Sequential Learning under Informational Ambiguity*

Jaden Yang Chen†

First version: November 2019 This version: April 2025

Abstract

This paper investigates a sequential social learning problem in which individuals face ambiguity about others' signal structures and have max-min expected utility preferences, thereby exhibiting ambiguity aversion. Unlike previous findings, which suggest that learning outcomes depend on the specifics of the learning environment, this study establishes information cascades as a robust outcome under ambiguity. With sufficient ambiguity, cascades arise almost surely, regardless of the statistical properties of signal structures. Moreover, standard results predicting the absence of cascades can easily break down: even minimal ambiguity can trigger cascades when signals are bounded and lead to incorrect herding 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, Yingni Guo, Yoram Halevy, Fei Li, Qingmin Liu, Erik Madsen, Suraj Malladi, Peter Norman, Seth Sanders, Todd Sarver, Mehdi Shadmehr, Peter Sørensen, Omer Tamuz, Jianrong Tian, Alex Wolitzky, and many conference participants for discussion and comments. I thank Hanqing Ye for excellent research assistance. 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

Social learning is a pervasive phenomenon across various contexts, such as financial markets, technological adoption, and social networks, in which individuals often base their actions on the observed behaviors of others.[1] In such settings, informational *ambiguity* naturally arises because individuals typically lack direct access to others' signals and cannot fully assess the informativeness of others' actions. For example, traders in financial markets may observe others' trading behavior but remain uncertain whether these tradings reflect high-quality private information or are driven by noise traders speculating without solid information (Avery and Zemsky, 1998; Banerjee and Green, 2015). Moreover, traders may not even know the distribution of other traders' information quality, making it difficult to interpret others' trading behavior.[2] Such ambiguity plays a crucial role in shaping decision-making, influencing whether individuals choose to follow, ignore, or deviate from observed actions.

In this paper, I investigate a social learning problem with informational ambiguity and provide a tractable analysis of how ambiguity influences learning dynamics. The main finding is that under sufficient ambiguity, *information cascades* emerge as a robust outcome, persisting regardless of the specifics of the society's data-generating process (DGP).[3] Here, an information cascade means that individuals eventually follow a possibly incorrect action and disregard their private signals. This finding stands in sharp contrast to the standard Bayesian case, where information cascades are considered fragile and highly sensitive to the precise details of the information environment. In the following, I explain the details of the model and the key mechanism driving this result.

This paper builds on the standard sequential social learning (SSL) framework (Banerjee, 1992; Bikhchandani et al., 1992). In that model, individuals act sequentially to match an unknown binary state of the world. One might think of decisions such as whether to buy or short a stock, which restaurant to visit, or whether to share a piece of content on social media. Each individual observes the actions of their predecessors and receives a private signal generated by a DGP. The key innovation of this paper is that individuals face ambiguity about their predecessors' DGPs when they consider a set of possible DGPs, rather than a single known process. This ambiguity characterizes a situation in which individuals

[1] See Golub and Sadler (2017) and Bikhchandani et al. (2024) for surveys on social learning.

[2] This type of ambiguity can also occur in many other scenarios. For example, in technology adoption, farmers may be uncertain whether widespread adoption reflects the genuine quality of the technology or mere herd behavior (Conley and Udry, 2010; Barham et al., 2014); in online reviews, consumers can observe highly rated products or viral posts but remain uncertain whether the ratings reflect true user experiences (Hu et al., 2006; Ifrach et al., 2019; Acemoglu et al., 2022).

[3] Robustness under ambiguity refers to the extension of an information cascade from the standard Bayesian setting to the ambiguity setting. That is, when ambiguity is introduced, an information cascade remains (or emerges if it was not present initially), whereas other learning outcomes do not persist.

lack sufficient knowledge to determine the exact DGP that explains others' actions or to form a prior over the set of DGPs.[4] In the context of social learning, ambiguity arises very naturally because the quality of individuals' information can vary significantly—some may be experts, while others are laymen. However, since individuals cannot directly observe others' signals, it is difficult for them to determine who has more precise information or even to form a prior belief about the information environment. The benchmark model assumes that individuals have max-min expected utility (MEU) preferences, as in Wald (1950) and Gilboa and Schmeidler (1989). Therefore, individuals are ambiguity-averse and seek to maximize utility in the worst-case scenario. When evaluating an action, individuals use the DGPs compatible with the public history that generate the lowest expected utility for that action.

Introducing ambiguity significantly alters standard learning dynamics. Without ambiguity, individuals adopt a fixed interpretation of public information based on a single DGP. With ambiguity, however, they perceive a set of DGPs and consider multiple interpretations simultaneously. Under MEU preferences, the worst-case DGPs become endogenous, that is, individuals interpret the same data differently depending on their observations or the actions they are considering. This marks a key departure from standard Bayesian learning, where information interpretation is often exogenous and fixed.

Despite the added complexity, ambiguity leads to a sharper theoretical prediction of social learning outcomes. Under sufficient ambiguity—when individuals consider a broad set of DGPs—information cascades arise with probability 1. This result reflects a deeper asymmetry in how ambiguity shapes the forces that drive or deter cascades. While different DGPs have varying implications for learning, this paper shows that the forces encouraging cascades and those preventing them are fundamentally *asymmetric*. As ambiguity increases, the cascade-inducing force dominates, making information cascades the robust outcome.

To illustrate the asymmetry, consider a scenario in which customers choose between two restaurants, A and B. Suppose all previous customers chose restaurant A, but the next customer receives a signal favoring B. If she chooses B, the worst-case scenario for her is that all predecessors were experts with highly precise signals, making their actions strongly favor A. Conversely, if she chooses A, the worst-case scenario for her is that all predecessors were laymen with uninformative signals, so their actions reveal no information. As ambiguity increases, the customer becomes more hesitant to deviate from the herd, since she cannot rule out the possibility of contradicting highly precise signals. In contrast, the risk of following the herd is more controllable, since she would be disregarding only her own signal, whose precision is known to her. This asymmetry in worst-case scenarios nudges her to follow the

[4] See Hansen et al. (2014), Marinacci (2015), and Hansen and Marinacci (2016) for surveys on ambiguity over DGPs.

herd, reinforcing the cascade force. Below is a specific example.

Example 1. The state space $\Theta = \{0, 1\}$. The underlying 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 zero 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 know their own signal precision but face ambiguity about others' precision, and they believe that every $\gamma_i \in \Gamma$ is possible. Suppose that the first individual's (his) action is $a_1 = 1$. Let $V_2(a)$ denote the infimum expected utility of the second individual (she) if she takes action a. It follows that

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

To see that, the worst-case scenario for $a_2 = 1$ is that individual 1 received only 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 infimum utility is zero. 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 more general framework. Theorem 1 considers a benchmark case in which individuals know only the support of society's DGPs and regard all DGPs on this support as possible. Under this form of ambiguity, an information cascade emerges almost surely, regardless of the statistical properties of the underlying DGPs. This result shows that cascades can arise when individuals have minimal knowledge about each other's signal structure.

Moreover, information cascades can arise in more general situations. The paper then provides sufficient conditions for cascades to emerge under general perceptions of DGPs. It first examines the case of bounded signals. Theorem 2 shows that an information cascade occurs almost surely whenever the perceived set of DGPs includes at least one adequately informative DGP—that is, a DGP that places sufficiently high weight on precise signals. The intuition follows from Example 1: the perception of a highly informative DGP would encourage individuals to follow the herd, creating an asymmetrically high cascade force which cannot be offset by the perception of any other DGPs. Notably, Theorem 2 even implies that non-cascade results are knife-edge cases in certain settings. Even a slight degree of ambiguity can almost surely induce an information cascade, even though cascades are absent in some standard cases (Smith and Sørensen, 2000).

The paper then discusses the cases in which signals are unbounded. It is worth noting that information cascades are highly demanding in this setting, as they require individuals to disregard arbitrarily informative signals. For a cascade to occur, individuals must consider arbitrarily informative DGPs—that is, DGPs in which an individual's action, based on a fixed interval of private signal realizations, can generate arbitrarily high likelihood ratios for one of the two states. This represents an extreme scenario in many cases. Given these stringent requirements, this paper then explores a weaker but qualitatively similar concept, *herding*—which refers to the phenomenon in which individuals eventually take the same action, which may be incorrect, yet they do not necessarily ignore their private signals. Theorem 3 provides sufficient conditions for herding to emerge. These conditions parallel those in Theorem 2, which require individuals to consider an adequately informative DGP, whose informativeness is characterized by the thickness of its tail.

The main model focuses on MEU preferences, but I then extend the insights to other ambiguity preferences, such as the smooth ambiguity preferences and α-max-min preferences. I show that the main qualitative finding—the occurrence of an information cascade—still persists with those preferences.

This paper is among the first to introduce ambiguity into a social learning framework. It sheds light on how individuals can learn from others when they have limited knowledge of society's signal structures. Beyond its conceptual contributions, the paper also makes some methodological advances. Analyzing social learning under ambiguity poses significant challenges: individuals form a set of posteriors, making belief updating complex, and standard martingale techniques are inapplicable. This paper provides a tractable approach, showing that under MEU preferences, a simple statistic—the average public likelihood ratio—serves as a sufficient statistic for social learning. It also develops non-martingale techniques to estimate herding probabilities by analyzing the local stability of this statistic, which depends on the tail properties of the perceived DGPs.

2 The Model

States and Actions. There are two possible states of the world, $\Theta = \{0, 1\}$. Without loss of generality, the underlying 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 made by all predecessors. Individuals get a payoff of 1 when their actions match the underlying state and a payoff of zero otherwise.

Information structures. Individuals do not know the underlying 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 $\left\{\mathbb{G}_1^\theta, \mathbb{G}_2^\theta, ...\right\}$, where $\mathbb{G}_i^\theta : \mathcal{S} \to [0,1]$ 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 $\mathbb{G}_i^0$ and $\mathbb{G}_i^1$ are mutually absolutely continuous. Following the convention, I introduce the normalized signal, λ_i, where $\lambda_i(s) = \frac{d\mathbb{G}_i^1(s)}{d\mathbb{G}_i^0(s)}$ denotes the likelihood ratio induced by signal s. The distribution of the likelihood ratio λ_i is denoted by $\mathbb{F}_i^\theta$, which must satisfy $\lambda = \frac{d\mathbb{F}_i^1(\lambda)}{d\mathbb{F}_i^0(\lambda)}$ almost everywhere. The rest of the paper focuses on the normalized signal, λ, and the normalized DGP, $\mathbb{F}_i^\theta$. 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 notational convenience, I assume: (i) all signals are continuous, that is, $\mathbb{F}_i^\theta$ is continuous for all i and θ; and (ii) signals are symmetric $\mathbb{F}_i^1(\lambda) = 1 - \mathbb{F}_i^0(1/\lambda)$ for all i and λ.[6] Let $\mathfrak{F}$ denote the set of all normalized DGPs with a typical element being $F = (F^0, F^1)$. Let Λ^∞ denote the set of all signal paths, where a typical element is $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, ...)$, and associate it with the σ-algebra $\sigma(\Lambda^\infty)$. Let $\mathbb{P}^*$ denote the probability measure on $(\Lambda^\infty, \sigma(\Lambda^\infty))$ induced by the signal distributions $\{\mathbb{F}_1^0, \mathbb{F}_2^0, ...\}$. This paper refers to $\mathbb{P}^*$ as the ***objective probability***, i.e., the probability conditional on the underlying state. Unless stated otherwise, all events are evaluated according to the objective probability.

Ambiguous Information. Individuals know their own DGPs and understand that all signals are independently distributed, but they may face ***ambiguity*** about others' DGPs and therefore consider a set of possible DGPs.[7] Specifically, individuals share a common set of DGPs, $\mathcal{F}_0 \subset \mathfrak{F}$, and believe that every other individual's DGP belongs to $\mathcal{F}_0$, though they do not know which is the actual DGP.

Updating Ambiguous Beliefs. 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 set of individual i—that is, her private

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

[6] Continuity avoids the need for separate notation handling discontinuity points, while symmetry allows the characterization of just one side of the distribution. The results extend to more general cases, because (i) discontinuous cases can be treated as limits of continuous ones, and (ii) asymmetry can be addressed by imposing analogous conditions on the other side of the distribution.

[7] The knowledge of their own DGPs is actually not necessary. Individuals are only required to use the normalized signal λ_i to update beliefs, but its origin is nonessential.

signal λ_i and history h_i. Let $\mathcal{I}_i$ be the set of all possible information sets of individual 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 governed by the probability measure $\mathbb{P}_{F_{-i}}(.|\theta; \sigma_{-i})$. Given history h_i and strategy profile σ_{-i}, let $\Pi(h_i, \sigma_{-i})$ denote the set of beliefs generated by DGPs in $\mathcal{F}_0$, which I refer to as the ***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 according to F_{-i}, and $\mathcal{F}_0^{i-1}$ denotes $i-1$ 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 the ***private belief set***. Individuals use the full Bayesian rule (axiomatized by Pires (2002)) to update beliefs, as shown here:

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

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 used in applications, but it faces two major criticisms: (i) the set of DGPs remains fixed even after new information is observed; and (ii) it can lead to dynamic inconsistency. These concerns are less relevant in this paper for two reasons. First, individuals observe only one action from each predecessor, so there is typically limited information to update beliefs about others' DGPs. Second, individuals make a once-in-a-lifetime decision, so dynamic inconsistency is not relevant here.

Equilibrium. Individuals have ***max-min expected utility*** (MEU) preferences, as in Gilboa and Schmeidler (1989). The equilibrium concept is defined as follows:

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

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

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 and private belief sets, respectively. I also assume that individuals follow some tie-breaking rule when indifferent.[8] By focusing on pure strategies, the paper implicitly assumes that individuals cannot be better off by playing mixed strategies. This assumption is further discussed in the next section.

3 Discussion of Modeling Assumptions

This section discusses several key assumptions in the baseline model. A key feature of the setup is the **ambiguity regarding others' DGPs**, which arises when individuals lack sufficient information to determine others' signal structures. This type of ambiguity naturally emerges in sequential social learning for two reasons: (i) individuals observe only actions, not other individuals' private signals, so multiple DGPs can be consistent with the observed public information, leading to identification problems; and (ii) each individual observes only one action from each predecessor, making it impossible to fully infer the true DGP through social learning.[9]

By contrast, this paper assumes that individuals face **no ambiguity about their own DGPs**. This serves as a reasonable benchmark, since individuals typically have a better understanding of their own signal-generating processes than those of others. Importantly, the paper's main results remain valid even when individuals face ambiguity about their own DGPs, as discussed in Section S7 of the Supplementary Materials.[10] In this case, social learning follows a structure similar to that of the baseline model. Specifically, I show that the average private likelihood ratio—the geometric mean of the highest and lowest likelihood ratios induced by the private signal—plays the same role as the normalized private signal in the baseline model.[11] This implies that individuals need only a rough sense of what their signals indicate, i.e., which state the private signal favors on average. For instance, in Example 1, it suffices for individuals to understand only that signal H supports state 1 and signal L supports state 0, whereas a precise understanding of their likelihoods is unnecessary. Consequently, the learning dynamics remain similar, and the main results in the paper continue to hold.[12]

[8]The choice of tie-breaking rule is not essential to the result, so I do not specify it in this paper. When signals are continuous, these tie cases happen with zero probability.

[9]The lack of identification is a common justification for ambiguity in the literature (Manski, 2000; Battigalli et al., 2015).

[10]Throughout this paper, the Supplementary Materials refer to Chen (2023).

[11]Suppose individual i receives a private signal s_i, but she faces ambiguity about the signal's implication and believes that the likelihood ratio can be any number between $[1/2, 8]$. Intuitively, her private signal should favor state 1, since its maximum likelihood for state 1 is much higher than for state 0. Formally, I show that this individual will behave as if she received a normalized private signal equal to $\sqrt{1/2 \times 8} = 2$.

[12]One situation in which the paper's analysis breaks down is when individuals interpret all signals symmetrically for both states. For example, private signals could be interpreted as any likelihood ratio between

This paper focuses on the **pure-strategy equilibrium**. In the literature, whether mixed strategies can be used to hedge against ambiguity is subjective and depends on individuals' inherent preferences for randomization (Saito, 2015; Ke and Zhang, 2020; Calford, 2021).[13] For simplicity, the baseline model focuses on the pure-strategy equilibrium, as in many other studies of ambiguity (Bose and Renou, 2014; Tillio et al., 2016; Libgober and Mu, 2021). An implicit assumption here is that ambiguity cannot be eliminated by randomization, so restricting attention to pure strategies is without loss of generality. In Section S5 of the Supplementary Materials, I explore an alternative assumption, namely, that individuals can use mixed strategies to hedge against ambiguity. This modification introduces a new equilibrium region in which individuals randomize between two actions when their signals are moderately precise. The asymptotic learning outcomes exhibit two key features: (i) an information cascade still emerges almost surely under sufficient ambiguity; and (ii) unlike the baseline model, the cascade now takes a mixed-strategy form, such that, after some point, individuals will mix two actions with equal probability, regardless of their private signals. This pattern arises from the additional incentive to hedge against ambiguity through randomization.

The paper also assumes **homogeneous preferences**, as is standard in the social learning literature. This assumption reflects scenarios in which society broadly agrees on which action is optimal in each state. For example, investors buy stocks when they anticipate future price increases and short them otherwise. The results extend to settings with heterogeneous preferences, as discussed in Section S9 of the Supplementary Materials. I show that an information cascade can still occur almost surely even when a positive fraction of individuals have opposite preferences, i.e., when they seek to mismatch the true state. One minor difference is that during a cascade, individuals may take different actions—not due to differences in information, but solely as a result of preference heterogeneity.[14]

$[1/2, 2]$. In this case, signals are effectively uninformative, since they do not favor any state. Thus, social learning also becomes trivial as there is nothing to learn.

[13] From Ke and Zhang (2020), it depends on individuals' subjective beliefs about the timing of nature's move. For example, if they believe that nature moves after the implementation of randomization, then a mixed strategy cannot provide a hedge against ambiguity; on the contrary, if they believe that nature moves before the implementation, then a mixed strategy can provide a hedge.

[14] For example, suppose that during a cascade, society believes that state 1 is very likely. Then, individuals who wish to match the state will choose action 1, whereas those who seek to mismatch it will choose action 0—both regardless of their private signals.

4 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.

4.1 Characterizations of Equilibrium Strategies

When individuals face ambiguity, it seems difficult to characterize the learning dynamics because individuals now form a set of posteriors instead of a single posterior. Fortunately, the MEU model enables us to extend the concept of the likelihood ratio and represent the posterior set using the average likelihood ratio of beliefs it contains. This property leads to a simple equilibrium characterization, which enhances tractability.

Definition 2. Let $L(h_i) = \left\{\frac{\pi(1)}{\pi(0)} : \pi \in \Pi\left(h_i\right)\right\}$ denote the set of public likelihood ratios in the equilibrium. Let $\underline{l}_i = \inf L(h_i)$ and $\overline{l_i} = \sup L(h_i)$, and define

$$r_i = \sqrt{\overline{l_i} \cdot \underline{l}_i},$$

which is 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 using the average public likelihood ratios.

Proposition 1. (Equilibrium Strategy) *There exists a pure-strategy max-min equilibrium. In the equilibrium, for any individual $i \in N$ and information set $I_i \in \mathcal{I}_i$, we have*

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

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

Proof. The existence of an equilibrium comes from standard induction arguments.[15] Define $\underline{\pi}_i\left(\theta\right) = \inf\left\{\pi\left(\theta\right) : \pi \in \Pi_i\left(I_i\right)\right\}$, then $a_i = 1$ if $\underline{\pi}_i\left(1\right) > \underline{\pi}_i\left(0\right)$. Note that

$$\underline{\pi}_i\left(1\right) = \frac{\lambda_i \underline{l}_i}{1 + \lambda_i \underline{l}_i} \quad \text{and} \quad \underline{\pi}_i\left(0\right) = \frac{1}{1 + \lambda_i \overline{l_i}}.$$

[15]Formally, given predecessors' strategies $(\sigma_1, ..., \sigma_{i-1})$, the maximization problem (1) in Section 2 has a solution for each $i \in N$ and each possible information set $I_i \in \mathcal{I}_i$; applying this argument inductively gives an equilibrium (Acemoglu et al., 2011).

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

The average public likelihood ratio is an extension of the public 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 components: private information, given by the private signal λ_i, and public information, captured by the average public likelihood ratio, r_i. When the product, $\lambda_i \cdot r_i$, is greater than 1—indicating that state 1 is more likely—individuals will choose action 1; otherwise, they choose action 0.

4.2 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 $\mathbb{P}^*(a_i = a | h_i) = 1$ for all $i \geq I$.

An information cascade occurs when, after some point, individuals choose a specific action regardless of their private signals. During a cascade, information aggregation ceases, and society may settle on an incorrect action, even in the presence of infinitely many informative signals. Using Proposition 1, information cascades can be characterized in terms of the average public likelihood ratio.

Lemma 1. *Let* $C_0 = \left[0, \frac{1}{\gamma}\right]$ *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 information favoring action a becomes so strong that individuals choose action a regardless of their private signals—thus, an information cascade takes place. In classical models with finite signals, an information cascade almost surely emerges (Banerjee, 1992; Bikhchandani et al., 1992). However, in many settings, a cascade is not always guaranteed. Several other outcomes are also possible, as described in Definition 4.

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 nonconvergence*** occurs if a_i fails to converge; and (iii) ***complete learning*** occurs if correct-action herding occurs $\mathbb{P}^*$-almost surely.

Herding and information cascades are different concepts. Herding occurs when all individuals take the same action, but they may still be responding to their private signals. In contrast, an information cascade arises when individuals completely ignore their private signals and follow the prevailing action.[16] A cascade always generates a herd, but the converse is not necessarily true. The distinction between these two concepts is important from the perspective of information aggregation: during herding, information may continue to accumulate, whereas during a cascade, information aggregation stops entirely.

The literature shows that cascades are a nongeneric property for continuous signals. In particular, when signals are continuous and satisfy the *increasing hazard ratio property* (IHRP), information cascades will not occur (Herrera and Hörner, 2012; Smith et al., 2021); when signals are unbounded, complete learning occurs, and society will eventually settle on correct actions (Smith and Sørensen, 2000). In some situations, action nonconvergence may emerge—for example, when individuals misspecify the true DGPs (Bohren and Hauser, 2021; Arieli et al., 2025).[17]

5 Main Result

This section presents the main result of the paper. To build the intuition, Section 5.1 first considers a benchmark case in which individuals consider all DGPs on the actual support, and shows that in this case, an information cascade occurs almost surely. The discussion is then extended to more general cases in Sections 5.2 and 5.3, which focus on situations in which the true DGPs are bounded and unbounded, respectively.

5.1 Benchmark Case: Cascades under Extreme Ambiguity

Let $\mathcal{F}$ denote the set of all DGPs on the actual support, $\Lambda = [1/\gamma, \gamma]$, where $\gamma \in (1, \infty]$ denotes the supremum signal. We then have the following theorem:

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

The condition $\mathcal{F}_0 = \mathcal{F}$ describes a setting in which individuals know only the support of signals and consider all DGPs on this support to be possible. Theorem 1 establishes the prevalence of information cascades in this benchmark case. In standard models, cascades depend on specific properties of the underlying DGPs. In contrast, Theorem 1 imposes no

[16] The theoretical distinction between these two concepts can be found in Smith and Sørensen (2000) and Bikhchandani et al. (2024). They are also studied in experimental works such as Anderson and Holt (1997), Çelen and Kariv (2004a), and Cai et al. (2009). In particular, Çelen and Kariv (2004a) find that information cascades still occur under a signal structure that implies the absence of a cascade in Bayesian models.

[17] Action nonconvergence can also emerge in general networks, especially when individuals have incomplete observations of past actions, e.g., when they observe only their most recent predecessors (Çelen and Kariv, 2004b).

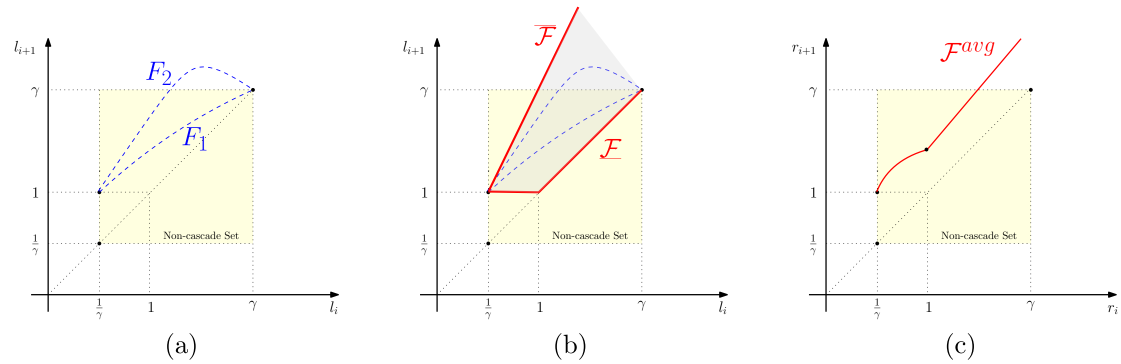

Figure 1: Information Cascades under Ambiguity

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

such restrictions and demonstrates that information cascades can occur under *all* possible DGPs on $[1/\gamma, \gamma]$. This result highlights the robustness of cascades when individuals face extreme ambiguity about others' signal structures. The intuition is explained below.

The Intuition behind Theorem 1

When signals are unbounded ($\gamma = \infty$), an information cascade occurs immediately after the first individual, as in Example 1. I now focus on the bounded-signal case ($\gamma < \infty$). Suppose that the first i individuals took action 1, and individual $i+1$ received a signal $\frac{1}{\gamma}$, the strongest signal favoring state 0. Suppose also that an information cascade did not occur when the first i individuals made decisions. Now, consider the decision problem of individual $i+1$. Since she has MEU preferences, 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, $\frac{1}{\gamma}$.

- If she does not follow the herd and takes action 0 instead, the worst case is that every predecessor's DGP has the most precise DGP—i.e., the DGP that generates only signals γ and $1/\gamma$. In this case, the predecessors' actions reveal that their signals must be γ.[18] Hence, by taking action 0, individual $i+1$ would act against i signals, each equal to γ.

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

As can be seen, the forces driving a cascade and those preventing it are **asymmetric**. As i increases, individual $i+1$ would act against an increasing number of γ-signals in the worst case if she didn't follow the herd; however, she would act against only 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 the herd, which creates an information cascade.

Graphic illustration. Figure 1a illustrates how the occurrence of a cascade depends on the actual DGPs when there is no ambiguity. In the figure, F_1 satisfies the IHRP, while F_2 does not. As can be seen, posteriors under F_1 remain trapped in the non-cascade set, so a cascade never occurs; in contrast, posteriors under F_2 can enter the cascade set, allowing a cascade to emerge.

Figure 1b illustrates the fact that, under ambiguity, the forces that either drive a cascade or discourage it are asymmetric. When beliefs are in the non-cascade set, the **upper envelope** of updating curves of all DGPs under $\mathcal{F}$ (marked by $\overline{\mathcal{F}}$) has a slope of γ, which means that observing an action 1 can increase the public likelihood ratio at most 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 action 1 cannot decrease the likelihood of state 1.[19] The asymmetry of these two curves corresponds to the asymmetric forces that drive a cascade and those that inhibit it. The worst-case scenario for not following the herd happens when beliefs are updated according to the upper envelope, in which case individuals would act against a sequence of signals, each equal to 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 utility when all predecessors have uninformative DGPs.

Figure 1c depicts the updating curve of the average public likelihood ratio under $\mathcal{F}$ (marked by $\mathcal{F}^{avg}$), which is obtained by averaging the two envelope curves.[20] As the figure shows, the average public likelihood curve extends to the cascade set, so an information cascade can occur. To fully prove Theorem 1, what remains to be shown is that the probability of a cascade is 1, and this proof is delegated to the Appendix.

[19] The upper envelope is achieved by the DGP that generates only the most precise signals, γ and $1/\gamma$. The lower envelope has a kink at 1: when $r_i > 1$, it is achieved by the uninformative DGP; when $r_i < 1$, individual i's prior favors state 0, so she must have received minimum information to take action 1, and the lower envelope is achieved by the DGP that generates only signals $1/r_i$ and r_i.

[20] The explicit formula can be found in Lemma 4 in the Appendix.

5.2 Information Cascades with Bounded Signals

The previous section shows how information cascades can emerge when individuals consider all DGPs on the actual support to be possible. To achieve an information cascade, we do not require individuals to face substantial ambiguity. This section provides more general conditions 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 either 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, and an incorrect cascade occurs with strictly positive $\mathbb{P}^*$-probability.

The two conditions say that individuals consider a DGP that assigns large weights to high-precision signals. With some abuse of language, I refer to DGPs that satisfy similar heavy-tail conditions as *highly informative*. Theorem 2 therefore says that (i) an information cascade will occur almost surely as long as individuals consider a highly informative DGP, and (ii) there are no other restrictions on the set of perceived DGPs.

Theorem 2 further highlights the robustness of information cascades. As discussed earlier, the occurrence of cascades in Bayesian models depends on fine details of individuals' DGPs. In contrast, Theorem 2 imposes few restrictions on individuals' DGPs, which may exhibit diverse properties and vary widely across individuals.[21] Furthermore, Theorem 2 suggests that cascades are robust to perceptions of arbitrary DGPs: whenever individuals consider a highly informative DGP, the perception of any other DGP does not jeopardize the occurrence of a cascade. The intuition behind this robustness can be explained by the following example:

Example 2. Let F_γ denote the DGP that generates only signals γ and $1/\gamma$, i.e., the most precise DGP. When individuals perceive only F_γ, a cascade will occur almost surely from standard results (Bikhchandani et al., 1992). Now suppose that individuals consider these two DGPs:

$$\mathcal{F}_0 = \left\{ F_\gamma, \hat{F} \right\},$$

[21]The limited dependence on the underlying DGPs is because the updating rule of r_i is primarily determined by individuals' perceived DGPs. The actual DGPs affect only the transition probability of r_i to each feasible r_{i+1}. Therefore, the actual DGPs influence the probability of each cascade, but not its almost sure occurrence, as long as they satisfy certain regularity conditions, as outlined in Section 2.

where $\text{supp}(\hat{F}) \subset [1/\gamma, \gamma]$. Theorem 2 says that an information cascade emerges almost surely, *regardless of* the statistical properties of $\hat{F}$. To grasp the intuition, let's still consider the case in which individual $i+1$ observed a sequence of action 1. In that case, we have the following:

- The minimum utility for breaking away from the herd is still obtained at F_γ, under which predecessors' actions reveal a sequence of signals, each equal to γ.
- The minimum utility for following the herd is obtained at $\hat{F}$, and this utility will inevitably depend on the properties of $\hat{F}$. The robustness of cascades comes from the fact that this utility can always be **bounded from below** by the utility that would arise when the first i individuals have uninformative DGPs. Formally, under $\hat{F}$, the likelihood ratio of observing individual k taking action 1 is:

$$\frac{\mathbb{P}^1_{\hat{F}}\left(a_k = 1|h_k\right)}{\mathbb{P}^0_{\hat{F}}\left(a_k = 1|h_k\right)} = \frac{1-\hat{F}^1(1/r_k)}{1-\hat{F}^0(1/r_k)}, \tag{2}$$

  where the equality comes from the equilibrium strategy in Proposition 1.[22] From the definition of normalized signals, it follows that $\hat{F}^0 \geq \hat{F}^1$ for all possible $\hat{F}$.[23] This implies that the right-hand side of (2) must be weakly greater than 1 for all possible $\hat{F}$. Intuitively, this means that observing an action cannot decrease the posterior belief in state 1 for any DGP, so the minimum utility for herding cannot be lower than the utility that would arise when predecessors have uninformative DGPs.[24]

In summary, the worst case for deviating from the herd is the same as in Theorem 1, while the worst case for following the herd is better than in Theorem 1, where it arises when predecessors have uninformative DGPs. Consequently, the cascading force is even stronger in this example, implying that a cascade must also emerge.

The intuition from the previous example extends to more general cases. To generate a cascade, it is not necessary for individuals to consider F_γ; any DGP satisfying the heavy-tail property in Theorem 2 serves the same role. Under such DGPs, individuals receive precise signals with high probability. As a result, if a herd persists, individuals' posteriors on the herding state will grow relatively quickly. Consequently, the worst-case utility for deviating from the herd steadily deteriorates, while the worst-case utility for following the herd remains

[22] That is, individual k will choose action 1 whenever her private signal λ_k is greater than $1/r_k$.

[23] Intuitively, individuals are more likely to receive higher signals in state 1 than in state 0, so F^1 must first-order stochastically dominate F^0. See Lemma 2 (1) in the Appendix.

[24] This property is also implicitly employed in the intuitive arguments under Theorem 1, which directly used the fact that the worst-case scenario for herding occurs when predecessors have uninformative DGPs.

bounded below by the uninformative case. This asymmetric force gives rise to an information cascade following the same logic as in Theorem 1.

Information Cascades under ε-Ambiguity

The conditions in Theorem 2 can easily hold in many scenarios, since (i) it requires $\mathcal{F}_0$ to contain only one highly informative DGP without restricting other structures of $\mathcal{F}_0$; and (ii) it imposes restrictions only on the tails without any restrictions in the middle. In many examples, Theorem 2 even implies that standard results featuring the absence of an information cascade represent knife-edge cases:

Example 3. (ε-contamination set) Suppose that the set $\mathcal{F}_0$ satisfies:

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

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 a Bayesian social learning model. In that case, we can have a variety of learning outcomes—such as cascades, herding, or persistent action oscillations—depending on the properties of G and its relation to the actual DGPs. In sharp contrast, for any $\varepsilon > 0$, an information cascade occurs almost surely for all possible choices of G and the actual DGPs. Hence, any non-cascade result is not robust to arbitrarily small perturbations in perceived signal structures.

Example 4. (ε-ambiguity ball) Suppose that the set $\mathcal{F}_0$ satisfies:

$$\mathcal{F}_0 = \{F \in \mathcal{F} : d\,(F, G) \leq \varepsilon\} \text{ for some } \varepsilon \geq 0,$$

where d is a metric on the space of DGPs, and $G \in \mathcal{F}$ has a positive density on $[1/\gamma, \gamma]$. Suppose also that d is consistent with weak convergence in the sense that $F_n^\theta \Rightarrow F^\theta$ implies $d\,(F_n, F) \to 0$. A canonical example is the sup-norm metric.[25] Then, for any $\varepsilon > 0$, and for all possible choices of G and the actual DGPs, an information cascade occurs almost surely.[26] Thus, even an arbitrarily small degree of ambiguity suffices to trigger a cascade.

[25]That is, $d\,(F, G) = \sup_x \left|F^\theta(x) - G^\theta(x)\right|$ for some fixed θ (the choice of θ is irrelevant for symmetric distributions; for asymmetric distributions, we can choose θ as the one that maximizes the distance). Other examples include total variation distance and the Lévy–Prokhorov metric.

[26]This follows from the fact that, under a metric consistent with weak convergence, any DGP can be approximated by a discrete DGP with positive mass on tail signals. Hence, condition (i) of Theorem 2 is satisfied.

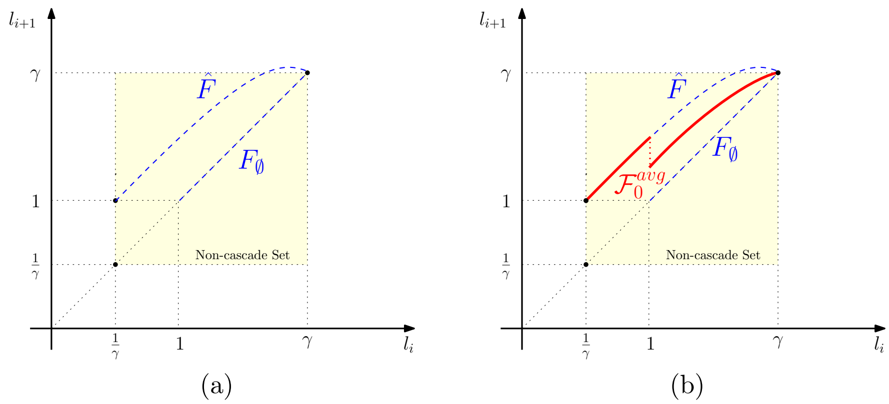

Figure 2: Necessary Conditions for Cascades

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 at least one DGP that can induce a cascade when perceived in isolation, such as F_2 in Figure 1a. This condition is violated, for instance, when all perceived DGPs satisfy the IHRP: in that case, the public likelihood ratio under each perceived DGP remains trapped in the non-cascade set. As a result, the average public likelihood ratio is also confined to that region, thus a cascade cannot occur.

This necessary condition, however, is not sufficient. In Figure 2a, the updating curve under $\hat{F}$ enters the cascade set, thus satisfying the necessary condition. However, when individuals consider $\mathcal{F}_0 = \left\{\hat{F}, F_\emptyset\right\}$, where $F_\emptyset$ denotes the uninformative DGP, a cascade does not occur, as shown in Figure 2b. In this case, $\hat{F}$ fails to meet the heavy-tail condition of Theorem 2; therefore, when it is perceived alongside a highly uninformative DGP, the force against cascades dominates, which prevents cascades from occurring.[27]

5.3 Incorrect Herding with Unbounded Signals

This section extends Theorem 1 to settings with unbounded signals. Note that an information cascade is a restrictive concept for unbounded signals and is unlikely to occur under a small degree of ambiguity.[28] Nonetheless, we can still obtain results that parallel those in the

[27] A simple necessary and sufficient condition for cascades has not yet been discovered, even in the Bayesian case. The closest condition in the literature is perhaps the IHRP or log-concavity condition (Smith et al. (2021)), which is necessary and sufficient for the *posterior monotonicity property*, that is, the posterior public likelihood ratio is increasing in the prior public likelihood ratio. This condition implies that cascades never occur. A similar condition under ambiguity is provided in Section S2 of the Supplementary Materials.

[28] When signals are unbounded, a cascade requires individuals to ignore arbitrarily strong signals. Thus, the perceived DGP set must also be "unbounded" in the sense that it includes arbitrarily informative DGPs, i.e., DGPs which can induce arbitrarily high likelihood ratios.

bounded-signal case by focusing on a weaker concept: herding. I show that, under moderate conditions, herding occurs almost surely, and an incorrect herd occurs with strictly positive probability. In some interesting situations, the complete learning result established by Smith and Sørensen (2000) no longer holds even when there is only a small degree of ambiguity. The following theorem provides a sufficient condition for herding.

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

Theorem 3 is a statement parallel to Theorem 2 for cases where signals are unbounded. The restriction $\mathbb{F}_i^0(x) \leq ax^\alpha$ means that every individual's DGP is bounded by some power function. This condition resembles a similar regularity condition in Rosenberg and Vieille (2019). It is a very mild technical assumption that can accommodate many interesting DGPs.[29] The condition $x^p = o(F^0(x))$ means that the tail of $F^0(x)$ is sufficiently thick—thicker than some power function x^p with $p \in (0, \alpha)$—which is parallel to the conditions in Theorem 2. Thus, Theorem 3 states that when individuals consider a highly informative DGP, herding occurs almost surely, and the herd 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 DGP, so an incorrect herd can emerge. Moreover, Theorem 3 requires only that $\mathcal{F}_0$ contains a specific DGP without imposing restrictions on other structures, so it can easily hold in many interesting cases, as demonstrated by the following corollary:

Corollary 1. *Suppose that signals are i.i.d. with $\mathbb{F}^0(x) = O(x^\alpha)$ with $\alpha > 0$ as $x \to 0$. If there exists some $F \in \mathcal{F}_0$ 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 strictly positive $\mathbb{P}^*$-probability.*

Corollary 1 states that if individuals' DGPs have a power tail, then arbitrarily small ambiguity in the power of $\mathbb{F}^0$ is sufficient to trigger an incorrect herd.[30] This suggests that, within this class of DGPs, complete learning is not robust. Below is a concrete example.

Example 5. (ε-ambiguity). For better exposition, this example focuses on the nominal signal s_i (instead of normalized signals). Consider the signal space $\mathcal{S} = (0, 1)$; signals are

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

[30] 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.

i.i.d., and the DGP takes the form of $g_m = (g_m^0, g_m^1)$, 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 actual 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 occurs if individuals precisely perceive the actual DGP. Now suppose that individuals face ambiguity and perceive a set $M_\varepsilon = [m_0 - \varepsilon, m_0 + \varepsilon] \subset \mathbb{R}_+$. Corollary 1 implies that for any $\varepsilon > 0$, complete learning no longer holds, and society will settle on an incorrect action with strictly positive probability.

Conditions for Complete Learning

Previous discussion highlights the fragility of complete learning under ambiguity. It is therefore natural to ask when complete learning does occur. Section 1 of the Supplementary Materials provides a *necessary and sufficient* condition for complete learning when DGPs have power tails. Specifically, I show that when the actual DGP and all perceived DGPs have power tails and satisfy some regularity conditions, complete learning occurs if and only if: (i) the power of every perceived DGP is weakly greater than the true power (i.e., the power of the actual DGP); and (ii) at least one perceived DGP has power strictly less than the true power plus one. That is,

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

where $\mathcal{P}^*$ denotes the power of the actual DGP, and $\min \mathcal{P}(\mathcal{F}_0)$ denotes the minimum power among the DGPs in $\mathcal{F}_0$.[31] The intuition is as follows: to achieve complete learning, we need to exclude two obstacles—incorrect herding and action nonconvergence. First, to exclude incorrect herding, $\mathcal{F}_0$ must not contain highly informative DGPs, since they introduce a strong cascading force that cannot be offset by other DGPs due to the asymmetric effect. Second, to exclude action nonconvergence, $\mathcal{F}_0$ must include at least one sufficiently informative DGP. Otherwise, individuals will perceive others' actions as uninformative, which prevents action convergence.

[31] A similar condition also appears in Arieli et al. (2025). They study sequential social learning with misspecification, which corresponds to the special case in which $\mathcal{F}_0$ is a singleton set.

6 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 preferences and the α-max-min expected utility preferences.

6.1 Smooth Ambiguity Model

The MEU model makes a restrictive assumption that decisions depend only on the worst cases. To relax this assumption, I consider an extension in which individuals have the ***smooth ambiguity preferences*** (Klibanoff et al., 2005; Denti and Pomatto, 2022). Suppose that the set of perceived DGPs can be parametrized such that $\mathcal{F}_0 = \{F(\cdot, \alpha)\}_{\alpha \in \mathcal{A}}$, where $\mathcal{A} = [\underline{\alpha}, \overline{\alpha}] \subset \overline{\mathbb{R}}$, and that the preferences satisfy:

$$V_i(a) = \phi^{-1}\left(\int_{\mathcal{A}^{i-1}} \phi\left[\mathbb{E}_{\alpha_1,...,\alpha_{i-1}}(U(a,\theta)|I_i)\right] d\mu(\alpha_1,...,\alpha_{i-1})\right), \tag{3}$$

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

Example 1'. (Cascades under smooth ambiguity preferences). Consider the same signal structures as in Example 1. Assume that signals are i.i.d.. That is, for all i, we have $g_i = g(s|\theta)$ for some signal distribution g. Individuals have constant relative ambiguity aversion (CRAA) preferences.[32] That is,

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

where ρ is the coefficient of relative ambiguity aversion, and μ satisfies some regularity conditions that will be specified later.[33] Define $g_\mu(s|\theta) = \int g_\gamma(s|\theta)d\mu(\gamma)$, where g_γ denotes the DGP with signal precision γ. Here, g_μ represents individuals' expected signal distribution under μ, referred to as society's perceived DGP. The following results hold:

[32] ϕ exhibits constant relative ambiguity aversion (CRAA) with coefficient ρ if it satisfies $\phi(x) = \frac{x^{1-\rho}}{1-\rho}$ when $\rho > 0$ and $\rho \neq 1$, and $\phi(x) = \ln(x)$ when $\rho = 1$.

[33] We require that either μ has bounded support (Proposition 2), or μ has unbounded support and satisfies some "thick-tail" condition as in Assumption 1 (Proposition 3).

- **Expected utility**. When $\rho = 0$, the model reduces to the expected-utility case in which individuals perceive the actual DGP as g_μ. There are two cases: (i) if $g_\mu = g$, individuals are *correctly specified*, so the problem degenerates to standard Bayesian social learning (Smith and Sørensen, 2000); and (ii) if $g_\mu \neq g$, individuals are *misspecified*, which corresponds to misspecified social learning (Bohren and Hauser, 2021; Arieli et al., 2025). In both cases, a variety of learning outcomes can emerge, such as cascades, complete learning, or action oscillations, depending on the statistical properties of g_μ and g.

- **Ambiguity aversion**. As $\rho \to +\infty$, the probability of an information cascade approaches 1, regardless of the learning outcomes that would have occured under expected utility. In other words, as individuals become increasingly ambiguity-averse, an information cascade gradually emerges as the *only* outcome. When $\rho = \infty$, individuals have MEU preferences as in the baseline model, in which case an information cascade occurs almost surely.

This example provides a clearer perspective by allowing for a continuum of ambiguity aversion. It illustrates how ambiguity aversion fundamentally shapes social learning: while expected-utility individuals exhibit diverse behaviors, individuals with stronger ambiguity aversion are systematically driven toward information cascades, making them the dominant outcome.

Information Cascades with Smooth Ambiguity Preferences

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

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

Proposition 2 states that an information cascade can occur with an arbitrarily large probability when individuals are sufficiently ambiguity-averse. Therefore, the result under the MEU model can serve as a benchmark for high ambiguity aversion. The intuition comes from two facts: (i) the smooth model approaches the MEU model as $\rho_\phi \to \infty$, and (ii) an information cascade occurs in finite time. By making ρ_ϕ sufficiently large, we can ensure that the belief dynamics under both the smooth model and the MEU model will be arbitrarily close, up to any finite time. As a consequence, the probability of a cascade can also be arbitrarily close to 1. Note that Proposition 2 requires bounded signals. I now show that when signals are unbounded, an information cascade can still occur under some conditions.

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

Note that $\chi(\alpha) = +\infty$ means that $F(x, \alpha)$ is perfectly informative, so Assumption 1 (i) says that perceived DGPs can be arbitrarily informative. Assumption 1 (ii) imposes a lower bound on the right tail of the second-order belief, which means that individuals believe that highly informative DGPs are realized with adequately high probability.

Proposition 3. *Suppose that Assumption 1 holds, and that* ϕ *is CRAA with coefficient* ρ. *There exists* $\rho_0 < +\infty$ *such that if* $\rho > \rho_0$, *an information cascade occurs* $\mathbb{P}^*$*-almost surely, and an incorrect cascade occurs with strictly positive* $\mathbb{P}^*$*-probability.*

Proposition 3 imposes no restriction on the underlying DGPs, so an information cascade occurs for all DGPs—regardless of whether signals are bounded or unbounded—if there is adequate ambiguity in the sense of Assumption 1, and if individuals are sufficiently ambiguity-averse in the CRAA sense. Interestingly, Proposition 3 also demonstrates that an information cascade with unbounded signals is less extreme than it might first appear. 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. However, with smooth ambiguity preferences, a cascade can occur in less extreme cases in which the perfectly informative DGP carries zero weight.

6.2 α-MEU Model

The occurrence of a cascade can go even 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

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

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 expected utility model. We have the following proposition.

Proposition 4. *All previous results under MEU preferences hold for* $\alpha-$*MEU preferences.*

Proof. It can be verified that the equilibrium strategy 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 otherwise, so all action dynamics are identical. $\square$

Proposition 4 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 asymmetry between the herding force and the contrarian force: Every action taken 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 away from a herd cannot exceed the case in which previous actions contain no information. To accommodate ambiguity-loving, a more general statement should be that an information cascade emerges when: (i) there is sufficient ambiguity (i.e., individuals perceive sufficiently many DGPs), and (ii) individuals are sufficiently ambiguity-sensitive (i.e., their decisions are sufficiently influenced by the best or the worst outcomes).

Remark 1. It is worth noting that the equivalence in dynamics between MEU and α-MEU depends critically on the binary action space. Suppose that we have a general action space; then ambiguity attitudes can affect which actions will be taken in the end. For example, if individuals are ambiguity-averse, society may settle on safe actions, while if individuals are ambiguity-loving, the society will select riskier actions.[34]

7 Discussion

In this section, I discuss how the paper's framework relates to misspecified learning and Bayesian learning when individuals form a prior over DGPs. I also discuss some extensions of the baseline model.

7.1 Model Misspecification

Let us first consider how the paper's framework differs from the literature on misspecified learning. While both approaches deviate from standard Bayesian updating, they do so in fundamentally different ways. In misspecified learning, individuals consider a single, *fixed* DGP when interpreting past actions. Under ambiguity, however, individuals consider multiple DGPs simultaneously, leading to more *flexible* interpretations of public information. In particular:

[34] 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. The extension to multiple-action space is explored in Section S3 of the Supplementary Materials.

- The interpretation of public information depends on the action being evaluated. For example, if an individual observes a sequence of action 1s, she will use the most informative DGP to evaluate the utility of taking action 0, while using an uninformative DGP to evaluate the utility of taking action 1.

- The interpretation of public information can also depend on the realization of the history itself. Consider a different history in which individuals in odd-numbered positions choose action 1 while those in even-numbered positions choose action 0. In this case, when evaluating action 1, an individual will use the most informative DGPs to interpret the actions of even-indexed predecessors and the least informative DGPs consistent with the history to interpret the actions of odd-indexed predecessors.[35]

These differences imply that the mechanisms driving herding and cascades under ambiguity differ from those under misspecification. In misspecified learning, individuals perceive a single DGP, so the learning outcome depends heavily on how this DGP is specified. For example, overestimating other individuals' signal precision encourages herding, while underestimating it beyond a certain threshold can lead to action nonconvergence (Chen, 2023; Arieli et al., 2025). Under ambiguity, however, individuals consider multiple DGPs simultaneously—they may overestimate others' signal precision under some DGPs while underestimating it under others—leading to less straightforward learning outcomes. The key driver of cascades in this setting is the inherent asymmetry in worst-case reasoning across DGPs under MEU preferences. This asymmetry, absent in the single-DGP framework, allows this paper to establish the robustness of information cascades.[36]

7.2 Bayesian Uncertainty over DGPs

An alternative way to model uncertainty over DGPs is to assume that individuals assign a prior over the set of possible DGPs and update their beliefs on DGPs in a Bayesian manner.[37] In this setup, the learning outcome depends on the prior specification, much like

[35]Note that uninformative DGPs are not consistent with this history. For instance, suppose individual 2 has an uninformative DGP; then she should follow the first individual and choose action 1. Thus, observing $a_2 = 0$ shows that individual 2 must have received an informative signal supporting state 0.

[36]To isolate the role of ambiguity attitudes from misspecification, Section S6 in the Supplementary Materials analyzes a setting in which individuals' DGP perceptions align with the objective probability (termed "consistent ambiguity"). Imagine that there is an ex ante stage in which each individual's DGP is drawn from a second-order distribution, and individuals consider only those DGPs that were possible ex ante. Moreover, all probabilities are assessed using this ex ante distribution. This scenario, which can be viewed as a special case of this paper's framework, eliminates the effects of misspecification and ensures that cascades arise solely from ambiguity attitudes.

[37]In this case, individuals are learning two objectives simultaneously—the underlying state of the world and others' DGPs. Examples include Acemoglu et al. (2016), Liang and Mu (2020), and Huang (2024).

in misspecified learning.

If the prior assigns strictly positive probability to the true DGP sequence—the "grain of the truth" condition in Kalai and Lehrer (1993)—and signals are unbounded, complete learning occurs almost surely, as if individuals perfectly understood the true DGPs.[38] In contrast, if the above condition doesn't hold, complete learning is not guaranteed. For example, if individuals believe that predecessors' DGPs are i.i.d. draws from some distribution μ, then observed actions provide no information about others' DGPs. This effectively reduces the problem to a misspecified learning setting, in which the perceived DGP is the expected DGP under μ, i.e., $F_\mu = \int F d\mu(F)$. Similarly, when signals are bounded, the emergence of information cascades also depends on the prior specification.

In summary, within the Bayesian framework, different DGPs still have different effects on social learning—some DGPs encouraging cascades, others preventing them—but these effects are filtered through individuals' beliefs, which in turn depend on their priors. In contrast, this paper adopts a prior-free approach, in which decisions are guided directly by the utility induced by different DGPs, rather than by beliefs about them. This represents a fundamental departure from the Bayesian perspective. Moreover, when the prior has high-dimensional support, Bayesian learning can become analytically intractable. In this sense, the prior-free framework also offers a more tractable way to study learning under substantial uncertainty.

7.3 Extensions

I also discuss several extensions in the Supplementary Materials. First, while this paper focuses on the standard setup in which the state and action space are binary, key insights still apply in cases with multiple states and actions.[39] Second, the paper assumes that all individuals share a common set of DGPs, and I discuss an extension in which individuals hold heterogeneous DGP sets. For example, some individuals may face less ambiguity than others by considering a smaller set of DGPs. Third, the benchmark model assumes that individuals update beliefs prior-by-prior, 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 perfectly understand the network structure. I present an extension

[38] This is because under the "grain of the truth" condition, subjective and objective probabilities merge in the limit; see also Blackwell and Dubins (1962). A formal proof in this context can be found in Appendix A.1.9 in Chen (2022b). However, it is worth noting that the "grain of the truth" condition is more restrictive than its name suggests, as the prior is defined over an infinite-dimensional space, making the assumption of assigning positive probability to a single DGP sequence quite strong.

[39] Arieli and Mueller-Frank (2021) and Kartik et al. (2024) extend the standard SSL framework to allow for general state and action space. They focus on Bayesian agents, so similar techniques cannot be applied here, which prevents this paper from achieving similar generality.

in which individuals also face ambiguity about the network structure.

8 Related Literature

This paper builds on the classical sequential social learning framework introduced by Banerjee (1992) and Bikhchandani et al. (1992). In this framework, herding occurs eventually; however, (i) the correctness of herding depends on whether signals are bounded (Smith and Sørensen, 2000), and (ii) the emergence of an information cascade depends on some statistical properties of the underlying DGP (Herrera and Hörner, 2012; Smith et al., 2021). This paper introduces ambiguity into the standard framework and shows that an information cascade is a robust result under ambiguity.

This paper is also related to the literature on social learning with misspecification. Bohren (2016) and Bohren and Hauser (2021) examine a sequential social learning problem in which individuals misspecify the learning environment. Bohren (2016) finds that different specifications can lead to different learning outcomes, including complete learning, incomplete learning, and cyclical actions. Bohren and Hauser (2021) incorporate these results into a more general framework and find that complete learning is robust with respect to small misspecifications.[40] Frick et al. (2020, 2023) establish the fragility of complete learning in social learning. Specifically, Frick et al. (2020) 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. (2023) propose a local martingale-based approach and show the fragility of complete learning with heterogeneous risk preferences. Arieli et al. (2025) show that the efficiency of social learning can be enhanced if individuals moderately underestimate the precision of others' signals. In contrast, the key feature of this paper is that individuals face ambiguity and entertain multiple DGPs simultaneously, so the resulting dynamics and mechanism are very different from misspecified learning, as discussed in Section 7.

This paper contributes to the growing literature on learning under ambiguity. Most works in this thread of literature focus on individual learning rather than social learning. Examples of the former include passive learning (Marinacci, 2002; Epstein and Schneider, 2007; Marinacci and Massari, 2019; Reshidi et al., 2025), active learning and experimentation (Battigalli et al., 2019; Auster et al., 2024), and biased learning (Fryer Jr et al., 2019; Chen, 2022a). This paper complements that literature by investigating a social learning problem in which informational ambiguity occurs very naturally. One relevant paper is by Ford et al.

[40]This finding stands in contrast to this paper's finding that complete learning is not robust. The difference is driven by their assumption that society has a positive fraction of "autarkic agents" who only act according to their private signals.

(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 behavior, whereas this paper shows an information cascade must occur under ambiguity, and the mechanism is quite different.[41] In addition to previous applications, learning under ambiguity is also examined in recent works in decision theory (Cheng, 2022; Li, 2022; Tang, 2022; Kovach, 2024) and in experimental studies (De Filippis et al., 2022; Epstein and Halevy, 2024; Galanis et al., 2024; Liang, 2024).

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, this can occur 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 information (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; as a result, this paper's deviation from the Bayesian paradigm is created mainly by ambiguity and ambiguity preferences.

9 Conclusion

This paper investigates a sequential social learning problem in which individuals face ambiguity regarding other people's DGPs. In contrast to previous studies showing that various learning outcomes can emerge depending on modeling details, this paper establishes information cascades as a robust outcome under ambiguity. Specifically, under sufficient ambiguity, an information cascade occurs almost surely—regardless of many features of the learning environment, such as the statistical properties of the actual signal-generating processes or the presence of a particular DGP specification. Interestingly, the paper also demonstrates that several non-cascade results in prior research are fragile to even small perturbations in ambiguity. While the focus here is on sequential social learning, an interesting direction for future research is to explore how ambiguity affects learning in broader settings, such as networked environments or repeated interactions.

[41] Their result employs 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 behavior.

A Proofs

A.1 Proof of Theorem 1

I first present some useful results:

Lemma 2. *For all normalized DGPs 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, and an incorrect cascade occurs with strictly positive $\mathbb{P}^*$-probability.

Proof. Suppose that for all $r_i \in \left(\frac{1}{\gamma}, \gamma\right)$, the ratio $\frac{r_{i+1}}{r_i}$ 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 then lead to a cascade of action 1. Also notice that

$$\frac{\mathbb{P}^*(\lambda_i > 1)}{1 - \mathbb{P}^*(\lambda_i > 1)} = \frac{1 - \mathbb{F}^0(1)}{\mathbb{F}^0(1)} = \frac{\mathbb{F}^1(1)}{\mathbb{F}^0(1)} = \frac{\int_{1/\gamma}^1 x d\mathbb{F}^0(x)}{\int_{1/\gamma}^1 d\mathbb{F}^0(x)} \geq \frac{1}{\gamma} \Rightarrow \mathbb{P}^*(\lambda_i > 1) \geq \frac{1}{1+\gamma}, \tag{4}$$

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

$$\mathbb{P}^*(\text{Cascade}|r_i \geq 1) \geq \mathbb{P}^*(\lambda_i, \lambda_{i+1}, ..., \lambda_{i+K-1} > 1 | r_i \geq 1) \geq \left(\frac{1}{1+\gamma}\right)^K > 0, \tag{5}$$

and similarly,

$$\mathbb{P}^*(\text{Cascade}|r_i < 1) \geq \left(\frac{\gamma}{1+\gamma}\right)^K > 0. \tag{6}$$

By Levy's 0-1 Law, as $i \to \infty$,

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

Equations (5) and (6) 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—that is, a cascade occurs almost surely. Because a cascade occurs after finitely many individuals, it can be incorrect with strictly positive probability. □

Lemma 4. (Dynamics of r_i) *For all* $r_i \in \left(\frac{1}{\gamma}, \gamma\right)$, *we have:*

$$r_{i+1} = \sqrt{\sup_{F_i \in \mathcal{F}_0} \phi_{F_i} (r_i, a_i) \times \inf_{F_i \in \mathcal{F}_0} \phi_{F_i} (r_i, a_i)} \times r_i, \tag{7}$$

where

$$\phi_{F_i} (r_i, 1) = \frac{1 - F_i^1 (1/r_i)}{1 - F_i^0 (1/r_i)} \text{ and } \phi_{F_i} (r_i, 0) = \frac{F_i^1 (1/r_i)}{F_i^0 (1/r_i)}.$$

Proof. From Proposition 1, individual i chooses action 1 if her private signal λ_i is greater than $1/r_i$, and action 0 if otherwise. Therefore, $\phi_{F_i}(r_i, a_i)$ represents the likelihood ratio of observing $a_i = 1$ in state 1 versus state 0 when individual i's DGP is F_i. Consequently, the highest and lowest public likelihood ratios satisfy:

$$\bar{l}_{i+1} = \sup_{F_i \in \mathcal{F}_0} \phi_{F_i} (r_i, a_i) \times \bar{l}_i \text{ and } \underline{l}_{i+1} = \sup_{F_i \in \mathcal{F}_0} \phi_{F_i} (r_i, a_i) \times \underline{l}_i,$$

where the equalities follow from the Cartesian structure of $\mathcal{F}_0^i$, which allows the supremum and infimum public likelihood ratios to be expressed recursively. Taking the geometric average of these two equations yields equation (7). □

Proof of Theorem 1

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

Lemma 5. *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. From Lemma 4, we know that if $a_i = 1$, then

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

Let F_γ be the DGP satisfying $\operatorname{supp}(F_\gamma) = \left\{\gamma, \frac{1}{\gamma}\right\}$, i.e., the "most informative" DGP that generates only signals with the most extreme signals. Then, for all $r_i \in \left(\frac{1}{\gamma}, \gamma\right)$, we have:

$$\sup_{F_i \in \mathcal{F}_0} \frac{1 - F_i^1\left(\frac{1}{r_i}\right)}{1 - F_i^0\left(\frac{1}{r_i}\right)} \geq \frac{1 - F_\gamma^1\left(\frac{1}{r_i}\right)}{1 - F_\gamma^0\left(\frac{1}{r_i}\right)} = \gamma. \tag{8}$$

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

$$\inf_{F_i \in \mathcal{F}_0} \frac{1 - F_i^1\left(\frac{1}{r_i}\right)}{1 - F_i^0\left(\frac{1}{r_i}\right)} \geq 1. \tag{9}$$

Combining (8) and (9), it follows that $r_{i+1} \geq \sqrt{\gamma} \times r_i$ for all $r_i \in \left(\frac{1}{\gamma}, \gamma\right)$, when $a_i = 1$. The argument for $a_i = 0$ is symmetric. □

Lemma 5 immediately implies Theorem 1. If signals are unbounded (i.e., $\gamma = \infty$), then by Lemma 5, $r_2 = \begin{cases} \infty & \text{if } a_1 = 1 \\ 0 & \text{if } a_1 = 0 \end{cases}$, so a cascade occurs immediately after the first action. If signals are bounded, then Lemma 5 guarantees that the condition in Lemma 3 is satisfied, which also implies that an information cascade occurs almost surely.

A.2 Proof of Theorem 2

Proof. **Proof of Theorem 2 (1):** Suppose that there exists $F \in \mathcal{F}_0$ that is discrete at γ. Let $p = F^0\left(\frac{1}{\gamma}\right) > 0$, which is the probability that F^0 places on $1/\gamma$. Suppose $a_i = 1$ and

$r_i \in \left(\frac{1}{\gamma}, \gamma\right)$. Then:

$$\bar{l}_{i+1} = \bar{l}_i \times \sup_{F_i \in \mathcal{F}_0} \frac{1 - F_i^1\left(\frac{1}{r_i}\right)}{1 - F_i^0\left(\frac{1}{r_i}\right)} \geq \bar{l}_i \times \frac{1 - F^1\left(\frac{1}{r_i}\right)}{1 - F^0\left(\frac{1}{r_i}\right)} \geq \bar{l}_i \cdot \left[\lim_{r \to \gamma} \frac{1 - F^1\left(\frac{1}{r}\right)}{1 - F^0\left(\frac{1}{r}\right)}\right] = \bar{l}_i \cdot \frac{1 - \frac{1}{\gamma} \cdot p}{1 - p}, \tag{10}$$

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

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

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

Proof of Theorem 2 (2): Suppose that there exists some $F \in \mathcal{F}_0$ that is continuously differentiable on $(\gamma - \varepsilon, \gamma)$ for some $\varepsilon > 0$, and that its density f satisfies $f^1(\gamma) > \frac{2}{\gamma - 1}$. If F is discrete at γ, then part (1) already implies that an information cascade occurs almost surely. So we now consider the case in which F is continuous at γ. Suppose $a_i = 1$. Then,

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

Let $I'(\gamma) \equiv \lim_{\delta \to 0} I'(\gamma - \delta)$, and let f^0 denote the density function of F^0. Then:

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

where the second equality follows from the fact that $f^0(\gamma) = \frac{1}{\gamma} f^1(\gamma)$. Since F^1 is continuously differentiable on $(\gamma - \varepsilon, \gamma)$, there exists $\varepsilon_0 > 0$ such that $I'(r) < 0$ for all $r \in [\gamma - \varepsilon_0, \gamma)$. And since $I(\gamma) = \gamma$, it follows that $I(r) \geq \gamma$ for all $r \in [\gamma - \varepsilon_0, \gamma]$. Now for all $r_i \in \left(\frac{1}{\gamma - \varepsilon_0}, \gamma - \varepsilon_0\right)$, and 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 $K < \infty$ such that after K actions of 1, $r_i \geq \gamma - \varepsilon_0$. Moreover, if $r_i \in [\gamma - \varepsilon_0, \gamma]$

and $a_i = 1$, we have $r_{i+1} \geq I(r_i) \geq \gamma$, which implies that $K+1$ consecutive actions of 1 will trigger a cascade of action 1. Similarly, $K+1$ consecutive actions of 0 will trigger a cascade of action 0. Applying the proof of Lemma 3 again, it follows 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 $r \in \mathbb{R}_{++}$ (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 sets Π_0 with $r_0 < r$ (or $r_0 > R$).

Here, $\mathbb{P}^*_{r_0}$ denotes the objective probability measure conditional on the average public likelihood ratio being r_0. Roughly speaking, state θ is locally stable if beliefs converge to δ_θ with strictly positive probability when public beliefs are close to δ_θ. We have the following results:

Lemma 6. *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 symmetry, I prove the result only for herding on action 1; the argument for action 0 is analogous. 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,

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

Hence $\{r_i\}$ 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 $F \in \mathcal{F}_0$ be a DGP with unbounded signals. Then:

$$r_{i+1} \geq r_i \times \sqrt{\frac{1-F^1(1/r_i)}{1-F^0(1/r_i)}}. \tag{11}$$

Taking limits on both sides of (11) gives $R \geq \sqrt{\frac{1-F^1(1/R)}{1-F^0(1/R)}} \times R$, which implies $\sqrt{\frac{1-F^1(1/R)}{1-F^0(1/R)}} \leq 1$.

But 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 7. *Suppose that $\mathcal{F}_0$ contains a DGP with unbounded signals. If both states 0 and 1 are locally stable, then: (i) both correct and incorrect herding occur with strictly positive $\mathbb{P}^*$-probability; and (ii) herding occurs $\mathbb{P}^*$-almost surely.*

Proof. (i) From the definition of local stability, Lemma 6, and the fact that $\{r_i\}$ is a Markov process, we know that both correct and incorrect herding occur with strictly positive probability when r_i is sufficiently small or large—that is, when $r_i \in C = \{r_i < r\} \cup \{r_i > R\}$ for some $r, R \in (0, +\infty)$. Outside of C, r_i is bounded away from 0 and $+\infty$, so there exists $K < \infty$ such that K identical actions can push r_i into C.[42] This further implies that both $\{r_i \to 0\}$ and $\{r_i \to \infty\}$ occur with strictly positive probability, i.e., both types of herding can occur. (ii) Let $H = \{r_i \to 0\} \cup \{r_i \to \infty\}$, which denotes the event that herding occurs (by Lemma 6). Levy's 0-1 Law implies that $\mathbb{P}^*(H|h_i) \to \mathbb{P}^*(H|h_\infty) = 1_H \in \{0, 1\}$. From part (i), there exists $\delta > 0$ such that $\mathbb{P}^*(H|h_i) > \delta$ for all i and h_i. Hence, $1_H = 1$ almost surely—that is, herding occurs almost surely. □

A.3.2 Formal Proof of Theorem 3

Lemma 8. $\sqrt{G_F(1/x)} = \sqrt{\frac{1-F^1(x)}{1-F^0(x)}} \sim 1 + \frac{1}{2}F^0(x)$ *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 $\sqrt{\frac{1-F^1(x)}{1-F^0(x)}} \sim \sqrt{1 + F^0(x)} = 1 + \frac{1}{2}F^0(x) + o(F^0(x))$, which proves the lemma. □

Lemma 9. *Under the conditions of Theorem 3, state 1 is locally stable.*

Proof. We 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 that $a_i = \theta$ for all i,

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

i.e., an action-θ herd starts from the first individual. Then:

$$\mathbb{P}_{r_0}^{*}\left(H_1\right)=\lim_{i\to\infty}\mathbb{P}_{r_0}^{0}\left(a_1=a_2=...a_i=1\right)=\prod_{i=1}^{\infty}\left[1-\mathbb{F}_i^0\left(\frac{1}{r_i}\right)\right]\geq\prod_{i=1}^{\infty}\left[1-a\times\left(\frac{1}{r_i}\right)^{\alpha}\right],\tag{12}$$

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

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

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

$$\begin{aligned}\lim_{r\to\infty}\frac{\sqrt{\frac{1-F^1(1/r)}{1-F^0(1/r)}}-1}{\left(1+\frac{1}{r^q}\right)^{1/q}-1}&=\lim_{r\to\infty}\frac{\sqrt{\frac{1-F^1(1/r)}{1-F^0(1/r)}}-1}{\frac{1}{r^q}}\times\lim_{r\to\infty}\frac{\frac{1}{r^q}}{\left(1+\frac{1}{r^q}\right)^{1/q}-1}\\&=\lim_{r\to\infty}\frac{\frac{1}{2}F^0\left(1/r\right)}{\frac{1}{r^q}}\times\lim_{r\to\infty}\frac{\frac{1}{r^q}}{\left(1+\frac{1}{r^q}\right)^{1/q}-1}\qquad(13)\\&>\lim_{r\to\infty}\frac{\frac{1}{2}\left(1/r\right)^p}{\frac{1}{r^q}}\times q=\infty,\end{aligned}$$

where (13) follows from Lemma 8. From Lemma 6, 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

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

which further implies that

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

After iterations, we obtain

$$r_i\geq\left(r_0^q+i\right)^{1/q},\qquad\forall i\geq1.\tag{14}$$

After substituting (14) into (12), we know that for all $r_0\geq R$ with R sufficiently large,

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

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

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

which establishes the local stability of state 1. □

Lemma 10. *Under the conditions of Theorem 3, state 0 is locally stable.*

Proof. The case for state 0 is symmetric to Lemma 9. Let r_i denote the average likelihood ratio after $h_i=(0,...,0)$. By symmetry, it follows that

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

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 9, there exists R such that $\mathbb{P}^*_{1/r_0}\left(H_1\right)\geq\varepsilon>0$ for all $1/r_0>R$. So we have $\mathbb{P}^*_{r_0}\left(H_0\right)\geq\mathbb{P}^*_{1/r_0}\left(H_1\right)\geq\varepsilon>0$ for all $r_0<1/R$, which establishes the local stability of state 0. □

Combining Lemmas 7 to 10, we know that herding occurs almost surely, and an incorrect herd occurs with strictly positive probability, so Theorem 3 is proved.

A.4 Proof of Proposition 2

Proof. Since ϕ is strictly increasing, the decision rule is monotonic in λ_i, so there exists a cutoff $r^\phi(h_i)$ (denoted 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 smooth ambiguity preferences converge to MEU preferences as $\underline{\rho}_\phi\to+\infty$ (see Proposition 3 in Klibanoff et al. (2005)). Under MEU, the decision cutoff is given by the average likelihood ratio r_i. Hence, 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 any $\epsilon>0$ and $I<\infty$, there exists some $\rho_0<\infty$ such that for all ϕ that satisfies $\underline{\rho}_\phi>\rho_0$, we have:

$$|r_{i+1}^\phi/r_i^\phi-r_{i+1}/r_i|<\epsilon\quad\text{for all } i\leq I. \tag{15}$$

Suppose that $\mathcal{F}_0$ satisfies condition (i) in Theorem 2, i.e., it contains a DGP that is discrete at γ.[43] From the proof of Theorem 2 (i), the average public likelihood ratios satisfy:

$$r_{i+1}/r_i \begin{cases} \geq \beta & a_i = 1 \\ \leq 1/\beta & a_i = 0 \end{cases} \text{ when } r_i \in (1/\gamma, \gamma), \tag{16}$$

for some $\beta > 1$. Combining (15) and (16), it follows that for all ϕ satisfying $\underline{\rho}_\phi > \rho_0$ and all $i \leq I$, we have:

$$r_{i+1}^\phi/r_i^\phi \begin{cases} \geq \beta - \epsilon & a_i = 1 \\ \leq \frac{1}{\beta} + \epsilon & a_i = 0 \end{cases} \text{ when } r_i^\phi \in (1/\gamma, \gamma). \tag{17}$$

This inequality implies that there exists some $K < \infty$ such that a cascade is triggered after at most K identical actions. Let N_i denote the event that $r_i^\phi \in \left(\frac{1}{\gamma}, \gamma\right)$. Then:

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

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

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

As a consequence,

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

Therefore, the probability that no cascade occurs before individual I is less than $\frac{1}{(1-q)I}$, which implies that the probability of an information cascade is greater than $1 - \frac{1}{(1-q)I}$. This probability can be made arbitrarily close to 1 as long as I is sufficiently large. □

[43] The proof for condition (ii) in Theorem 2 is analogous. The only change is that (16) and (17) are satisfied when r_i and r_i^ϕ are in $\left(\frac{1}{\gamma - \varepsilon_0}, \gamma - \varepsilon_0\right)$.

A.5 Proof of Proposition 3

Proof. Suppose that $a_1 = 1$, and that individual 2 receives 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 $F(\cdot, \alpha_1)$. Therefore, 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 1, 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

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

which implies $V_2(0) = 0$. The utility of choosing action 1 satisfies

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

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

Bibliography

Acemoglu, Daron, Victor Chernozhukov, and Muhamet Yildiz (2016) "Fragility of asymptotic agreement under Bayesian learning," *Theoretical Economics*, 11 (1), 187–225. 25

Acemoglu, Daron, Munther A Dahleh, Ilan Lobel, and Asuman Ozdaglar (2011) "Bayesian learning in social networks," *The Review of Economic Studies*, 78 (4), 1201–1236. 10

Acemoglu, Daron, Ali Makhdoumi, Azarakhsh Malekian, and Asuman Ozdaglar (2022) "Learning from reviews: The selection effect and the speed of learning," *Econometrica*, 90 (6), 2857–2899. 2

Anderson, Lisa R and Charles A Holt (1997) "Information cascades in the laboratory," *American Economic Review*, 847–862. 12

Arieli, Itai, Yakov Babichenko, Stephan Müller, Farzad Pourbabaee, and Omer Tamuz (2025) "The Hazards and Benefits of Condescension in Social Learning," *Theoretical Economics*, 20 (1). 12, 20, 22, 25, 27

Arieli, Itai and Manuel Mueller-Frank (2021) "A general analysis of sequential social learning," *Mathematics of Operations Research*, 46 (4), 1235–1249. 26

Auster, Sarah, Yeon-Koo Che, and Konrad Mierendorff (2024) "Prolonged learning and hasty stopping: the wald problem with ambiguity," *American Economic Review*, 114 (2), 426–461. 27

Avery, Christopher and Peter Zemsky (1998) "Multidimensional uncertainty and herd behavior in financial markets," *American Economic Review*, 724–748. 2

Banerjee, Abhijit V (1992) "A simple model of herd behavior," *Quarterly Journal of Economics*, 107 (3), 797–817. 2, 11, 27

Banerjee, Snehal and Brett Green (2015) "Signal or noise? Uncertainty and learning about whether other traders are informed," *Journal of Financial Economics*, 117 (2), 398–423. 2

Barham, Bradford L, Jean-Paul Chavas, Dylan Fitz, Vanessa Ríos Salas, and Laura Schechter (2014) "The roles of risk and ambiguity in technology adoption," *Journal of Economic Behavior & Organization*, 97, 204–218. 2

Battigalli, Pierpaolo, Simone Cerreia-Vioglio, Fabio Maccheroni, and Massimo Marinacci (2015) "Self-confirming equilibrium and model uncertainty," *American Economic Review*, 105 (2), 646–77. 8

Battigalli, Pierpaolo, Alejandro Francetich, Giacomo Lanzani, and Massimo Marinacci (2019) "Learning and self-confirming long-run biases," *Journal of Economic Theory*, 183, 740–785. 27

Bikhchandani, Sushil, David Hirshleifer, Omer Tamuz, and Ivo Welch (2024) "Information cascades and social learning," *Journal of Economic Literature*, 62 (3), 1040–1093. 2, 12

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, 11, 15, 27

Blackwell, David and Lester Dubins (1962) "Merging of opinions with increasing information," *The Annals of Mathematical Statistics*, 33 (3), 882–886. 26

Bohren, J Aislinn (2016) "Informational herding with model misspecification," *Journal of Economic Theory*, 163, 222–247. 27

Bohren, J Aislinn and Daniel N Hauser (2021) "Learning with heterogeneous misspecified models: Characterization and robustness," *Econometrica*, 89 (6), 3025–3077. 12, 22, 27

Bose, Subir and Ludovic Renou (2014) "Mechanism design with ambiguous communication devices," *Econometrica*, 82 (5), 1853–1872. 9

Cai, Hongbin, Yuyu Chen, and Hanming Fang (2009) "Observational learning: Evidence from a randomized natural field experiment," *American Economic Review*, 99 (3), 864–882. 12

Calford, Evan M (2021) "Mixed strategies and preference for randomization in games with ambiguity averse agents," *Journal of Economic Theory*, 197, 105326. 9

Çelen, Boğaçhan and Shachar Kariv (2004a) "Distinguishing informational cascades from herd behavior in the laboratory," *American Economic Review*, 94 (3), 484–498. 12

——— (2004b) "Observational learning under imperfect information," *Games and Economic behavior*, 47 (1), 72–86. 12

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. 28

Chen, Jaden Yang (2022a) "Biased learning under ambiguous information," *Journal of Economic Theory*, 105492. 27

——— (2022b) *Essays on Learning under Model Uncertainty*: Cornell University. 26

——— (2023) "Supplement to "Sequential Learning under Informational Ambiguity"," *Available at SSRN 4616856*. 8, 25

Cheng, Xiaoyu (2022) "Relative maximum likelihood updating of ambiguous beliefs," *Journal of Mathematical Economics*, 99, 102587. 28

Conley, Timothy G and Christopher R Udry (2010) "Learning about a new technology: Pineapple in Ghana," *American Economic Review*, 100 (1), 35–69. 2

Dasaratha, Krishna and Kevin He (2020) "Network structure and naive sequential learning," *Theoretical Economics*, 15 (2), 415–444. 28

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. 28

DeMarzo, Peter M, Dimitri Vayanos, and Jeffrey Zwiebel (2003) "Persuasion bias, social influence, and unidimensional opinions," *Quarterly Journal of Economics*, 118 (3), 909–968. 28

Denti, Tommaso and Luciano Pomatto (2022) "Model and predictive uncertainty: A foundation for smooth ambiguity preferences," *Econometrica*, 90 (2), 551–584. 21

Epstein, Larry G and Yoram Halevy (2024) "Hard-to-interpret signals," *Journal of the European Economic Association*, 22 (1), 393–427. 28

Epstein, Larry G and Martin Schneider (2007) "Learning under ambiguity," *Review of Economic Studies*, 74 (4), 1275–1303. 26, 27

Eyster, Erik and Matthew Rabin (2010) "Naive herding in rich-information settings," *American Economic Journal: Microeconomics*, 2 (4), 221–43. 28

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. 27

Frick, Mira, Ryota Iijima, and Yuhta Ishii (2020) "Misinterpreting others and the fragility of social learning," *Econometrica*, 88 (6), 2281–2328. 27

——— (2023) "Belief convergence under misspecified learning: A martingale approach," *The Review of Economic Studies*, 90 (2), 781–814. 27

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. 27

Galanis, Spyros, Christos A Ioannou, and Stelios Kotronis (2024) "Information aggregation under ambiguity: theory and experimental evidence," *Review of Economic Studies*, 91 (6), 3423–3467. 28

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, 7

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. 28

Hansen, Lars Peter et al. (2014) "Uncertainty outside and inside economic models," *Journal of Political Economy*, 122 (5), 945–87. 3

Hansen, Lars Peter and Massimo Marinacci (2016) "Ambiguity aversion and model misspecification: An economic perspective," *Statistical Science*, 31 (4), 511–515. 3

Herrera, Helios and Johannes Hörner (2012) "A necessary and sufficient condition for information cascades." 12, 27

Hu, Nan, Paul A Pavlou, and Jennifer Zhang (2006) "Can online reviews reveal a product's true quality? Empirical findings and analytical modeling of online word-of-mouth communication," in *Proceedings of the 7th ACM conference on Electronic commerce*, 324–330. 2

Huang, Wanying (2024) "Learning about informativeness," *arXiv preprint arXiv:2406.05299*. 25

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

Ifrach, Bar, Costis Maglaras, Marco Scarsini, and Anna Zseleva (2019) "Bayesian social learning from consumer reviews," *Operations Research*, 67 (5), 1209–1221. 2

Kalai, Ehud and Ehud Lehrer (1993) "Rational learning leads to Nash equilibrium," *Econometrica*, 1019–1045. 26

Kartik, Navin, SangMok Lee, Tianhao Liu, and Daniel Rappoport (2024) "Beyond unbounded beliefs: How preferences and information interplay in social learning," *Econometrica*, 92 (4), 1033–1062. 26

Ke, Shaowei and Qi Zhang (2020) "Randomization and ambiguity aversion," *Econometrica*, 88 (3), 1159–1195. 9

Klibanoff, Peter, Massimo Marinacci, and Sujoy Mukerji (2005) "A smooth model of decision making under ambiguity," *Econometrica*, 73 (6), 1849–1892. 21, 36

Kovach, Matthew (2024) "Ambiguity and partial Bayesian updating," *Economic Theory*, 78 (1), 155–180. 28

Li, Jian (2022) "Updating uncertainty-averse preferences," *Working paper*. 28

Liang, Annie and Xiaosheng Mu (2020) "Complementary information and learning traps," *The Quarterly Journal of Economics*, 135 (1), 389–448. 25

Liang, Yucheng (2024) "Learning from unknown information sources," *Management Science*. 28

Libgober, Jonathan and Xiaosheng Mu (2021) "Informational robustness in intertemporal pricing," *The Review of Economic Studies*, 88 (3), 1224–1252. 9

Manski, Charles F (2000) "Identification problems and decisions under ambiguity: Empirical analysis of treatment response and normative analysis of treatment choice," *Journal of Econometrics*, 95 (2), 415–442. 8

Marinacci, Massimo (2002) "Learning from ambiguous urns," *Statistical Papers*, 43 (1), 143. 27

——— (2015) "Model uncertainty," *Journal of the European Economic Association*, 13 (6), 1022–1100. 3

Marinacci, Massimo and Filippo Massari (2019) "Learning from ambiguous and misspecified models," *Journal of Mathematical Economics*, 84, 144–149. 27

Molavi, Pooya, Alireza Tahbaz-Salehi, and Ali Jadbabaie (2018) "A theory of non-Bayesian social learning," *Econometrica*, 86 (2), 445–490. 28

Pires, Cesaltina (2002) "A rule for updating ambiguous beliefs," *Theory and Decision*, 53 (2), 137–152. 7

Reshidi, Pëllumb, Joao Thereze, and Mu Zhang (2025) "Asymptotic learning with ambiguous information," *American Economic Journal: Microeconomics*, Forthcoming. 27

Rosenberg, Dinah and Nicolas Vieille (2019) "On the efficiency of social learning," *Econometrica*, 87 (6), 2141–2168. 19, 34

Saito, Kota (2015) "Preferences for flexibility and randomization under uncertainty," *American Economic Review*, 105 (3), 1246–1271. 9

Smith, Lones and Peter Sørensen (2000) "Pathological outcomes of observational learning," *Econometrica*, 68 (2), 371–398. 5, 12, 19, 22, 27, 29

Smith, Lones, Peter Norman Sørensen, and Jianrong Tian (2021) "Informational herding, optimal experimentation, and contrarianism," *Review of Economic Studies*, 88 (5), 2527–2554. 12, 18, 27

Tang, Rui (2022) "A Theory of Contraction Updating," *Available at SSRN*. 28

Tillio, Alfredo di, Nenad Kos, and Matthias Messner (2016) "The design of ambiguous mechanisms," *The Review of Economic Studies*, 84 (1), 237–276. 9

Wald, Abraham (1950) *Statistical Decision Functions*: Wiley. 3