Learning by Slow Unlearning

Jaden Yang Chen* Yuyi Li†

December 21, 2024

Abstract

This paper examines a sequential social learning model with finite signals, where individuals update beliefs using a power Bayes' rule that allows heterogeneous weighting of public and private information. We establish that, under certain regularity conditions, a correct informational cascade almost surely emerges in the limit if and only if society gradually reduces its reliance on public information at a rate slower than $1/i$, where i denotes the individual's position in the sequence. If society abandons public information too rapidly, information aggregation fails, and individuals rely solely on their private signals in the limit. Conversely, if public information is never fully discarded, both correct and incorrect cascades may persist.

JEL Classification: D81, D83, C72

Keywords: Social learning, informational cascades, non-Bayesian learning, overconfidence.

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

†University of North Carolina at Chapel Hill, E-mail: yuyil123@ad.unc.edu

1 Introduction

Bayes' rule forms the foundation of most social learning studies, where individuals are assumed to assign equal weight to different information sources. However, numerous economic experiments, beginning with Grether (1980, 1992), reveal that people often assign varying weights to distinct sources when forming probabilities, and these weights may differ across individuals. This raises an important question: how do social learning dynamics change when individuals have the flexibility to weigh different information sources differently?

This paper investigates this problem within the classical sequential social learning model of Banerjee (1992) and Bikhchandani et al. (1992). In this model, individuals receive i.i.d. private signals from a finite signal space and observe the actions of their predecessors. We extend this framework by introducing the "power-Bayesian" rule, where individuals assign a weight as an exponent to the Bayesian term of public information. This weight, referred to as the Bayesian weight, reflects the relative importance an individual places on public versus private information. Depending on its magnitude, individuals may demonstrate overconfidence or underconfidence in their private signals. Heterogeneity in Bayesian weights further allows confidence levels to vary across the population; for example, individuals later in the sequence might assign less weight to public information, perceiving accumulated public actions as increasingly noisy and less informative.

This paper characterizes learning dynamics under the power-Bayesian rule. We show that, under certain regularity conditions, society achieves an asymptotically correct informational cascade with probability 1 if and only if the Bayesian weight gradually decays to 0 at a rate slower than $1/i$, where i denotes the individual's position in the sequence. In contrast, if the Bayesian weight decays at faster than $1/i$, no informational cascade will emerge asymptotically, and individuals will only follow their private signals in the limit. Conversely, if the Bayesian weight does not converge to 0, both correct and incorrect cascades may arise in the limit, similar to the standard model.

Next, we provide an intuitive explanation of our main result. The key concept is the informational cascade, which occurs when individuals base their decisions entirely on the actions of others, ignoring their private signals. Once a cascade occurs, the accumulation of social information halts, and society remains trapped in the cascade indefinitely. However, this is not the case in our model. When the Bayesian weights converge to zero, individuals progressively place less value on public information, so the influence of public information on individuals' choices diminishes over time. When an informational cascade occurs, the influence of public information is gradually reduced until it no longer overwhelms private signals, causing the cascade to end. As a result, society can always escape a cascade and resume accumulating information. Over time, individuals can gain access to an unbounded amount of public information.

However, the discarding of public information alone is insufficient to guarantee efficient information aggregation. If individuals discard their social observations too quickly, they would fail to adequately utilize the available public information. In such cases, private signals quickly dominate decision-making, leading individuals to act solely on their private signals and resulting in the failure of information aggregation. In contrast, discarding information too slowly will, perhaps surprisingly, not jeopardize information aggregation, provided that society fully discards all public information in the limit. While slower decay increases the likelihood of temporary cascades, as long as Bayesian weights eventually decay to 0, society cannot remain trapped in a cascade indefinitely, and social learning resumes. This paper identifies the critical decay rate as $1/i$: any Bayesian weights decaying at a rate slower than $1/i$ produce correct limit cascades. Conversely, if individuals' Bayesian weights do not decay to 0, there will always be a minimum weight on public information, enabling incorrect cascades to arise, as in the standard model.

This paper contributes to the social learning literature by introducing a tractable model to study how varying confidence in public information influences social learning outcomes. Under the power-Bayesian rule, reliance on public information is represented by Bayesian weights, enabling a precise and simple characterization of learning dynamics. The paper also identifies a novel dynamic in social learning. Traditional models with bounded signals typically exhibit one of the following outcomes: path-dependent dynamics, which can generate both correct and incorrect cascades (e.g., Banerjee (1992); Bikhchandani et al. (1992)) or complete learning, where individuals become perfectly confident in one state (e.g., Goeree et al. (2006)).[1] In contrast, this paper demonstrates that a correct informational cascade emerges almost surely in the limit—a dynamic that departs from established patterns in the literature. Specifically, (i) it is not path-dependent, as social learning asymptotically supports the correct action, and (ii) it does not constitute complete learning, as sufficiently strong private signals can still prompt deviations from the cascade. This paper also develops a new proof technique to analyze learning dynamics in environments where the martingale property does not hold, which may motivate broader applications in social learning.

[1]Another typical dynamic is non-convergence of beliefs or actions. It can happen when observations are incomplete (Çelen and Kariv (2004) and Acemoglu et al. (2011)) or when there is model misspecification (Bohren (2016) and Arieli et al. (2025)). This is also quite different from the dynamic in this paper.

1.1 Related Literature

The literature has shown that overconfidence can enhance social learning. Bernardo and Welch (2001) analyze a sequential social learning model where a proportion of agents are overconfident, underweighting public information. They show that moderate overconfidence improves social learning outcomes. Arieli et al. (2025) study a sequential social learning model with unbounded signals, where individuals misinterpret their predecessors' data-generating process. Their results demonstrate that mild condescension—underestimating the precision of predecessors—can improve information aggregation. In particular, society can achieve efficient learning, which is stronger than the complete learning result in standard sequential learning model with unbounded signals.[2]

A common feature in these studies and our paper is that individuals can exhibit overconfidence in their private information. However, previous works assume homogeneous confidence levels, leaving open the question of how heterogeneous confidence levels affect social learning. As noted in a recent survey by Bikhchandani et al. (2024):

> "A possible direction for future research is understanding conditions under which overconfidence harms social learning instead of helping. For example, if overconfidence were growing rapidly with later agents, all agents would act based only on their private signals and there would be no information aggregation."

This paper addresses this gap by introducing a tractable framework for analyzing how varying confidence levels impact social learning. Proposition 4 in this paper directly echoes the conjecture in the above quote. Specifically, we show that overconfidence harms social learning if society's confidence level, as reflected by the Bayesian weights, grows faster than a critical threshold related to $1/i$, thereby confirming the conjecture in Bikhchandani et al. (2024).[3] Conversely, overconfidence helps social learning if the confidence level grows more slowly than that threshold, resulting in limit correct cascades. Thus, a key contribution of this paper is to extend the literature's focus on the ***magnitude*** of confidence levels by demonstrating that the ***rate*** of change in confidence also plays a crucial role in social learning.

This paper assumes heterogeneous weighting of public information, which also relates to Goeree et al. (2006) who study a sequential learning model where agents have idiosyncratic

[2]Efficient learning means that the expected number of incorrect actions is finite, as defined in Rosenberg and Vieille (2019). While Smith and Sørensen (2000) shows that complete learning occurs with unbounded signals, it does not necessarily ensure efficiency.

[3]Here, we interpret the Bayesian weight as the confidence level in public information. Thus, the decreasing speed of Bayesian weights reflects the growing speed of society's confidence level over time. Suppose we use the reciprocal of the Bayesian weight to represent society's confidence level, then social learning is harmful when the confidence level grows faster than i.

tastes for actions. They show that under certain conditions, social beliefs converge to the true state even with bounded signals. This result arises because diverse preferences ensure that all actions are taken with positive probability in every state, continually revealing information. However, their mechanism is fundamentally different from ours. Besides, their model leads to complete learning, whereas ours leads to correct limit cascades but still incomplete learning.

There is substantial experimental and empirical evidence showing that individuals weigh private and public information differently. For instance, Grether (1980, 1992), Anderson and Holt (1997), Weizsäcker (2010), and Duffy et al. (2021) document that people tend to overweight their own information relative to others'.[4] In a recent experimental paper, Conlon et al. (2022) find that individuals exhibit a clear bias toward their own information, perceiving it as more precise or relevant.

Finally, this paper contributes to the broader literature on non-Bayesian social learning. A prominent strand in this literature explores naive social learning, where individuals follow heuristic rules when learning from others (DeGroot (1974), DeMarzo et al. (2003), Golub and Jackson (2010)). In this paper, individuals can also be interpreted as using a heuristic rule, assigning varying weights to public information based on their position in the sequence.[5]

2 Model

Preliminaries. The setup follows the traditional sequential social learning literature. The true state of the world is $\theta \in \{1, 0\}$. Without loss of generality, we assume that the true state is 0. An infinite sequence of individuals $I = \{1, 2, 3, \dots\}$ act in an exogenously given order, trying to match their actions with the state. Specifically, each individual i is supposed to choose an action from $A = \{1, 0\}$. Her payoff is 1 if her action a_i matches the true state, and otherwise, her payoff is 0. Individuals do not know the true state and hold a flat prior. Before taking an action, each individual i privately receives a signal s_i from a finite signal space S. All signals are independently and identically distributed with conditional c.d.f. G^θ. Following the convention, we work directly with the normalized signal $\lambda(s) = \frac{dG^1(s)}{dG^0(s)}$, which is the likelihood ratio induced by signal s. Let F^θ denote the conditional c.d.f. of the normalized signal. Let $\Lambda = \{\lambda^1, \lambda^2, \dots, \lambda^n\}$ denote the normalized signal space with $\frac{1}{\gamma} = \lambda^1 < \lambda^2 < \cdots < \lambda^n = \gamma$, where $\gamma \in (1, +\infty)$ denotes the highest likelihood ratio

[4] For more experimental evidence on overweighting or underweighting of private information relative to social information, see Morin et al. (2021) for an exhaustive survey. Benjamin (2019) introduces different types of behavioral bias that can lead to deviations from Bayesian updating.

[5] Other examples of non-Bayesian learning include Ellison and Fudenberg (1993, 1995), Bala and Goyal (1998), Epstein et al. (2010), Jadbabaie et al. (2012), Guarino and Jehiel (2013), Arieli et al. (2021), Molavi et al. (2018), Eyster and Rabin (2010, 2014), Dasaratha and He (2020), Chen (2019), Frick et al. (2024).

induced by private signals. We assume that all signals are informative, so $\lambda^j \neq 1$ for all $j = 1, 2, \ldots, n$. Other than the private signal, each individual i is able to observe the ordered action history of all her predecessors, $h_i = \{a_1, a_2, \ldots, a_{i-1}\}$.

Updating rule. The key feature of this paper is that individuals do not update beliefs in a strictly Bayesian manner. When forming posteriors, we allow individuals to value their private (λ_i) and public information (h_i) differently, which is modeled by a non-negative weight on the public information. Specifically, we assume that all individuals follow the ***power Bayes' rule***:

$$\pi_i(\theta|\lambda_i, h_i) \propto \mathbb{P}(\theta) \times \mathbb{P}(\lambda_i|\theta) \times \mathbb{P}^{\alpha_i}(h_i|\theta), \tag{1}$$

where $\alpha_i \geq 0$ is referred to as the ***Bayesian weight*** of individual i. When $\alpha_i = 1$, the rule reduces to Bayes' rule. When $\alpha_i < 1$ (or > 1), individual i underweights (or overweights) the public information relative to private information. In the extreme cases of $\alpha_i = 0$, the individual completely disregards social observations; conversely, when $\alpha_i = +\infty$, individual i places full confidence in the state where the public history is most likely to be observed. The paper aims to explore how the asymptotic learning outcome depends on the Bayesian weights sequence, $\boldsymbol{\alpha} = \{\alpha_i\}_{i=1}^n$. To simplify the analysis, we impose following assumptions:

Assumption 1. (Monotonicity) α_i *is monotonic in* i.

Assumption 2. (Power rate) $\alpha_i = O(i^{-\rho})$ *as* $i \to +\infty$.[6]

Assumption 1 can be relaxed as that α_i is monotonic when i is adequately large. Its goal is to eliminate the case where α_i oscillates infinitely often, which makes the discussion tedious.[7] Assumption 2 is the key assumption in the paper, which requires α_i decreases or increases at a power rate $i^{-\rho}$ in the limit. These two assumptions simplify the analysis. As will be shown later, for any sequence $\boldsymbol{\alpha}$ satisfying them, it is sufficient to focus on a single statistic—the rate ρ—to characterize asymptotic learning dynamics. There are three important cases: (i) when $\rho > 0$, α_i converges to 0, implying that tail individuals completely ignore public information; (ii) when $\rho < 0$, α_i grows to infinity, implying that tail individuals attach infinite weight to public information; (iii) when $\rho = 0$, α_i converges to a positive constant.

2.1 Discussion of Modeling Assumptions

This section discusses the key assumptions underlying the power Bayes' rule and their role in modeling sequential social learning. Traditional models often assume that individuals

[6] More formally, it means there exists some positive constant $c > 0$ such that $\alpha_i \times i^{\rho} = c$ as $i \to +\infty$.
[7] For a simple discussion of general Bayesian weights, see Section 7.4.

update their beliefs using Bayes' rule, treating private and public information symmetrically. In contrast, the power Bayes' rule in this paper captures the phenomenon that individuals exhibit different levels of confidence in information from different sources, weighing them differently when forming a posterior.

In the framework, confidence in public information is captured by Bayesian weights, which are heterogeneous across individuals. This heterogeneity reflects the reality that individuals at different stages in the sequence may encounter varying levels of noisiness and reliability in their public observations. For example, individuals later in the sequence may perceive public information as noisier due to accumulated inaccuracies, prompting them to rely more heavily on private signals. Conversely, it is also plausible that individuals later in the sequence believe the "wisdom of crowds" by assigning greater weight to public information. This flexibility allows us to study how varying confidence in public information affects social learning outcomes, which has not been explored in the literature.[8]

We also assume that Bayesian weights are commonly known, which is a potentially strong assumption. To justify this, we can consider all individuals as homogeneous in their decision-making process, adjusting their reliance on public information according to a common heuristic rule that depends solely on their position in the sequence. In this sense, the heterogeneity in Bayesian weights arises purely from differences in individuals' positions, while the common knowledge assumption reflects a universally understood adjustment rule. This interpretation aligns with the monotonicity assumption, which requires Bayesian weights to increase or decrease systematically with the number of observations. This suggests that society shares a simple rule for adjusting reliance on public information.

There are some other assumptions imposed on the Bayesian weights. First, they are exogenously given, determined independent of the information individuals receive. Second, we assume that each individual only distinguishes between two information sources—public and private—but assigns a homogeneous weight to each predecessor's action when interpreting the public information. While it may be more realistic to allow Bayesian weights to be endogenous or vary for each predecessor, this paper proposes the simplest possible model to capture variations in confidence levels without introducing unnecessary complexities. As will be demonstrated, even this simplified model generates dynamics that differ significantly from those in standard frameworks. Future work can build on this foundation by exploring

[8]In terms of real-world scenarios this model might be able to capture, consider the development of culture across human generations. As culture evolves, younger generations are more often encouraged to express their own feelings and ideas. As a result, people put more faith in themselves and depend less on public history. Another case this model might potentially cover is how the advancement of technology changes the way people learn. As people access different sources of information more easily, they are more likely to form their own opinions rather than blindly following the predecessors. In other words, self-confidence grows.

more realistic extensions once the core dynamics of this model are better understood.

3 Decision Criterion and Learning Concepts

In this section, we present a characterization of individuals' decision criterion in the power sequential social learning problem. Later, we define informational cascades and limit cascades.

3.1 The Power Public Likelihood Ratio

Definition 1. For each individual i and each history h_i, we define

$$x_i\left(h_i\right) = \left[\frac{\mathbb{P}\left(h_i|\theta=1\right)}{\mathbb{P}\left(h_i|\theta=0\right)}\right]^{\alpha_i},$$

which is called the ***power public likelihood ratio*** of individual i based on history h_i.

The power public likelihood ratio x_i can be viewed as a generalization of the public likelihood ratio as in the Bayesian sequential social learning problem. It is equal to the public likelihood ratio with an exponent equal to the Bayesian weight α_i, which reflects how individual i evaluates the likelihood ratio between states 1 and 0 after reweighting public information. The following proposition characterizes each individual's decision rule using the power public likelihood ratio:

Proposition 1. (Characterization of decision rule) *For all $i \in I$ and all feasible history h_i, we have:*

$$a_i = \begin{cases} 1, & \lambda_i > \frac{1}{x_i} \\ 0, & \lambda_i < \frac{1}{x_i} \end{cases}, \tag{2}$$

where x_i is individual i's power public likelihood ratio based on history h_i.

Proof. From the power Bayesian rule, we have:

$$\frac{\pi_i\left(1|\lambda_i,h_i\right)}{\pi_i\left(0|\lambda_i,h_i\right)} = \frac{\mathbb{P}\left(\lambda_i|\theta=1\right)}{\mathbb{P}\left(\lambda_i|\theta=0\right)} \times \left[\frac{\mathbb{P}\left(h_i|\theta=1\right)}{\mathbb{P}\left(h_i|\theta=0\right)}\right]^{\alpha_i} = \lambda_i \times x_i.$$

Thus, individual i chooses action 1 (or 0) if the ratio is larger (or smaller) than 1. □

When there is an indifference, we impose a tie-breaking rule that individuals follow the power public likelihood ratio, i.e., they choose action 1 whenever $x_i \geq 1$.[9] From Proposition

[9] That is, we assume individuals always follow the power public likelihood ratio. This tie-breaking rule makes the discussion a bit easier, but the choice of tie-breaking rule will not affect asymptotic dynamics.

1, the power public likelihood ratio serves as a sufficient statistic for the learning problem. Therefore, despite the complexity brought by the heterogeneity in Bayesian weights, the new learning problem can be highly tractable—we only need to keep track of the dynamics of power public likelihood ratio to understand the learning dynamics.

3.2 Informational Cascades and Limit Cascades

When the power public likelihood ratio x_i exceeds the highest signal, γ, or falls below the lowest signal, $1/\gamma$, individual i will choose a particular action irregardless of their private signals. We introduce the following concepts:

Definition 2. We say that: (i) individual i is in an ***information cascade*** if $x_i \in C \equiv [0, 1/\gamma] \cup [\gamma, +\infty]$, and (ii) a ***limit cascade*** occurs if $x_\infty \in C$, where $x_\infty = \lim_{i \to +\infty} x_i$, assuming that a limit exists.

An informational cascade occurs when the power public likelihood ratio is so large or small such that no signal can make the individual choose a different action. Here, the set C is called the ***cascade set***. The complement set, $(1/\gamma, \gamma)$, is called the ***non-cascade set***. We define $C_1 = [\gamma, +\infty]$ and $C_0 = [0, 1/\gamma]$, where C_a is called the ***cascade set of action*** a. Individual i will always choose action a if $x_i \in C_a$, regardless of her private signals.

Under power Bayes' rule, the occurrence of a cascade for individual i may not imply the occurrence of a cascade for all followers, as will be discussed in the next section. A more relevant concept is the limit cascade, in which the social information outweights private information asymptotically. We say that a ***correct limit cascade*** occurs if $x_\infty \in C_0$, that is, x_i enters the cascade set of the correct action in the limit.

4 Discussion of Learning Dynamics

In this section, we discuss the learning dynamics under the power-Bayesian rule and explain how they differ from those in standard Bayesian models. By Definition 1 and Proposition 1, the power public likelihood ratios $\{x_i\}$ satisfy the recursive expression:

$$x_{i+1} = x_i^{\frac{\alpha_{i+1}}{\alpha_i}} \times \left[\frac{\phi^1(a_i, x_i)}{\phi^0(a_i, x_i)}\right]^{\alpha_{i+1}}, \tag{3}$$

where

$$\phi^\theta(a, x) = \mathbb{P}(a|h, \theta) = \begin{cases} 1 - F^\theta\left(\frac{1}{x}\right), & \text{if } a = 1, \\ F^\theta\left(\frac{1}{x}\right), & \text{if } a = 0, \end{cases}$$

represents the probability of taking action a given the power public likelihood ratio x.

4.1 Bayesian Benchmark

If $\alpha_i = 1$ for all i, the power-Bayesian model reduces to the standard Bayesian model. In this case, equation (3) simplifies to:

$$x_{i+1} = x_i \times \frac{\phi^1(a_i, x_i)}{\phi^0(a_i, x_i)}, \tag{4}$$

where x_i is the public likelihood ratio. An important property of x_i in this setting is that it is a ***martingale***.[10] By the martingale convergence theorem, x_i almost surely converges to a limit x_∞. This result has two critical implications for social learning:

First, an informational cascade occurs almost surely. This comes from the fact that x_∞ must be a stationary point of (4), so it must satisfy:

$$\frac{\phi^1(a_\infty, x_\infty)}{\phi^0(a_\infty, x_\infty)} = 1, \tag{5}$$

where a_∞ is the limit action. As a result, the limit action conveys no information, which implies $x_\infty \in C$, consistent with the emergence of an informational cascade.[11]

Second, both correct and incorrect informational cascades occur with positive probability. The martingale property implies:

$$\mathbb{E}(x_\infty) = x_0 \in (1/\gamma, \gamma).$$

Consequently, x_∞ cannot always exceed γ or remain below $1/\gamma$. Instead, society may experience both correct and incorrect cascades. Once in a cascade—whether correct or incorrect—x_i remains unchanged, and society cannot escape an informational cascade even if it is incorrect, as illustrated in Figure 1.

4.2 Power-Bayesian Case

Under the power-Bayesian rule, the learning dynamics described by equation (3) differ fundamentally from the Bayesian case:

First, x_i may no longer be a martingale under the power-Bayesian rule. Consequently, the martingale convergence theorem does not apply, and it becomes unclear whether x_i

[10] To be more precise, it is a martingale under the probability measure conditional on the true state.
[11] When signals are finite, x_i will enter the cascade set in finite time, so a cascade must occur.

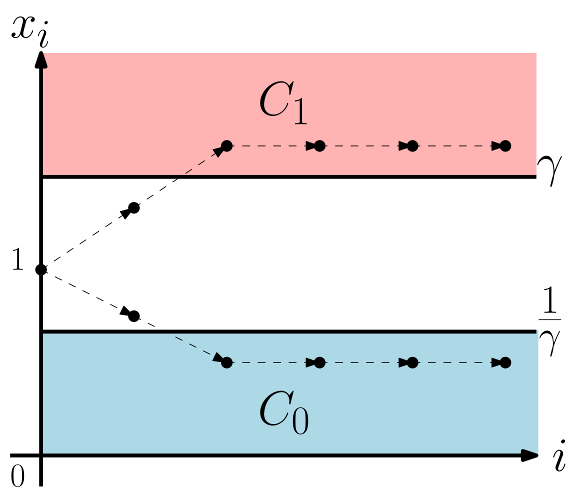


Figure 1: Bayesian case

converges. Even if we establish the convergence of x_i, it is also not obvious what limit it converges to. For instance, if $\alpha_i = 1/i$ and x_i converges, the limit x_∞ must satisfy:

$$\left[\frac{\phi^1(a_\infty, x_\infty)}{\phi^0(a_\infty, x_\infty)}\right]^{\alpha_\infty} = 1. \tag{6}$$

Since $\alpha_i \to 0$, this condition is consistent with ***any*** value of x_∞. Therefore, we are unable to determine the limit by investigating the stationary points as in the Bayesian case.

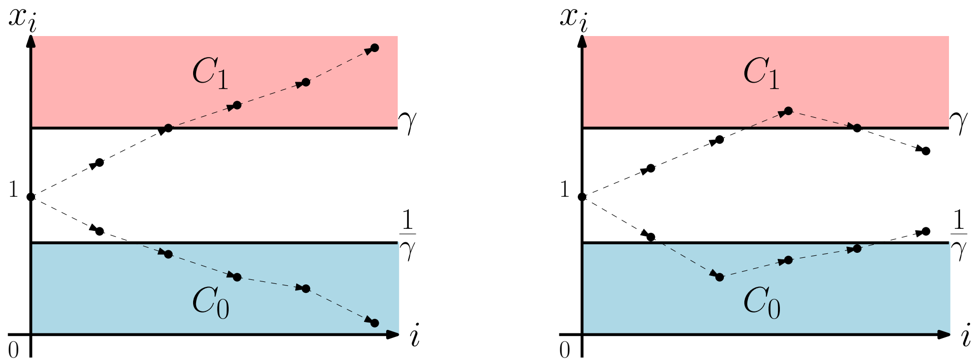
Second, informational cascades no longer imply the cessation of information aggregation. Suppose an informational cascade occurs for individual i, meaning $x_i \in C$. Then the next power public likelihood ratio is given by:

$$x_{i+1} = x_i^{\alpha_{i+1}/\alpha_i}. \tag{7}$$

In the Bayesian case, $\alpha_{i+1} = \alpha_i = 1$, so $x_{i+1} = x_i$. Once x_i enters the cascade set, it remains unchanged, and information aggregation halts permanently. However, in the power-Bayesian case, the Bayesian weight α_i influences x_{i+1}, as illustrated in Figure 2:

- If $\alpha_{i+1} > \alpha_i$, x_{i+1} moves deeper into the cascade set;
- if $\alpha_{i+1} < \alpha_i$, x_{i+1}moves closer to the boundary of the cascade set or exits it entirely.

In the latter case, x_i exiting the cascade set allows information aggregation to resume. Thus, society may experience multiple rounds of halting and restarting information aggregation. This dynamic makes the analysis of the power-Bayesian rule considerably more complex compared to the Bayesian case.



(a) α_i is increasing (b) α_i is decreasing

Figure 2: Dynamics with different α_i

5 Main Results

This section characterizes learning dynamics under the power-Bayesian rule. With some abuse of notation, we henceforth use $\mathbb{P}$ to denote the objective probability measure, under which all events are evaluated.[12] In the classical model with bounded signals, both correct and incorrect cascades occur with positive probability. However, this is no longer the case with power-Bayesian individuals:

Theorem 1. *Suppose Assumptions 1 and 2 hold. Then, a correct limit cascade happens $\mathbb{P}$-almost surely if $\rho \in (0,1)$ and only if $\rho \in (0,1]$.*

Theorem 1 provides an almost necessary and sufficient condition for a correct limit cascade to emerge almost surely.[13] There are two key insights in this result: First, the condition $\rho > 0$ means that α_i decreases and gradually approaches zero, which reflects individuals' declining confidence in the public history. Therefore, a correct limit cascade requires society to discard ***all*** public information in the limit. Second, the condition $\rho \leq 1$ requires that the decay of α_i must not be faster than $1/i$. Thus, a correct limit cascade also requires society to discard its reliance on public information at a sufficiently ***slow*** rate, i.e., sub-reciprocal rate. Next, we focus on why these two conditions are essential for ensuring a correct limit cascade. For simplicity, we assume that Assumptions 1 and 2 hold throughout this section without explicitly restating them.

5.1 Bayesian weights not decreasing to zero ($\rho \leq 0$)

To understand why the Bayesian weight must decay to 0, we present the following result:

[12] The objective probability measure refers to the probability measure conditional on the true state, 0.

[13] When $\rho = 1$, we have $\alpha_i \sim c \times i^{-1}$ for some constant $c > 0$. In this case, the existence of a correct limit cascade depends on the magnitude of c.

Proposition 2. *If $\rho \leq 0$, an informational cascade occurs $\mathbb{P}$-almost surely, and both correct and incorrect cascades occur with $\mathbb{P}$-strictly positive probability.*

The condition $\rho \leq 0$ encompasses two cases: (i) $\rho < 0$, in which the Bayesian weight increases to infinity; (ii) $\rho = 0$, in which the Bayesian weight converges to a non-zero constant, but it may increase or decrease with i. Proposition 2 establishes that if society does not eventually discard public history or increases reliance on it, an informational cascade must occur, and such cascades can be either correct or incorrect. This result arises because, if α_i does not converge to 0, there exists a minimum weight on public information. In this case, finitely many identical actions will trigger an informational cascade, just as in the standard case. If α_i is weakly increasing, once a cascade is triggered, all successors will also remain in that cascade because they value public information at least as much as the first individual in the cascade; thus, an informational cascade can persist. Because only finitely many signals are required to trigger a cascade, both correct and incorrect cascades can arise with positive probability.

Even if α_i is decreasing but does not converge to 0—which covers part of the case $\rho = 0$—an informational cascade can still persist. In such cases, however, it is possible for x_i to exit the cascade set, allowing information aggregation to resume. This dynamic resembles the case in Theorem 1, but it does not preclude the possibility of an incorrect cascade. Figure 3 provides an illustration, where $x_i \in C_1$ but $x_{i+1} \notin C_1$. When i is adequately large, we have $\alpha_{i+1} \approx \alpha_i$, so x_{i+1}—satisfying (7)—is also very close to the boundary of C_1, which is γ. Now suppose individual $i+1$ receives a signal supporting state 1, such as signal γ. Since society always places some minimum weight on the public information, the next-period power public likelihood ratio, x_{i+2}, enters the interior of C_1 and is bounded away from γ. When i is large, all future x_{i+k}'s remain close to x_{i+2}, thus staying bounded away from γ as well (i.e., they remain above $\gamma + \epsilon$ in Figure 3).[14] As a result, society is trapped in the cascade set, and an incorrect informational cascade persists.

The case of constant α_i

Previous discussion shows that if society places a minimum weight on public information, an incorrect cascade can emerge. To further understand why α_i must decrease to 0, it is helpful to consider a special case where α_i is constant:

Proposition 3. *Suppose $\alpha_i = \alpha > 0$ for all i. Then: (i) An informational cascade occurs $\mathbb{P}$-almost surely, and both correct and incorrect cascades occur with $\mathbb{P}$-strictly positive*

[14] To see this, (7) implies $x_{i+k} = x_{i+2}^{\alpha_{i+k}/\alpha_{i+2}}$ if x_{i+k} is in the cascade set. When i is large, α_{i+k} is very close to α_{i+2}, so x_{i+k} is very close to x_{i+2}.

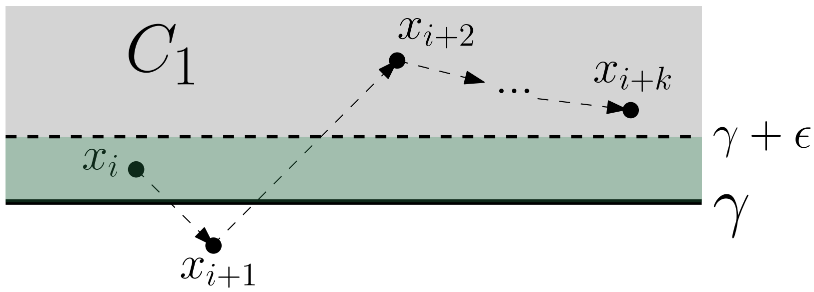


Figure 3: Power-Bayesian with $\rho = 0$ and decreasing α_i

probability; (ii) Let $p(\alpha)$ denote the probability of a correct cascade, then $p(\alpha) \to 1$ as $\alpha \to 0$.

Part (i) of Proposition 3 follows directly from Proposition 2. In fact, when α_i is constant, the public likelihood ratio becomes a martingale, and dynamics of x_i are identical to the standard case depicted in Figure 1. Proposition 3 (ii) provides an interesting comparative static: as society's common weight on public information decays to 0, the probability of a correct cascade approaches 1. The intuition is that when α is smaller, it takes more identical actions to form an informational cascade. This decay allows society to experience more information aggregation, increasing the probability of a correct cascade. For very small α, society observes many informative actions, making the probability of a correct cascade approach 1.

Proposition 3 implies that reducing Bayesian weights can improve information aggregation. However, there is a discontinuity at $\alpha = 0$: when $\alpha = 0$, individuals completely discard public information and rely solely on their private signals, resulting in $p(0) = 0$. For positive α, while the probability of an incorrect cascade can be made arbitrarily small, it remains strictly positive, leading to learning dynamics qualitatively similar to the standard Bayesian case. A natural conjecture to restore continuity is to allow α_i to gradually decay to 0. By doing so, society can place arbitrarily small weight on public information in the limit, potentially achieving better information aggregation than in the constant-α case. This conjecture also motivates why α_i is required to decay to 0 in Theorem 1.

5.2 Bayesian weights decreasing too fast ($\rho > 1$)

After explaining why the Bayesian weight must decay to 0, we now turn to discussing why the decay speed cannot be too fast.

Definition 3. We say that *no information aggregation* occurs in the limit if $x_i \to 1$.

From Proposition 1, $x_i \to 1$ implies that individuals rely only on their private signals in the limit. In this case, social learning does not influence the decisions of tail individuals, and

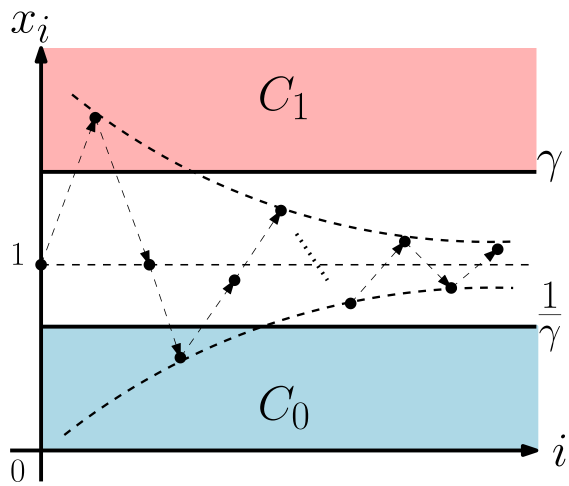


Figure 4: Power-Bayesian with $\rho > 1$

we say there is no information aggregation. We have the following result:

Proposition 4. *If $\rho > 1$, no information aggregation occurs in the limit $\mathbb{P}$-almost surely.*

Proposition 4 formalizes the conjecture in Bikhchandani et al. (2024). It states that if the society eventually discards all public information adequately fast—at a super-reciprocal speed—then no information aggregation is achieved in the limit. The intuition is as follows: when reliance on public information declines too rapidly, the speed at which new information accumulates cannot keep pace. Consequently, individuals predominantly rely on their private signals, preventing the formation of an informational cascade. The dynamics are depicted in Figure 4, in which the trajectory of x_i is bounded by two curves converging to 1.[15] As a result, we have $x_i \to 1$, and no information aggregation is achieved.

The above discussion explains why the discard speed cannot be too fast but does not yet address why the critical speed is precisely $1/i$. A deeper analysis of the underlying dynamics is required to establish this result, which is deferred to Section 6.

5.3 Summary: Learning by slow unlearning ($0 < \rho < 1$)

The previous discussion highlights the delicate balance required for successful information aggregation. Excessive reliance on public information prevents individuals from effectively conveying their private information, potentially trapping society in an incorrect informational cascade (Proposition 2). Conversely, insufficient reliance on public information can also harm social learning, as the speed of discarding public information outpaces the speed of

[15] As shown in the Appendix, the upper curve is $\gamma^{i\alpha_{i+1}}$ and the lower curve is $1/\gamma^{i\alpha_{i+1}}$. When $\rho > 1$, we have $i\alpha_{i+1} \to 0$ as $i \to +\infty$, so both curves converge to 1 in the limit.

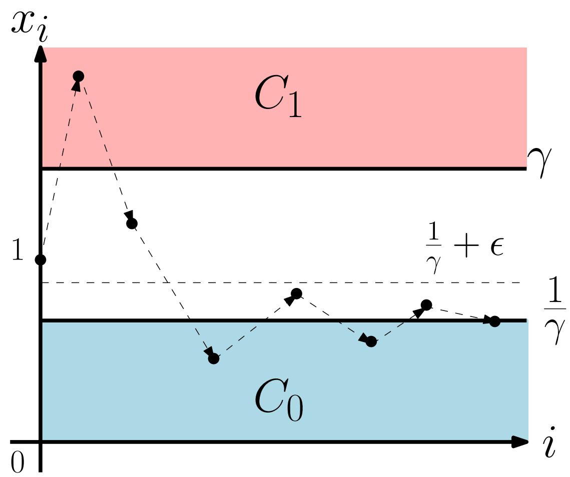


Figure 5: Power-Bayesian with $0 < \rho < 1$

information accumulation, leading to no information aggregation in the limit (Proposition 4). Thus, successful information aggregation can only be achieved when society carefully regulates the rate at which it discards public information.

Figure 5 illustrates the dynamics of x_i when $\rho \in (0, 1)$. Since $\rho > 0$, individuals assign zero weight to public information in the limit, ensuring that society avoids being trapped in any cascade set and retains access to an unbounded amount of information. On the other hand, because $\rho < 1$, individuals do not discard public information too quickly, allowing their decisions to reflect a sufficient fraction of group wisdom. These two mechanisms jointly ensure that x_i almost surely converges to the correct cascade set C_0. As shown in Figure 5, x_i undergoes multiple rounds of entering and exiting cascade sets before eventually stabilizing below $1/\gamma + \varepsilon$, where $\varepsilon > 0$ can be arbitrarily small. Importantly, this outcome is fundamentally different from the Bayesian case or the case of constant α_i, as depicted in Figure 1, where the asymptotic social learning outcome is path-dependent.

6 Proof Sketch

This section provides a sketch of the proof for Theorem 1. To illustrate the key ideas, we focus on the simplest case, where:

- Signal is binary: $\Lambda = \{1/\gamma, \gamma\}$;
- The Bayesian weight is given by $\alpha_i = 1/i^\rho$, where $0 < \rho < 1$.

From (3), the learning dynamics in this simplified case can be explicitly characterized as:

$$\ln(x_i) = \frac{1}{i^\rho} \times \underbrace{\sum_{j=1}^{i-1} \ln(\lambda_j) \mathbf{1}\{x_j \notin C\}}_{\text{public LLR}}, \tag{8}$$

where the dynamics are determined by two components: (i) the Bayesian weight α_i, and (ii) the public log likelihood ratio (LLR) based on history h_i. For the binary-signal structure, the public LLR follows a simple updating manner:

- if j is not in the cascade set, her action perfectly reveals the private signal (γ or $1/\gamma$);
- if j is in the cascade set, her action is uninformative, and public LLR stops updating.

The most challenging part of formula (8) arises from the indicator function, which depends on whether x_j is in the cascade set. This dependency introduces a key difficulty: the components of the public LLR are no longer independent. Addressing this dependence is one of the central hurdles in our proof. We outline the proof in the following five steps:

STEP 1: x_i cannot be trapped in the non-cascade set.

To solve the aforementioned difficulty, we first notice that if x_i remains outside of the cascade set, the indicator function can be eliminated. In this case, the public LLR reduces to the sum of independent signals, allowing the application of standard techniques. Guided by this observation, we focus on the event $E = \{x_i \in (1/\gamma, \gamma) \text{ for all } i \geq 1\}$, where x_i consistently stays within the non-cascade set. Under E, we have:

$$\ln(x_i) = \frac{1}{i^\rho} \times \sum_{j=1}^{i-1} \ln(\lambda_j).$$

By the Strong Law of Large Numbers (SLLN), we know:

$$\frac{1}{i} \sum_{j=1}^{i-1} \ln(\lambda_j) \to \mathbb{E} \ln(\lambda) < 0 \quad \mathbb{P}\text{-a.s.} \tag{9}$$

Since $0 < \rho < 1$, the Bayesian weight $1/i^\rho$ decays slower than $1/i$. Consequently, we have: $\ln(x_i) \to -\infty$ or equivalently, $x_i \to 0$ almost surely on E, a contradiction. This result implies that E must be a null event. By a similar argument, we can further conclude that x_i cannot stay in the non-cascade set for all sufficiently large i. In other words, once x_i enters the non-cascade set, it must exit after finitely many individuals.

Remark 1. This step also provides insight into why the Bayesian weight α_i cannot decay faster than $1/i$. Specifically, consider the case where $\rho > 1$. As $i \to +\infty$, we have:

$$\ln(x_i) = \frac{1}{i^{\rho-1}} \times \left(\frac{1}{i} \sum_{j=1}^{i-1} \ln(\lambda_j) \right) \to 0,$$

or equivalently $x_i \to 1$ almost surely on E, which is consistent with Proposition 4.

STEP 2: x_i cannot be trapped in the cascade set.

From Step 1, we know that x_i must almost surely enter a cascade set. However, once x_i enters a cascade set, it cannot remain there indefinitely. Suppose there is an informational cascade of length k starting from individual i, meaning $x_i, ..., x_{i+k-1} \in C$. Then:

$$\ln(x_{i+k}) = \frac{1}{(i+k)^{\rho}} \times \sum_{j=1}^{i-1} \ln(\lambda_j) \mathbf{1}\{x_j \notin C\}.$$

The right-hand side converges to 0 as $k \to +\infty$. This implies that an informational cascade cannot persist indefinitely, and therefore, x_i cannot remain in a cascade set forever.

Combining Steps 1 and 2, we know that x_i must undergo infinitely many rounds of entering and exiting cascade sets.

STEP 3: x_i enters the correct cascade set C_0 infinitely often.

Next, we demonstrate that x_i must enter the ***correct*** cascade set C_0 infinitely often. The intuition is as follows: (i) Step 2 implies that x_i enters the non-cascade set infinitely often; (ii) Step 1 implies that once x_i is in the non-cascade set, information is perfectly revealing. The SLLN (9) implies that there is a systematic tendency for x_i to decrease until it reaches C_0. To formalize this, consider the following event:

$$\mathcal{E} = \left\{ \frac{1}{k} \sum_{i=1}^{k} \ln(\lambda_i) < -\varepsilon \text{ for all } k \geq 1 \right\}, \text{ for some } \varepsilon > 0,$$

which represents the scenario where the sample average of all individuals' log signals is consistently negative. We show that on this event, x_i almost surely enters C_0. The intuition is that under $\mathcal{E}$, the public information consistently favors state 0. This prevents x_i from increasing sufficiently to reach C_1, and since x_i cannot remain in the non-cascade set forever (Step 1), it must eventually enter C_0. (iii) Using the theory of large deviations, we establish that $\mathbb{P}(\mathcal{E}) > 0$, which ensures that x_i enters C_0 with a positive probability. Similar arguments can show that for any i and conditional on any history h_i, the probability that x_j enters C_0 for some $j \geq i$ is strictly positive and must be greater than $\mathbb{P}(\mathcal{E})$. This fact implies that x_i enters C_0 infinitely often by standard Borel-Cantelli arguments.

STEP 4: x_i stays in the ε-neighborhood of C_0 with positive probability.

Once x_i enters C_0, we demonstrate that it remains in an ε-neighborhood of C_0 with positive probability. However, as x_i approaches the cascade set, its dynamics become increasingly complex due to the intermittent cessation and resumption of information aggregation. The proof hinges on the observation that for x_i to escape the ε-neighborhood of C_0, it must experience an upcrossing from $1/\gamma$ to $1/\gamma + \varepsilon$ without returning to C_0. This process is depicted in Figure 7 in the Appendix. The advantage of focusing on upcrossings is that during an upcrossing, x_i remains in the non-cascade set, allowing us to eliminate the indicator function as in Step 1.

Our goal is to show that such upcrossings cannot occur almost surely when i is sufficiently large. First, we show that as i increases, the length of an upcrossing that concludes at individual i grows polynomially with i and has the same order as $1/\alpha_i$. In other words, the number of individuals involved in such upcrossing has the same order as i^ρ. This fact allows us to estimate the probability of each upcrossing:

$$\mathbb{P}\left(\text{An upcrossing finishes at } i+k | x_i \in C_0\right) \leq (i+k)\exp\left(-D(i+k)^\rho\right), \tag{10}$$

for some constant $D > 0$. Denote by $\mathcal{G}_i$ as the event that an upcrossing occurs after i. Equation (10) implies:

$$\mathbb{P}\left(\mathcal{G}_i | x_i \in C_0\right) \leq \sum_k (i+k)\exp\left(-D(i+k)^\rho\right) \equiv \mathcal{Q}(i), \tag{11}$$

The right-hand side of (11) is a convergent series, and we have $\mathcal{Q}(i) \to 0$ as $i \to +\infty$. As a consequence, when i is large enough, it happens with a strictly positive probability that x_i never escapes the ε-neighborhood of C_0, as illustrated in Figure 5.

STEP 5: x_i converges to $1/\gamma$ almost surely.

Lastly, we show that x_i must converge to a limit within the correct cascade set, specifically to the boundary point, $1/\gamma$. This result comes from two facts: (i) From Step 3, x_i enters C_0 infinitely often. Moreover, we can demonstrate that for sufficiently large i, the value of x_i at its first entry into C_0 is very close $1/\gamma$, which implies $x_i \geq 1/\gamma + \varepsilon$ when i; (ii) From Step 4, x_i remains below $\frac{1}{\gamma} + \varepsilon$ for all large i with a strictly positive probability. These two facts jointly imply that $x_i \in [1/\gamma - \varepsilon, 1/\gamma + \varepsilon]$ for sufficiently large i almost surely. Since this statement holds for any $\varepsilon > 0$, it follows that $x_i \to 1/\gamma$ almost surely.

7 Discussion

In this section, we discuss some important implications of our results, including learning completeness, information aggregation, and limit actions. Then, we talk about potential extensions of our model.

7.1 Learning incompleteness

Our findings demonstrate that when $0 < \rho < 1$, the society can almost surely have a correct limit cascade. However, this outcome is ***not*** the complete learning result in Smith and Sørensen (2000) and Goeree et al. (2006), where individuals' subjective posteriors converge to assigning probability 1 on the correct state. In our case, $x_i \to 1/\gamma$, which corresponds to the subjective belief:

$$\text{As } i \to +\infty : \quad \pi_i\left(0|\lambda_i, h_i\right) \to \pi_\infty\left(0|\lambda_i\right) = \frac{\gamma}{\gamma + \lambda_i} \quad \mathbb{P} - \text{a.s..}$$

Notably, the limit belief $\pi_\infty\left(0|\lambda_i\right) \geq 1/2$ for all λ_i between $1/\gamma$ and γ. This indicates that, in the limit, individuals believe the true state is the most likely state for almost all signals. However, individuals never achieve full confidence in the true state: their subjective beliefs remain strictly less than 1. Moreover, if tail individuals could have received a signal stronger than γ, they could switch to the incorrect state.

7.2 Information aggregation

Although learning is incomplete, information accumulation in social learning continues indefinitely. Due to the infinite cycles of exiting and re-entering the cascade set, society reveals an unbounded amount of information over time. More precisely, consider a ***Bayesian outside observer***. For such an observer, the belief about the true state satisfies:

$$\text{As } i \to +\infty : \quad \pi^B\left(0|h_i\right) \to 1 \quad \mathbb{P} - \text{a.s..}$$

In other words, a Bayesian outside learner can completely learn the true state based on the public information. The co-existence of incomplete learning and continuous information aggregation is another interesting feature of our model. This feature is also present in sequential social learning models with continuous signals (Smith and Sørensen (2000) and Herrera and Hörner (2012)). In those models, learning remains incomplete because the speed of information aggregation decreases over time. As a result, even though information continues to accumulate, the informativeness of society becomes bounded in the limit. In contrast, in our model, this phenomenon arises because the weight on public information decays to zero, yet the speed of information accumulation remains steady, allowing society to achieve unbounded informativeness in the Bayesian sense asymptotically.

7.3 Limit actions

Note that we only have limit cascades instead of persistent informational cascades, so actions do not converge, and there is no herding behavior in the limit. To see it, as shown in Theorem 1, the power public likelihood ratio x_i converges to $1/\gamma$—the boundary of the correct cascade set—and oscillates around it indefinitely. As such, there are infinitely periods where x_i is in the non-cascade set. During those periods, individuals may act based on their own signals. In fact, they will choose action 1 if and only if they receive the strongest signal, γ. Since signals are discrete, individuals in the non-cascade set take action 1 with probability $\delta := \mathbb{P}(\lambda_i = \gamma | 0) > 0$. This implies that incorrect actions are taken infinitely many times. However, the total fraction of incorrect actions in the population must be bounded by $\delta < 1/2$, so the majority of actions taken by the population are always correct. This result has important implications in situations like sequential voting, where standard models often predict that the voting outcomes are path-dependent and heavily influenced by early voters (Ali and Kartik (2008), Knight and Schiff (2010)). In contrast, our model suggests that the majority always takes the correct action except for null events. Furthermore, as the true data-generating process has an increasingly thin tail $\delta \to 0$, the fraction of incorrect actions also approaches 0, and nearly all individuals act correctly. A natural conjecture is that for continuous signals with adequately thin tails, the fraction of incorrect actions might equal zero. However, analyzing continuous signals requires techniques beyond the scope of this paper and may be an interesting direction for future research.

7.4 Extensions

This paper provides a comprehensive characterization for monotonic α_i. A natural extension is to explore the implications of non-monotonic α_i. An extreme case is that there are infinitely many non-social learners with $\alpha_i = 0$, while the remaining individuals are Bayesian social learners with $\alpha_i = 1$. In this case, non-social learners' actions are independent and informative, so complete learning can be achieved for Bayesian social learners in the limit. Another interesting direction is to consider a more general framework where individuals assign weights to public information based not only on their own position in the decision sequence but also on the positions of their predecessors. This approach captures scenarios where individuals place greater weight on the actions of predecessors who are temporally or socially closer to them. Additional extensions could involve stochastic Bayesian weights, where weights vary randomly, or endogenous Bayesian weights, where individuals strategically decide how much weight to assign to each predecessor.

8 Conclusion

Understanding the ways in which people integrate information from external sources is crucial for analyzing social learning outcomes. This paper introduces a power-Bayesian social learning model, which allows individuals to assign different weights to public information based on their position in the decision sequence.

We show that, even with bounded private signals, correct informational cascades can emerge almost surely if public information is gradually discarded at a controlled pace. However, completely discarding public history too quickly or failing to fully ignore public information in the limit can be detrimental. This paper simplifies the analysis by restricting attention to finite signals and a common weight for every predecessor, so extending it to a general model is a direction for future research.

Bibliography

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. 3

Ali, S Nageeb and Navin Kartik (2008) "Social learning in elections," *University of California, San Diego. Working Paper*. 21

Anderson, Lisa R and Charles A Holt (1997) "Information cascades in the laboratory," *American Economic Review*, 847–862. 5

Arieli, Itai, Yakov Babichenko, and Manuel Mueller-Frank (2021) "Sequential naive learning," *arXiv preprint arXiv:2101.02897*. 5

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). 3, 4

Bala, Venkatesh and Sanjeev Goyal (1998) "Learning from neighbours," *The review of economic studies*, 65 (3), 595–621. 5

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

Benjamin, Daniel J (2019) "Errors in probabilistic reasoning and judgment biases," *Handbook of Behavioral Economics: Applications and Foundations 1*, 2, 69–186. 5

Bernardo, Antonio E and Ivo Welch (2001) "On the evolution of overconfidence and entrepreneurs," *Journal of Economics & Management Strategy*, 10 (3), 301–330. 4

Bikhchandani, Sushil, David Hirshleifer, Omer Tamuz, and Ivo Welch (2024) "Information cascades and social learning," *Journal of Economic Literature*, 62 (3), 1040–1093. 4, 15

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, 3

Bohren, J Aislinn (2016) "Informational herding with model misspecification," *Journal of Economic Theory*, 163, 222–247. 3

Çelen, Boğaçhan and Shachar Kariv (2004) "Observational learning under imperfect information," *Games and Economic behavior*, 47 (1), 72–86. 3

Chen, Jaden Yang (2019) "Sequential learning under informational ambiguity," *Available at SSRN 3480231*. 5

Conlon, John J, Malavika Mani, Gautam Rao, Matthew W Ridley, and Frank Schilbach (2022) "Not learning from others,"Technical report, National Bureau of Economic Research. 5

Dasaratha, Krishna and Kevin He (2020) "Network structure and naive sequential learning," *Theoretical Economics*, 15 (2), 415–444. 5

DeGroot, Morris H (1974) "Reaching a consensus," *Journal of the American Statistical association*, 69 (345), 118–121. 5

DeMarzo, Peter M, Dimitri Vayanos, and Jeffrey Zwiebel (2003) "Persuasion bias, social influence, and unidimensional opinions," *Quarterly Journal of Economics*, 118 (3), 909–968. 5

Duffy, John, Ed Hopkins, and Tatiana Kornienko (2021) "Lone wolf or herd animal? Information choice and learning from others," *European Economic Review*, 134, 103690. 5

Ellison, Glenn and Drew Fudenberg (1993) "Rules of thumb for social learning," *Journal of political Economy*, 101 (4), 612–643. 5

——— (1995) "Word-of-mouth communication and social learning," *The Quarterly Journal of Economics*, 110 (1), 93–125. 5

Epstein, Larry G, Jawwad Noor, and Alvaro Sandroni (2010) "Non-bayesian learning," *The BE Journal of Theoretical Economics*, 10 (1), 0000102202193517041623. 5

Eyster, Erik and Matthew Rabin (2010) "Naive herding in rich-information settings," *American Economic Journal: Microeconomics*, 2 (4), 221–43. 5

——— (2014) "Extensive imitation is irrational and harmful," *The Quarterly Journal of Economics*, 129 (4), 1861–1898. 5

Frick, Mira, Ryota Iijima, and Yuhta Ishii (2024) "Welfare comparisons for biased learning," *American Economic Review*, 114 (6), 1612–1649. 5

Goeree, Jacob K, Thomas R Palfrey, and Brian W Rogers (2006) "Social learning with private and common values," *Economic theory*, 28, 245–264. 3, 4, 20

Golub, Benjamin and Matthew O Jackson (2010) "Naive learning in social networks and the wisdom of crowds," *American Economic Journal: Microeconomics*, 2 (1), 112–49. 5

Grether, David M (1980) "Bayes rule as a descriptive model: The representativeness heuristic," *The Quaterly Journal of Economics*, 95 (3), 537–557. 2, 5

——— (1992) "Testing Bayes rule and the representativeness heuristic: Some experimental evidence," *Journal of Economic Behavior & Organization*, 17 (1), 31–57. 2, 5

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

Herrera, Helios and Johannes Hörner (2012) "A necessary and sufficient condition for information cascades." 21

Jadbabaie, Ali, Pooya Molavi, Alvaro Sandroni, and Alireza Tahbaz-Salehi (2012) "Non-Bayesian social learning," *Games and Economic Behavior*, 76 (1), 210–225. 5

Knight, Brian and Nathan Schiff (2010) "Momentum and social learning in presidential primaries," *Journal of political economy*, 118 (6), 1110–1150. 21

Molavi, Pooya, Alireza Tahbaz-Salehi, and Ali Jadbabaie (2018) "A theory of non-Bayesian social learning," *Econometrica*, 86 (2), 445–490. 5

Morin, Olivier, Pierre Olivier Jacquet, Krist Vaesen, and Alberto Acerbi (2021) "Social information use and social information waste," *Philosophical Transactions of the Royal Society B*, 376 (1828), 20200052. 5

Rosenberg, Dinah and Nicolas Vieille (2019) "On the efficiency of social learning," *Econometrica*, 87 (6), 2141–2168. 4

Smith, Lones and Peter Sørensen (2000) "Pathological outcomes of observational learning," *Econometrica*, 68 (2), 371–398. 4, 20, 21, 34

Weizsäcker, Georg (2010) "Do we follow others when we should? A simple test of rational expectations," *American Economic Review*, 100 (5), 2340–2360. 5

A Proofs

A.1 Proof of Theorem 1

To extend the intuition in Section 6 from binary-signal case to multi-signal case, we divide the interval $(1/\gamma, \gamma)$ into $n-1$ disjoint zones, with each divider being a signal:[16]

$$Z_1 = \left(1/\gamma, \lambda^2\right), Z_2 = \left(\lambda^2, \lambda^3\right), ..., Z_{n-2} = \left(\lambda^{n-2}, \lambda^{n-1}\right), Z_{n-1} = \left(\lambda^{n-1}, \gamma\right).$$

Proposition 1 implies that individual i chooses action 1 when her private signal is above $1/x_i$ and action 0 otherwise. Suppose $1/x_i \in Z_t = (\lambda^t, \lambda^{t+1})$ for some $t \in \{1, 2, \ldots, n-1\}$, then individual i will choose action 1 if her private signal $\lambda_i \in S_1^t \equiv \{\lambda^{t+1}, \ldots, \gamma\}$, and she will choose action 0 if $\lambda_i \in S_0^t \equiv \left\{\frac{1}{\gamma}, \ldots, \lambda^t\right\}$. The proof for Theorem 1 proceeds as follows:

Step 1: $\mathbb{P}\left(x_i \leq \frac{1}{\gamma} \text{ happens infinitely often}\right) = 1$.

The proof idea is to show that $1/x_i$ can transit in the following trajectory:

$$C_0 \to Z_1 \to Z_2 \to ... \to Z_{n-1} \to C_1,$$

with every step happening with some probability bounded away from 0, as illustrated in Figure 6.[17] We first show that $x_i < \gamma$ happens infinitely often (i.e., $C_0 \to Z_1$ always happens). Note that when $x_i \geq \gamma$, individual i chooses action 1 regardless of the signal she receives, and hence $\frac{\phi^1(a_i, x_i)}{\phi^0(a_i, x_i)} = 1$. However, since the Bayesian weight is decreasing, x_i will keep decreasing until it eventually exits C_1, so $x_i < \gamma$ must happen infinitely often. In the following, we focus on the case where $x_i < \gamma$, i.e., $1/x_i > 1/\gamma$. Next, suppose $1/x_i \in Z_1 = (1/\gamma, \lambda^2)$.[18] We then show $1/x_i$ will transit from $Z_1 \to Z_2$ and then eventually to C_1 with positive probability.

To facilitate discussion, we introduce some related random varaibles and events: For all i and all $t \in \{1, 2, \ldots, n-1\}$, we define a sequence of random variables $\left\{X_i^t, X_{i+1}^t, ...\right\}$, where

$$X_{i+k}^t = \begin{cases} \ln\left(\frac{\mathbb{P}\left(S_1^t|\theta=1\right)}{\mathbb{P}\left(S_1^t|\theta=0\right)}\right) & \lambda_{i+k} \in S_1^t \\ \ln\left(\frac{\mathbb{P}\left(S_0^t|\theta=1\right)}{\mathbb{P}\left(S_0^t|\theta=0\right)}\right) & \lambda_{i+k} \in S_0^t \end{cases}.$$

[16] Dividers themselves are excluded due to triviality, and they can easily be nested under pre-specified tie-breaking rules.

[17] For convenience, the proof focuses on the trajectory of $1/x_i$ in the proof. Figure 6 depicts the trajectory of x_i. Note that $1/x_i \in C_\theta$ means $x_i \in C_{1-\theta}$, so equivalently, we want to show that x_i can transit from C_1 to C_0 with each step happening with a strictly positive probability.

[18] Here, we first consider the case where $\lambda^2 < 1$ to accommodate the multiple signal structure better.

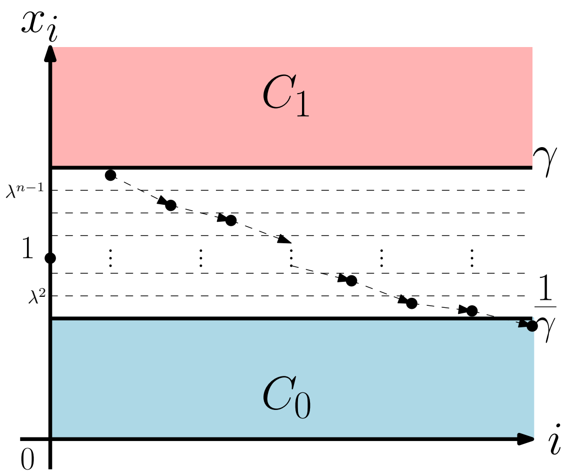


Figure 6: General signal space

Intuitively, X_{i+k}^t describes the increment of the log public LLR for individual $i+k$ when $1/x_{i+k} \in Z_t$. We define event:

$$E_i^t = \left\{ \frac{1}{k+1} \sum_{j=0}^{k} X_{i+k}^t < -c, \text{ for all } k = 0, 1, ... \right\},$$

where $c > 0$ is an adequately small constant. We have the following lemma:

Lemma 1. *For all i and $t \in \{1, 2, \dots, n-1\}$, there exists $\delta > 0$ such that $\mathbb{P}(E_i^t) > \delta$.*

Proof. For all t, i, $\{X_{i+k}^t\}$ is a sequence of i.i.d. random variables. Besides,

$$\mathbb{E}\left(X_{i+k}^t\right) = \mathbb{P}\left(S_1^t|\theta = 0\right) \ln\left(\frac{\mathbb{P}\left(S_1^t|\theta = 1\right)}{\mathbb{P}\left(S_1^t|\theta = 0\right)}\right) + \mathbb{P}\left(S_0^t|\theta = 0\right) \ln\left(\frac{\mathbb{P}\left(S_1^t|\theta = 1\right)}{\mathbb{P}\left(S_1^t|\theta = 0\right)}\right) < 0,$$

where the inequality comes from Jensen's inequality. Lemma 8 in Appendix B implies that for all $\varepsilon > 0$, we have:

$$\mathbb{P}\left(\frac{1}{k+1} \sum_{j=0}^{k} X_{i+k}^t \leq \mathbb{E}\left(X_{i+k}^t\right) + \varepsilon, \forall k \geq 0\right) \equiv \delta_t > 0,$$

and δ_t is independent of i since all individuals' signals are i.i.d. We can choose ε such that $\mathbb{E}\left(X_{i+k}^t\right) + \varepsilon < 0$ and let $c = \min\left(-\mathbb{E}\left(X_{i+k}^t\right) - \varepsilon\right)$ and $\delta = \min_t \delta_t$, then we have $\mathbb{P}\left(E_i^t\right) > \delta$ for all t, i. □

We define a stopping time $T_t(i) = \inf\{k : 1/x_{i+k} \geq \lambda^t\}$. We have the following lemma:

Lemma 2. *For all i, given $1/x_i \in Z_1$, we have $\{T_2(i) < \infty\}$ on E_i^1.*

Proof. Suppose $1/x_i \in Z_1$. Then on E_i^1, we have:

$$\ln\left(\frac{1}{x_{i+1}}\right) = \frac{\alpha_{i+1}}{\alpha_i} \times \ln\left(\frac{1}{x_i}\right) - \alpha_{i+1} \times \ln\left(\frac{\phi^1(a_i, x_i)}{\phi^0(a_i, x_i)}\right)$$
$$> \ln\left(\frac{1}{\gamma}\right) - \alpha_{i+1} \times \ln\left(\frac{\phi^1(a_i, x_i)}{\phi^0(a_i, x_i)}\right) \geq \ln\left(\frac{1}{\gamma}\right) + \alpha_{i+1} \times c > \ln\left(\frac{1}{\gamma}\right), \tag{12}$$

where: (i) the first inequality comes from $1/x_i \in Z_1$, $\ln\left(\frac{1}{x_i}\right) < 0$ and $\alpha_{i+1} < \alpha_i$, and (ii) the second inequality comes from the definition of E_i^1 and $1/x_i \in Z_1$, which implies

$$\ln\left(\frac{\phi^1(a_i, x_i)}{\phi^0(a_i, x_i)}\right) = \ln\left(\frac{\mathbb{P}\left(S_{a_i}^1 | \theta = 1\right)}{\mathbb{P}\left(S_{a_i}^1 | \theta = 0\right)}\right) = X_i^1 < -c.$$

Thus, we either have $\frac{1}{x_{i+1}} \geq \lambda^2$, or $\frac{1}{x_{i+1}} \in Z_1$. In the former case, we have $T_2(i) = 1 < \infty$, so the claim is true. Suppose that the latter happens, then

$$\ln\left(\frac{1}{x_{i+2}}\right) = \frac{\alpha_{i+2}}{\alpha_i} \times \ln\left(\frac{1}{x_i}\right) - 2\alpha_{i+2} \times \frac{1}{2}\sum_{j=i}^{i+1} \ln\left(\frac{\phi^1(a_j, x_j)}{\phi^0(a_j, x_j)}\right) > \ln\left(\frac{1}{\gamma}\right),$$

where the inequality comes from $\frac{1}{2}\sum_{j=i}^{i+1} \ln\left(\frac{\phi^1(a_j,x_j)}{\phi^0(a_j,x_j)}\right) = \frac{1}{2}\sum_{k=0}^{1} X_{i+k}^t < -c$. Similarly, we either have $\frac{1}{x_{i+2}} \geq \lambda^2$, or $\frac{1}{x_{i+2}} \in Z_1$. In the former case, we have $T_2(i) = 2 < \infty$; in the latter case, we repeat the same argument. By induction, suppose $T_2(i) > k-1$, i.e., $\frac{1}{x_i}, \frac{1}{x_{i+1}}, ..., \frac{1}{x_{i+k-1}} \in Z_1$, then we have:

$$\ln\left(\frac{1}{x_{i+k}}\right) = \frac{\alpha_{i+k}}{\alpha_i} \times \ln\left(\frac{1}{x_i}\right) - k\alpha_{i+k} \times \frac{1}{k}\sum_{j=i}^{i+k-1} \ln\left(\frac{\phi^1(a_j, x_j)}{\phi^0(a_j, x_j)}\right) > \ln\left(\frac{1}{\gamma}\right).$$

However, it is impossible for $\frac{1}{x_{i+k}} \in Z_1$ for all $k \geq 0$. This is because as $k \to +\infty$, we almost surely have: (i) $k\alpha_{i+k} \approx k/(i+k)^\rho \to +\infty$, and (ii)

$$\frac{1}{k}\sum_{j=i}^{i+k-1} \ln\left(\frac{\phi^1(a_j, x_j)}{\phi^0(a_j, x_j)}\right) \to \mathbb{E}\left(X_{i+k}^1\right) < 0.$$

Thus, $\ln\left(\frac{1}{x_{i+k}}\right) \to +\infty$ except for null events, suggesting that $\frac{1}{x_{i+k}}$ must escape from Z_1. As a consequence, we must almost surely have $\{T_2(i) < \infty\}$ on E_i^1. □

Lemma 3. *For all $t \in \{1, 2, ..., n-1\}$ and all i su ciently large, given $1/x_i \in Z_t$, we have $\{T_{t+1}(i) < \infty\}$ on E_i^t .*

Proof. The argument in Lemma 2 can be directly applied to any case where $1/x_i \in Z_t = (\lambda^t, \lambda^{t+1})$, where $\lambda^{t+1} < 1$, i.e. $1/x_i$ is in the lower half of the interval $(1/\gamma, \gamma)$. The situation becomes a bit different for the rest zones, but the idea does not change.

Suppose $1/x_i \in Z_q = (\lambda^q, \lambda^{q+1})$ where $\lambda^{q+1} > 1$. Similarly, we have

$$\ln\left(\frac{1}{x_{i+1}}\right) = \frac{\alpha_{i+1}}{\alpha_i} \times \ln\left(\frac{1}{x_i}\right) - \alpha_{i+1} \times \ln\left(\frac{\phi^1(a_i, x_i)}{\phi^0(a_i, x_i)}\right) > \ln(\lambda^q) \tag{13}$$

if the following holds:[19]

$$\ln\left(\frac{\phi^1(a_i, x_i)}{\phi^0(a_i, x_i)}\right) \leq \left(\frac{1}{\alpha_i} - \frac{1}{\alpha_{i+1}}\right) \ln\left(\frac{1}{x_i}\right).$$

Notice that when i is sufficiently large, we have

$$\left(\frac{1}{\alpha_i} - \frac{1}{\alpha_{i+1}}\right) \ln\left(\frac{1}{x_i}\right) \approx \left(\frac{1}{Ci^\rho} - \frac{1}{C(i+1)^\rho}\right) \ln\left(\frac{1}{x_i}\right) \approx 0.$$

On E_i^q, we have $\ln\left(\frac{\phi^1(a_i,x_i)}{\phi^0(a_i,x_i)}\right) < -c$, which implies that $\ln\left(\frac{1}{x_{i+1}}\right) > \ln(\lambda^q)$ when i is sufficiently large. By induction, suppose $T_{q+1}(i) > k-1$, i.e., $\frac{1}{x_i}, \frac{1}{x_{i+1}}, ..., \frac{1}{x_{i+k-1}} \in Z_q$, then we have:

$$\ln\left(\frac{1}{x_{i+k}}\right) = \frac{\alpha_{i+k}}{\alpha_i} \times \ln\left(\frac{1}{x_i}\right) - k\alpha_{i+k} \times \frac{1}{k}\sum_{j=i}^{i+k-1} \ln\left(\frac{\phi^1(a_j, x_j)}{\phi^0(a_j, x_j)}\right) > \ln(\lambda^q)$$

if

$$\frac{1}{k}\sum_{j=i}^{i+k-1} \ln\left(\frac{\phi^1(a_j, x_j)}{\phi^0(a_j, x_j)}\right) \leq \frac{1}{k}\left(\frac{1}{\alpha_i} - \frac{1}{\alpha_{i+k}}\right) \ln\left(\frac{1}{x_i}\right)$$
$$\approx \left(\frac{1}{Ci^\rho} - \frac{1}{C(i+k)^\rho}\right) \ln\left(\frac{1}{x_i}\right) \approx 0 \text{ when } i \text{ is sufficiently large.}$$

Applying identical arguments in the proof of Lemma 2's proof, we can then show that $\{T_{q+1}(i) < \infty\}$ on E_i^q when i is large.

There is a remaining case: $\frac{1}{x_i} \in (\lambda^o, \lambda^{o+1})$ where $\lambda^o < 1$ and $\lambda^{o+1} > 1$. This case can be incorporated into previous two cases by applying the proof to regions $(\lambda^o, 1)$ and $(1, \lambda^{o+1})$ separately. □

[19]Note that when $1/x_i \in Z_q = (\lambda^q, \lambda^{q+1})$ and $\lambda^{q+1} > 1$, we have $\ln\left(\frac{1}{x_i}\right) > 0$, so we cannot use $\frac{\alpha_{i+1}}{\alpha_i} \times \ln\left(\frac{1}{x_i}\right) > \ln\left(\frac{1}{x_i}\right)$ as in (12).

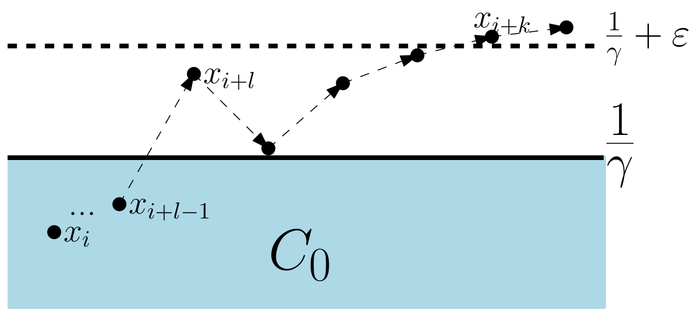


Figure 7: Illustration of an upcrossing

Now, we can finish the proof of Step 1. From previous lemmas, there exists $\varepsilon > 0$ such that when i is sufficiently large,

$$\mathbb{P}\left(\exists i' \geq i \text{ s.t. } x_{i'} \leq \frac{1}{\gamma} | x_i < \gamma\right) > \varepsilon.$$

Further, since $x_i < \gamma$ happens infinitely often, we have

$$\sum_i \mathbb{P}\left(\exists i' \geq i \text{ s.t. } x_{i'} \leq \frac{1}{\gamma} | \mathcal{F}_i\right) = +\infty,$$

where $\mathcal{F}_i$ represents any information set at i. By Levy's extension of Borel-Cantelli Lemma, we have $x_i \leq \frac{1}{\gamma}$ happens infinitely often with probability 1.

Step 2: For large i, $\mathbb{P}\left(x_{i+k} < \frac{1}{\gamma} + \epsilon \text{ for all } k \geq 0 | x_i \leq \frac{1}{\gamma}\right) > 0$, $\forall \epsilon > 0$.

Next, x_i will remain in the ϵ-neighborhood of C_0. To show it, we notice that to escape from the ϵ-neighborhood, x_i must experience some ***upcrossing*** from $\frac{1}{\gamma}$ to $\frac{1}{\gamma} + \epsilon$ during which it never falls back into C_0. In other words, there exist some uninterrupted periods in which x_i first jumps above $\frac{1}{\gamma}$, and keeps moving outside of C_0 (but still within the ϵ-neighborhood of $\frac{1}{\gamma}$) until it touches upon $\frac{1}{\gamma} + \epsilon$.

Here, we focus on the first upcrossing following x_i, and we use x_{i+l} $(l > 0)$ to denote the starting point, and x_{i+k} $(k > l)$ for the endpoint.[20] An upcrossing that finishes at $i + k$ is called ***an*** $i+k$***-upcrossing***, and $k - l$ is called the ***length*** of that upcrossing. The concepts of upcrossing and its length are illustrated in Figure 7. The following lemma provides an estimation of the length:

Lemma 4. $0 < \liminf\,(k - l)\,\alpha_{i+k} \leq \limsup\,(k - l)\,\alpha_{i+k} < +\infty$ *except for null events.*

[20] When i is sufficiently large, given $x_i \leq \frac{1}{\gamma}$, if $x_{i+1} > \frac{1}{\gamma}$, then $x_{i+1} - \frac{1}{\gamma}$ is extremely small (no larger than $\frac{\epsilon}{2}$). Therefore, each upcrossing consists of at least one movement of the threshold.

Proof. (1) We first show $\limsup_{i\to+\infty}(k-l)\,\alpha_{i+k}<+\infty$ almost surely. Notice that

$$\begin{aligned}\ln(x_{i+k}) &= \frac{\alpha_{i+k}}{\alpha_{i+l}}\ln(x_{i+l})+\alpha_{i+k}\sum_{j=i+l}^{i+k-1}\ln\left(\frac{\phi^1(a_j,x_j)}{\phi^0(a_j,x_j)}\right)\\ &= \frac{\alpha_{i+k}}{\alpha_{i+l}}\ln(x_{i+l})+(k-l)\,\alpha_{i+k}\times\frac{1}{k-l}\sum_{j=i+l}^{i+k-1}\ln\left(\frac{\phi^1(a_j,x_j)}{\phi^0(a_j,x_j)}\right).\end{aligned}$$

Suppose $\limsup_{i\to+\infty}(k-l)\,\alpha_{i+k}=+\infty$, then we almost surely have:

$$\lim_{i\to+\infty}\inf\left[(k-l)\,\alpha_{i+k}\times\frac{1}{k-l}\sum_{j=i+l}^{i+k-1}\ln\left(\frac{\phi^1(a_j,x_j)}{\phi^0(a_j,x_j)}\right)\right]=-\infty,$$

so x_i will be driven downward and return back to C_0, contradicting the definition of an upcrossing. Therefore, we must have $\limsup(k-l)\,\alpha_{i+k}<+\infty$.

(2) Next, we show $0<\liminf(k-l)\,\alpha_{i+k}$ almost surely. We first note that when i is large enough,

$$\begin{aligned}\ln(x_{i+k}) &= \frac{\alpha_{i+k}}{\alpha_{i+l}}\ln(x_{i+l})+\alpha_{i+k}\sum_{j=i+l}^{i+k-1}\ln\left(\frac{\phi^1(a_j,x_j)}{\phi^0(a_j,x_j)}\right)\\ &\le \frac{\alpha_{i+k}}{\alpha_{i+l}}\ln\left(\frac{1}{\gamma}+\frac{\epsilon}{3}\right)+\alpha_{i+k}\sum_{j=i+l}^{i+k-1}\ln\left(\frac{\phi^1(a_j,x_j)}{\phi^0(a_j,x_j)}\right)\\ &\le \left(\frac{i+l}{i+k}\right)\ln\left(\frac{1}{\gamma}+\frac{\epsilon}{3}\right)+\alpha_{i+k}\sum_{j=i+l}^{i+k-1}\ln\left(\frac{\phi^1(a_j,x_j)}{\phi^0(a_j,x_j)}\right)\\ &\le \ln\left(\frac{1}{\gamma}+\frac{\epsilon}{2}\right)+\alpha_{i+k}\sum_{j=i+l}^{i+k-1}\ln\left(\frac{\phi^1(a_j,x_j)}{\phi^0(a_j,x_j)}\right),\end{aligned}$$

where: (i) the first inequality comes from the fact that when i is sufficiently large, $x_{i+l}-\frac{1}{\gamma}$ is extremely small. This is because when $x_{i+l-1}\le\frac{1}{\gamma}$ and $x_{i+l}>\frac{1}{\gamma}$,

$$\ln(x_{i+l})=\frac{\alpha_{i+l}}{\alpha_{i+l-1}}\ln(x_{i+l-1})+\alpha_{i+l}\ln\left(\frac{\phi^1(a_{i+l-1},x_{i+l-1})}{\phi^0(a_{i+l-1},x_{i+l-1})}\right)=\frac{\alpha_{i+l}}{\alpha_{i+l-1}}\ln(x_{i+l-1}).$$

Then, $\lim\limits_{i\to+\infty}\frac{\ln(x_{i+l})}{\ln(x_{i+l-1})}=\lim\limits_{i\to+\infty}\frac{\alpha_{i+l}}{\alpha_{i+l-1}}=1$, implying that

$$\lim_{i\to+\infty}\ln(x_{i+l})=\ln\left(\frac{1}{\gamma}\right),$$

and (ii) the second inequality is based on a fact that

$$\frac{\alpha_{i+k}}{\alpha_{i+l}} \approx \left(\frac{i+l}{i+k}\right)^{\rho} > \frac{i+l}{i+k},$$

and $\ln\left(\frac{1}{\gamma}+\epsilon\right) < 0$, and (iii) the third inequality comes from $\limsup\left(k-l\right)\alpha_{i+k} < +\infty$, which implies that

$$\liminf \frac{i+l}{i+k} = 1 - \limsup \frac{k-l}{i+k} = 1 - \limsup \frac{k-l}{(i+k)^{\rho}} \frac{(i+k)^{\rho}}{i+k} = 1.$$

We let $\eta \equiv \ln\left(\frac{\frac{1}{\gamma}+\epsilon}{\frac{1}{\gamma}+\frac{\epsilon}{2}}\right) > 0$. Then, we have:

$$\ln\left(\frac{1}{\gamma}+\epsilon\right) - \ln\left(x_{i+k}\right) \geq \eta - \alpha_{i+k} \sum_{j=i+l}^{i+k-1} \ln\left(\frac{\phi^{1}\left(a_j, x_j\right)}{\phi^{0}\left(a_j, x_j\right)}\right).$$

Since $x_{i+k} \geq \frac{1}{\gamma}+\epsilon$, we have:

$$\alpha_{i+k} \sum_{j=i+l}^{i+k-1} \ln\left(\frac{\phi^{1}\left(a_j, x_j\right)}{\phi^{0}\left(a_j, x_j\right)}\right) \geq \eta,$$

which implies that after an upcrossing, the increment of public likelihood ratio (with weight) must be greater than some fixed constant. The intuition is that to finish an upcrossing, the power public LLR must go through some fixed distance.

For $i+l \leq j \leq i+k-1$, $\frac{1}{x_j} \in Z_{n-1} = \left(\lambda^{n-1}, \gamma\right)$. Thus, $S_0^{n-1} = \left\{\frac{1}{\gamma}, \ldots, \lambda^{n-1}\right\}$, and $S_1^{n-1} = \{\gamma\}$, and

$$\ln\left(\frac{\phi^{1}\left(a_j, x_j\right)}{\phi^{0}\left(a_j, x_j\right)}\right) = \begin{cases} \ln\left(\frac{\mathbb{P}\left(S_0^{n-1}|\theta=1\right)}{\mathbb{P}\left(S_0^{n-1}|\theta=0\right)}\right) & \text{if } \lambda_j \in S_0^{n-1} \\ \ln\left(\frac{\mathbb{P}\left(S_1^{n-1}|\theta=1\right)}{\mathbb{P}\left(S_1^{n-1}|\theta=0\right)}\right) & \text{if } \lambda_j \in S_1^{n-1} \end{cases}.$$

Let $\xi = \max\left\{\ln\left(\frac{\mathbb{P}\left(S_0^{n-1}|\theta=1\right)}{\mathbb{P}\left(S_0^{n-1}|\theta=0\right)}\right), \ln\left(\frac{\mathbb{P}\left(S_1^{n-1}|\theta=1\right)}{\mathbb{P}\left(S_1^{n-1}|\theta=0\right)}\right)\right\}$, then we have:

$$\left(k-l\right)\xi \geq \sum_{j=i+l}^{i+k-1} \ln\left(\frac{\phi^{1}\left(a_j, x_j\right)}{\phi^{0}\left(a_j, x_j\right)}\right) \geq \frac{\eta}{\alpha_{i+k}} \Rightarrow \left(k-l\right)\alpha_{i+k} \geq \eta/\xi > 0,$$

which establishes $\liminf\left(k-l\right)\alpha_{i+k} > 0$. □

We then define event $E_{i+k}^{l} = \{\text{an } i+k\text{-upcrossing starts at } i+l\}$. Then, when i is sufficiently large, we have

$$\begin{aligned}
\mathbb{P}\left(E_{i+k}^{l}|x_i \leq \frac{1}{\gamma}\right) &\leq \mathbb{P}\left(\sum_{j=i+l}^{i+k-1} \ln\left(\frac{\phi^1(a_j,x_j)}{\phi^0(a_j,x_j)}\right) \geq \frac{\eta}{\alpha_{i+k}}|x_i \leq \frac{1}{\gamma}\right) \\
&= \mathbb{P}\left(\frac{1}{k-l}\sum_{j=i+l}^{i+k-1} \ln\left(\frac{\phi^1(a_j,x_j)}{\phi^0(a_j,x_j)}\right) \geq \frac{\eta}{(k-l)\,\alpha_{i+k}}\right) \\
&\leq \mathbb{P}\left(\frac{1}{k-l}\sum_{j=i+l}^{i+k-1} \ln\left(\frac{\phi^1(a_j,x_j)}{\phi^0(a_j,x_j)}\right) \geq 0\right) \\
&\approx \exp\left(-C(k-l)\right) \text{ for some constant } C>0 \\
&\leq \exp\left(-D(i+k)^\rho\right) \text{ for some constant } D>0,
\end{aligned} \tag{14}$$

where the approximation comes from the theory of large deviation, and the last inequality comes from Lemma 4. Further, define event $E_{i+k} = \{\text{an } i+k\text{-upcrossing starts}\}$. From (14), we have:

$$\mathbb{P}\left(E_{i+k}|x_i \leq \frac{1}{\gamma}\right) \leq (i+k)\exp\left(-D(i+k)^\rho\right).$$

Finally, define event $E_i = \{\text{an upcrossing starts after } i\}$. Then, we get

$$\mathbb{P}\left(E_i|x_i \leq \frac{1}{\gamma}\right) \leq \sum_k \mathbb{P}\left(E_{i+k}|x_i \leq \frac{1}{\gamma}\right) \leq \sum_k (i+k)\exp\left(-D(i+k)^\rho\right) \to 0$$

as $i \to +\infty$. Therefore, there exists some $\sigma > 0$ such that when i is large enough, we have

$$\mathbb{P}\left(E_i|x_i \leq \frac{1}{\gamma}\right) \leq 1-\sigma.$$

As a consequence, we have

$$\mathbb{P}\left(x_{i+k} < \frac{1}{\gamma}+\epsilon, \forall k \geq 0 | x_i \leq \frac{1}{\gamma}\right) > \sigma,$$

and $\epsilon > 0$ is arbitrarily chosen.

Step 3: x_i converges to $\frac{1}{\gamma}$ almost surely.

Define event $G_\epsilon = \left\{x_i < \frac{1}{\gamma}+\epsilon \text{ for all } i \text{ sufficiently large}\right\}$. From previous steps, we know that $\mathbb{P}(G_\epsilon|\mathcal{F}_i) > \sigma > 0$. Then, from Levy's 0-1 law, $\mathbb{P}(G_\epsilon) = 1$. Next, we want to show that $\liminf x_i = \frac{1}{\gamma}$. From the dynamics of x_i, we know that whenever x_i falls below $\frac{1}{\gamma}$, it immediately starts moving upward until getting out of the cascade. Thus, the smallest x_i's $(x_i \leq \frac{1}{\gamma})$ are the ones that follow $x_{i-1} \geq \frac{1}{\gamma}$, and individual i chooses action 0. We have:

$$\ln(x_i) = \frac{\alpha_i}{\alpha_{i-1}} \ln(x_{i-1}) + \alpha_i \ln\left(\frac{\phi^1(a_{i-1}, x_{i-1})}{\phi^0(a_{i-1}, x_{i-1})}\right) \leq \ln\left(\frac{1}{\gamma}\right),$$

which implies that

$$\ln(x_{i-1}) \leq \frac{\alpha_{i-1}}{\alpha_i} \ln\left(\frac{1}{\gamma}\right) - \alpha_{i-1} \ln\left(\frac{\phi^1(a_{i-1}, x_{i-1})}{\phi^0(a_{i-1}, x_{i-1})}\right).$$

When $i \to +\infty$, the right-hand side approaches $\ln\left(\frac{1}{\gamma}\right)$. Therefore, $\lim_{i\to+\infty} \ln(x_{i-1}) = \ln\left(\frac{1}{\gamma}\right)$, which also implies $\lim_{i\to+\infty} \inf \ln(x_i) = \ln\left(\frac{1}{\gamma}\right)$. As a consequence,

$$\frac{1}{\gamma} = \liminf_{i\to+\infty} x_i \leq \limsup_{i\to+\infty} x_i < \frac{1}{\gamma} + \epsilon, \quad \forall \epsilon > 0,$$

so we have $x_i \to \frac{1}{\gamma}$ almost surely.

A.2 Proof of Proposition 2

Lemma 5. *There exists $K < \infty$ and $\delta > 0$ such that $\mathbb{P}(x_{i+K} \in C | x_i \notin C) \geq \delta$ for all i.*

Proof. Suppose $x_i \notin C$, then the strongest signal γ (or $1/\gamma$) will induce individual i to choose action 1 (or 0). Given $x_i \geq 1$ and that $a_i = ... = a_{i+k-1} = 1$, we have

$$\begin{aligned} x_{i+k} &= x_i^{\frac{\alpha_{i+k}}{\alpha_i}} \times \left[\prod_{j=i}^{i+k-1} \frac{\phi^1(1, x_j)}{\phi^0(1, x_j)}\right]^{\alpha_{i+k}} \geq 1 \times \left[\prod_{j=i}^{i+k-1} \frac{\phi^1(1, x_j)}{\phi^0(1, x_j)}\right]^{\underline{\alpha}} \\ &\geq \left(\frac{1 - F^1(1/\gamma)}{1 - F^0(1/\gamma)}\right)^{k\underline{\alpha}} \quad \text{if } x_i, ..., x_{i+k-1} \notin C, \end{aligned}$$

where $\underline{\alpha} = \inf_i \alpha_i > 0$,[21] and the first inequality comes from $x_i \geq 1$, and the second inequality comes from

$$\frac{\phi^\theta(1, x_j)}{\phi^\theta(1, x_j)} = \frac{1 - F^1(1/x_i)}{1 - F^0(1/x_i)}$$

is decreasing in x_i (see Lemma A.1 in Smith and Sørensen (2000)), so we have

$$\frac{\phi^\theta(1, x_j)}{\phi^\theta(1, x_j)} \geq \frac{1 - F^1(1/\gamma)}{1 - F^0(1/\gamma)} \equiv \Omega_1 > 1.$$

[21]Suppose α_i is increasing in i—which includes $\rho < 0$ and part of $\rho = 0$—then $\underline{\alpha} = \alpha_1$. Suppose $\rho = 0$ and α_i is decreasing, we can let $\underline{\alpha} = \lim_{i\to+\infty} \alpha_i > 0$ (by definition, $\alpha_i \cdot i^0 = c > 0$ as $i \to +\infty$).

Therefore, after at most $K_1 = \lceil \frac{1}{\underline{\alpha}} \log_{\Omega_1}^{\gamma} \rceil$ steps, we have $x_{i+K_1} \geq \gamma$, so a cascade of action 1 is triggered. Symmetrically, given $x_i \leq 1$ and that $a_i = ... = a_{i+k-1} = 0$, we have

$$x_{i+k} = x_i^{\frac{\alpha_{i+k}}{\alpha_i}} \times \left[\prod_{j=i}^{i+k-1} \frac{\phi^1(0, x_j)}{\phi^0(a_j, x_j)} \right]^{\alpha_{i+k}}$$
$$\leq 1 \times \left[\prod_{j=i}^{i+k-1} \frac{\phi^1(0, x_j)}{\phi^0(0, x_j)} \right]^{\underline{\alpha}} \leq \Omega_0^{k\underline{\alpha}} \text{ if } x_i, ..., x_{i+k-1} \notin C,$$

where $\Omega_0 = \frac{F^1(\gamma-)}{F^0(\gamma-)} < 1$. Then, after at most $K_0 = \lceil \frac{1}{\underline{\alpha}} \log_{\Omega_0}^{1/\gamma} \rceil$ steps, we have $x_{i+K_0} \leq 1/\gamma$, so a cascade of action 0 is triggered. Note than when $x_i \notin C$, a sequence of consecutive $K = \max\{K_1, K_2\}$ signals γ (or $1/\gamma$) will trigger a sequence of action 1 (or 0). Let the probability of realizing K signals being δ, then we prove the lemma. □

Lemma 6. *If α_i is increasing, an informational cascade occurs almost surely, and both correct and incorrect cascasdes occur with positive probability.*

Proof. Suppose α_i is increasing. When $x_i \in C$, then

$$x_{i+1} = x_i^{\frac{\alpha_{i+1}}{\alpha_i}} \times \left[\frac{\phi^1(a_i, x_i)}{\phi^0(a_i, x_i)} \right]^{\alpha_{i+1}} = x_i^{\frac{\alpha_{i+1}}{\alpha_i}} \begin{cases} \geq x_i & x_i \in C_1 \\ \leq x_i & x_i \in C_0 \end{cases},$$

so $x_{i+1} \in C$ as well, which implies that the power public LLR will remain in the cascade set. Lemma 5 implies that the probability of entering a cascade set is uniformly bounded away from 0. Standard Borel-Cantelli arguments show that an informational cascade occurs almost surely. Finitely many signals can trigger a cascade, so both cascades will emerge with positive probability. □

Lemma 7. *Suppose $\rho = 0$ and α_i is decreasing, the result in Lemma 6 still holds.*

Proof. The difficulty is that with decreasing α_i, we can no longer guarantee that $x_i \in C$ implies $x_{i+1} \in C$. However, we notice that:

$$x_{i+1} - x_i = x_i^{\frac{\alpha_{i+1}}{\alpha_i}} - x_i \to 0 \text{ as } i \to +\infty.$$

Therefore, when i is adequately large, even if x_{i+1} falls outside of the cascade set, it is very close to the boundary, γ or $1/\gamma$. Without loss, suppose $x_{i+1} > \gamma - \varepsilon$, where $\varepsilon > 0$ can be arbitrarily small. Suppose that $\lambda_{i+1} = \gamma$. When i is sufficiently large, we have:

$$x_{i+2} = x_{i+1}^{\frac{\alpha_{i+2}}{\alpha_{i+1}}} \times \left[\frac{1 - F^1(1/\gamma)}{1 - F^0(1/\gamma)} \right]^{\alpha_{i+1}} \approx x_{i+1} \times \left[\frac{1 - F^1(1/\gamma)}{1 - F^0(1/\gamma)} \right]^{\underline{\alpha}} > (\gamma - \varepsilon) \times \Omega_1^{\underline{\alpha}}.$$

Therefore, we have:

$$x_{i+2} - \gamma > \left(\Omega_1^{\alpha} - 1\right)\gamma - \varepsilon\Omega_1^{\alpha} \equiv \Delta.$$

Note that $\Omega_1^{\alpha} > 1$, so we can choose $\varepsilon > 0$ such that $\Delta > 0$, then $x_{i+2} > \gamma + \Delta$, and

$$x_{i+3} = x_{i+2}^{\frac{\alpha_{i+3}}{\alpha_{i+2}}} > (\gamma + \Delta)^{\frac{\alpha_{i+3}}{\alpha_{i+2}}}.$$

Since $\alpha_i \to \underline{\alpha} > 0$, for all $\xi > 0$, there exists $I < \infty$ such that $|\alpha_m/\alpha_n - 1| < \xi$ whenever $m, n > I$. We choose $\xi = 1 - \log_{\gamma+\Delta}^{\gamma}$, then when i is sufficiently large, we have:

$$x_{i+3} > (\gamma + \Delta)^{\frac{\alpha_{i+3}}{\alpha_{i+2}}} > (\gamma + \Delta)^{1-\xi} = \gamma,$$

so $x_{i+3} \in C_1$. Suppose $x_{i+3}, ..., x_{i+k-1} \in C_1$, then we have:

$$x_{i+k} = x_{i+2}^{\frac{\alpha_{i+k}}{\alpha_{i+2}}} > (\gamma + \Delta)^{\frac{\alpha_{i+k}}{\alpha_{i+2}}} > (\gamma + \Delta)^{1-\xi} = \gamma,$$

so x_i stays inside the cascade set forever. In summary, when i is sufficiently large, even if x_{i+1} falls outside of C, another signal can bring it back to the cascade set and make it stay there forever. Formally, we define *ϵ-cascade sets* as follows:

$$C_0^{\epsilon} = [0, 1/\gamma + \epsilon], \quad C_1^{\epsilon} = [\gamma + \epsilon, +\infty],$$

and $C^{\epsilon} = C_0^{\epsilon} \cup C_1^{\epsilon}$. From previous discussion, there exists some $\epsilon > 0$ such that when i is sufficiently large, we have one of the following three cases:

1. If $x_i \in C^{\epsilon}$, then x_{i+k} remains in the cascade set C for all $k \geq 0$;
2. If $x_i \in C$, but $x_{i+1} \notin C$, then a single signal will bring x_{i+2} to C^{ϵ};
3. If $x_i \notin C$, finitely many signals will bring x_i to C, and we have either Case 1 or 2.

In summary, there exists $K < \infty$ such that when i is sufficiently large, the probability of x_{i+K} entering C and staying forever is bounded away from 0. Standard arguments yield the result in Lemma 6. □

A.3 Proof of Proposition 3

Proof. Suppose $\alpha_i = \alpha$ for all i. Then, the public likelihood ratio $l_i = x_i^{1/\alpha}$ is a martingale, so it almost surely converges to a finite random variable l_∞. Therefore, x_i also converges to some limit x_∞ almost surely, and we have

$$\mathbb{E}x_\infty^{1/\alpha} = \mathbb{E}l_\infty = \lim_{i\to+\infty} \mathbb{E}\left(l_i\right) = 1,$$

where the interchange of limit and expectation comes from the boundedness of l_i. Note that we must have $x_\infty \in C_0$ or $x_\infty \in C_1$ in the limit. Therefore,

$$1 = \int_{\{x_\infty \in C_0\}} x_\infty^{1/\alpha} d\mathbb{P} + \int_{\{x_\infty \in C_1\}} x_\infty^{1/\alpha} d\mathbb{P} \geq 0 + \mathbb{P}\left(x_\infty \in C_1\right) \times \gamma^{1/\alpha},$$

so $\mathbb{P}\left(x_\infty \in C_1\right) \leq \gamma^{-1/\alpha}$. From Proposition 2, an information cascade almost surely occurs when $\rho = 0$, so we have:

$$p\left(\alpha\right) = \mathbb{P}\left(x_\infty \in C_0\right) \geq 1 - \gamma^{-1/\alpha},$$

that is, the probability of a correct cascade is greater than $1 - \gamma^{-1/\alpha}$. As $\alpha \to 0$, we have $1 - \gamma^{-1/\alpha} \to 1$, thus $p\left(\alpha\right) \to 1$. □

A.4 Proof of Proposition 3

Proof. For any i, we have

$$x_{i+1} = x_1^{\frac{\alpha_{i+1}}{\alpha_1}} \left[\prod_{k=1}^{i} \frac{\phi^1\left(a_k, x_k\right)}{\phi^0\left(a_k, x_k\right)}\right]^{\alpha_{i+1}} = \left[\prod_{k=1}^{i} \frac{\phi^1\left(a_k, x_k\right)}{\phi^0\left(a_k, x_k\right)}\right]^{\alpha_{i+1}},$$

where the equality comes from the flat prior assumption, so $x_1 = 1$.[22] Since $\lambda_i \in [1/\gamma, \gamma]$, and

$$\phi^\theta\left(a, x\right) = \begin{cases} 1 - F^\theta\left(\frac{1}{x}\right), & a = 1 \\ F^\theta\left(\frac{1}{x}\right), & a = 0 \end{cases},$$

so we have $\frac{\phi^1(a_i,x_i)}{\phi^0(a_i,x_i)} \in \left[\frac{1}{\gamma}, \gamma\right]$ for all i, and hence

$$x_{i+1} = \left[\prod_{k=1}^{i} \frac{\phi^1\left(a_k, x_k\right)}{\phi^0\left(a_k, x_k\right)}\right]^{\alpha_{i+1}} \in \left[\frac{1}{\gamma^{i\alpha_{i+1}}}, \gamma^{i\alpha_{i+1}}\right].$$

Since $\rho > 1$, we have $\lim_{i\to+\infty} i\alpha_{i+1} = 0$, which implies that $x_i \to 1$ as $i \to +\infty$. Thus, in the limit, all individuals will only act according to their own signal, i.e., they choose action

[22] This is non-essential. For general x_1, we have $x_1^{\frac{\alpha_{i+1}}{\alpha_1}} \approx 1$ when i is large enough, and the same argument applies.

1 whenever $\lambda_i > 1$ and action 0 otherwise, so there is no information aggregation. □

B A Maximum Inequality

In this section, we present a maximum inequality that we used in the proof:

Lemma 8. *Suppose that $X_1, X_2, \dots$ are bounded i.i.d. random variables. Then,*

$$\forall \varepsilon > 0 : \quad \mathbb{P}\left(\overline{X}_n - \mathbb{E}(X) \le \varepsilon \text{ for all } n \ge 1\right) > 0,$$

where $\overline{X}_n$ denotes the sample mean of random variables $X_1, \dots, X_n$.

Remark 2. This lemma looks similar to many famous theorems in large deviation theory. Differently, we require the inequality holds for ***all*** n instead of sufficiently large n.

Proof. From Hoeffding's inequality, we have:

$$\mathbb{P}\left(\overline{X}_n - \mathbb{E}(X) > \varepsilon\right) \le \exp(-Cn) \text{ for some } C > 0.$$

Therefore,

$$\sum_n \mathbb{P}\left(\overline{X}_n - \mathbb{E}(X) > \varepsilon\right) \le \sum_n \exp(-Cn) < \infty,$$

which implies $\mathbb{P}\left(\overline{X}_n - \mathbb{E}(X) > \varepsilon \text{ i.o.}\right) = 0$ from the Borel–Cantelli lemma. Therefore,

$$\mathbb{P}\left(\cup_{k=1}^{\infty}\left\{\overline{X}_n - \mathbb{E}(X) > \varepsilon \text{ for all } n \ge k\right\}\right) = 1.$$

The continuity from below of the probability measure implies that there exists some k such that

$$\mathbb{P}\left(\overline{X}_n - \mathbb{E}(X) \le \varepsilon \text{ for all } n \ge k\right) > 0, \tag{15}$$

and we choose k to be the smallest number that satisfies the above inequality.

Notice that equation (15) is very similar to Lemma 8. Suppose $k = 1$, Lemma 8 holds automatically. Next we want to show that k cannot be greater than 1. We proceed by contradiction. Suppose that $k > 1$, and we define:

$$E = \{\boldsymbol{x} \in [a, b]^{\infty} : \overline{x}_n - \mathbb{E}(X) \le \varepsilon \text{ for all } n \ge k\} \subset \mathbb{R}^{\infty},$$

where $\overline{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i$ and $[a, b]$ is the range of X. For any sequence $\boldsymbol{x} = (x_1, x_2, \dots) \in \mathbb{R}^{\infty}$, we define $\mathrm{Proj}_{k+}(x) \equiv (x_k, x_{k+1}, \dots)$, which is the subsequence of x starting from the k-th term. We further define:

$$H = \left\{ \text{Proj}_{k+}\left(\boldsymbol{x}\right) : \boldsymbol{x} \in E \right\}$$

and

$$G = \left\{ (x_1, .., x_{k-1}) \in [a,b]^{k-1} : \overline{x}_n - \mathbb{E}\left(X\right) \leq \varepsilon \text{ for all } n = 1, 2, ..., k-1 \right\}.$$

For any $g = (g_1, ..., g_{k-1}) \in G$ and $h = (h_1, h_2, ...) \in H$, we construct a new sequence:

$$\boldsymbol{x} = (g_1, g_2, ..., g_{k-1}, h_1, h_2, ...) \in G \times H.$$

By construction, whenever $n \leq k-1$, we have:

$$\overline{x}_n - \mathbb{E}\left(X\right) = \overline{g}_n - \mathbb{E}\left(X\right) \leq \varepsilon.$$

Suppose that $n \geq k$. Then, by the definition of H, h is associated with some $(\hat{x}_1, ..\hat{x}_{k-1}) \in [a,b]^{k-1}$ such that the combined sequence $\hat{\boldsymbol{x}} = (\hat{x}_1, ..\hat{x}_{k-1}, h)$ satisfies: (i) $\overline{\hat{x}_n} - \mathbb{E}\left(X\right) \leq \varepsilon$ for all $n \geq k$, and (ii) $\overline{\hat{x}}_{k-1} - \mathbb{E}\left(X\right) > \varepsilon$. Therefore, we have:

$$\begin{aligned} \overline{x}_n - \mathbb{E}\left(X\right) &= \frac{k-1}{n}\overline{x}_{k-1} + \frac{1}{n}\sum_{i=k}^{n} x_i \\ &< \frac{k-1}{n}\overline{\hat{x}}_{k-1} + \frac{1}{n}\sum_{i=k}^{n} x_i = \overline{\hat{x}_n} - \mathbb{E}\left(X\right) \leq \varepsilon, \end{aligned}$$

where the first inequality comes from the fact that $\overline{x}_{k-1} - \mathbb{E}\left(X\right) \leq \varepsilon$, and hence $\overline{x}_{k-1} < \overline{\hat{x}}_{k-1}$. To sum up, whenever $\boldsymbol{x} \in G \times H$, we must have $\overline{x}_n - \mathbb{E}\left(X\right) \leq \varepsilon$ for all $n \geq 1$. From the independence assumption,

$$\begin{aligned} \mathbb{P}\left(X_1, X_2, ... \in G \times H\right) &= \mathbb{P}\left(X_1, X_2, ...X_{k-1} \in G\right) \times \mathbb{P}\left(X_k, X_{k+1}, ... \in H\right) \\ &\geq \mathbb{P}\left(X_1, X_2, ...X_{k-1} \in G\right) \times \mathbb{P}\left(X_1, X_2, ... \in E\right). \end{aligned}$$

Note that G is a non-empty set, and it is finitely dimensional, so $\mathbb{P}\left(X_1, X_2, ...X_{k-1} \in G\right) > 0$. From (15), we know that $\mathbb{P}\left(X_1, X_2, ... \in E\right) > 0$. Therefore, $\mathbb{P}\left(X_1, X_2, ... \in G \times H\right) > 0$, implying that $\overline{x}_n - \mathbb{E}\left(X\right) \leq \varepsilon$ for all $n \geq 1$ is a positive probability event, contradicting the presumption that $k > 1$. So, the lemma is proved. □