1-F^1(1/R)}{1-F^0(1/R)}} \leq 1$. Since F has unbounded signals, Lemma 2 (1) implies that $\sqrt{\frac{1-F^1(1/R)}{1-F^0(1/R)}} > 1$, which is a contradiction. As a consequence, $r_i \to \infty$. □

Lemma 6. *Suppose that $\mathcal{F}_0$ contains a DGP with unbounded signals. If both 0 and 1 are locally stable, then (i) both correct and incorrect herding occur with a $\mathbb{P}^*$-strictly positive probability and (ii) herding occurs $\mathbb{P}^*$-almost surely.*

Proof. (i) From the definition of local stability, Lemma 5, and the fact that $\{r_i\}$ is a Markov process, we know that both correct and incorrect herding occur with a strictly positive probability when r_i is sufficiently large or small—that is, $r_i \in C = \{r_i < r\} \cup \{r_i > R\}$ for some $r, R \in (0, +\infty)$. Outside of C, r_i is bounded away from 0 and $+\infty$, so there exists some $K < \infty$ such that K identical actions can bring r_i into C.[31] This further implies that $\{r_i \to 0\}$ and $\{r_i \to \infty\}$ both occur with a strictly positive probability—i.e., both types of herding occur with a positive probability. (ii) Denote by $H = \{r_i \to 0\} \cup \{r_i \to \infty\}$, which denotes the event of herding by Lemma 5. Levy's 0-1 Law implies that $\mathbb{P}^* (H|h_i) \to \mathbb{P}^* (H|h_\infty) = 1_H \in \{0, 1\}$. The arguments in (i) imply that we can find a constant $\delta > 0$ such that for all possible history h_i, $\mathbb{P}^* (H|h_i) > \delta$, so $1_H = 1$ almost surely—i.e., herding almost surely occurs. □

A.3.2 Formal Proof of Theorem 3

Lemma 7. $\sqrt{G_F (1/x)} = \sqrt{\frac{1-F^1(x)}{1-F^0(x)}} \sim 1 + \frac{1}{2} F^0 (x)$ *as* $x \to 0$.

Proof. Rosenberg and Vieille (2019) show that

$$\frac{1 - F^1 (x)}{1 - F^0 (x)} = 1 + F^0 (x) + o \left(F^0 (x)\right),$$

or equivalently, $\frac{1-F^1(x)}{1-F^0(x)} \sim 1 + F^0 (x)$, so $\sqrt{\frac{1-F^1(x)}{1-F^0(x)}} \sim \sqrt{1 + F^0 (x)} = 1 + \frac{1}{2} F^0 (x) + o \left(F^0 (x)\right)$, which proves the lemma. □

Lemma 8. *Under the conditions of Theorem 3, state* 1 *is locally stable.*

Proof. We want to show that there exists some $R < \infty$ such that for all $r_0 \geq R$, the probability of an action-1 herd is greater than some $\varepsilon > 0$. Let H_θ denote the event in which $a_i = \theta$ for all i, i.e., an action-θ herd. We have

$$\mathbb{P}^*_{r_0} (H_1) = \lim_{i\to\infty} \mathbb{P}^0_{r_0} (a_1 = a_2 = ...a_i = 1) = \prod_{i=1}^{\infty} \left[1 - \overline{F}^0_i \left(\frac{1}{r_i}\right)\right] \geq \prod_{i=1}^{\infty} \left[1 - a \times \left(\frac{1}{r_i}\right)^\alpha\right], \tag{14}$$

where r_i represents the average public likelihood ratio after $h_i = (1, 1, ..., 1)$. Recall that

$$r_{i+1} = r_i \times \sqrt{\sup_{F_i\in\mathcal{F}_0} \frac{1 - F^1_i (1/r_i)}{1 - F^0_i (1/r_i)} \times \inf_{F_i\in\mathcal{F}_0} \frac{1 - F^1_i (1/r_i)}{1 - F^0_i (1/r_i)}} \geq r_i \times \sqrt{\frac{1 - F^1 (1/r_i)}{1 - F^0 (1/r_i)}},$$

[31] Suppose that $r_i \in [r, R]$. Let F be any unbounded DGP contained in $\mathcal{F}_0$. Then when $a_i = 0$, we have $r_{i+1} \leq r_i \times \sqrt{\frac{F^1(1/r_i)}{F^0(1/r_i)}} \leq r_i \times \sqrt{\frac{F^1(1/r)}{F^0(1/r)}}$. Because $r \in (0, \infty)$, we have $r_{i+1}/r_i \leq \sqrt{\frac{F^1(1/r)}{F^0(1/r)}} \equiv \beta < 1$, so after $K = \left\lceil \log_\beta^{r/R} \right\rceil + 1$ consecutive action 0s, we have $r_{i+K} < r$. Similarly, K consecutive action 1s will result in $r_{i+K} > R$.

where F denotes the DGP in $\mathcal{F}_0$ such that $x^p = o\left(F^0(x)\right)$. Let $q \in (p, \alpha)$, then we have

$$\begin{aligned}
\lim_{r\to\infty} \frac{\sqrt{\frac{1-F^1(1/r)}{1-F^0(1/r)}} - 1}{\left(1+\frac{1}{r^q}\right)^{1/q} - 1} &= \lim_{r\to\infty} \frac{\sqrt{\frac{1-F^1(1/r)}{1-F^0(1/r)}} - 1}{\frac{1}{r^q}} \times \lim_{r\to\infty} \frac{\frac{1}{r^q}}{\left(1+\frac{1}{r^q}\right)^{1/q} - 1} \\
&= \lim_{r\to\infty} \frac{\frac{1}{2}F^0(1/r)}{\frac{1}{r^q}} \times \lim_{r\to\infty} \frac{\frac{1}{r^q}}{\left(1+\frac{1}{r^q}\right)^{1/q} - 1} \qquad (15) \\
&> \lim_{r\to\infty} \frac{\frac{1}{2}(1/r)^p}{\frac{1}{r^q}} \times q = \infty,
\end{aligned}$$

where (15) follows from Lemma 7. From Lemma 5, we know that $\{r_i\}$ is increasing during an action-1 herd, so $r_i \geq r_0$ for all i. Therefore, when r_0 is sufficiently large, we have

$$\sqrt{\frac{1-F^1(1/r_i)}{1-F^0(1/r_i)}} \geq \left(1+\frac{1}{r_i^q}\right)^{1/q},$$

for all $i \geq 1$, which further implies that

$$r_{i+1} \geq r_i \times \sqrt{\frac{1-F^1(1/r_i)}{1-F^0(1/r_i)}} \geq r_i \times \left(1+\frac{1}{r_i^q}\right)^{1/q} = \left(r_i^q+1\right)^{1/q}.$$

After iterations, we can obtain

$$r_i \geq \left(r_0^q + i\right)^{1/q}, \qquad \forall i \geq 1. \qquad (16)$$

After substituting (16) into (14), we know that for all $r_0 \geq R$ with R sufficiently large,

$$\mathbb{P}^*_{r_0}(H_1) \geq \prod_{i=1}^{\infty}\left[1 - a \times \left(\frac{1}{r_i}\right)^{\alpha}\right] \geq \prod_{i=1}^{\infty}\left[1 - a \times \frac{1}{\left(r_0^q+i\right)^{\alpha/q}}\right] \geq \prod_{i=1}^{\infty}\left[1 - a \times \frac{1}{\left(R^q+i\right)^{\alpha/q}}\right].$$

Here, we choose R to be sufficiently large such that $1 - a \times \frac{1}{R^\alpha} > 0$, so $1 - a \times \frac{1}{(R^q+i)^{\alpha/q}} \in (0,1)$ for all $i \geq 1$. The infinite product $\prod_{i=1}^{\infty}\left[1 - a \times \frac{1}{(R^q+i)^{\alpha/q}}\right] > 0$ if and only if the infinite series $\sum a \times \frac{1}{(R^q+i)^{\alpha/q}} < \infty$. Since $q < \alpha$, we know that $\sum a \times \frac{1}{(R^q+i)^{\alpha/q}} < \infty$, so

$$\mathbb{P}^*_{r_0}(H_1) \geq \prod_{i=1}^{\infty}\left[1 - a \times \frac{1}{\left(R^q+i\right)^{\alpha/q}}\right] \equiv \varepsilon > 0 \text{ for all } r_0 \geq R,$$

which establishes the local stability of state 1. □

Lemma 9. *Under the conditions of Theorem 3, state* 0 *is locally stable.*

Proof. The case for state 0 is symmetric to Lemma 8. Let r_i denote the average likelihood ratio after $h_i = (0, ..., 0)$. From symmetry, we have

$$\mathbb{P}^*_{r_0}(H_0) = \prod_{i=1}^{\infty} F^0\left(\frac{1}{r_i}\right) = \prod_{i=1}^{\infty}\left[1 - F^1(r_i)\right] \geq \prod_{i=1}^{\infty}\left[1 - F^0(r_i)\right] = \mathbb{P}^*_{1/r_0}(H_1),$$

which says that the probability of a correct herd with prior r_0 is higher than that of an incorrect herd with prior $1/r_0$. From Lemma 8, there exists R such that $\mathbb{P}^*_{1/r_0}(H_1) \geq \varepsilon > 0$ for all $1/r_0 > R$. So we have $\mathbb{P}^*_{r_0}(H_0) \geq \mathbb{P}^*_{1/r_0}(H_1) \geq \varepsilon > 0$ for all $r_0 < 1/R$, which establishes the local stability of state 0. □

Combining Lemmas 6 to 9, we know that herding occurs almost surely, and an incorrect herd occurs with a strictly positive probability, so Theorem 3 is proved.

A.4 Proof of Proposition 3

The proof can be decomposed into the following lemmas.

Lemma 10. *Proposition 3 holds if*

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

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

Proof. From the discussion in Lemma 8, we know that an incorrect herd occurs with a strictly positive probability (i.e., state 1 is locally stable) if $\sum_{i=1}^{\infty} \overline{F}^0(1/r_i) < \infty$, 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,

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

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 $\beta = \frac{C(\underline{\alpha})\underline{\alpha}}{3}$, and we have

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

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 $t \geq 1$. Note that $\overline{F}^0(x)$ is an increasing function, so to show $\sum_{i=1}^{\infty} \overline{F}^0(1/r_i) < \infty$, it suffices to show that $\sum_{t=1}^{\infty} \overline{F}^0 \left(1/\left(\beta t+1\right)^{1/\underline{\alpha}}\right) < \infty$. Similar to Lemma 9, the local stability of state 1 implies that of state 0, so we can further show that herding occurs almost surely. $\square$

Lemma 11. $\sum_{t=1}^{\infty} \overline{F}^0 \left(1/\left(\beta t+1\right)^{1/\underline{\alpha}}\right) < \infty$.

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

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

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

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

(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 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

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

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

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

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.5 Proof of Proposition 4

Proof. Since ϕ is strictly increasing, the decision rule is monotonic in λ_i, so there exists a cutoff $r^\phi(h_i)$ (I denote it by r_i^ϕ henceforth) such that individual i will choose action 1 if $\lambda_i \cdot r_i^\phi > 1$ and action 0 if otherwise. Note that the smooth ambiguity preference approaches that the max-min preference as $\underline{\rho}_\phi \to +\infty$ (see Proposition 3 in Klibanoff et al. (2005)), and the cutoff value under the max-min is r_i, where r_i represents the average likelihood ratio. So, we have $|r_i^\phi - r_i| \to 0$ as $\underline{\rho}_\phi \to +\infty$, which further implies that $|r_{i+1}^\phi/r_i^\phi - r_{i+1}/r_i| \to 0$ as $\underline{\rho}_\phi \to +\infty$. As a consequence, for all $\epsilon > 0$ and $I < \infty$, there exists some $\rho_0 < \infty$ such that for all ϕ that satisfies $\underline{\rho}_\phi > \rho_0$, we have

$$|r_{i+1}^\phi/r_i^\phi - r_{i+1}/r_i| < \epsilon \quad \text{for all } i \leq I. \tag{17}$$

Suppose that $\mathcal{F}_0$ satisfies Theorem 2 (i) by containing a DGP discrete at γ.[32] From the proof of Theorem 2 (i), we know that the average likelihood ratios satisfy

$$r_{i+1}/r_i \begin{cases} \geq \beta & a_i = 1 \\ \leq 1/\beta & a_i = 0 \end{cases} \text{when } r_i \in (1/\gamma, \gamma). \tag{18}$$

(17) and (18) imply that for all ϕ satisfying $\underline{\rho}_\phi > \rho_0$ and when $i \leq I$, we have

$$r_{i+1}^\phi/r_i^\phi \begin{cases} \geq \beta - \epsilon & a_i = 1 \\ \leq \frac{1}{\beta-\epsilon} & a_i = 0 \end{cases} \text{when } r_i^\phi \in (1/\gamma, \gamma). \tag{19}$$

So an information cascade will be triggered after at most $K \equiv \left\lceil \log_{\beta-\epsilon}^{\gamma} \right\rceil + 1$ identical actions. Let N_i denote the event that $r_i^\phi \in \left(\frac{1}{\gamma}, \gamma\right)$, then we have

$$\frac{\mathbb{P}^*(N_{i+K})}{\mathbb{P}^*(N_i)} = \frac{\mathbb{P}^*(N_{i+K} \cap N_i)}{\mathbb{P}^*(N_i)} = \mathbb{P}^*(N_{i+K}|N_i) \leq 1 - \left(\frac{1}{1+\gamma}\right)^K \equiv q < 1,$$

where the first equality comes from the fact that $N_{i+K} \subset N_i$, and the last inequality comes from the proof of Theorem 1. Therefore, the expected number of individuals before I who do not face a cascade is

$$\mathbb{E}^*\left(\sum_{i \leq I} 1_{N_i}\right) = \sum_{i \leq I} \mathbb{P}^*(N_i) \leq \mathbb{P}^*(N_1) \times \left(1 + q + ... + q^{I-1}\right) < \frac{1}{1-q} < \infty.$$

As a consequence,

$$\mathbb{P}^*(N_1 = ... = N_I = 1) \times I \leq \mathbb{E}^*\left(\sum_{i \leq I} 1_{N_i}\right) \leq \frac{1}{1-q} \Rightarrow \mathbb{P}^*(N_1 = ... = N_I = 1) \leq \frac{1}{(1-q)I}.$$

It means that the probability of no cascade before individual I is less than $\frac{1}{(1-q)I}$, which implies that the probability of an information cascade is greater than $1 - \frac{1}{(1-q)I}$. When I becomes arbitrarily large, the probability of a cascade can be arbitrarily close to 1. □

[32] The proof for $\mathcal{F}_0$ satisfies Theorem 2 (ii) is almost identical. The only change is that (18) and (19) are satisfied when r_i and r_i^ϕ are in $\left(\frac{1}{\gamma-\varepsilon_0}, \gamma - \varepsilon_0\right)$.

A.6 Proof of Proposition 5

Proof. Suppose that $a_1 = 1$. Suppose that individual 2 received a signal λ_2. We first note that

$$\mathbb{P}_{\alpha_1}(\theta = 0 | a_1, \lambda_2) = \frac{1 - F^0(1, \alpha_1)}{1 - F^0(1, \alpha_1) + \lambda_2 (1 - F^1(1, \alpha_1))} = \frac{F^1(1, \alpha_1)}{F^1(1, \alpha_1) + \lambda_2 F^0(1, \alpha_1)},$$

$$\mathbb{P}_{\alpha_1}(\theta = 1 | a_1, \lambda_2) = \frac{\lambda_2 F^0(1, \alpha_1)}{F^1(1, \alpha_1) + \lambda_2 F^0(1, \alpha_1)},$$

where $\mathbb{P}_{\alpha_1}$ denotes individual 2's posterior when individual 1's DGP is $F(\cdot, \alpha_1)$. So, individual 2's utility of choosing action 0 is

$$V_2(0) = \left[\int_{\underline{\alpha}}^{\overline{\alpha}} \mathbb{P}_{\alpha_1}^{1-\rho}(\theta = 0 | a_1, \lambda_2)\, \mu(\alpha_1)\, d\alpha_1\right]^{\frac{1}{1-\rho}} = \left[\int_{\underline{\alpha}}^{\overline{\alpha}} \left[\frac{1}{1 + \lambda_2 \chi(\alpha_1)}\right]^{1-\rho} \mu(\alpha_1)\, d\alpha_1\right]^{\frac{1}{1-\rho}}.$$

Suppose that $\rho > 1$. By Assumption 3, there exists some $R < \overline{\alpha}$ such that

$$\begin{aligned} V_2(0) &\leq \left[\int_{\underline{\alpha}}^{R} \left[\frac{1}{1 + \lambda_2 \chi(\alpha_1)}\right]^{1-\rho} \mu(\alpha_1)\, d\alpha_1 + \int_{R}^{\overline{\alpha}} \left[\frac{1}{1 + \lambda_2 \chi(\alpha_1)}\right]^{1-\rho} C \cdot \chi^{-k}(\alpha_1)\, d\alpha_1\right]^{\frac{1}{1-\rho}} \\ &\leq \left[\int_{\underline{\alpha}}^{R} \left[\frac{1}{1 + \lambda_2 \chi(\alpha_1)}\right]^{1-\rho} \mu(\alpha_1)\, d\alpha_1 + C \times \int_{R}^{\overline{\alpha}} \frac{[1 + \lambda_2 \chi(\alpha_1)]^{\rho-1}}{\chi^k(\alpha_1)} d\alpha_1\right]^{\frac{1}{1-\rho}}. \end{aligned}$$

Note that if $\rho > k + 1$, we have

$$\int_{R}^{\overline{\alpha}} \frac{[1 + \lambda_2 \chi(\alpha_1)]^{\rho-1}}{\chi^k(\alpha_1)} d\alpha_1 \geq \int_{R}^{\overline{\alpha}} \lambda_2^{\rho-1} \cdot \chi^{\rho-k-1}(\alpha_1)\, d\alpha_1 = +\infty,$$

so $V_2(0) = 0$. The utility of action 1

$$V_2(1) = \left[\int_{\underline{\alpha}}^{\overline{\alpha}} \left[\frac{\lambda_2 \chi(\alpha_1)}{1 + \lambda_2 \chi(\alpha_1)}\right]^{1-\rho} \mu(\alpha_1)\, d\alpha_1\right]^{\frac{1}{1-\rho}} \geq \frac{\lambda_2}{\lambda_2 + 1} > 0 = V_2(0),$$

so individual 2 will choose to follow the herd regardless of her private signal, i.e., an information cascade occurs almost surely. □