Observational Learning Under Imperfect Information*

Boğaçhan Çelen† New York University

Shachar Kariv‡ New York University

January 23, 2003

Abstract

The analysis explores 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, in contrast, beliefs and actions cycle forever. However, despite the 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 increasingly rare switches. These results suggest that imperfect information premise provides a better theoretical description of fads and fashions. (*JEL* D82, D83).

*We are grateful to Douglas Gale for his guidance, to an Associate Editor and an anonymous referee for their comments and to William J. Baumol who read carefully through the manuscript, and made invaluable suggestions on exposition. We also acknowledge helpful discussions of Jean-Pierre Benoît, Alberto Bisin, Colin F. Camerer, Andrew Caplin, Amil Dasgupta, Eric S. Maskin, Roger B. Myerson, Benjamin Polak, Matthew Rabin, Debraj Ray, Andrew Schotter, Peter Sørensen, S. R. Srinivasa Varadhan and the participants of seminars at Cornell University, CREED, Harvard University, New York University, University of Cologne, Yale University and 2001 South-East Economic Theory & International Economics Conference.

†Department of Economics, New York University, 269 Mercer St., 7th Floor, New York, NY, 10003 (e-mail: bc319@nyu.edu, url: http://home.nyu.edu/~bc319).

‡Department of Economics, New York University, 269 Mercer St., 7th Floor, New York, NY, 10003 (e-mail: sk510@nyu.edu, url: http://home.nyu.edu/~sk510).

1 Introduction

In the last decade a number of studies have explored the process of observational learning. Banerjee (1992) and Bikhchandani, Hirshleifer and Welch (1992) introduced the basic concepts and stimulated further research in this area. The literature analyzes an economy where a sequence of Bayesian decision-makers (*dm*s) make a once-in-a-lifetime decision under incomplete and asymmetric information. The typical conclusion is that, despite asymmetry of information, eventually *dm*s will imitate their predecessor's behavior even if it conflicts with their private information.

A central assumption of the previous 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 entire history of actions that have been taken before them. The *dm* thus compares her information with that of a large (in the limit, unboundedly large) number of other *dm*s. In reality, *dm*s have *imperfect information.* For greater realism, our model relaxes the perfect-information assumption, dealing instead with the case in which each *dm* observes only her immediate predecessor's decision. Our goal is to understand behavior under such an imperfect information structure.

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 zero payoff, the payoff from investment is a random variable with expected value zero. Each *dm* receives an informative private signal and observes only her immediate predecessor. We describe the *dm*s' optimal strategies recursively; they in turn, characterize the dynamics of learning and actions.

Smith and Sørensen (2000) make a clear distinction between learning dynamics and action dynamics. They emphasize the difference between *informational cascades* and *herd behavior*, two notions introduced by Banerjee (1992) and Bikhchandani, Hirshleifer and Welch (1992) to address the same phenomenon. *Informational cascades* occur when, after some finite time, all *dm*s ignore their private information when choosing an action, while *herd behavior* occurs when, after some finite time, all *dm*s choose the same action, not necessarily ignoring their private information.

Hence, an informational cascade implies herd behavior but herding is not necessarily the result of an informational cascade. When acting in a herd, *dm*s choose the same action, but they could have acted differently from one another if the realization of their private signals had been different. In an informational cascade, a *dm* considers it optimal to follow the behavior of her predecessors without regard to her private signal since her belief is so strongly held that no signal can outweigh it.

We replicate the results of the literature under perfect information and use them as a benchmark. The main difference between perfect and imperfect information is that learning under perfect information has the martingale property that permits establishment of convergence of beliefs and actions. Under imperfect information, by contrast, the learning process does not have the martingale property. The important implication is that beliefs and actions are not convergent but cycle forever. Despite this instability, over time, private information is increasingly ignored and *dm*s become increasingly likely to imitate their predecessors (Theorem 1). Consequently, we observe longer and longer periods of uniform behavior, punctuated by increasingly rare switches (Theorem 2). In other words, under perfect information, social learning eventually ceases as individual behavior becomes purely imitative and hence is uninformative. Under imperfect information, by contrast, individuals become more and more likely to imitate because the behavior of their immediate predecessors remains informative and, at the same time, behavior fails to converge.

These results suggest that the kind of episodic instability that is characteristic of socioeconomic behavior in the real world makes more sense in the imperfect-information model. The key economic phenomenon that imperfect information captures is a succession of fads starting suddenly, expiring rather easily, each replaced by another fad. As such, the imperfect-information model gives insight into phenomena such as manias, fashions, crashes and booms, and better answers such questions as: 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 the ‘in place’, and so on?

The paper is organized as follows. In the next section, we provide a discussion of closely related literature. The model is outlined in section 3, analyzed under some distribution specification in section 4 and for general distributions in section 5. We conclude in section 6.

2 Related literature

Observation of an immediate predecessor’s action is a particular form of imperfect information. Smith and Sørensen (1996) relax perfect information by assuming that each *dm* observes a random unordered sample of her predecessors' actions. This approach is in fact the finite-*dm* counterpart of Banerjee and Fudenberg (1995)'s continuum-*dm* model. Smith and Sørensen (1996) provide a thorough characterization for the case of *unbounded beliefs*, but their results are not exhaustive for the case of *bounded beliefs*[1]. With unbounded beliefs, Smith and Sørensen (1996) agree with Banerjee and Fudenberg (1995) that learning leads to correct decisions, and with bounded beliefs, they show that what is "learned" can be incorrect.

Taking off from Smith and Sørensen (1996), we assume that each *dm* samples her immediate predecessor's choice with probability one. This assumption captures, in an extreme format, the idea that more recent predecessors are more likely to be observed. Aside from modeling choices, the present paper differs from Smith and Sørensen (1996) in two ways. First, we show that behavior can be radically different under perfect and imperfect information. Second, we are able to describe not only the asymptotic outcomes but also the behavior in case of divergence.

In another study, Smith and Sørensen (1997) develop an example in which each *dm* observes her immediate predecessor's decision. But, their signal distribution assumes unbounded beliefs and their focus is on different properties of learning, so results are not directly comparable.

3 The model

3.1 Preliminaries

Our economy consists of a finite number of Bayes-rational *dm*s indexed by $n = 1, 2, \cdots, N$. Each *dm* n makes a once-in-a-lifetime decision, to invest or not to invest, indicated by $x_n = 1$ and $x_n = 0$ respectively. Decisions are made sequentially in an exogenously determined order. The preferences of the *dm*s are assumed to be identical and represented by the risk neutral vN-M utility function

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

where the value of investment Θ is a random variable defined by

$$\Theta = \sum_{n=1}^{N} \theta_n$$

and θ_n is *dm* n's private signal about Θ. We assume that the θ_n's are identically and independently distributed with *c.d.f.* F over a compact support

[1] Private beliefs are said to be *bounded* when there is no private signal that can perfectly reveal the true state of the world, and to be *unbounded* otherwise.

$\mathcal{S}$ with convex hull $[a, b]$, such that $\mathbb{E}[\theta] = 0$. Further, F satisfies symmetry when $a + b = 0$ and $F(\theta) = 1 - F(-\theta)\ \forall\theta \in [a, -a]$.

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* $\Theta \geq 0$ and *low* $\Theta < 0$. 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 the realized event, i.e., $P(\theta \geq 0 \,|\Theta \geq 0\,) > \frac{1}{2}$ and $P(\theta < 0 \,|\Theta < 0\,) > \frac{1}{2}$. Yet, private beliefs are bounded for any $N > 2$.

The summation version of Θ makes the model nicely tractable, but some clarifications are in order. Notice that we define a finite economy, yet we are interested in the behavior when the size of the economy N is arbitrarily large. In other words, we analyze the limit behavior of a sequence of economies indexed by N. With infinite N, the problem is not well formulated since Θ may not be summable as defined. Further, if we define Θ as the average signal rather than the sum of signals, then Θ is (with probability one) equal to zero by the law of large numbers, so *dm*s are always indifferent between different actions. However, with any finite N, the model is robust to any scaling of Θ, and thus, there are no practical problems since we are interested in limit theorems rather than theorems in the limit. That is to say, we are interested in approximating the behavior when the economy size increases. By the same token, even though the information of a *dm* about Θ is not constant across different sized economies, the underlying decision problem, the optimal decision rule, and hence our results are independent of N as in the traditional social learning models.

We refer to a perfect information economy $\mathcal{E}_F = \{F, x_n, u_n, I_n\}_{n=1}^N$, 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.,

$$I_n = \{\theta_n, (x_i)_{i=1}^{n-1}\} \in \mathcal{S} \times \{0, 1\}^{n-1}.$$

An imperfect information economy $\mathcal{E}'_F = \{F, x_n, u_n, I'_n\}_{n=1}^N$ differs in that each *dm* $n > 1$ observes only her immediate predecessor's action, i.e.,

$$I'_n = \{\theta_n, x_{n-1}\} \in \mathcal{S} \times \{0, 1\}.$$

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.

3.2 The decision problem

A *dm*'s strategy is a mapping from her information set into the set of actions. Next, we provide a definition that will be useful in characterizing the optimal strategy.

Definition 1 *dm n follows a* cutoff strategy *if her decision rule is defined by*

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

for some cutoff $\tilde{\theta}_n \in [a, b]$[2].

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

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

which yields the optimal decision rule

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

Since $\mathcal{I}_n$ does not provide any information about the content of successors' signals, we obtain

$$x_n = 1 \text{ if and only if } \theta_n \geq -\mathbb{E}\left[\sum_{i=1}^{n-1} \theta_i \mid \mathcal{I}_n\right].$$

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

Proposition 1 *For any n, the optimal strategy is the cutoff strategy*

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

where

$$\hat{\theta}_n = -\mathbb{E}\left[\sum_{i=1}^{n-1} \theta_i \mid \mathcal{I}_n\right] \tag{1}$$

is the optimal history-contingent cutoff.

[2] Notice that the tie-breaking assumption is such that $x_n = 1$ when $\theta_n = \tilde{\theta}_n$. One may assume different tie-breaking rules, but since these are probability zero events, the analysis does not alter.

The optimal cutoff $\hat{\theta}_n$ contains all the information that *dm* n acquires from the history and thus 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 $\{\hat{\theta}_n\}_{n=1}^N$ characterizes the behavior of the economy. Henceforth, we take $\{\hat{\theta}_n\}_{n=1}^N$ as the object of our analysis and refer to it as a cutoff process or learning process interchangeably.

3.3 Definitions

Next, we define some key concepts to which we refer throughout the paper. To economize on notation, whenever we take a limit over n we allow N to accommodate n by taking a double limit as $N \to \infty$ and $n \to \infty$.

Definition 2 (Informational cascade) *An informational cascade on action $x = 1$ ($x = 0$) occurs when $\exists n$ such that $\hat{\theta}_k \in (-\infty, a]$ ($\hat{\theta}_k \in [b, \infty)$) $\forall k \geq n$. Analogously, a limit-cascade on action $x = 1$ ($x = 0$) occurs when the process of cutoffs $\{\hat{\theta}_n\}$ converges almost surely to a random variable $\hat{\theta}_\infty = \lim_{n\to\infty} \hat{\theta}_n$, with $supp(\hat{\theta}_\infty) \subseteq (-\infty, a]$ ($supp(\hat{\theta}_\infty) \subseteq [b, \infty)$).*

Hence, a cascade occurs in the limit when all but finitely many *dm*s are almost surely convinced about which of the events will take place. Further, we call a finite sequence of *dm*s who act alike a *finite herd* and, we let

$$l_n^N \equiv \#\{x_k = x_n,\ n \leq k \leq N\}$$

denote the length of a finite herd following *dm* n in an economy of size N. Herd behavior is said to occur if *dm*s eventually settle on an action, i.e., action convergence almost surely obtains.

Definition 3 (Herd behavior) *Herd behavior occurs when $\exists n$ such that $\lim_{N\to\infty} l_n^N / N = 1$.*

Thus, *dm* n acts in a herd but does not follow a cascade when $\hat{\theta}_n \in (a, b)$, indicating that for some signal she is willing to make either decision, but when her private signal is realized she acts as her predecessors did.

4 The uniform case

In order to illustrate the model we study a simple symmetric example where private signals are distributed with uniform distribution U, over the support $[-1, 1]$.

4.1 The case of perfect information

According to (1), in the perfect information economy, $\mathcal{E}_U = \{U, x_n, u_n, I_n\}_{n=1}^N$, *dm* n's optimal history-contingent cutoff rule is

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

Since with perfect information any history of actions is public information shared by all succeeding *dm*s, all the information revealed by the history $(x_i)_{i=1}^{n-2}$ is already accumulated in *dm* $(n-1)$'s cutoff. Therefore, *dm* n's cutoff is altered only by the new information revealed by *dm* $(n-1)$'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

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

Equations (2) and (3) yield the following cutoff process:

Proposition 2 *In $\mathcal{E}_U$, the cutoff dynamics follows the stochastic process*

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

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

The impossibility of an informational cascade follows immediately since $\mid \hat{\theta}_n \mid < 1, \forall n$. Thus, in making a decision, any *dm* takes her private signal into account in a non-trivial way. Moreover, the learning process $\{\hat{\theta}_n\}$ has the martingale property $\mathbb{E}[\hat{\theta}_{n+1} \mid \hat{\theta}_n] = \hat{\theta}_n$. So, by the Martingale Convergence Theorem, it converges almost surely to a random variable $\hat{\theta}_\infty = \lim_{n\to\infty} \hat{\theta}_n$. Hence, it is stochastically stable in the neighborhood of the fixed points, -1 and 1, meaning that there is a limit-cascade. Finally, since convergence of the cutoff process implies convergence of actions, behavior can not overturn forever. In other words, behavior settles down in some finite time and is consistent with the limit learning. In conclusion, we agree with Smith and Sørensen (2000) that a cascade need not arise but a limit-cascade and herd behavior must.

4.2 The case of imperfect information

In the imperfect information economy, $\mathcal{E}'_U = \{U, x_n, u_n, I'_n\}_{n=1}^N$, the action of a *dm* is the only source of information available to her successor to indicate the nature of all past signals. 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 readily be noted that $\hat{\theta}_n$ can take two different values, conditional on $x_{n-1} \in \{0, 1\}$. That is,

$$\hat{\theta}_n = \begin{cases} \overline{\theta}_n \text{ if } x_{n-1} = 1 \\ \underline{\theta}_n \text{ if } x_{n-1} = 0 \end{cases} \tag{5}$$

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 derivation of the cutoff rule rests on three basic observations (For proofs, see Çelen and Kariv (2001)). First, the 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, that is, for any n,

$$P(x_{n-1} = 0 \mid x_n = 1) = P(x_{n-1} = 1 \mid x_n = 0). \tag{6}$$

Second, for each *dm* n both actions are *ex ante* equally probable,

$$P(x_n = 1) = \frac{1}{2}. \tag{7}$$

And, finally, the cutoff rule is symmetric,

$$\overline{\theta}_n + \underline{\theta}_n = 0. \tag{8}$$

These observations help us to derive a closed form solution of $\hat{\theta}_n$ recursively. Note that if *dm* n observes $x_{n-1} = 1$, she can determine the probabilities that $x_{n-2} = 1$ or $x_{n-2} = 0$ conditional on this information. If $x_{n-2} = 1$ then the actual cutoff of *dm* $(n-1)$ is $\overline{\theta}_{n-1}$, which already inherits all the information accumulated in the history. Moreover, the expected value of her signal θ_{n-1} is computable conditional on $\overline{\theta}_{n-1}$ and $x_{n-1} = 1$. An analogous argument also applies if $x_{n-2} = 0$. Thus, the law of motion for $\overline{\theta}_n$ is given by

$$\begin{aligned}\overline{\theta}_n = P(x_{n-2} = 1|x_{n-1} = 1)\left\{\overline{\theta}_{n-1} - \mathbb{E}\left[\theta_{n-1} \mid x_{n-2} = 1\right]\right\} \\ + P(x_{n-2} = 0|x_{n-1} = 1)\left\{\underline{\theta}_{n-1} - \mathbb{E}\left[\theta_{n-1} \mid x_{n-2} = 0\right]\right\}.\end{aligned}$$

Using observations (6) and (7) it simplifies to

$$\overline{\theta}_n = \frac{1-\overline{\theta}_{n-1}}{2}\left[\overline{\theta}_{n-1} - \frac{1+\overline{\theta}_{n-1}}{2}\right] + \frac{1-\underline{\theta}_{n-1}}{2}\left[\underline{\theta}_{n-1} - \frac{1+\underline{\theta}_{n-1}}{2}\right]. \quad (9)$$

Using (8), this leads to the following proposition.

Proposition 3 *In $\mathcal{E}'_U$, the cutoff dynamics follows the cutoff process*

$$\hat{\theta}_n = \begin{cases} -\frac{1+\hat{\theta}^2_{n-1}}{2} & \text{if} \quad x_{n-1} = 1 \\ \frac{1+\hat{\theta}^2_{n-1}}{2} & \text{if} \quad x_{n-1} = 0 \end{cases} \quad (10)$$

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

The impossibility of an informational cascade in $\mathcal{E}'_U$ follows immediately since, as in $\mathcal{E}_U$, $\mid \hat{\theta}_n \mid< 1$ $\forall n$. However, as we illustrate in Figure 1, *dm* n's cutoff rule partitions the signal space into 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 $\theta_n \in [-1, \overline{\theta}_n)$ *dm* n follows her private signal and takes action $x_n = 1$ or $x_n = 0$ respectively. In the intermediate subset $[\overline{\theta}_n, \underline{\theta}_n)$, which we call the *imitation set*, private signals are ignored in making a decision and *dm*s imitate their immediate predecessor's action.

[Figure 1 here]

Furthermore, since $\{\overline{\theta}_n\}$ ($\{\underline{\theta}_n\}$) is a decreasing (increasing) sequence bounded by -1 (1) and must converge, imitation 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, the imitation set becomes an attractor in the limit. Hence, over time, *dm*s tend to rely more on the information revealed by the predecessor's action rather than their private signal.

Note, however, that this does not imply convergence of the cutoff process $\{\hat{\theta}_n\}$. In fact, a simple analysis shows that the cutoff process (10) is not convergent[3]. Hence, a limit-cascade never arises since the cutoff process is not stable near any of the fixed points -1 and 1. Further, since $\{\hat{\theta}_n\}$ is not stable, it is obvious that convergence of actions in the standard herding manner is impossible. This is to say that the divergence of cutoffs implies divergence of actions.

As herd behavior is impossible, one might ask what the expected length of a finite herd starting from some finite *dm* n, $\mathbb{E}[l_n^N]$, is. Note that when a deviation occurs, the cutoff process switches from a point close to one of the fixed points to a point even closer to the other fixed point, therefore $\mathbb{E}[l_n^N]$ is increasing in n for an economy size N large enough. This can be shown using direct calculations. Hence, along the line of *dm*s, behavior is typified by monotonically longer lasting finite herds. Furthermore, a comparison test with $\Sigma_n \frac{1}{n}$ shows that for any n, $\lim_{N\to\infty} \mathbb{E}[l_n^N] = \infty$. Thus, one aspect of herding is preserved in $\mathcal{E}'_U$: as $N \to \infty$, the expected number of successors who will imitate any *dm* tends to infinity.

4.3 Perfect versus imperfect information

To understand the dissimilarities between $\mathcal{E}_U$ and $\mathcal{E}'_U$, consider a finite herd followed by a deviator. In both $\mathcal{E}_U$ and $\mathcal{E}'_U$, the deviator becomes a leader to her successors. Yet, there is substantial difference. In $\mathcal{E}_U$, 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 meagerly dominates the accumulated public information. Thus, her successor will be slightly in favor of joining the deviation. This is referred to by Smith and Sørensen (2000) as the *overturning principal*.

On the other hand, in $\mathcal{E}'_U$, one can not tell whether her predecessor is an imitator or a deviator. Thus, a deviator's action is her successor's sole source of information about the entire history of previous actions. Consequently, one who immediately follows a deviator can be expected to replicate the deviation. Moreover, most likely the deviation will turn out to be followed by a longer lasting finite herd.

To illustrate, assume that a long finite herd of investment precedes some dm n. Then, her cutoff is close to -1, for example $\hat{\theta}_n = -1+\varepsilon$ for some small $\varepsilon > 0$ in $\mathcal{E}_U$ and $\hat{\theta}_n = -1+\delta$ for some small $\delta > 0$ in $\mathcal{E}'_U$. Now, suppose that *dm* n does not invest because she receives an extreme contrary signal, say $\theta_n = -1$. In $\mathcal{E}_U$, her deviation reveals clear-cut information that $\theta_n \in [-1, -1+\varepsilon)$, and thus, having observed the deviation, *dm* $(n+1)$ overturns;

[3] 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 induction, it is not difficult to show that $\left(\bar{\theta}_n + 1\right) \geq \frac{1}{n}$ for all n.

yet her cutoff is close to zero, specifically $\hat{\theta}_{n+1} = \frac{\varepsilon}{2}$. In $\mathcal{E}'_U$, by contrast, since the deviation is not observed by dm $(n+1)$, she overturns dramatically by setting her cutoff even closer to 1, specifically $\hat{\theta}_{n+1} = 1 - \delta + \frac{\delta^2}{2}$.

As to the welfare properties of the equilibria, the likelihoods of correct decisions in $\mathcal{E}_U$ and $\mathcal{E}'_U$ can not be found analytically since conditional on Θ, θ_ns are negatively correlated. However, simulations show certain directional effects, which, to the extent that we can cover finite economy sizes, we conjecture that they are robust. In particular, in $\mathcal{E}_U$ the process is concentrated more often than in $\mathcal{E}'_U$ on the correct decision. And, in both $\mathcal{E}_U$ and $\mathcal{E}'_U$, the *ex ante* probability that dm n makes a correct decision increases in n for a given N. Figure 2 summarizes simulations that were carried out for economies $\mathcal{E}_U$ and $\mathcal{E}'_U$ of size $N = 10$.

[Figure 2 here]

5 The general case

All of our results to this point relied on the assumption that the signal distribution is uniform. In what follows, we show that the results obtained so far hold for any symmetric signal distribution. Since the perfect information case is studied in a general setting by Smith and Sørensen (2000), we concentrate on the imperfect information economy $\mathcal{E}'_F$.

5.1 The symmetric case

We consider an imperfect information economy $\mathcal{E}'_F$ where F satisfies symmetry. Without loss of generality, assume that $a = -1$. Additionally, for technical reasons we assume that there is no probability mass on any of the cutoff points, which is a set of measure zero.

One can show that observations (6), (7) and (8) for the uniform case hold for any F. Using observations (6) and (7), the law of motion for $\overline{\theta}_n$ is given by

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

where $E^{+}(\xi) \equiv E\left[\theta | \theta \geq \xi\right]$. Using symmetry (8) and direct calculations,

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

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

$$\begin{aligned}
\overline{\theta}_{n+1} &= \overline{\theta}_n - 2F(\overline{\theta}_n)\overline{\theta}_n + 2\int_{-1}^{\overline{\theta}_n} \theta dF \qquad (11)\\
&\geq \overline{\theta}_n - 2F(\overline{\theta}_n)\overline{\theta}_n - 2F(\overline{\theta}_n) \\
&\geq \overline{\theta}_n - (\overline{\theta}_n + 1) \\
&= -1
\end{aligned}$$

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 a decreasing sequence bounded by -1 and must converge. In fact, from (11) 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 $\varphi(\underline{\theta}) > \underline{\theta}$ for any $\underline{\theta} < 1$ and $\underline{\theta}_n \nearrow 1$ as $n \to \infty$.

The impossibility of informational cascades follows immediately since $\mid \hat{\theta}_n \mid < 1$, $\forall n$. Furthermore, it can be readily noted that imitation sets $[\overline{\theta}_n, \underline{\theta}_n)$ monotonically increase in n and converge to an attractor in the limit. We have already observed that when signals are uniformly distributed the cutoff process $\{\hat{\theta}_n\}$ is not stable. We, now, extend this result to any F.

Proposition 4 *In $\mathcal{E}'_F$, $\{\hat{\theta}_n\}$ is unstable near -1 and 1.*

Proof. Without loss of generality, we show that $\{\hat{\theta}_n\}$ unstable near -1, *i.e,* for any $k < \infty$, $\prod_{n=k}^{\infty}(1 - F(\overline{\theta}_n)) = 0$. First, note that it holds trivially whenever $F(-1) \neq 0$ since there is always a positive probability of deviation. When $F(-1) = 0$, by (11)

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

Let $\mu_n = 1 + \overline{\theta}_n$, then

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

Since $\prod_{n=k}^{\infty} \frac{\mu_{n+1}}{\mu_n} = 0$, $\prod_{n=k}^{\infty} \left(1 - 2F(\overline{\theta}_n)\right) = 0$ which implies that $\sum_{n=k}^{\infty} F(\overline{\theta}_n) = \infty$, and thus $\prod_{n=k}^{\infty}(1 - F(\overline{\theta}_n)) = 0$. ■

The next theorem summarizes the results on learning dynamics.

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

As to action dynamics, the impossibility of herd behavior follows immediately from the instability of the cutoff process $\{\hat{\theta}_n\}$. That is, since a deviation occurs with probability 1, action convergence in the standard herding manner is impossible.

Notwithstanding with the impossibility of herd behavior, when F has no mass on the boundaries of the signal support, *i.e,* $F(-1) = 0$ the expected length of a finite herd following any *dm* n, given by

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

need not be bounded, as in the uniform case. However, we know of no necessary condition on the primitive F that guarantees that $\lim_{N\to\infty} \mathbb{E}[l_n^N] = \infty$. The obvious difficulty is to determine the finiteness of a series whose terms are merely described in a difference equation.

A simple easily checked sufficient condition is that $\overline{\theta}_n$ converges fast enough such that $F(\overline{\theta}_n)$ converges at a rate faster than $\frac{1}{n}$. The proof of this result is not difficult and is omitted. It follows, with the help of induction, from a comparison test for which the divergent sequence is $\sum_k F(\overline{\theta}_{n+k})$.

On the other hand, when F has a mass on the boundaries, *i.e,* $F(-1) \neq 0$, the expected length of a finite herd following any *dm* n is bounded. To see this suppose that $F(-1) = \delta > 0$, then

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

and the inequality follows since $F(\overline{\theta}_{n+2}) \geq F(\overline{\theta}_{n+k})$ for all $k \geq 2$ and $\delta \geq 1 - F(\overline{\theta}_{n+k})$ for all $k \geq 1$. But since $\sum_{k=1}^{\infty} k\delta^k = \frac{\delta}{(1-\delta)^2}$, $\lim_{N\to\infty} \mathbb{E}[l_n^N] < \infty$.

The next theorem summarizes the results on action dynamics.

Theorem 2 (Behavior) *In $\mathcal{E}'_F$, (i) Herd behavior does not occur. (ii) The expected length of a finite herd following any dm n, $\mathbb{E}[l_n^N]$, need not be bounded, i.e., $\lim_{N\to\infty} \mathbb{E}[l_n^N] = \infty$, when $F(-1) = 0$. When $F(-1) \neq 0$, $\lim_{N\to\infty} \mathbb{E}[l_n^N] < \infty$.*

5.2 The asymmetric case

Next we consider an imperfect information economy, $\mathcal{E}'_F$, where private signals are distributed with F such that $\mathbb{E}[\theta] = 0$. Let $p_n \equiv P(x_n = 1)$. Then,

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

for every n 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

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

Notice that,

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

The inequality follows because

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

and

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

and the inequality is strict as long as $\overline{\theta}_{n+1} > a$ or $\underline{\theta}_{n+1} < b$. It is 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, $\underline{\theta}_k > b, \overline{\theta}_k < a, \forall k > n$, 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 $\{\overline{\theta}_n\}$ and $\{\underline{\theta}_n\}$ must exit (a, b) in finite time or in the limit. If this happens in finite time, an informational cascade arises, and if $\{\hat{\theta}_n\}$ is stabilized asymptotically, a limit-cascade arises. An immediate corollary of these convergence results is that

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

which means, *ex ante*, that the limit *dm* chooses $x = 1$ with probability $\frac{b}{b-a}$ and $x = 0$ with probability $\frac{-a}{b-a}$.

Although we do not have a full characterization of the possibility of informational cascades, we know that they may arise causing a herd on the corresponding action. In Çelen and Kariv (2001), we provide a sufficient condition for the impossibility of informational cascades, and give an example where a cascade arises.

6 Concluding remarks

The perfect- and imperfect-information versions of the model share the conclusion that *dm*s can, for a long time, make the same choice. The important difference is that, whereas in the perfect-information model a herd is an absorbing state, in the imperfect-information model, there are continued, occasional and sharp shifts in behavior. These results suggest that the imperfect information premise illuminates socioeconomic behavior that typically exhibits long-lasting but finite episodes of mass behavior. In particular, we argue that the imperfect information premise provides a better theoretical description of fads and fashions.

It is natural to ask about the robustness of the results when the number of most recent actions that a *dm* observes exceeds one. Our analysis does not properly address 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 are available. The cutoff rule becomes richer since further inferences based on the frequency of past actions can be obtained. More specifically, *dm*s are then able to identify deviators and imitators, and the information revealed by a deviation can be incorporated into their decision rule. Since the amount of information is increasing in the number of predecessors observed, a successor of a deviator is still inclined to follow the deviation but with less enthusiasm as this number increases.

Whether an increase in the number of predecessors observed would lead to sharply different results is not clear, since all the decision rules would have to be changed to reflect the new environment. Obviously, different information structures may lead to different outcomes. This remains a subject for further research.

References

[1] Banerjee, A. (1992) "A Simple Model of Herd Behavior." *Quarterly Journal of Economics, 107*(3), pp. 797-817.

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

[3] 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*(5), pp. 992-1026.

[4] Çelen, B. and S. Kariv (2001) "Observational Learning Under Imperfect Information." C.E.S.S. Working Papers #02-03, New York University.

[5] Gale, D. (1996) "What Have We Learned from Social Learning?" *European Economic Review, 40*(3-5), pp. 617-28.

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

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

[8] Smith, L. and P. Sørensen (2000) "Pathological Outcomes of Observational Learning." *Econometrica, 68*(2), pp. 371-398.

Figure 1: The partition of the signal space

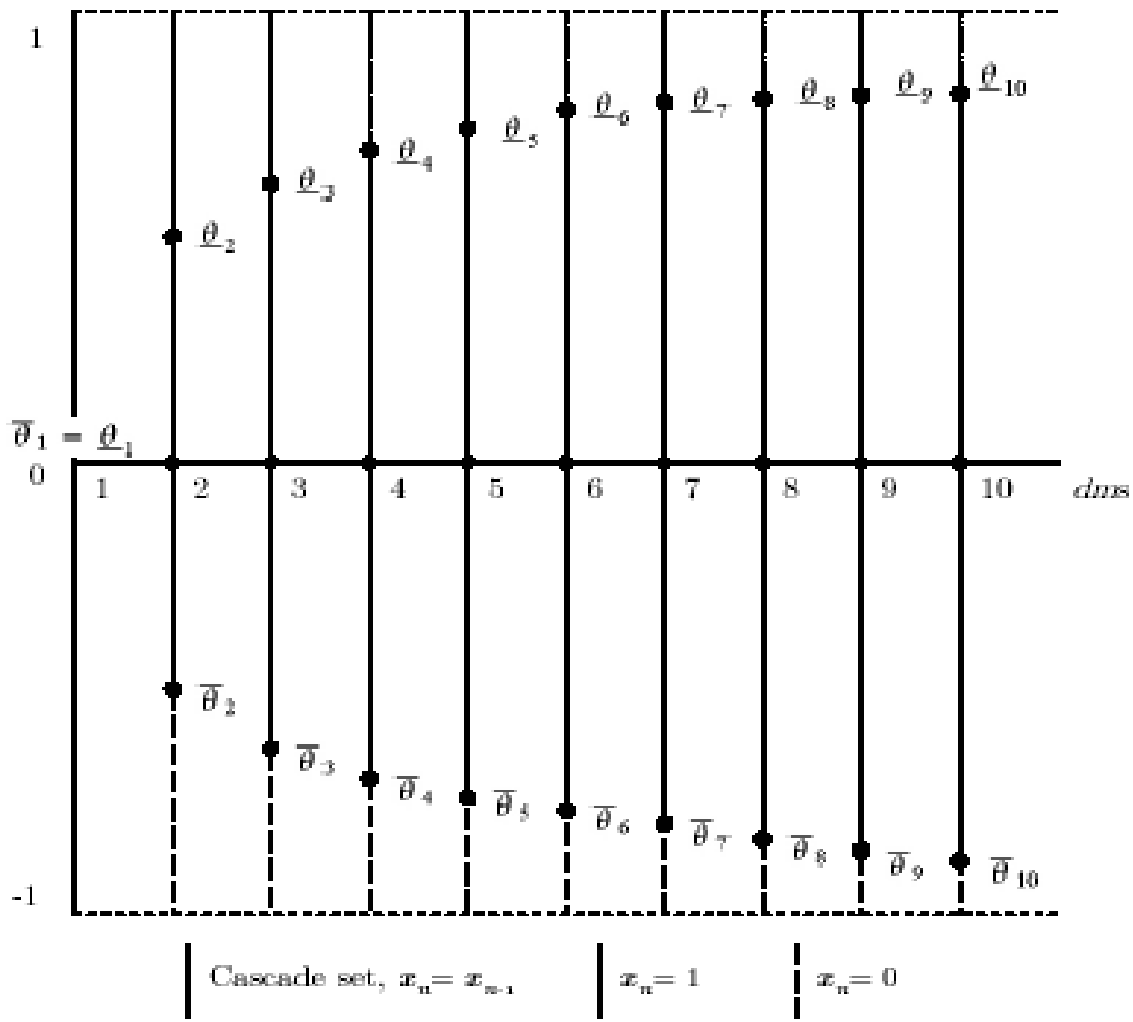
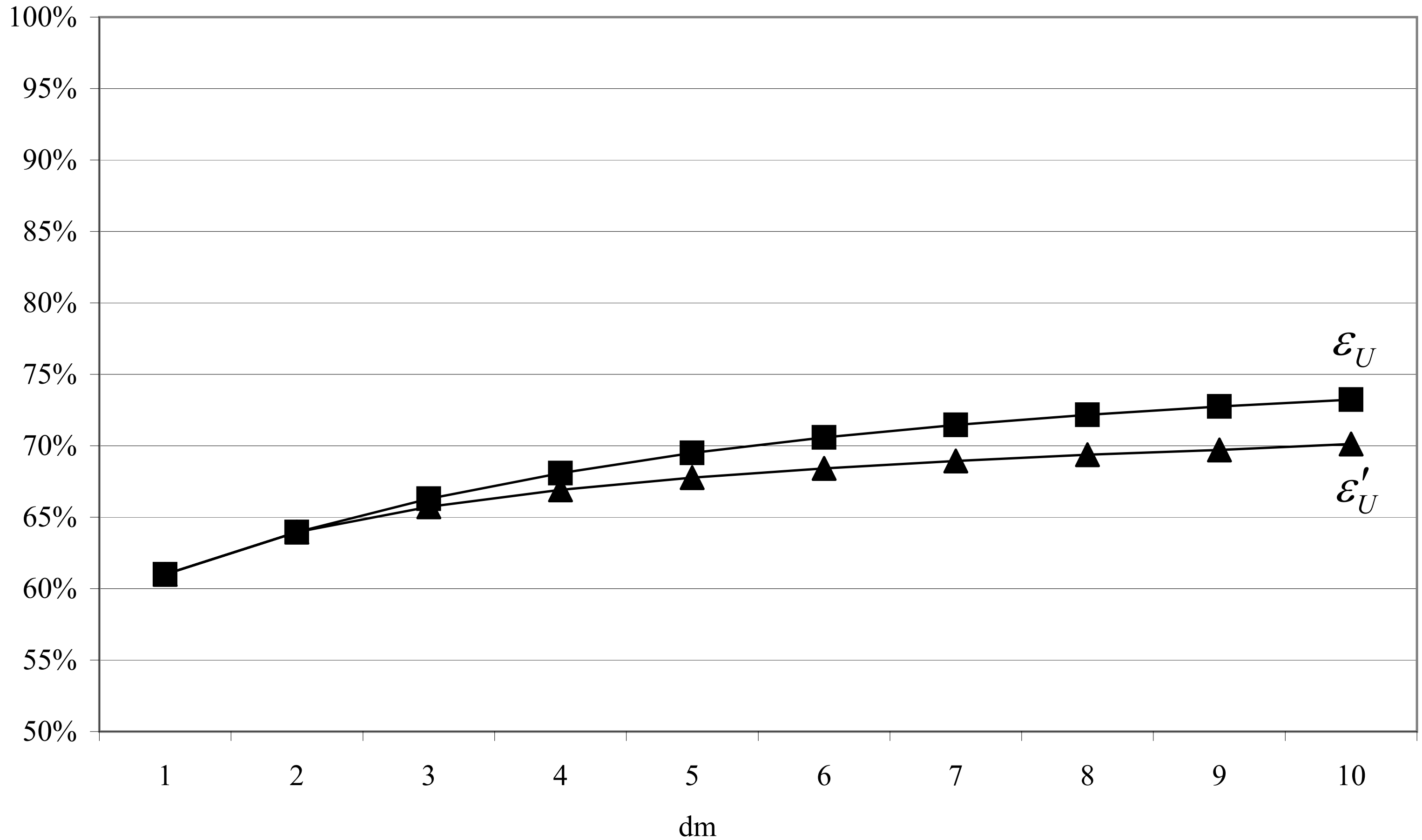


Figure 2: The *ax ante* probability of correct decision by turn
(perfect and imperfect information economies with N=10)



Numerical simulations. The likelihoods of correct decisions under perfect and imperfect information.