Observational Learning Under Imperfect Information[1]

Boğaçhan Çelen
Department of Economics, New York University
269 Mercer St., 7th Floor, New York, NY, 10003-6687.
E-mail: bc319@nyu.edu
Url: http://home.nyu.edu/~bc319

and

Shachar Kariv
Department of Economics, New York University
269 Mercer St., 7th Floor, New York, NY, 10003-6687.
E-mail: sk510@nyu.edu
Url: http://home.nyu.edu/~sk510

Version: November 25, 2001

We explore Bayes-rational sequential decision making in a game with pure information externalities, where each decision maker observes only her predecessor's binary action. Under perfect information the martingale property of the stochastic learning process is used to establish convergence of beliefs and actions. Under imperfect information, by contrast, beliefs and actions cycle forever. However, despite the stochastic instability, over time the private information is ignored and decision makers become increasingly likely to imitate their predecessors. Consequently, we observe longer and longer periods of uniform behavior, punctuated by switches increasingly rare. *Journal of Economic Literature* Classification Numbers: D82, D83

Key Words: Asymmetric Information, Herd Behavior, Informational Cascade.

1. INTRODUCTION

Observational learning describes a situation in which decision-makers (*dm*s) learn by observing the behavior of others. The behavior of a *dm* who faces a decision problem under uncertainty is typically correlated with her

[1] We are grateful to Douglas Gale for his guidance. We also acknowledge helpful discussions of Jean-Pierre Benoît, Alberto Bisin, Andrew Caplin, Amil Dasgupta, Eric S. Maskin, Roger B. Myerson, Ben Polak, Matthew Rabin, Debraj Ray, Andrew Schotter, Peter Sørensen, Srinivasa Varadhan and the participants of seminars at Cornell, NYU and Yale and 2001 South-East Economic Theory & International Economics Conference.

private information. If this behavior is observed by other *dm*s who face the same problem, then they will be rationally tempted to revise their beliefs. This situation, in which the behavior of an individual indirectly affects the welfare of others via revealed information, is referred as an *information externality*.

In the last decade, a number of studies explored observational learning phenomena. Banerjee (1992) and Bikhchandani, Hirshleifer and Welch (1992) (BHW) introduced the basic concepts and stimulated further research in this area. The basic concern of these papers is to understand the behavior of an economy composed of Bayes-rational *dm*s in the presence of a *pure informational externality*, *i.e.,* where a *dm*'s payoff is independent of others' actions. The literature analyzes an economy where a sequence of *dm*s are supposed to make a once-in-a-lifetime decision under incomplete and asymmetric information. The common conclusion is that, despite asymmetric information, eventually *dm*s will imitate their predecessor's behavior even if it conflicts with their private information. Despite their simplicity, these models give insight into the rationale behind mass behaviors such as manias, fads, fashion, crashes and booms and their fragility.

The most comprehensive study on observational learning under perfect information is provided by Smith and Sørensen (2000) (SSb). In addition to their robust results, they introduce martingale techniques which facilitate the analysis. Moreover, SSb highlight the difference between *informational cascades* and *herd behavior*, two notions introduced by Banerjee (1992) and BHW to address the same phenomenon. *Informational cascade* takes place when, after some finite time, all *dm*s ignore their private information when choosing an action and *herd behavior* takes place when, after some finite time, all *dm*s choose the same action, not necessarily ignoring their private information. Hence, an informational cascade implies a herd but herd is not necessarily the result of an informational cascade. In a herd every *dm* chooses the same action, but they might have chosen a different sequence of actions if the realization of their private signals had been different.

A common feature of these models is that all *dm*s are assumed to be able to observe all the decisions that have previously been made, *i.e.,* they have *perfect information* about the history of decisions. In such a situation, previous decisions reveal a good deal of information to the subsequent *dm* who will make a decision. Suppose that a *dm* observes a history in which all the previous decisions are identical. In that case, she will deduce that each one of her predecessors either has private information that favors this particular decision or has some unfavorable information that is not convincing enough to lead her to deviate to another decision. This reasoning becomes stronger as the number of identical decisions that one observes increases. Therefore, unless a *dm*'s private information is very convincing relative to the decisions she observes, she will join the herd. This is the main intuition behind the results of the literature.

Now, consider an economy with incomplete information where everyone faces an identical decision problem, one *dm* after the other. A *dm*, when it is her turn to make a decision, knows nothing but her private information and her observation of decisions by those who recently faced the same problem. In other words, the *dm* has *imperfect information* about the history of decisions. In such a situation, the best she can do is to make a Bayesian inference about the entire history, based on these few recent observations. To illustrate, think of a Ph.D. student who is on the job market and has to decide which universities to apply to. She is aware that many students faced the same problem in previous years but she is able to observe only the decisions of the recent years' job market candidates.

Our goal is to understand behavior under such an imperfect information structure, in particular when each *dm* observes only her immediate predecessor's decision. This information structure not only captures numerous social learning situations in which recent data are much likely to be available to *dm*s but also helps us to question how well-grounded are the results of the literature under conventional perfect information assumption.

The model which we analyze builds on Gale (1996). Each *dm* is faced with a once-in-a-lifetime binary choice, say, an investment decision. While non-investment is a safe action yielding a payoff zero, payoff from investment is a random variable with expected value zero. Each *dm* receives an informative signal, that can not be observed by others, drawn from a distribution over a compact support. We describe the *dm*s' optimal strategies in a recursive way which, in turn, characterize the dynamics of learning and actions.

Under perfect information, the martingale character of the learning process establishes the results on learning and thus on actions. A cascade can not arise, yet a limit-cascade and thus herd do (Theorem 1 and Theorem 2). Under imperfect information, however, learning everlastingly cycles although, over time, private information is gradually ignored. Our striking result is that despite instability of learning dynamics, eventually all *dm*s tend to ignore their private signals (Theorem 5). That is, learning cycles forever between contrary confident beliefs about states of the world, but eventually no one is self-reliant. Consequently, the behavior falls short of convergence of actions in the standard herding manner, but exhibits ever longer lasting clusters of *dm* acting alike (Theorem 6). To be precise, in some finite time, there will be a *dm* who, in spite of her predisposition to imitate the action of her predecessor, will receive a conflicting private signal decisive enough to lead her to deviate. The reason behind this decision is the reversal of her belief which dominates her predisposition. In contrast, if this *dm* were to receive a non-conflicting signal, she would not only imitate her predecessor but she would also be even more confident about her decision than her predecessor was.

In a nutshell, under perfect information,

> Learning converges to an informational pooling equilibrium in which the complete history of actions provides a decisive guide regarding to what the state of the world is.
>
> As a result, the behavior settles down in some finite time and is consistent with the limit learning. That is to say that a herd arises.

By contrast, under imperfect information,

> An informational pooling equilibrium never exists. That is, learning is everlastingly flanked by conviction in contrary states of the world.
>
> Consequently, the standard herding outcome is impossible. Nonetheless, the behavior is typified by longer and longer periods of uniform behavior, punctuated by increasingly more rare switches.

The intuition for these results is simple. When *dm*s observe the entire history of actions, they know the reason why their predecessors choose the way they did. Therefore, once a long cluster of identical actions forms, late *dm*s lose confidence in their private information. It is swamped by the information contained in the history. However, if the entire history of actions is not available to a Bayesian *dm*, she draws probabilistic conclusions about the history based only on the limited information available to her. Since history is providing less information in this setting, *dm*s tend to rely more on their own signal and hence there are always some *dm*s who follow their own signals.

Our results lead us to argue that imperfect information brings about observational learning outcomes which are atypical and more extreme than those described by the standard perfect-information inference. The key socioeconomic phenomenon that imperfect information captures is a succession of fads starting suddenly, expiring rather easily, and each replaced by another fad. We believe that an imperfect information model better answers the questions: Why do markets move from boom to crash without settling down? Why is a technology adopted by a wide range of users more rapidly than expected and then, suddenly, replaced by an alternative? What makes a restaurant fashionable over night and equally unexpectedly unfashionable, while another becomes *in*, and so on?

Smith and Sørensen (1996) (SSa) study an observational learning model in which all *dm*s observe an unordered random sample of actions from the history. They provide a thorough characterization for the case of unbounded signals, but their results are not exhaustive for the case of bounded signals. In particular, they show that with bounded signals, learning might be incorrect. We discuss the related literature later in the paper.

The paper is organized as follows. In the next section, we introduce the model and illustrate it with an example. The model is analyzed under some distribution specification in section 3 and for general distributions in section 4. We discuss our results in section 5. The related literature is discussed in section 6. We conclude in section 7.

2. THE MODEL

2.1. Preliminaries

The economy consists of a finite number of Bayes-rational *dm*s, $\mathcal{N} = \{1, 2, \cdots, N\}$. Each *dm* n makes a once-in-a-lifetime decision, to invest or not to invest, denoted by $x_n = 1$ and $x_n = 0$ respectively. Decision making is assumed to be sequential in an exogenous order.

The preferences of the *dm*s are assumed to be identical and represented by a risk neutral vN-M utility function. If a *dm* decides not to invest the payoff is 0. Yet, the payoff from investment is determined by a random variable Θ with expected value 0. One interpretation is that while staying intact is a safe action, investing is a risky one which can result in loss as well as gain.

$$u(x_n) = \begin{cases} \Theta & \text{if} \quad x_n = 1 \\ 0 & \text{if} \quad x_n = 0 \end{cases}$$

It is immediate that the range of Θ defines the set of the states of the world. Moreover, since the risk-free action $x_n = 0$ constitutes a benchmark for decision making, the payoff-relevant states are partitioned into two decision-relevant events, *high* states $\Theta \geq 0$ and *low* states $\Theta < 0$.

As it is customary in the literature, *dm*s are assumed to have a private information about the realization of the state of the world. To this end we assume

$$\Theta = \sum\nolimits_{n \in \mathcal{N}} \theta_n$$

where each θ_n is distributed by F over a compact support with convex hull $[a, b]$, $-\infty < a < 0 < b < \infty$, such that $\mathbb{E}[\theta_n] = 0$ and that θ_ns are *i.i.d.*. We, then, assume that each *dm* gets to observe one of the variables. Therefore, each *dm* n receives a signal θ_n about the unknown value of investment Θ[2].

Notice that the signal structure is informative in the sense that, conditional on the true state of the world, one is more likely to receive a signal favoring this state, *i.e.,* $P(\theta \geq 0 \,|\Theta \geq 0\,) > \frac{1}{2}$ and $P(\theta < 0 \,|\Theta < 0\,) > \frac{1}{2}$. Yet, for $N > 2$ there is no private signal for which a *dm* can resolve the uncertainty by herself. In other words, one's private information alone can never perfectly reveal the state of the world. This is referred in the literature as *bounded beliefs*.

[2] We thank Eric S. Maskin who provided us with the following example. Consider an auction for an oil field. In order to acquire information auctioneers get samples and conduct tests on the samples. Each auctioneer forms his valuation according to his own test result. However the real value of the field is reflected by the aggregate of the results of all samples.

Thus far, the model depicted a decision problem under incomplete and asymmetric information. That is, *dms* are uncertain about the value of the investment and the information regarding it is shared asymmetrically among them. Moreover, since there is no strategic interaction, *dms* may benefit from the observation of past decisions solely because they reveal some information about the signals on which they were based upon.

We refer to a perfect information economy $\mathcal{E}_F = \{F, x_n, u_n, I_n\}_{n \in \mathcal{N}}$, as an economy where the information set of each *dm* n consists of her private signal as well as the entire history of actions, *i.e.,*

$$I_n = \{\theta_n, (x_i)_{i=1}^{n-1}\} \in [a,b] \times \{0,1\}^{n-1}$$

An imperfect information economy $\mathcal{E}'_F = \{F, x_n, u_n, I'_n\}_{n \in \mathcal{N}}$ differs in that *dms* are imperfectly informed about the decisions that have been previously taken. In particular, we assume that any *dm* $n > 1$ observes only her immediate predecessor's action, *i.e.,*

$$I'_n = \{\theta_n, x_{n-1}\} \in [a,b] \times \{0,1\}$$

Finally, we assume that the structure of any *dm*'s information set is common knowledge. Thus, every *dm* knows whose actions each *dm* observes as well as all the decision rules.

2.2. The decision problem

Given her information set, *dm* n's strategy is a mapping from the space of private signals and the set of histories of actions that she can observe into the set of actions. Next, we provide a definition that will be useful in characterizing the optimal strategy.

Definition 1. Given her information set i_n, *dm* n follows a *cutoff strategy* if her decision rule is defined by a cutoff $\tilde{\theta}_n$:

$$x_n(i_n) = \begin{cases} 1 & if \quad \theta_n \geq \tilde{\theta}_n \\ 0 & if \quad \theta_n < \tilde{\theta}_n \end{cases}$$

for some $\tilde{\theta}_n \in [-1, 1]$.

The decision problem of *dm* n is to choose $x_n \in \{0,1\}$ to maximize her expected utility given the information set i_n. That is,

$$\underset{x_n \in \{0,1\}}{Max} \; x_n \mathbb{E}\left[\Theta \mid i_n\right]$$

which yields the optimal decision rule

$$x_n = 1 \text{ if and only if } \mathbb{E}\left[\Theta \mid i_n\right] \geq 0$$

Since i_n does not provide any information about the content of the signals θ_i for any $i > n$, we obtain

$$x_n = 1 \text{ if and only if } \mathbb{E}\left[\sum\nolimits_{i=1}^{n} \theta_i \mid i_n\right] \geq 0$$

thus,

$$x_n = 1 \text{ if and only if } \theta_n \geq -\mathbb{E}\left[\sum\nolimits_{i=1}^{n-1} \theta_i \mid i_n\right]$$

It readily follows that the optimal decision takes the form of the *cutoff strategy.* We state this in the next proposition.

PROPOSITION 1. *For any $n \in \mathcal{N}$, the optimal strategy is the cutoff strategy*

$$x_n(i_n) = \begin{cases} 1 & if \quad \theta_n \geq \hat{\theta}_n \\ 0 & if \quad \theta_n < \hat{\theta}_n \end{cases}$$

where

$$\hat{\theta}_n = -\mathbb{E}\left[\sum\nolimits_{i=1}^{n-1} \theta_i \mid i_n\right] \tag{1}$$

is the optimal history-contingent cutoff.

The optimal cutoff $\hat{\theta}_n$ inherits all the information that *dm* n learns from the history that she observes and it determines the minimum private signal for which she optimally decides to invest. Hence, $\hat{\theta}_n$ is sufficient to characterize *dm* n's behavior, and $\{\hat{\theta}_n\}_{n\in\mathcal{N}}$ characterizes the behavior of the economy. Henceforth, we take $\{\hat{\theta}_n\}_{n\in\mathcal{N}}$ as the object of our analysis and refer to it as a cutoff process or learning process interchangeably.

Next, we define some key concepts that we refer to throughout the paper.

2.3. Definitions

An informational cascade is said to take place when, from some finite *dm* on, the history of decisions does not provide any additional information since the *dm*s' actions are independent of their signals. Formally,

DEFINITION 2. An informational cascade occurs if there exists some $n \in \mathcal{N}$ such that either $\forall k \geq n$ $\hat{\theta}_k = 1$ or $\forall k \geq n$ $\hat{\theta}_k = -1$. Analogously, a limit-cascade occurs when either $\hat{\theta}_n \to 1$ as $n \to \infty$ or $\hat{\theta}_n \to -1$ as $n \to \infty$.

We shall refer to learning as complete, but not necessarily correct, whenever an informational cascade or a limit-cascade occurs. That is, when all but finitely many *dm*s are almost surely convinced about the realization of the state of the world.

We define a *potential herd* as a finite cluster of *dm*s who act alike and we let l_n denote the length of the potential herd starting from *dm* n. Herd behavior is said to take place when, from some finite dm on, an infinite cluster of dms act alike. An informational cascade implies a herd but herd is not necessarily the result of an informational cascade. In a herd every dm chooses the same action, but they might have chosen a different sequence of actions if the realization of their private signals had been different. Formally,

DEFINITION 3. Herd behavior occurs when, for some finite $n \in \mathcal{N}$, $P(l_n = \infty) > 0$ and $P(l_n = \infty) \to 1$ as $n \to \infty$.

Now, let us illustrate the basic features of the model with a simple example.

2.4. A simple illustration

We illustrate the underlying differences between perfect and imperfect information economies by a three-dm example where the signals, θ_n, are distributed uniformly over the support $[-1, 1]$, *i.e.*, $-a = b = 1$. The first dm's decision is based solely on her private signal, thus she takes action $x_1 = 1$ if and only if $\theta_1 \geq 0$ and action $x_1 = 0$ otherwise. Since the second dm observes the first's decision, according to (1), it is optimal for her to invest if and only if $\theta_2 \geq -\mathbb{E}[\theta_1 \mid x_1]$. Thus, she invests for $\theta_2 \geq -1/2$ or $\theta_2 \geq 1/2$ conditional on observing $x_1 = 1$ or $x_1 = 0$ respectively. Note that the second dm might imitate the first even though she would have made a contrary decision had she based her decision solely on her own signal.

When it is the third dm's turn to make a decision, perfect and imperfect information lead to different cutoff rules due to the different information sets that the third dm has *i.e.,* under perfect information $I_3 = \{\theta_3, (x_i)_{i=1,2}\}$ whereas under imperfect information $I'_3 = \{\theta_3, x_2\}$. Under perfect information, since the first dm's action is public information known to both successors, the third knows the observation on which the second based her decision. By (1), a simple computation yields the third dm's cutoff rule,

$$\hat{\theta}_3 = \begin{cases} \frac{3}{4} & \text{if} \quad x_1 = 0, x_2 = 0, \\ \frac{1}{4} & \text{if} \quad x_1 = 1, x_2 = 0, \\ -\frac{1}{4} & \text{if} \quad x_1 = 0, x_2 = 1, \\ -\frac{3}{4} & \text{if} \quad x_1 = 1, x_2 = 1. \end{cases}$$

Notice that any deviation of the second dm reveals that her private signal is contrary to and stronger than the third's expectation of the first's private signal. So, when the third observes a deviation, her cutoff is more sensitive to the second's action.

Under imperfect information, the third dm does not observe the first's decision. Still, she can draw a probabilistic conclusion about it by Bayesian reasoning. That is, by observing x_2, she can assign probability $P(x_1 \mid x_2)$ to the actions that the first dm could have taken. By incorporating the decision of the first dm in this way, according to (1) a simple calculation yields

$$\hat{\theta}_3 = \begin{cases} \frac{5}{8} & \text{if} \quad x_2 = 0, \\ -\frac{5}{8} & \text{if} \quad x_2 = 1. \end{cases}$$

Thus, the third dm invests for any signal $\theta_3 \geq -5/8$ or $\theta_3 \geq 5/8$ given that the second invests or not respectively.

To summarize, notice that under imperfect information, less information is revealed than under perfect information, but it may still be enough to alter the successors' decisions. The central difference between the third dm's decision under perfect and imperfect information is that the former lacks the information embodied in a deviation. Under perfect information, the third dm knows upon what observation the second made her decision. Thus, by observing a deviation she knows that the second dm had a signal persuasive enough to dominate her observation. On the other hand, under imperfect information, the best she can obtain is a Bayesian inference about the decision of the first dm. Thus, by the time it is her turn to make a decision, the information inherent in the first's decision is suppressed. However, along the line of dms, more information is accumulated in the last dm's decision.

In what follows we analyze the case of uniform distribution in more detail to obtain our main results, which we later generalize to any symmetric distributions.

3. THE UNIFORM CASE

For the ease of exposition and analysis, we first analyze the case where each signal θ_n is distributed with uniform distribution U, over $[-1,1]$.

3.1. Perfect information

In the perfect information economy, $\mathcal{E}_U = \{U, x_n, u_n, I_n\}_{n\in\mathcal{N}}$, dm n's information set is given by $I_n = \{\theta_n, (x_i)_{i=1}^{n-1}\}$. Thus, according to (1) her optimal history-contingent cutoff rule is,

$$\hat{\theta}_n = -\mathbb{E}\left[\sum\nolimits_{i=1}^{n-1} \theta_i \mid (x_i)_{i=1}^{n-1}\right]$$

Under perfect information, any history of actions $(x_i)_{i=1}^{n-2}$ is public information shared by all dms $i \geq n-1$ which results in a common expectation of $\sum_{i=1}^{n-2} \theta_i$. Since all the information revealed by the history $(x_i)_{i=1}^{n-2}$ is already accumulated in dm $(n-1)$'s cutoff, dm n's cutoff alters only by the new information revealed from $(n-1)$'s action. To be exact, $\hat{\theta}_n$ is different from $\hat{\theta}_{n-1}$ only by $\mathbb{E}[\theta_{n-1} \mid x_{n-1}, \hat{\theta}_{n-1}]$. As a result, the cutoff rule exhibits the following recursive structure,

$$\hat{\theta}_n = \hat{\theta}_{n-1} - \mathbb{E}[\theta_{n-1} \mid x_{n-1}, \hat{\theta}_{n-1}] \tag{2}$$

where

$$\mathbb{E}\left[\theta_{n-1} \mid x_{n-1}\right] = \begin{cases} \frac{1+\hat{\theta}_{n-1}}{2} & \text{if} \quad x_{n-1}=1 \\ \frac{-1+\hat{\theta}_{n-1}}{2} & \text{if} \quad x_{n-1}=0 \end{cases} \tag{3}$$

Equations (2) and (3) yield the stochastic cutoff process which we state in the next proposition.

Proposition 2. *In $\mathcal{E}_U$, cutoff dynamics abides by the following stochastic process,*

$$\hat{\theta}_n = \begin{cases} \frac{-1+\hat{\theta}_{n-1}}{2} & \textit{if} \quad x_{n-1}=1 \\ \frac{1+\hat{\theta}_{n-1}}{2} & \textit{if} \quad x_{n-1}=0 \end{cases} \tag{4}$$

where $\hat{\theta}_1 = 0$.

The following two corollaries are immediate from Proposition 2.

Corollary 1. *In $\mathcal{E}_U$, $\mid \hat{\theta}_n \mid < 1$, $\forall n \in \mathcal{N}$.*

Proof. Suppose $\exists n \in \mathcal{N}$ such that $\mid \hat{\theta}_n \mid = 1$. By applying (4) repeatedly we obtain $\mid \hat{\theta}_1 \mid = 1$. A contradiction. ∎

Corollary 2. *In $\mathcal{E}_U$, the cutoff process (4) has the Martingale property.*

Proof. $\mathbb{E}[\hat{\theta}_n \mid \hat{\theta}_{n-1}] = \frac{1-\hat{\theta}_{n-1}}{2}\frac{-1+\hat{\theta}_{n-1}}{2} + \frac{1+\hat{\theta}_{n-1}}{2}\frac{1+\hat{\theta}_{n-1}}{2} = \hat{\theta}_{n-1}$. ∎

The impossibility of an informational cascade, in $\mathcal{E}$, follows immediately from Corollary 1. That is, in making a decision, every *dm* takes into account her private signal in a non-trivial way. By virtue of Corollary 2, the learning process $\{\hat{\theta}_n\}$ has the Martingale property. So, by the Martingale Convergence Theorem it converges to a random variable $\hat{\theta}_\infty$ almost surely as $n \to \infty$. Hence, it is stochastically stable in the neighborhood of the fixed points, namely -1 and 1, meaning there is a limit-cascade and thus complete learning obtains. Summarizing:

Theorem 1 (Learning). *In $\mathcal{E}_U$,*
(*i*) *An informational cascade never arises.*
(*ii*) *A limit-cascade obtains.*

Next, we shall argue that convergence of the cutoff process implies convergence of actions. To begin with, we analyze by (4) how the behavior changes in the presence of a deviation. Suppose that a potential herd is followed by a *dm* who deviates since she receives a decisive signal favoring the contrary action. Since such a deviation reveals fairly clear-cut information regarding the signal it has been based upon, the newly revealed information slightly dominates the information that had already accumulated. That is, all the accumulated information favoring a particular state of the world is offset by the newly revealed information. Thus, a deviator induces her successor dm to be slightly in favor of joining the deviation. This is referred to by SSb as the *overturning principle*. To illustrate, suppose that dm $(n-1)$ is a part of a potential herd of investment and her cutoff is $\hat{\theta}_{n-1} = -1 + 2\varepsilon$ for arbitrarily small $\varepsilon > 0$. Thus, according to (4) dm n's cutoff is $\hat{\theta}_n = -1 + \varepsilon$. Now, if dm n has an extreme contrary signal, say $\theta_n = -1$, she deviates by choosing not to invest. Moreover, having observed the deviation, dm $(n+1)$'s overturns yet be close to zero, specifically $\hat{\theta}_{n+1} = \frac{\varepsilon}{2}$.

Yet, since $\{\hat{\theta}_n\}$ is stochastically stable near -1 and 1, the behavior can not overturn forever. That is, convergence of the cutoff process implies convergence of actions. In other words, the behavior settles down in some finite time and is consistent with the limit learning. Concluding,

THEOREM 2 (Behavior). *In $\mathcal{E}_U$, a herd arises.*

In what follows, we assume imperfect information and analyze the model accordingly.

3.2. Imperfect information

In the imperfect information economy, $\mathcal{E}'_U = \{U, x_n, u_n, I'_n\}_{n\in\mathcal{N}}$, where the signals are distributed by uniform distribution U, dm n's information set is given by $I_n = \{\theta_n, x_{n-1}\}$. The action of a dm is the only source of information to disclose the nature of all past signals for her immediate successor. Thus, according to (1) dm n's history-contingent cutoff rule is,

$$\hat{\theta}_n = -\mathbb{E}\left[\sum\nolimits_{i=1}^{n-1} \theta_i \mid x_{n-1}\right]$$

It can be readily noted that $\hat{\theta}_n$ can take two different values conditional on $x_{n-1} \in \{0,1\}$. That is,

$$\hat{\theta}_n = \begin{cases} \overline{\theta}_n & \text{if } x_{n-1} = 1, \\ \underline{\theta}_n & \text{if } x_{n-1} = 0. \end{cases}$$

where,

$$\begin{aligned} \overline{\theta}_n &= -\mathbb{E}\left[\sum\nolimits_{i=1}^{n-1} \theta_i \mid x_{n-1} = 1\right], \\ \underline{\theta}_n &= -\mathbb{E}\left[\sum\nolimits_{i=1}^{n-1} \theta_i \mid x_{n-1} = 0\right]. \end{aligned}$$

The following three lemmata will help us to derive the cutoff process $\{\hat{\theta}_n\}$. First, we observe that Bayesian inference of any dm is symmetric in the sense that upon observing the predecessor's action the probability assigned to a deviation (imitation) is independent of the actual action taken.

LEMMA 1. *$P(x_{n-1} = 0 \mid x_n = 1) = P(x_{n-1} = 1 \mid x_n = 0)$, $\forall n \in \mathcal{N}$.*

Proof. See Appendix. ∎

In words, upon observing x_n the probability that *dm* $(n+1)$ assigns to the event $x_n \neq x_{n-1}$ is the same for each $x_n \in \{0, 1\}$. The next lemma simply captures the symmetry of the cutoff rule.

LEMMA 2. *$\overline{\theta}_n + \underline{\theta}_n = 0$, $\forall n \in \mathcal{N}$.*

Proof. See Appendix. ∎

The last lemma states that *ex ante* any *dm* $n \in \mathcal{N}$ may take both actions equally likely.

LEMMA 3. *$P(x_n = 1) = \frac{1}{2}$, $\forall n \in \mathcal{N}$.*

Proof. See Appendix. ∎

These lemmata yield the cutoff dynamics stated in the next proposition.

PROPOSITION 3. *In $\mathcal{E}'_U$, the cutoff dynamics abides by the following stochastic process,*

$$\hat{\theta}_n = \begin{cases} -\frac{1+\hat{\theta}^2_{n-1}}{2} & \textit{if} \quad x_{n-1} = 1, \\ \frac{1+\hat{\theta}^2_{n-1}}{2} & \textit{if} \quad x_{n-1} = 0 \end{cases} \tag{5}$$

where $\hat{\theta}_1 = 0$.

Proof. See Appendix. ∎

We immediately obtain the following corollaries.

COROLLARY 3. *In $\mathcal{E}'_U$, $\mid \hat{\theta}_n \mid < 1$, $\forall n \in \mathcal{N}$.*

Proof. Suppose $\exists n \in \mathcal{N}$ such that $\mid \hat{\theta}_n \mid = 1$. By applying (5) repeatedly we obtain $\mid \hat{\theta}_1 \mid = 1$. A contradiction. ∎

COROLLARY 4. *In $\mathcal{E}'_U$, $[\overline{\theta}_n, \underline{\theta}_n) \supset [\overline{\theta}_{n-1}, \underline{\theta}_{n-1})$, $\forall n \in \mathcal{N}$ and $[\overline{\theta}_n, \underline{\theta}_n) \to [-1, 1]$ as $n \to \infty$.*

Proof. It follows immediately from (5) since $\overline{\theta}_n < \overline{\theta}_{n-1}$ and $\underline{\theta}_n > \underline{\theta}_{n-1}$. ∎

COROLLARY 5. *In $\mathcal{E}'_U$, the cutoff process $\{\hat{\theta}_n\}$ is not convergent.*

Proof. See Appendix. ∎

Now we are ready to address the behavior of the economy under imperfect information. We shall address the two fundamental questions.

What are the long run learning outcomes?

How does the behavior alter along the line of dms?

The impossibility of informational cascade follows immediately from Corollary 3, *i.e., dm*s forever rely on their signals in a non-trivial way in making a decision. Furthermore, Corollary 5 points out that also a limit-cascade never arises since the cutoff process is unstable near any of the fixed points -1 and 1. However, as we illustrate in Figure 1, *dm n*'s cutoff rule partitions the signal space to three subsets: $[-1, \overline{\theta}_n), [\overline{\theta}_n, \underline{\theta}_n)$ and $[\underline{\theta}_n, 1]$. For high-value signals $\theta_n \in [\underline{\theta}_n, 1]$ and low-value signals $\theta_n \in [-1, \overline{\theta}_n)$ she follows her private signal and decides to invest or not respectively. In the intermediate subset $[\overline{\theta}_n, \underline{\theta}_n)$, which we call a *cascade set,* private signals are ignored in making a decision and *dm*s imitate their immediate predecessor's action. Furthermore, by Corollary 4, cascade sets monotonically increase in n regardless of the actual history of actions and converge to the entire signal space in the limit. That is to say that, in the limit, the cascade set is an attractor. Hence, over time, *dm*s tend to rely more on the information revealed by the predecessor's action rather than their private signal and the limit *dm*s rely on it exclusively. We summarize these results in the next theorem.

THEOREM 3 (Learning). *In* $\mathcal{E}'_U$,
(i) Learning is incomplete, i.e., neither an informational cascade nor a limit-cascade arises.
(ii) The cascade set $[\overline{\theta}_n, \underline{\theta}_n)$ *is increasing in* n *and is an attractor in the limit, i.e.,* $[\overline{\theta}_n, \underline{\theta}_n) \supset [\overline{\theta}_{n-1}, \underline{\theta}_{n-1}), \forall n \in \mathcal{N}$ *and* $[\overline{\theta}_n, \underline{\theta}_n) \to [-1, 1]$ *as* $n \to \infty$.

Note that, unlike in $\mathcal{E}$ where the martingale character of the cutoff process rules out limit cycles, in $\mathcal{E}'$ the process is unstable near any of the fixed points and it exhibits an extreme jump character. That is, infinitely often, due to a contrary decisive private signal the cutoff swings from a point close to one of the fixed points to a point even closer to the other fixed point. Consequently, consecutive *dm*s, based on what they observe, may have such conflicting beliefs that the former is confident that the state is *high* and the latter relatively more confident that the state is *low*. As a result, the long run learning outcome is such that in spite of divergence in cutoffs, the limit *dm* acts irrespective of her signal. That is, learning forever cycles between confidence that the state is *high* and *low*, but eventually no one is self-reliant.

Next, we show that the divergence of cutoffs implies divergence of actions. In other words, the standard herd behavior is impossible. However, the length of a potential herd starting from any *dm n*, is a random variable which its expected value is increasing in n and tends to infinity for an arbitrarily large population. We state these results formally in the following theorem.

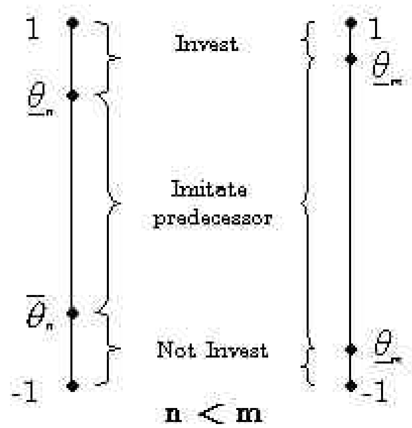

FIG. 1 The partition of the signal space: When *dm* n receives a signal in the cascade set $[\overline{\theta}_n, \underline{\theta}_n)$, she imitates her predecessor. Moreover, the cascade set becomes larger and converges to $[-1, 1]$ as n increases.

THEOREM 4 (Behavior). *In* $\mathcal{E}'_U$,
(i) Herd behavior does not arise, i.e., $P(l_n = \infty) = 0$ *for* N *arbitrarily large.*
(ii) Expected length of the potential herds, $\mathbb{E}[l_n]$, *is increasing in* n.
(iii) Expected length of a potential herd starting from dm n *is unbounded, i.e.,* $\mathbb{E}[l_n] \to \infty$ *as* $N \to \infty$ $\forall n \in \mathcal{N}$.

Proof. *(i)* Follows immediately from Corollary 5. *(ii)* Follows immediately from Corollary 4. *(iii)* See Appendix. ∎

Addressing the overturning principle under imperfect information may well clarify. From Corollary 5 it follows that the behavior overturns forever. That is, there is always one who deviates since she receives a decisive private signal incompatible with the information revealed from her predecessor's action. Furthermore, no one can tell whether her predecessor is an imitator or a deviator. Consequently, one who is subsequent to a deviator would be very enthusiastic to join the deviation. That is to say that because of the cutoff process' extreme jump character, the behavior in the presence of a deviation alters extremely. Since the cutoff process tends to one of the fixed points, the action consistent with this fixed point is potentially herded upon. Further, by virtue of the overturning principle, a deviation causes the cutoff process to jump closer to the other fixed point. As a result, it is likely that the contrary action would turn out to be potentially herded upon. Thus, along the line of *dm*s behavior is typified by monotonically longer lasting potential herds. To sum, notwithstanding with the impossibility of action convergence in the standard herding manner, The process of cutoffs converges fast enough so that the length of an identical-action-string following any *dm* n is infinite when the size of the population is arbitrarily large.

So far, we have focused our attention to unacceptably restrictive assumption of uniformity regarding the distribution of the signals. In what follows, we show that all the results obtained so far, in fact, hold for any symmetric distributions.

4. THE GENERAL CASE

We concentrate only on imperfect information economy $\mathcal{E}'_F$ since perfect information economy $\mathcal{E}_F$ is well-understood in the literature, and because the martingale character of the cutoff process readily follows from (2)

$$\hat{\theta}_n = \hat{\theta}_{n-1} - \mathbb{E}[\theta_{n-1} \mid x_{n-1}, \hat{\theta}_{n-1}].$$

Therefore, Theorem 1 and Theorem 2 generalize immediately to any distribution.

4.1. The symmetric case

In this part we consider a class of distributions $\mathcal{F}$ whose generic element $F \in \mathcal{F}$ satisfies symmetry, *i.e.*, $a+b=0$ and $\forall \theta \in [a, -a]$, $F(\theta) = 1 - F(-\theta)$.

Now, we assume an imperfect information economy, $\mathcal{E}'_F$, such that private signals are distributed with $F \in \mathcal{F}$. Without loss of generality, we assume that $b = 1$.

Symmetry implies that Bayesian inference of any *dm* is symmetric, *i.e.*, upon observing x_n the probability that *dm* $(n+1)$ assigns to a deviation (imitation) is independent of the actual action taken, the symmetry of the cutoff rule, *i.e.*, $\overline{\theta}_n + \underline{\theta}_n = 0$, and that *ex ante* any *dm* n may take both actions equally likely. These results can be shown by induction arguments similar to those in the proofs of lemmata 1-3. Using these results, the law of motion for $\overline{\theta}_{n+1}$ is given by

$$\overline{\theta}_{n+1} = -\frac{\frac{1}{2}\left[1 - F(\overline{\theta}_n)\right]\left[E^+(\overline{\theta}_n) - \overline{\theta}_n\right] + \frac{1}{2}\left[1 - F(\underline{\theta}_n)\right]\left[E^+(\underline{\theta}_n) - \underline{\theta}_n\right]}{\frac{1}{2}\left[1 - F(\overline{\theta}_n)\right] + \frac{1}{2}\left[1 - F(\underline{\theta}_n)\right]}$$

where $E^+(\xi) \equiv E[\theta|\theta \geq \xi]$. Using symmetry and direct calculation,

$$\begin{aligned}
\overline{\theta}_{n+1} &= -\left[1 - F(\overline{\theta}_n)\right]\left[E^+(\overline{\theta}_n) - \overline{\theta}_n\right] - \left[1 - F(\underline{\theta}_n)\right]\left[E^+(\underline{\theta}_n) - \underline{\theta}_n\right] \\
&= \overline{\theta}_n - 2F(\overline{\theta}_n)\overline{\theta}_n - \int_{\overline{\theta}_n}^{1} \theta_t dF - \int_{\underline{\theta}_n}^{1} \theta_t dF \\
&= \overline{\theta}_n - 2F(\overline{\theta}_n)\overline{\theta}_n + 2\int_{-1}^{\overline{\theta}_n} \theta_t dF \\
&\leq \overline{\theta}_n .
\end{aligned}$$

The inequality is strict as long as $\overline{\theta}_n > -1$. The same expression yields,

$$\begin{aligned}
\overline{\theta}_{n+1} &= \overline{\theta}_n - 2F(\overline{\theta}_n)\overline{\theta}_n + 2\int_{-1}^{\overline{\theta}_n} \theta_n dF \qquad (6) \\
&\geq \overline{\theta}_n - 2F(\overline{\theta}_n)\overline{\theta}_n - 2F(\overline{\theta}_n) \\
&= \overline{\theta}_n - 2F(\overline{\theta}_n)(1 + \overline{\theta}_n) \\
&\geq \overline{\theta}_n - (1 + \overline{\theta}_n) = -1
\end{aligned}$$

as long as $-1 \leq \overline{\theta}_n \leq 0$ and the inequality is strict as long as $\overline{\theta}_n > -1$. Hence, $\{\overline{\theta}_n\}$ is an decreasing sequence bounded by -1 and must converge. In fact, from (6) the relation $\overline{\theta}_{n+1} = \varphi(\overline{\theta}_n)$ is continuous on $[-1, -1+\varepsilon]$ for some $\varepsilon > 0$, so $\varphi(\overline{\theta}) < \overline{\theta}$ for any $\overline{\theta} > -1$ implies that $\overline{\theta}_n \searrow -1$ as $n \to \infty$. An analogous analysis yields that $\varphi(\underline{\theta}) > \underline{\theta}$ for any $\underline{\theta} < 1$ and that $\underline{\theta}_n \nearrow 1$ as $n \to \infty$.

The impossibility of informational cascade follows immediately since $|\hat{\theta}_n| < 1$, $\forall n \in \mathcal{N}$. Furthermore, it can be readily noted that cascade sets monotonically increase in n and converge to be an attractor in the limit, *i.e.*, $[\overline{\theta}_n, \underline{\theta}_n) \supset [\overline{\theta}_{n-1}, \underline{\theta}_{n-1})$, $\forall n \in \mathcal{N}$ and $[\overline{\theta}_n, \underline{\theta}_n) \to [-1, 1]$ as $n \to \infty$.

We have already observed that when signals are uniformly distributed the cutoff process $\{\hat{\theta}_n\}$ is not stable. We, now, show that the result extends to any $F \in \mathcal{F}$. That is:

PROPOSITION 4. *In $\mathcal{E}'_F$, where $F \in \mathcal{F}$, $\{\hat{\theta}_n\}$ is not convergent.*

Proof. See Appendix. ▯

Having said all these, it turns out that Theorem 3 holds for any $F \in \mathcal{F}$. For completeness let us state:

THEOREM 5 (Learning). *In $\mathcal{E}'_F$, where $F \in \mathcal{F}$,*
(i) Learning is incomplete, i.e., neither an informational cascade nor a limit-cascade arises.
(ii) The cascade set $[\overline{\theta}_n, \underline{\theta}_n)$ is increasing in n and is an attractor in the limit, i.e., $[\overline{\theta}_n, \underline{\theta}_n) \supset [\overline{\theta}_{n-1}, \underline{\theta}_{n-1})$, $\forall n \in \mathcal{N}$ and $[\overline{\theta}_n, \underline{\theta}_n) \to [-1, 1]$ as $n \to \infty$.

Now we are after the generalization of our results on action dynamics. The fact that herd behavior does not arise is an immediate consequence of instability of the cutoff process $\{\hat{\theta}_n\}$. Since the existence of a deviator along the time is an event that occurs with probability 1, action convergence is impossible.

Note that expected length of a potential herd starting from dm n is

$$\mathbb{E}[l_n] = \sum_{k=1}^{\infty} kP(x_n = x_{n+1} = \cdots = x_{n+t-1} \neq x_{n+k}).$$

Without loss of generality, expected length of a potential herd of investment is

$$\mathbb{E}[l_n] = F(\overline{\theta}_{n+1}) + \sum_{k=2}^{\infty} k(1 - F(\overline{\theta}_{n+1}))\cdots(1 - F(\overline{\theta}_{n+k-1}))F(\overline{\theta}_{n+k}).$$

First let us observe that when $F(-1) \neq 0$, $\mathbb{E}[l_n]$ is finite. To see this, suppose that $F(-1) = \delta > 0$

$$\mathbb{E}[l_n] \leq F(\overline{\theta}_{n+1}) + F(\overline{\theta}_{n+1}) \sum_{k=2}^{\infty} k\delta^{k-1}$$

since $F(\overline{\theta}_{n+1}) \geq F(\overline{\theta}_{n+k})$ and $(1 - F(\overline{\theta}_{n+k})) \leq \delta$ for each $k \geq 1$. Moreover a simple computation yields

$$\sum_{k=2}^{\infty} k\delta^{k-1} = \frac{\delta}{(1-\delta)^2}$$

which gives

$$\mathbb{E}[l_n] \leq F(\overline{\theta}_{n+1}) \frac{\delta^2 - \delta + 1}{(1-\delta)^2}.$$

On the other hand, it is not trivial to see that $\mathbb{E}[l_n]$ is unbounded when $F(-1) = 0$. The proof of this claim is in the appendix. Now we are ready to state generalization of Theorem 4.

Theorem 6 (Behavior). *In $\mathcal{E}'_F$, where $F \in \mathcal{F}$,*
(i) Herd behavior does not arise, i.e., $P(l_n = \infty) = 0$ for N arbitrarily large.
(ii) Expected length of the potential herds, $\mathbb{E}[l_n]$, is increasing in n.
(iii) Expected length of a potential herd starting from dm n is unbounded when $F(-1) = 0$, i.e., $\mathbb{E}[l_n] \to \infty$ as $N \to \infty$ $\forall n \in \mathcal{N}$. When $F(-1) \neq 0$, $\mathbb{E}[l_n] < \infty$.

4.2. The asymmetric case

In this part, we assume an imperfect information economy, $\mathcal{E}'_F$, where private signals are distributed with F over a compact support with a convex hull $[a,b]$, such that $\mathbb{E}[\theta]=0$. Let p_n denote the probability that $x_n=1$, *i.e.*, $p_n \equiv P(x_n=1)$. Then,

$$p_n\overline{\theta}_{n+1}+(1-p_n)\underline{\theta}_{n+1}=0$$

for every $n\in\mathcal{N}$ and the state of the dynamic system can be represented by the ordered pair $(p_n,\overline{\theta}_{n+1})$. The law of motion for $(p_n,\overline{\theta}_{n+1})$ is given by

$$p_{n+1}=p_n(1-F(\overline{\theta}_{n+1}))+(1-p_n)(1-F(\underline{\theta}_{n+1}))$$

and

$$\begin{aligned}\overline{\theta}_{n+2} &= -\frac{p_n(1-F(\overline{\theta}_{n+1}))[-\overline{\theta}_{n+1}+E^+(\overline{\theta}_{n+1})]}{p_n(1-F(\overline{\theta}_{n+1}))+(1-p_n)(1-F(\underline{\theta}_{n+1}))}\\ &\quad -\frac{(1-p_n)(1-F(\underline{\theta}_{n+1}))[-\underline{\theta}_{n+1}+E^+(\underline{\theta}_{n+1})]}{p_n(1-F(\overline{\theta}_{n+1}))+(1-p_n)(1-F(\underline{\theta}_{n+1}))}\end{aligned}$$

Notice that,

$$\begin{aligned}p_{n+1}\overline{\theta}_{n+2} &= -p_n(1-F(\overline{\theta}_{n+1}))[-\overline{\theta}_{n+1}+E^+(\overline{\theta}_{n+1})]\\ &\quad -(1-p_n)(1-F(\underline{\theta}_{n+1}))[-\underline{\theta}_{n+1}+E^+(\underline{\theta}_{n+1})]\\ &= -p_n(1-F(\overline{\theta}_{n+1}))[-\overline{\theta}_{n+1}+E^+(\overline{\theta}_{n+1})]\\ &\quad +(1-p_n)(1-F(\underline{\theta}_{n+1}))\left[\frac{p_n\overline{\theta}_{n+1}}{1-p_n}+E^+(\underline{\theta}_{n+1})\right]\\ &= p_n\overline{\theta}_{n+1}+p_n\overline{\theta}_{n+1}(F(\underline{\theta}_{n+1})-1-F(\overline{\theta}_{n+1}))\\ &\quad -p_n\int_{\overline{\theta}_{n+1}}^{b}\theta dF-(1-p_n)\int_{\underline{\theta}_{n+1}}^{b}\theta dF\\ &\leq p_n\overline{\theta}_{n+1}+p_n\overline{\theta}_{n+1}[F(\underline{\theta}_{n+1})-1-F(\overline{\theta}_{n+1})]\\ &\quad +p_n\overline{\theta}_{n+1}F(\overline{\theta}_{n+1})+p_n\overline{\theta}_{n+1}[1-F(\overline{\theta}_{n+1})]\\ &= p_n\overline{\theta}_{n+1}\end{aligned}$$

The inequality follows because

$$\begin{aligned}-p_n\int_{\overline{\theta}_{n+1}}^{b}\theta dF &= p_n\int_{a}^{\overline{\theta}_{n+1}}\theta dF\\ &\leq p_n\overline{\theta}_{n+1}F(\overline{\theta}_{n+1})\end{aligned}$$

and

$$\begin{aligned}-(1-p_n)\int_{\underline{\theta}_{n+1}}^{b}\theta dF &\leq -(1-p_n)\underline{\theta}_{n+1}[1-F(\underline{\theta}_{n+1})]\\ &= p_n\overline{\theta}_{n+1}[1-F(\underline{\theta}_{n+1})]\end{aligned}$$

and the inequality is strict as long as $\overline{\theta}_{n+1} > a$ or $\underline{\theta}_{n+1} < b$.

It can be readily noted that the sequence $\{p_n\overline{\theta}_{n+1}\}$ is bounded. To see this, note that if $\{p_n\overline{\theta}_{n+1}\}$ is unbounded then $\{(1-p_n)\underline{\theta}_{n+1}\}$ is unbounded too and for some finite n

$$\underline{\theta}_i > b, \overline{\theta}_i < a, \forall i > n,$$

which implies that

$$p_{n+1}\overline{\theta}_{n+2} = p_n\overline{\theta}_{n+1},$$

a contradiction.

As to learning and actions dynamics, the boundedness of $\{p_n\overline{\theta}_{n+1}\}$ implies that $\{\overline{\theta}_{n+1}\}$ and $\{\underline{\theta}_{n+1}\}$ must exit the open interval (a, b) in finite time or in the limit as $n \to \infty$ for N arbitrarily large. If this happens in finite time, an informational cascade arises. In this case, convergence of the cutoff process implies convergence of actions. If it happens asymptotically, then $\hat{\theta}_n \in (a, b)$ for all n, and $\overline{\theta}_n \to a$ and $\underline{\theta}_n \to b$ as $n \to \infty$. An immediate corollary of these convergence results is that

$$\lim_{n\to\infty} p_n = \lim_{n\to\infty} \frac{\underline{\theta}_{n+1}}{\underline{\theta}_{n+1} - \overline{\theta}_{n+1}} = \frac{b}{b-a}$$

This means, *ex ante*, a limit dm decides to invest with probability $\frac{b}{b-a}$ and not to invest with probability $\frac{-a}{b-a}$.

Next, we shall give an example to show that an informational cascade may arise when one considers asymmetry. Then, we provide a sufficient condition for the impossibility of informational cascades.

4.2.1. An example

Consider the following structure (See figure 2): suppose private signals are distributed over $[-1, 4]$ with $p.d.f.$

$$f(\theta) = \begin{cases} -3(2\theta+1) & \text{if} \quad -1 \le \theta \le -\frac{1}{2} \\ \frac{1}{81}(2\theta+1) & \text{if} \quad -\frac{1}{2} \le \theta \le 4 \end{cases}$$

Notice that $E[\theta] = 0$. The first dm invests for any $\theta_1 \ge 0$ and the second dm, who observes an investment, invests for any $\theta_2 \ge -E[\theta_1|\theta_1 > 0]$. Note that,

$$\begin{aligned} E[\theta_1|\theta_1 \quad &> \quad 0] = \frac{\int_0^4 (\frac{2}{81}\theta^2 + \frac{1}{81}\theta)d\theta}{\int_0^4 (\frac{2}{81}\theta + \frac{1}{81})d\theta} \\ &= \quad 2.53 \end{aligned}$$

meaning that the second dm invests for any $\theta_2 \ge -2.53$. Hence, an informational cascade occurs.

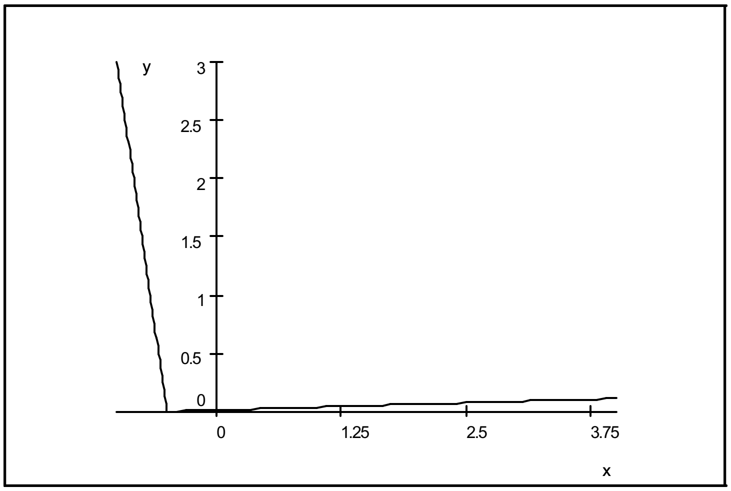

FIG. 2 If the signals are distributed by $f(\theta)$ then when the second dm observes an investment she sets her cutoff to $\hat{\theta}_2 \geq -2.53$. That is, an informational cascade starts.

Note that in the example, the value of investment ranges in the interval $[-N, 4N]$. That is, in case of investment, the loss is relatively smaller than the gain. On the other hand the probability of loss is much higher than the probability of gain. Yet, when the second dm observes an investment she is 'sure' that investment is profitable. There are two interrelated reasons. One is that despite poor possibility of a positive return on investment, an investment by the first dm reveals a big deal of information. The other reason is since possible gain is much higher than possible loss, a risk neutral dm is easily convinced about the profitability of investment.

The asymmetric case, in particular above example may be thought as an investment in a high-tech project. There is a high probability of a failure and a small probability of being extremely successful, therefore signals are drawn from an asymmetric distribution with a longer tail of positive realizations.

4.2.2. A sufficient condition for the impossibility of informational cascades

The sufficient condition which guarantees impossibility of informational cascades is of the same spirit of our example. We state the proposition for only one side since the condition for the other side is similar.

PROPOSITION 5. *Suppose private signals are distributed with continuous p.d.f.* f *over* $[a, b]$ *such that* $\int_a^b \theta dF = 0$. *If* $F(0) < 1/2$ *and* f *is* *increasing on* $[a, 0]$ *then* $\overline{\theta}_n > a$ *for all* n.

Proof. See Appendix. ▯

4.2.3. Herds

In the asymmetric case, although we do not have a full characterization on the possibility of informational cascades we know that informational cascade may arise, and thus a herd consistent with the beliefs in cascade, may arise.

We also know that $\overline{\theta}_n \to a$ and $\underline{\theta}_n \to b$ as $n \to \infty$. Having this, if a limit cascade on one action occurs with positive probability, then instability on the other action is a zero probability event. To see this keep in mind that $p_n\overline{\theta}_{n+1} + (1 - p_n)\underline{\theta}_{n+1} = 0$. Now, for a contradiction suppose that there is limit-cascade on one action say $x = 1$ and not on the other $x = 0$. This means that with probability one there will always be a dm who will deviate to $x = 1$. On the other hand, we know that $x_k = 1$ for all $k \geq n^*$ almost surely for some n^*. Therefore probability of $x_n = 0$ as $n \to \infty$ tends to zero. However, from $p_n\overline{\theta}_{n+1} + (1 - p_n)\underline{\theta}_{n+1} = 0$ we have $1 > \Pr(x_n = 0) > 0$ as $n \to \infty$. A contradiction.

If a limit cascade does not arise then by using the same argument that we used in symmetric case we can show that the expected length of a potential herd is infinity when there is no mass of probability on the boundaries of the signal support.

5. DISCUSSION

Our analysis shows that for $F \in \mathcal{F}$ the outcomes in $\mathcal{E}_F$ and $\mathcal{E}'_F$ are radically different. The dissimilarities are based on two related grounds. First, in $\mathcal{E}_F$ any history of actions is shared as public information by all successors, and thus everyone can infer perfectly what each of her predecessors has learned. On the other hand, in $\mathcal{E}'_F$, all learn from their immediate predecessor's action. As a result, no subset of the history of actions is shared as public information, and thus everyone makes different inferences about what predecessors have learned. Second, while in $\mathcal{E}_F$ the valuable information revealed by the frequency of past actions is available, in $\mathcal{E}'_F$ no one can tell if her predecessor is a deviator or part of a cluster of dms acting alike.

Thus, in $\mathcal{E}'_F$ one's Bayesian inference induces a probability measure over all possible histories conditional on her immediate predecessor's action. Evidently, the information embedded in the history is suppressed in a way that a significant weight is given to the event in which all predecessors acted as the immediate predecessor. As a result, dms' attempts to capture the content of all predecessors' signals by using their immediate predecessor's action lead to a severe failure in internalizing the information externality.

As to learning dynamics, in $\mathcal{E}_F$, the stability of the cutoff process near either -1 or 1 assures that learning is complete. That is, eventually history provides everyone a decisive lesson about the realization of the decision-relevant state of the world, and hence the presence of a limit-cascade follows. However, our puzzling result is that in $\mathcal{E}'_F$ eventually no one is self-reliant even though cutoffs diverge. More precisely, although the cutoff process is unstable nearby any of its fixed points, and thus learning is forever incomplete, the limit dm acts irrespective of her private information. Thus, learning has a gripping character; even though a predecessor's action never provides a decisive lesson, eventually decisions are based exclusively upon it.

In a nutshell, in $\mathcal{E}_F$ learning converges to an informational pooling equilibrium in which the complete history of actions provides a decisive guide regarding to what the state of the world is, and thus eventually all ignore their signals. Conversely, in $\mathcal{E}'_F$, an informational pooling equilibrium never exists. That is, learning is everlastingly flanked by conviction that the state is exceedingly *high* and *low*. However, although a predecessor's action is never a decisive guide, in the limit dms rely exclusively on it.

Regarding the action dynamics, while in $\mathcal{E}_F$ herd arises in $\mathcal{E}'_F$ the standard herding outcome is impossible basically since learning forever cycles between confidence in contrary states of the world. Nonetheless, since the cutoff process tends to an informational outcome in which all base their decisions upon their predecessor's action, imitation behavior becomes more likely along the line of dms. Moreover, since in due course learning overturns, most likely that the action consistent with the innovative learning will turned out to be potentially herded upon. Thus, action dynamics exhibit monotonically longer lasting potential herds.

The behavior in $\mathcal{E}_F$ and $\mathcal{E}'_F$ is considerably different also in the medium-run. According to the overturning principle, consider a potential herd followed by a deviator. In both $\mathcal{E}_F$ and $\mathcal{E}'_F$, the deviator becomes a leader to her successors. Yet, there is substantial difference. In $\mathcal{E}_F$, the deviator can be identified since previous actions are publicly known. As a result, her deviation reveals clear cut information regarding her private signal that basically meagerly dominates the accumulated public information. Thus, her successor will be slightly in favor of joining the deviation. On the other hand, in $\mathcal{E}'_F$, one can not tell whether her predecessor is an imitator or a deviator. Thus, a deviator's action is her successor's statistic to infer the entire history of actions. Consequently, one who is subsequent to a deviator would be very enthusiastic to join the deviation.

Finally, let us argue that the behavior in $\mathcal{E}'_F$ deserves attention since socioeconomic behavior more typically exhibits long-lasting but finite episodes of mass behavior. For instance, our result captures what makes a fad to start suddenly, to expire rather easily, and to being replaced by another fad.

A natural question is the robustness of the results when the number of

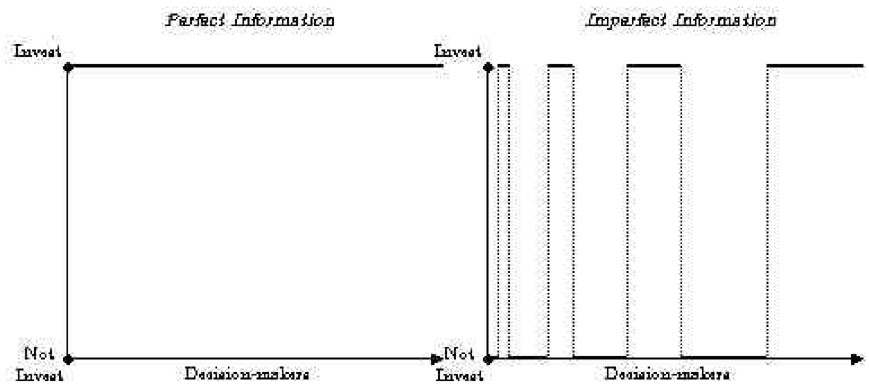

FIG. 3 A computer simulation with 50,000 *dm*s. Under perfect information all the *dm*s are in a herd of investment. Under imperfect information, with the same vector of signals the behavior perpetuates between investment and not investment.

most recent actions that a *dm* observes exceeds one *i.e.,* any *dm* n observes $x_{n-1}, x_{n-2}, \cdots, x_{n-k}$ whenever available. Our analysis falls short of properly addressing this issue since for any observation of histories larger than one the recursive structure of the cutoff dynamics is extremely involved. However, some key insights can be rapidly noted. Evidently, for any finite k-predecessor histories, any *dm* n can observe $2^{\min\{n,k\}-1}$ possible histories of actions. The cutoff rule becomes richer since further inference based on the frequency of past actions can be acquired. More specifically, *dm*s are able to identify the deviators and imitators, and the information revealed by a deviation is incorporated in their decision rule. Since the amount of the mentioned information is increasing in k, the magnitude of the overturning principle varies with k. That is, a successor of a deviator is still more inclined to follow the deviation but less enthusiastic as k increases. However, whether increasing the number of predecessor observed would lead to sharply different results is not clear, since all the decision rules would change to reflect the new environment.

6. RELATED LITERATURE

Observation of the immediate predecessor's action is a particular form of imperfect information that captures the idea that *dm*s are able to observe the actions of the most recent *dm*s and to identify who took which action. Another approach that formulizes imperfect information is in the form of observing unordered random samples from the history. What follows is a discussion to underline the similarities and the dissimilarities between the present paper and the random sampling approach.

An observational learning model in which *dm*s observe an unordered random sample of actions from the history has been studied by SSa. Their approach is in fact the finite *dm* counterpart of Banerjee and Fudenberg (1995) (BF)'s continuum-*dm* model. BF presuppose a continuum of *dm*s where at each period a cohort of a positive measure is replaced by a new generation of the same measure. Among other results, they show that even with bounded private beliefs for a sample size larger than one, complete and correct learning[3] occurs. The continuum of *dm*s approach is criticized by SSa on the basis that it produces a fundamentally different learning dynamics than a discrete *dm* sequential entry setup. In particular, in a continuum-*dm* model all the information is available in the first period. While with unbounded private beliefs SSa corroborate with BF that complete and correct learning may arise, with bounded beliefs they show that in their model it is impossible. SSa provide a thorough characterization for the case of unbounded signals, but their results are not exhaustive for the case of bounded signals. In particular, they show that with bounded signals, learning might be incorrect.

Unlike the approach taken by SSa, we assume a particular sampling mechanism in which each *dm* samples her immediate predecessor with probability one. It captures, although in an extreme manner, the idea that more recent predecessors are more likely to be observed. Aside from modeling choices, our analysis differs from the analysis of SSa in several respects. First, we describe the learning process even when learning is incomplete. Second, we show that the behavior of the model is radically different under perfect and imperfect information. Third, as was pointed out by Fudenberg and Levine (1998), most studies focus on asymptotic outcomes, which tend to be easier to obtain. Yet, undoubtedly the medium-run properties of learning models are worth study. In contrast to the existing literature, we are able to describe not only the asymptotic outcomes but also the medium run dynamics of the model. Finally, from a technical point of view, our model differs from SSa in that given the state of the world, as opposed to SSa signals are assumed to be conditionally dependent hence the results are not directly comparable.

In another independent study, Smith and Sørensen (1997) study an example of a social learning model where each *dm* observes her immediate predecessor. However, their signal distribution assumes unbounded beliefs and their focus is on different properties of learning.

[3] Although learning is complete, correct learning is criticized by SSa since initial conditions are given arbitrarily rather than informatively produced.

7. CONCLUDING REMARKS

Our results under imperfect information are noticeably different from those under perfect information mainly since the learning process exhibits divergence. However, we were able to provide a clear-cut description of the long-run learning and behavior outcomes. Moreover we described the noteworthy medium-run features of the model. We showed that even if everyone learns only from the immediate predecessors' decision, and thus learning is forever incomplete, over time, *dm*s tend to rely more on the information revealed by the predecessor's action rather than their private signal. Consequently, potential herds are expected to form and are expectedly longer lasting (Theorem 5 and Theorem 6). Hence we made a case that imperfect information is more applicable in providing a theoretical prediction of fads and fashions.

Recently, a series of experimental studies, such as Anderson and Holt (1997), test discrete-signal observational learning models under perfect information. They find that subjects frequently acts against their private signals when history dictates otherwise. However, the experimental literature still falls short in providing tests of continuous signal space and imperfect information observational learning models. In recent experimental projects (Çelen and Kariv (2001a, 2001b)), we test our present model under perfect and imperfect information.

Obviously, different information structures may lead to different outcomes. This remains a subject for further theoretical and experimental research on observational learning.

APPENDIX A: APPENDIX

A.1. Auxiliary results

Denote,

$$\begin{aligned}\overline{q}_n &\equiv P(x_n = 1 \mid x_{n+1} = 1)\\ \underline{q}_n &\equiv P(x_n = 1 \mid x_{n+1} = 0)\end{aligned}$$

and

$$p_n \equiv P(x_n = 1)$$

Recall that,

$$\begin{aligned}\overline{\theta}_n &\equiv -\mathbb{E}\left[\sum_{i=1}^{n-1} \theta_i \mid x_{n-1} = 1\right]\\ \underline{\theta}_n &\equiv -\mathbb{E}\left[\sum_{i=1}^{n-1} \theta_i \mid x_{n-1} = 0\right]\end{aligned}$$

Rewriting,

$$\begin{aligned}\overline{\theta}_n &= -\mathbb{E}\left[\sum\nolimits_{i=1}^{n-1}\theta_i \mid \theta_{n-1} \geq -\mathbb{E}[\sum\nolimits_{i=1}^{n-2}\theta_i \mid x_{n-2}]\right] \\ &= \overline{q}_{n-2}\left(\frac{\overline{\theta}_{n-1}-1}{2}\right) + (1-\overline{q}_{n-2})\left(\frac{\underline{\theta}_{n-1}-1}{2}\right) \qquad \text{(A)}\end{aligned}$$

and similarly,

$$\begin{aligned}\underline{\theta}_n &= -\mathbb{E}\left[\sum\nolimits_{i=1}^{n-1}\theta_i \mid \theta_{n-1} < -\mathbb{E}[\sum\nolimits_{i=1}^{n-2}\theta_i \mid x_{n-2}]\right] \\ &= \underline{q}_{n-2}\left(\frac{\overline{\theta}_{n-1}+1}{2}\right) + (1-\underline{q}_{n-2})\left(\frac{\underline{\theta}_{n-1}+1}{2}\right) \qquad \text{(B)}\end{aligned}$$

In addition, by Bayes' rule we get $\overline{q}_n$ and $\underline{q}_n$ as,

$$\begin{aligned}\overline{q}_n &= \frac{P(x_{n+1}=1 \mid x_n=1)P(x_n=1)}{\sum\limits_{i=0}^{1} P(x_{n+1}=1 \mid x_n=i)P(x_n=i)} \\ &= \frac{(1-\overline{\theta}_{n+1})p_n}{p_n(\underline{\theta}_{n+1}-\overline{\theta}_{n+1}) + (1-\underline{\theta}_{n+1})} \qquad \text{(C)}\end{aligned}$$

and,

$$\begin{aligned}\underline{q}_n &= \frac{P(x_{n+1}=0 \mid x_n=1)P(x_n=1)}{\sum\limits_{i=0}^{1} P(x_{n+1}=0 \mid x_n=i)P(x_n=i)} \\ &= \frac{(1+\overline{\theta}_{n+1})p_n}{p_n(\overline{\theta}_{n+1}-\underline{\theta}_{n+1}) + (1+\underline{\theta}_{n+1})} \qquad \text{(D)}\end{aligned}$$

Fact 1. If $p_n = \frac{1}{2}$ and $\overline{\theta}_{n+1} + \underline{\theta}_{n+1} = 0$ then $\overline{q}_n + \underline{q}_n = 1$, $\forall n \in \mathcal{N}$.

Proof. Suppose $p_n = \frac{1}{2}$ and $\overline{\theta}_{n+1} + \underline{\theta}_{n+1} = 0$. By (C) and (D),

$$\overline{q}_n + \underline{q}_n = \frac{(1-\overline{\theta}_{n+1})p_n}{p_n(\underline{\theta}_{n+1}-\overline{\theta}_{n+1}) + (1-\underline{\theta}_{n+1})} + \frac{(1+\overline{\theta}_{n+1})p_n}{p_n(\overline{\theta}_{n+1}-\underline{\theta}_{n+1}) + (1+\underline{\theta}_{n+1})}$$

substituting $p_n = \frac{1}{2}$ and $\overline{\theta}_{n+1} = -\underline{\theta}_{n+1}$ yields the result,

$$\begin{aligned}\overline{q}_n + \underline{q}_n &= \frac{(1-\overline{\theta}_{n+1})\frac{1}{2}}{\frac{1}{2}(-2\overline{\theta}_{n+1}) + (1+\overline{\theta}_{n+1})} + \frac{(1+\overline{\theta}_{n+1})\frac{1}{2}}{\frac{1}{2}(2\overline{\theta}_{n+1}) + (1-\overline{\theta}_{n+1})} \\ &= (1-\overline{\theta}_{n+1})\frac{1}{2} + (1+\overline{\theta}_{n+1})\frac{1}{2} = 1.\end{aligned}$$

∎

Fact 2. If $\overline{q}_n + \underline{q}_n = 1$ and $\overline{\theta}_{n+1} + \underline{\theta}_{n+1} = 0$ then $\overline{\theta}_{n+2} + \underline{\theta}_{n+2} = 0$, $\forall n \in \mathcal{N}$.

Proof. Suppose $\overline{q}_n + \underline{q}_n = 1$ and $\overline{\theta}_{n+1} + \underline{\theta}_{n+1} = 0$. By (A) and (B),

$$\begin{aligned}
\overline{\theta}_{n+2} + \underline{\theta}_{n+2} &= \overline{q}_n \left(\frac{\overline{\theta}_{n+1} - 1}{2} \right) + (1 - \overline{q}_n) \left(\frac{\underline{\theta}_{n+1} - 1}{2} \right) \\
&\quad + \underline{q}_n \left(\frac{\overline{\theta}_{n+1} + 1}{2} \right) + (1 - \underline{q}_n) \left(\frac{\underline{\theta}_{n+1} + 1}{2} \right) \\
&= \overline{\theta}_{n+1} \left(\frac{\overline{q}_n + \underline{q}_n}{2} \right) + \underline{\theta}_{n+1} \left(1 - \frac{\overline{q}_n + \underline{q}_n}{2} \right)
\end{aligned}$$

and substituting $\overline{q}_n + \underline{q}_n = 1$ yields the result,

$$\overline{\theta}_{n+2} + \underline{\theta}_{n+2} = \frac{1}{2} \left(\overline{\theta}_{n+1} + \underline{\theta}_{n+1} \right) = 0.$$

∎

Fact 3. If $p_n = \frac{1}{2}$ and $\overline{\theta}_{n+1} + \underline{\theta}_{n+1} = 0$ then $p_{n+1} = \frac{1}{2}$, $\forall n \in \mathcal{N}$.

Proof. Suppose $p_n = \frac{1}{2}$ and $\overline{\theta}_{n+1} + \underline{\theta}_{n+1} = 0$. Note that the probability of observing action $x_{n+1} = 1$ following the action $x_n = 1$ or $x_n = 0$ is $\frac{1 - \overline{\theta}_{n+1}}{2}$ or $\frac{1 - \underline{\theta}_{n+1}}{2}$ respectively. Therefore,

$$p_{n+1} = p_n \frac{1 - \overline{\theta}_{n+1}}{2} + (1 - p_n) \frac{1 - \underline{\theta}_{n+1}}{2}$$

substituting $p_n = \frac{1}{2}$ and $\overline{\theta}_{n+1} = -\underline{\theta}_{n+1}$ yields the result, $p_{n+1} = \frac{1}{2}$. ∎

A.2. Proofs of Lemmata 1-3

Proof of Lemma 1. We prove by induction.

For $n = 1$, a straightforward computation yields $\overline{q}_1 + \underline{q}_1 = 1$.

Suppose the hypothesis holds for all $n \leq k - 1$.

For $n = k$, by induction on Fact 2 using the induction hypothesis and $\overline{\theta}_1 = \underline{\theta}_1 = 0$,

$$\overline{\theta}_{n+1} + \underline{\theta}_{n+1} = 0, \ \forall n = 1, \cdots, k \qquad (*)$$

By induction on Fact 3 using equation $(*)$ and $p_1 = \frac{1}{2}$,

$$p_{n+1} = \frac{1}{2}, \ \forall n = 1, \cdots, k \qquad (**)$$

By Fact 1, equations $(*)$ and $(**)$ yield the result. ∎

Proof of Lemma 2. We prove by induction.

For $n = 1$, $\overline{\theta}_1 = \underline{\theta}_1 = 0$ trivially holds.

Suppose the hypothesis holds for all $n \leq k - 1$.

For $n = k$, by induction on Fact 3 using the induction hypothesis and $p_1 = \frac{1}{2}$,

$$p_{n-1} = \frac{1}{2},\ \forall n = 1, \cdots, k \tag{$*$}$$

By induction on Fact 1 using equation ($*$) and the induction hypothesis,

$$\overline{q}_{n-2} + \underline{q}_{n-2} = 1,\ \forall n = 1, \cdots, k \tag{$**$}$$

Equation ($**$) and the induction hypothesis yield the result by Fact 2. ∎

Proof of Lemma 3. We prove by induction.

For $n = 1$, we readily get $p_1 = \frac{1}{2}$ since $\overline{\theta}_1 = \underline{\theta}_1 = 0$. First notice that a simple calculation yields $\overline{\theta}_2 + \underline{\theta}_2 = 0$.

Suppose the hypothesis holds for all $n \leq k - 1$.

For $n = k$, by induction on Fact 1 and Fact 2 recursively using the induction hypothesis and $\overline{\theta}_2 + \underline{\theta}_2 = 0$,

$$\overline{q}_{n-1} + \underline{q}_{n-1} = 1,\ \forall n = 1, \cdots, k \tag{$*$}$$

Again by induction on Fact 2 using equation ($*$) and $\overline{\theta}_2 + \underline{\theta}_2 = 0$ we get,

$$\overline{\theta}_{n+1} + \underline{\theta}_{n+1} = 0,\ \forall n = 1, \cdots, k \tag{$**$}$$

Equation ($**$) and the induction hypothesis yield the result by Fact 3. ∎

A.3. Proof of Proposition 3

Proof. We have already observed that,

$$\overline{\theta}_n = \overline{q}_{n-2}\left(\frac{\overline{\theta}_{n-1} - 1}{2}\right) + (1 - \overline{q}_{n-2})\left(\frac{\underline{\theta}_{n-1} - 1}{2}\right)$$

By Lemma 2,

$$\begin{aligned} \overline{\theta}_n &= \overline{q}_{n-2}\left(\frac{\overline{\theta}_{n-1} - 1}{2}\right) + (1 - \overline{q}_{n-2})\left(\frac{-\overline{\theta}_{n-1} - 1}{2}\right) \\ &= \frac{2\overline{q}_{n-2}\overline{\theta}_{n-1} - \overline{\theta}_{n-1} - 1}{2} \end{aligned} \tag{$*$}$$

We also know that,

$$\overline{q}_{n-2} = \frac{(1 - \overline{\theta}_{n-1})p_{n-2}}{p_{n-2}(\underline{\theta}_{n-1} - \overline{\theta}_{n-1}) + (1 - \underline{\theta}_{n-1})}$$

which by Lemma 2 and Lemma 3 simplifies to,

$$\overline{q}_{n-2} = \frac{(1 - \overline{\theta}_{n-1})\frac{1}{2}}{\frac{1}{2}(-2\overline{\theta}_{n-1}) + (1 + \overline{\theta}_{n-1})} = \frac{(1 - \overline{\theta}_{n-1})}{2} \tag{$**$}$$

then substituting equation $(**)$ in equation $(*)$, we obtain the result,

$$\overline{\theta}_n = -\frac{1+\overline{\theta}^2_{n-1}}{2}$$

Analogous analysis yields,

$$\underline{\theta}_n = \frac{1+\underline{\theta}^2_{n-1}}{2}.$$

▯

A.4. Proof of Corollary 5

Proof. In order to show that $\{\hat{\theta}_n\}$ does not converge, we show that for any $k \in \mathcal{N}$, $\prod_{n=k}^{\infty} \frac{1-\bar{\theta}_n}{2} = 0$. That is, the probability that all, $WLOG$, invest after any finite dm is zero. Since the same argument applies to any k, we show only for $n = 1$. Note that, $\prod_{n=1}^{\infty} \frac{1-\bar{\theta}_n}{2} = 0$ if and only if $\sum_{n=1}^{\infty} \left(\bar{\theta}_n + 1\right)$ does not converge. By comparison test we will show that

$$\sum\nolimits_{n=1}^{\infty} \left(\bar{\theta}_n + 1\right) \geq \sum\nolimits_{n=1}^{\infty} \frac{1}{n}$$

By induction we show $\left(\bar{\theta}_n + 1\right) \geq \frac{1}{n}$, $\forall n \geq 1$.
It trivially holds for $n = 1$.
Suppose that it holds for $n = k$, that is

$$\bar{\theta}_k \geq \frac{1-k}{k} \qquad (*)$$

Now we show that $\bar{\theta}_{k+1} \geq -\frac{k}{k+1}$. Recall that $\bar{\theta}_{k+1} = -\frac{1+\bar{\theta}_k^2}{2}$, so by equation $(*)$ we have

$$\frac{1+\bar{\theta}_k^2}{2} \leq \frac{1+(\frac{1-k}{k})^2}{2}$$

and we need to show that

$$\frac{1+\bar{\theta}_k^2}{2} \leq \frac{k}{k+1}$$

Hence, it is sufficient to show that

$$\frac{1+(\frac{1-k}{k})^2}{2} \leq \frac{k}{k+1}$$

but it follows immediately from a simple algebra,

$$\begin{aligned} (k+1)(k-1)^2 &\leq (k-1)k^2 \\ k^2 - 1 &\leq k^2. \end{aligned}$$

▯

A.5. Proof of Theorem 4 ii-iii

Proof. First, notice that

$$\begin{aligned}
\mathbb{E}[l_n] &= \sum_{k=1}^{\infty} kP(x_n = x_{n+1} = \cdots = x_{n+t-1} \neq x_{n+k}) \\
&= \sum_{k=1}^{\infty} k \frac{1-\mid \hat{\theta}_{n+k} \mid}{2} \left\{ \prod_{m=1}^{k-1} \frac{1+\mid \hat{\theta}_{n+m} \mid}{2} \right\}
\end{aligned}$$

where, by convention, $\prod_{m=1}^{0} \frac{1+|\hat{\theta}_{n+m}|}{2} = 1$.

Clearly, since over time *dm*s become increasingly likely to imitate their predecessor, $\mathbb{E}[l_n]$ is increasing in n. Thus, it would be enough to show that the sequence $\mathbb{E}[l_1]$ does not converge. To this end, we use comparison test for which the divergent sequence is $\sum_{n=1}^{\infty} \frac{1}{n}$ and we prove by induction that $n\frac{1+\bar{\theta}_{1+n}}{2} \left\{ \prod_{m=1}^{n-1} \frac{1-\bar{\theta}_{1+n}}{2} \right\} \geq \frac{1}{n}, \forall n \geq 4$.

For $n = 4$, a trivial computation yields the result.

Suppose that the hypothesis holds for $n = k-1$. Then,

$$\begin{aligned}
(k-1)\frac{1+\bar{\theta}_k}{2} \left(\frac{1-\bar{\theta}_{k-1}}{2} \frac{1-\bar{\theta}_{k-2}}{2} \cdots \frac{1-\bar{\theta}_2}{2} \right) &\geq \frac{1}{k-1} \\
\left(\frac{1-\bar{\theta}_{k-1}}{2} \frac{1-\bar{\theta}_{k-2}}{2} \cdots \frac{1-\bar{\theta}_2}{2} \right) &\geq \frac{2}{(k-1)^2} \frac{1}{1+\bar{\theta}_k}
\end{aligned}$$

Multiplying both sides by $k\frac{1+\bar{\theta}_{k+1}}{2} \frac{1-\bar{\theta}_k}{2}$,

$$k\frac{1+\bar{\theta}_{k+1}}{2} \left(\frac{1-\bar{\theta}_k}{2} \frac{1-\bar{\theta}_{k-1}}{2} \cdots \frac{1-\bar{\theta}_2}{2} \right) \geq (1+\bar{\theta}_{k+1}) \frac{1-\bar{\theta}_k}{1+\bar{\theta}_k} \frac{k}{2(k-1)^2}$$

Thus, it is sufficient to show that $(1+\bar{\theta}_{k+1})\frac{1-\bar{\theta}_k}{1+\bar{\theta}_k} \frac{k}{2(k-1)^2} \geq \frac{1}{k}$. Now, recall that $\bar{\theta}_{k+1} = -\frac{1+\bar{\theta}_k^2}{2}$. Substituting,

$$\begin{aligned}
(1 - \frac{1+\bar{\theta}_k^2}{2}) \frac{1-\bar{\theta}_k}{1+\bar{\theta}_k} \frac{k}{2(k-1)^2} &\geq \frac{1}{k} \\
\frac{1-\bar{\theta}_k^2}{2} \frac{1-\bar{\theta}_k}{1+\bar{\theta}_k} \frac{k}{2(k-1)^2} &\geq \frac{1}{k} \\
\left(\frac{1-\bar{\theta}_k}{2} \right)^2 &\geq \left(\frac{k-1}{k} \right)^2
\end{aligned}$$

Hence, the problem reduces to show that

$$\begin{aligned}\frac{1-\bar{\theta}_k}{2} &\geq \frac{k-1}{k}\\ -\bar{\theta}_k &\geq 1-\frac{2}{k}\\ \underline{\theta}_k &\geq 1-\frac{2}{k}\end{aligned}$$

By induction this can be shown to hold for each $k \geq 1$.

For $k = 1$ it is trivial.

Suppose that it holds for $k = m-1$, *i.e.*, $\underline{\theta}_{m-1} \geq 1 - \frac{2}{m-1}$.

For $k = m$, we have to show that $\underline{\theta}_m \geq 1 - \frac{2}{m}$. But since $\underline{\theta}_m = \frac{1+\underline{\theta}^2_{m-1}}{2}$, the question reduces to

$$\begin{aligned}\frac{1+\underline{\theta}^2_{m-1}}{2} &\geq 1-\frac{2}{m}\\ \underline{\theta}^2_{m-1} &\geq 1-\frac{4}{m}\end{aligned}$$

By the induction step, $\underline{\theta}^2_{m-1} \geq \left(1 - \frac{2}{m-1}\right)^2$. Hence, it is sufficient to show that $\left(1 - \frac{2}{m-1}\right)^2 \geq 1 - \frac{4}{m}$. But re-arranging,

$$\begin{aligned}(m^2-6m+9)m &\geq (m^2-5m+4)(m-1)\\ 0 &\geq -4.\end{aligned}$$

$\square$

A.6. Proof of Proposition 4

Proof. By symmetry, it will be enough to show that the cutoff process is unstable near one of the fixed points. Without loss of generality we show that it is unstable near -1. That is for any $k < \infty$, $\prod_{n=k}^{\infty}(1 - F(\overline{\theta}_n)) = 0$. Equivalently $\sum_{n=k}^{\infty} F(\overline{\theta}_n)$ does not converge.

First note that if $F(-1) \neq 0$, *i.e.* there is a probability mass on the boundaries, then since there is always a positive probability of switching $\prod_{n=k}^{\infty}(1 - F(\overline{\theta}_n)) = 0$ holds trivially. So we restrict attention to distributions $F \in \mathcal{F}$ such that $F(-1) = 0$.

The law of motion of $\overline{\theta}_n$ is given by (6)

$$\overline{\theta}_{n+1} = \overline{\theta}_n - 2F(\overline{\theta}_n)\overline{\theta}_n + 2\int_{-1}^{\overline{\theta}_n} \theta dF$$

Let us write $x_n = 1 + \overline{\theta}_n$, then

$$\begin{aligned} x_{n+1} &= x_n - 2F(\overline{\theta}_n)x_n + 2\int_{-1}^{\overline{\theta}_n}(1+\theta)dF \\ \frac{x_{n+1}}{x_n} &\geq 1 - 2F(\overline{\theta}_n) \end{aligned}$$

Since $\prod_{n=k}^{\infty}\frac{x_{n+1}}{x_n} = 0$, $\sum_{n=k}^{\infty}F(\overline{\theta}_n)$ does not converge, hence $\prod_{n=k}^{\infty}(1 - F(\overline{\theta}_n)) = 0$. ∎

A.7. Proof of Theorem 6 iii

Proof. We show, by induction, that $\mathbb{E}[l_n]$ is unbounded when $F(-1) = 0$. Wlog we show for a potential herd of investment. Let us write $F_k = F(\overline{\theta}_k)$. Now suppose that

$$(k-1)F_k(1-F_{k-1})(1-F_{k-2})\cdots(1-F_1) \geq \frac{1}{k-1}$$

We need to show that

$$kF_{k+1}(1-F_k)(1-F_{k-1})(1-F_{k-2})\cdots(1-F_1) \geq \frac{1}{k}$$

Multiply both sides of the induction hypothesis by $kF_{k+1}(1-F_k)$. We get

$$kF_{k+1}(1-F_k)(1-F_{k-1})(1-F_{k-2})\cdots(1-F_1) \geq \frac{k}{(k-1)^2}\frac{(1-F_k)}{F_k}F_{k+1}$$

So, we need to show that

$$\frac{k}{(k-1)^2}\frac{(1-F_k)}{F_k}F_{k+1} \geq \frac{1}{k}$$

Recall from the divergence of cutoffs $\sum_{n=k}^{\infty}F(\overline{\theta}_n)$ is divergent so that $F_k \geq \frac{1}{k}$ for all k.

It will be enough to show that

$$\begin{aligned} \frac{k}{(k-1)^2}\frac{(1-F_k)}{F_k}F_{k+1} &\geq \frac{1}{k-1} \\ \frac{k}{k-1}\frac{(1-F_k)}{F_k}F_{k+1} &\geq 1 \end{aligned}$$

Thus,

$$\frac{k}{k^2-1} \geq \frac{F_k}{(1-F_k)}$$

Rearranging,

$$F_k \geq \frac{k}{k^2+k-1}$$

Hence, it is sufficient to show that

$$\frac{1}{k} \geq \frac{k}{k^2+k-1}$$

but, indeed

$$1 \geq \frac{k^2}{k^2+k-1}$$

∎

A.8. Proof of Proposition 5

Proof. We will show that the cutoff $\tilde{\theta}_n = -\mathbb{E}[\Theta|x_1 = x_2 = \cdots = x_{n-1} = 1] > a$ since $\tilde{\theta}_n < \bar{\theta}_n$ $\forall n$.

Note that from (2) we have

$$\begin{aligned} -\tilde{\theta}_{n+1} &= -\tilde{\theta}_n + \frac{\int_{\tilde{\theta}_n}^{b} \theta dF}{1-F(\tilde{\theta}_n)} \\ &= -\tilde{\theta}_n - \frac{\int_{a}^{\tilde{\theta}_n} \theta dF}{1-F(\tilde{\theta}_n)} \\ &\leq -\tilde{\theta}_n - a\frac{F(\tilde{\theta}_n)}{1-F(\tilde{\theta}_n)} \end{aligned}$$

Now since $F(0) < 1/2$, we have

$$\begin{aligned} \frac{F(0)}{1-F(0)} &< 1 \\ \frac{F(0)}{-a-0} &< \frac{1-F(0)}{-a} \end{aligned}$$

and that $\frac{F(\tilde{\theta}_n)}{\tilde{\theta}_n - a} < \frac{1-F(\tilde{\theta}_n)}{-a}$ at each $\tilde{\theta}_n \in [a,0]$ for f increasing in $[a,0]$. To see this observe that

$$\frac{\partial \frac{F(x)}{-a+x}}{\partial x} = \frac{F'(x)(-a+x) - F(x)}{(-a-x)^2}$$

but if f is increasing in $[a,0]$ we have

$$F'(x) > \frac{F(x)}{(-a+x)}$$

yielding

$$\frac{\partial \frac{F(x)}{-a+x}}{\partial x} > 0$$

Also, observe that
$$\frac{\partial \frac{1-F(x)}{-a}}{\partial(x)} = \frac{F'(x)}{a} < 0$$
Hence, the inequality holds at $x = 0$ and as x gets smaller the LHS gets smaller and the RHS gets larger, meaning it holds for all $x \in [a, 0]$.

Since $\frac{F(\tilde{\theta}_n)}{\tilde{\theta}_n - a} < \frac{1-F(\tilde{\theta}_n)}{-a}$ for all n, we have
$$\begin{aligned} \frac{F(\tilde{\theta}_n)}{1-F(\tilde{\theta}_n)} &< \frac{\tilde{\theta}_n - a}{-a} \\ -\tilde{\theta}_n - a\frac{F(\tilde{\theta}_n)}{1-F(\tilde{\theta}_n)} &< -a \\ -\tilde{\theta}_{n+1} &< -a \end{aligned}$$
for all n. ∎

REFERENCES

Anderson, L and Holt, C (1997) "Information Cascades in the Library", *American Economic Review,* **87,** 847-862.

Banerjee, A. (1992) "A Simple Model of Herd Behavior", *Quarterly Journal of Economics* **107**, 797-817.

Banerjee, A. and D. Fudenberg (1995) "Word-of-Mouth Learning", MIT mimeo.

Bikhchandani, S., D. Hirshleifer and I. Welch (1992) "A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascade", *Journal of Political Economy* **100**, 992-1026.

Çelen, B. and S. Kariv (2001a) "Distinguishing Informational Cascades from Herd Behavior in the Laboratory", mimeo.

Çelen, B. and S. Kariv (2001b) "An Experimental Test of Observational Learning under Imperfect Information", mimeo.

Fudenberg, D. and D. K. Levine (1998), The Theory of Learning in Games, The MIT Press, Cambridge.

Gale, D. (1996) "What Have We Learned from Social Learning?", *European Economic Review* **40**, 617-628.

Smith, L. and P. Sørensen (1996) "Rational Social Learning by Random Sampling", mimeo.

Smith, L. and P. Sørensen (1997) "Martingales, Momentum or Mean Revision? On Bayesian Conventional Wisdom", Nuffield College, mimeo.

Smith, L. and P. Sørensen (2000) "Pathological Outcomes of Observational Learning", *Econometrica* **68**, 371-398.