The Value of Noise for Informational Cascades

Tho Ngoc Le, Vijay G. Subramanian, Randall A. Berry
EECS Department
Northwestern University, Evanston, IL 60208
Email: thole2012@u.northwestern.edu, v-subramanian@northwestern.edu, rberry@eecs.northwestern.edu

***Abstract*—*Informational cascades* are said to occur when rational agents ignore their own private information and blindly follow the actions of other agents. Models for such cascades have been well studied for Bayesian agents, who observe perfectly the actions of other agents. In this paper, we investigate the impact of errors in these observations; the errors are modelled via a binary symmetric channel (BSC). Using a Markov chain model, we analyze the net payoff of each agent as a function of his signal quality and the crossover error probability in the channel. Our main result is that a lower error level does not always lead to a higher payoff when the number of agents is large.**

I. Introduction

Consider a recommendation system where agents sequentially decide whether to buy an item, for which they have some prior knowledge of its quality/utility. Later agents benefit from the information obtained by observing their predecessors' choices. *Herding* or an *informational cascade* occurs when it is optimal for the agents to ignore their own signals and follow the actions of others. In addition to the possibility of herding to the wrong conclusion, an informational cascade results in a loss of information about the private signals held by all the agents following the onset of herding.

The study of herding was initiated in the seminal papers [2] and [3]. In these papers, each individual can observe exactly the actions of the previous agents. This assumption leads to herding eventually happening, with positive probability of it being in error. In our paper, the observations process is assumed to have imperfections. We assume that information of the history of past actions can be in error as it is received via a BSC, with crossover probability ϵ, by the subsequent agents. For example, this could model a setting where agents are asked to report their decision on a website and agents occasionally misreport. The objective is to study the effect of such noise in the observation process. Our main results, shown in Fig. 1, demonstrate a counter-intuitive phenomenon for the behavior of the asymptotic welfare: For both low and high prior signal quality p, while having no observation noise ϵ maximizes the asymptotic welfare, it is not monotonically decreasing in the noise. Thus, for certain cases, an agent's total payoff can be increased by increasing the noisiness of the observation process.

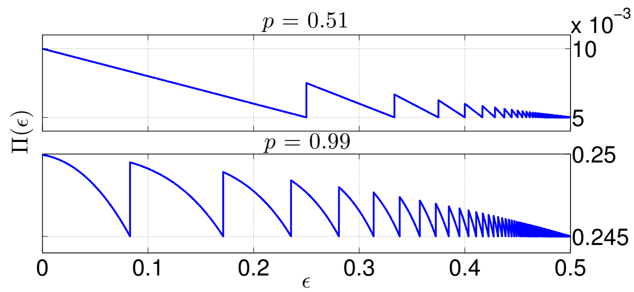

Figure 1: Limiting welfare for low and high signal quality.

Herding is typically studied as a form of Bayesian learning, where each agent has a private signal and observes the actions of others. As in [6], these observations may depend on an underlying network structure. Here, we focus on a simple structure, as in [2], [3], [4], where individuals take actions sequentially and observe all past actions (though in our case these observations are noisy).

In [4], the authors consider a different form of noise by allowing a fraction of individuals to randomly choose their actions. This leads to occasional irrational actions in the history. The authors prove that such irrational actions would be ignored by later individuals in an informational cascade and thus this type of noise does not qualitatively affect the final herding behavior. Our model, on the other hand, assumes that all actions are rational, with noise uniformly introduced through the observation process.

This paper is organized as follows. In Section II we first develop a noisy version of the model in [2]. We study the effect of the added noise in Section III and model this as a Markov chain. In Section IV, we discuss the agent welfare for high and low signal qualities and noise levels. We conclude in Section V.

II. Model

We consider a model similar to [2] in which there is a countable population of agents, indexed $i = 1, 2, \ldots$ with the index reflecting the time and order of actions of the agents. Each agent i has an action choice A_i of saying either Yes (Y) or No (N) to a new item. The true value (V) of the item can be either 0 (bad) or 1 (good); both possibilities are assumed to be equally likely. The agents are Bayes-rational utility maximizers whose payoff structure is based on the agent's choice of action and the true value of the item. If an agent chooses N, his payoff is 0. On the other hand, if he chooses Y, he faces a cost of $C = 1/2$ and two possibilities of outcomes: his gain is 0 if $V = 0$ and 1 if $V = 1$. Thus, the *ex-ante* payoff of each agent is $E[V] - C = 0$. To reflect the agents' prior knowledge about the true value of the item, we assume that each agent i receives a private signal S_i through a BSC with crossover probability $1 - p$, where $1/2 < p < 1$. (See Fig. 2.) Thus, the private signals are informative, but not revealing. We further assume that each agent i makes a one-time action A_i based on his own private signal S_i and the observations $O_1, \dots, O_{i-1}$ of all previous agents' actions $A_1, \dots, A_{i-1}$.

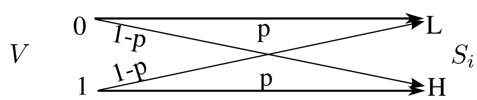

Figure 2: The BSC over which agents receive signals.

It was shown in [2] and [5] when observations are perfect (i.e., $O_i = A_i$), this model exhibits a *herding* phenomenon. *Herding* happens when an agent chooses to follow the majority of his predecessors' actions regardless of his own private signal. Here, we instead consider a model where later agents' observations are noisy versions of their predecessors' actions. For simplicity, our model assumes that each agent reports his action to a public database which is available to all successors. The errors in this process are modelled by passing every action A_i through another BSC with crossover probability $\epsilon \in (0, 1/2)$. This means with probability $1 - \epsilon$, $O_i = A_i$, and with probability ϵ, $O_i = \overline{A_i}$, where $\overline{A_i}$ is the opposite action of A_i. This assumption reduces the dependence of every agent's decision on the predecessors' choices and drives him toward using his own signal.

III. Herding in noisy observations

A. Herding properties

In this section we outline some basic properties of herding with noisy observations. These naturally extend properties for the noiseless case shown in [2], [5] and so we omit detailed derivations.

The first agent always follows his private signal since no observation history is available. Starting from the second agent, every agent i considers his private signal S_i and the observations $O_1, ..., O_{i-1}$. Let the information set of agent i be $I_i = \{S_i, O_1, ..., O_{i-1}\}$. Based on I_i, agent i will update his posterior probability denoted as $\gamma_{i,I_i} = Pr[V = 1|I_i]$ using Bayes' formula. If this posterior probability is greater than the cost C, the agent will choose Y. If γ_{i,I_i} is less than C, the agent i will declare N. Thus, agent i is said to *herd* $Y(N)$ if $\gamma_{i,I_i} > C$ $(< C)$ for all $S_i \in \{H, L\}$. Finally, if γ_{i,I_i} equals the cost, then agent i follows his private signal.[1]

Property 1. *Until herding occurs, each agent's Bayesian update depends only on their private signal and the difference in the number of Y's and N's in the observation history.*

In other words, the difference in the number of Y's and N's is a sufficient statistic for the observation history; we denote this quantity by $H_n = \#Y's - \#N's$ for agent n. This follows from the symmetry of the signal quality and the channel noise, which enables each agent to "cancel out" opposite observations.

Property 2. *Once herding happens, it lasts forever.*

The reason for this phenomenon is that when herding starts, agents stop using their private signals and thus provide no more information to their successors. The successors are left in the same situation as the first agent who started the herding and thus have the same optimal action choice.

B. Error thresholds

In contrast to the model in [2], where the second agent has 50% of the chance creating a herd, in our model he always follows his own signal. However, depending on the noise, this will not be the case starting from the third agent. As the amount of error, ϵ, introduced into the observation is increased, each observation provides less information. Therefore, every agent does not herd unless the magnitude of the sufficient statistic, $|H_n|$, is sufficiently high as shown in the following lemma.

Lemma 1. *An agent n with $|H_n| < k$, and $k \geq 2$, will never herd if $\epsilon \geq \epsilon^*(k, p)$, where:*

$$\epsilon^*(k,p) = \frac{1 - \left(\frac{1-p}{p}\right)^{\frac{k-2}{k-1}}}{1 - \left(\frac{1-p}{p}\right)^{\frac{k-2}{k-1}} + \left(\frac{1-p}{p}\right)^{\frac{-1}{k-1}} - \left(\frac{1-p}{p}\right)}. \quad (1)$$

[1] This differs from [2], where it is assumed that indifferent agents randomly choose one action. Our assumption simplifies the analysis but does not qualitatively change the conclusions.

A simple consequence of this Lemma is that agents with index $n \leq k$ will never herd when $\epsilon \geq \epsilon^*(k,p)$. The proof of this follows from direct calculation of γ_{i,I_i}. Fig. 3 shows the thresholds $\epsilon^*(k,p)$ for different values of k and $p \in (1/2, 1)$. For $k = 2$, this Lemma yields $\epsilon^*(2,p) = 0$ which is the case of noiseless observations as in [2].

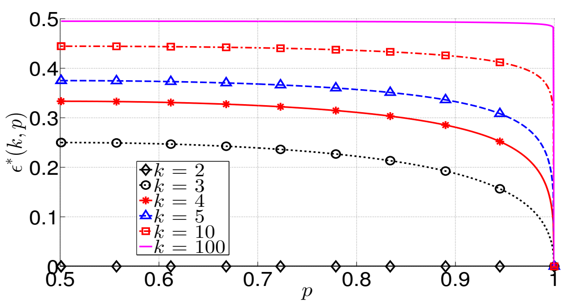

Figure 3: Thresholds for ϵ

From (1) we can obtain some useful insights about the properties of the threshold $\epsilon^*(k,p)$. First, $\epsilon^*(k,p)$ is an increasing function of k, and as $k \to \infty$, $\epsilon^*(k,p) \to 1/2$ for all values of signal quality $p \in (1/2, 1)$. Later agents have a higher likelihood of herding and such effects can be countered if the channel is noisier. In the limit, if the channel flips the bits half of the time, no information is passed through. Thus agents only use their own signals and herding is prevented. However, no learning occurs either and the *ex-post* payoff of each agent remains $\frac{2p-1}{4}$. Secondly, $\epsilon^*(k,p)$ is a decreasing function in $p \in (1/2, 1)$; as $p \to 1$, $\epsilon^*(k,p) \to 0$. This agrees with the intuition that the more accurate the private signal, the less likely it is for herding to occur in the "wrong" direction. Notice that the threshold curves are relatively flat for a wide interval of p and only drop quickly when p is sufficiently close to 1. This means that even if the private signal quality is very high, with an intermediate level of noise herding may still occur for most agents.

By Property 1, 2 and Lemma 1, for a given observation error $\epsilon \in [\epsilon^*(k,p), \epsilon^*(k+1,p))$, herding would happen for agent n if and only if $H_n \geq k$. This helps establish a simple finite-state birth-death Markov chain for our model as presented in the next section.

C. Markov analysis of herding

By the symmetry of the model, first consider the case $V = 1$. From the previous section, for an arbitrary agent n who has not herded, the observations history can be summarized by $H_n = \#Y's - \#N's$. Thus, viewing each agent as a time-epoch, we can consider the agent's observation as a state of a discrete time Markov chain. Each state i represents values of H_n that an arbitrary agent n may see before making his decision. Note that the first agent starts at state 0 when no observation history is available.

Assume $\epsilon^*(k,p) \leq \epsilon < \epsilon^*(k+1,p)$, so that an agent will not herd unless he observes the counts of one action dominate the other by at least k. Since herding lasts forever once it starts, this Markov Chain is finite-state with the state space $\{-k, -k+1, ..., 0, ..., k-1, k\}$, with states $\pm k$ being absorbing. The two events N and Y herding translate into hitting the left $(-k)$ and the right (k) walls, respectively. The probability of moving one step to the right is the probability that one more Y is added to the observation history, i.e., $a = Pr[O_i = Y|V = 1] = (1-\epsilon)p + \epsilon(1-p) > 1/2$. Likewise, the probability of moving one-step to the left is $1-a$. Hence this Markov chain is a simple random walk with a drift to the right as shown in Fig. 4.

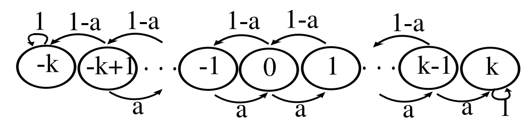

Figure 4: Transition diagram of the random walk when V=1.

We will use methods developed in [1, Sections XIV.4-5] to calculate the probability of being at each absorbing state at an arbitrary time. Assume that the process starts at state i. Let $u^*_{i,n}, v^*_{i,n}$ be the probabilities of being at the left wall, $-k$, and the right wall, k, at the n^{th} step, respectively; i.e., the probability of herding of the n^{th} agent. Let $u_{i,n}, v_{i,n}$ be the probabilities of hitting the left wall and the right wall for the first time at the n^{th} step, respectively; i.e., the probability that the n^{th} agent is the first to herd. If $n - i - k$ is an odd number, the chain cannot be at either wall for the first time, thus $u_{i,n} = v_{i,n} = 0$. Therefore, the absorption probabilities at steps n and $n-1$ are identical, i.e., $v^*_{i,n} = v^*_{i,n-1}$ and $u^*_{i,n} = u^*_{i,n-1}$. Moreover, as agent 1 starts at step 0, agent $n+1$ cannot herd if $n \leq k-1$, i.e., $u^*_{i,n} = v^*_{i,n} = 0$ for $1 \leq n \leq k-1$. For $n \geq k$, the probabilities of agent $n+1$ herding the wrong and correct way are, respectively:

$$u^*_{0,n} = \sum_{\substack{m=k \\ (m-k) even}}^{n} u^*_{-k,n-m} u_{0,m} = \sum_{\substack{m=k \\ (m-k) even}}^{n} u_{0,m}, \tag{2}$$

$$v^*_{0,n} = \sum_{\substack{m=k \\ (m-k) even}}^{n} v^*_{k,n-m} v_{0,m} = \sum_{\substack{m=k \\ (m-k) even}}^{n} v_{0,m}, \tag{3}$$

since $u^*_{-k,n-m} = v^*_{k,n-m} = 1$ (once agent m is the first one to herd, the subsequent agents $m+1,...,n$ will herd with probability 1). The next lemma gives explicit expressions for the terms on the right-hand side in (2) and (3).

Lemma 2.

$$u_{0,n} = \begin{cases} 0, & n-k \text{ odd}, \\ \frac{1}{k} 2^n a^{\frac{n-k}{2}} (1-a)^{\frac{n+k}{2}} A_k, & n-k \text{ even}, \end{cases} \tag{4}$$

$$v_{0,n} = \begin{cases} 0, & n-k \text{ odd}, \\ \frac{1}{k} 2^n a^{\frac{n+k}{2}} (1-a)^{\frac{n-k}{2}} A_k, & n-k \text{ even}, \end{cases} \tag{5}$$

where

$$A_k = \sum_{\substack{\nu=1 \\ \nu \text{ odd}}}^{\nu<k} \cos^{n-1}\left(\frac{\nu\pi}{2k}\right) \sin\left(\frac{\nu\pi}{2k}\right) (-1)^{\frac{\nu-1}{2}}. \tag{6}$$

Proof. The proof follows using techniques from [1]. Let $\tau_{-k,i}$ and $\tau_{k,i}$ be random variables denoting the first time the Markov chain hits the absorbing states $-k$ and k, respectively, starting from state i. Let $U_i(s), V_i(s)$ be the corresponding probability generating functions. We have:

$$u_{i,n} = P\left[\tau_{-k,i} = n\right], \, v_{i,n} = P\left[\tau_{k,i} = n\right], \tag{7}$$

$$U_i(s) = E[s^{\tau_{-k,i}}] = \sum_{n=0}^{\infty} u_{i,n} s^n, \tag{8}$$

$$V_i(s) = E[s^{\tau_{k,i}}] = \sum_{n=0}^{\infty} v_{i,n} s^n. \tag{9}$$

With probabilities a and $1-a$, respectively, the state one step after state i is $i+1$ or $i-1$. Thus we obtain the following the difference equations:

$$U_i(s) = asU_{i+1}(s) + (1-a)sU_{i-1}(s), \tag{10}$$

$$V_i(s) = (1-a)sV_{i+1}(s) + asV_{i-1}(s), \tag{11}$$

where $-k < i < k$, with the boundary conditions:

$$U_{-k}(s) = 1, U_k(s) = 0, V_{-k}(s) = 0, V_k(s) = 1. \tag{12}$$

The solutions to the above equations are:

$$U_i(s) = \frac{\lambda_1^{i+k}(s)\lambda_2^{2k}(s) - \lambda_1^{2k}(s)\lambda_2^{i+k}(s)}{\lambda_2^{2k}(s) - \lambda_1^{2k}(s)}, \tag{13}$$

$$V_i(s) = \frac{\lambda_1^{i+k}(s) - \lambda_2^{i+k}(s)}{\lambda_1^{2k}(s) - \lambda_2^{2k}(s)}, \tag{14}$$

where

$$\lambda_{1,2}(s) = \left[1 \pm \sqrt{1-4a(1-a)s^2}\right] / (2as). \tag{15}$$

Considering that our Markov chain starts at state $i=0$, $u_{0,n}$ and $v_{0,n}$ can be written as:

$$u_{0,n} = \frac{\mathrm{d}^n U_0(s)}{n!(\mathrm{d}s)^n}|_{s=0}, \; v_{0,n} = \frac{\mathrm{d}^n V_0(s)}{n!(\mathrm{d}s)^n}|_{s=0}, \tag{16}$$

which can be written in closed-form expressions as in (4) and (5). □

By symmetry, for the case $V=0$, the first time hitting probabilities are $\tilde{v}_{0,n} = u_{0,n}, \tilde{u}_{0,n} = v_{0,n}$, thus $\tilde{v}^*_{0,n} = u^*_{0,n}, \tilde{u}^*_{0,n} = v^*_{0,n}$.

The probability of wrong herding eventually, $\lim_{n\to\infty} u^*_{0,n}$, is shown in Fig. 5. At the threshold where k changes, this probability discontinuously decreases. For an observation error ϵ between thresholds, the probability of wrong herding increases with more noise.

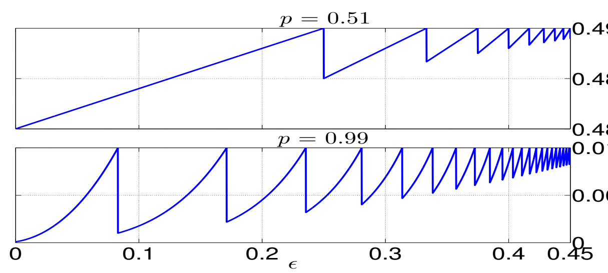

Figure 5: Probability of wrong herding eventually

IV. Agent Welfare

Let π_i be the payoff or welfare of agent i. From Section II we have that $\pi_i = 0$ if $A_i = N$, while if $A_i = Y$, π_i is either $1/2$ or $-1/2$ corresponding to $V=1$ or $V=0$, respectively. All agents i from 1 to k use their own signals, thus they all have the same welfare given by:

$$\begin{aligned} E\left[\pi_i\right] &= \frac{1}{4}\left\{P\left[A_i = Y|V=1\right] - P\left[A_i = Y|V=0\right]\right\} \\ &= \frac{2p-1}{4} > 0, \text{ since } p > 1/2. \end{aligned} \tag{17}$$

For agents $i \geq k+1$:

$$\begin{aligned} E\left[\pi_i\right] &= \frac{1}{4}\left[v^*_{0,i-1} + p(1 - v^*_{0,i-1} - u^*_{0,i-1})\right] \\ &- \frac{1}{4}\left[\tilde{v}^*_{0,i-1} + (1-p)(1 - \tilde{v}^*_{0,i-1} - \tilde{u}^*_{0,i-1})\right] \\ &= F + \frac{1-p}{2} \sum_{\substack{j=k \\ (j-k) even}}^{i-1} v_{0,j} - \frac{p}{2} \sum_{\substack{j=k \\ (j-k) even}}^{i-1} \tilde{v}_{0,j}, \end{aligned} \tag{18}$$

where $F = \frac{2p-1}{4}$ is the fixed welfare of the first k agents.

Theorem 1. *With the same signal quality p and k satisfying $\epsilon^*(k,p) \leq \epsilon < \epsilon^*(k+1,p)$, we have:*

1) *The welfare for each agent is at least equal to the welfare of his predecessors. Thus* $E[\pi_i] \geq F$ *and is non-decreasing in* i.
2) $\lim_{i\to\infty} E[\pi_i]$ *exists and equals:*

$$\Pi(\epsilon) = \frac{2p-1}{4} + \frac{1}{2}\left[\frac{1}{1+\left(\frac{1-a}{a}\right)^k} - p\right], \quad (19)$$

where $a = (1-\epsilon)p + \epsilon(1-p)$.

3) $\Pi(\epsilon)$ *decreases continuously as* ϵ *increases over a range where* k *is fixed. More specifically:*

$$\lim_{\epsilon\downarrow\epsilon^*(k,p)} \Pi(\epsilon) > \Pi(\epsilon) > \lim_{\epsilon\uparrow\epsilon^*(k+1,p)} \Pi(\epsilon) = F. \quad (20)$$

Furthermore, the maximum value of $\Pi(\epsilon)$ *for each value of* k *is decreasing in* k. *More specifically:*

$$\lim_{\epsilon\downarrow\epsilon^*(k,p)} \Pi(\epsilon) > \lim_{\epsilon\downarrow\epsilon^*(k+1,p)} \Pi(\epsilon). \quad (21)$$

Proof. An outline of the proof is as follows: 1), 3) and (21) are proved by noting that $\epsilon^*(k,p) \leq \epsilon < \epsilon^*(k+1,p)$ leads to $0 < \left(\frac{1-a}{a}\right)^k < \frac{1-p}{p} < 1$; and 2) is proved by using $\lim_{i\to\infty} E[\pi_i] - F = \frac{1}{2}[(1-p)V_0(1) - pU_0(1)]$. □

This result, illustrated in Fig. 1, implies that zero error probability, $\epsilon = 0$, yields the maximum asymptotic welfare $\Pi(\epsilon)$. But at the thresholds where k changes, Property 3 of the theorem shows that $\Pi(\epsilon)$ will discontinuously increase. Comparing Fig. 5 to Fig. 1 and noting that the discontinuities occur at the same place, this suggests that for ϵ near these values, adding more noise to the observations reduces the probability of wrong herding which leads to increased overall welfare. Our proof does not apply to an arbitrary agent. However, by plotting (17) and (18) numerically, as shown in Fig. 6 and Fig. 7, the same behaviors are seen for agents when the index is small.[2]

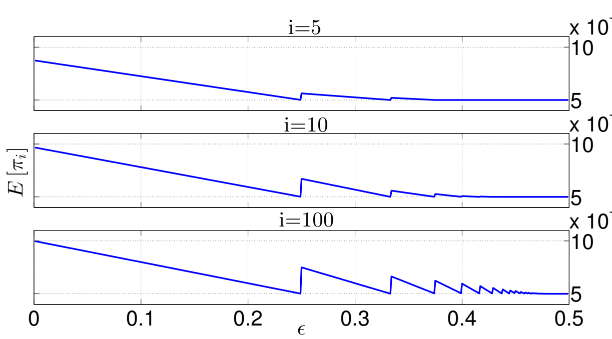

Figure 6: Agent welfare for low signal quality p=0.51.

[2]Also by summing these welfare across the agents, it can be seen that the same insights apply to the total welfare.

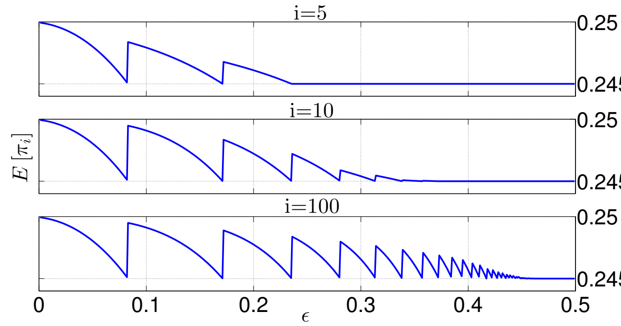

Figure 7: Agent welfare for high signal quality p=0.99.

V. Conclusions and Future Work

This paper studied the effect of noise in a simple informational cascade model. By assuming that the agents observe the actions of others through a BSC, and using a Markov chain based analysis, we determined the probabilities of herding for an arbitrary agent and used these to calculate the agents' welfare based on the given signal quality and the crossover probability in the BSC. Our main result shows that even though an error-free channel is the optimal case, a lower noise channel does not guarantee a higher total welfare as the number of agents get large. As the noise level approaches specific thresholds, there is benefit from adding a controlled amount of noise into the observations in order to achieve the next local maximum. Studying the possibility of decision errors by the agents and heterogeneous private signals are possible directions for future work.

Acknowledgment

The authors would like to thank Dr. Thanh Nguyen for useful comments and suggestions. We are grateful for financial support from NSF under grant IIS-1219071 and the Vietnam Education Foundation (VEF) Fellowship.

References

[1] W. Feller, *An introduction to Probability Theory and Its Applications*, Vol. I, 3rd ed., New York Wiley, 1968.
[2] S. Bikhchandani, D. Hirshleifer, and I. Welch. *A Theory of Fads, Fashion, Custom and Cultural Change as Informational Cascades*, J. Polit. Econ., vol. 100, No. 5, pp. 992-1026, 1992.
[3] A. Banerjee, *A Simple Model of Herd Behavior*, The Quarterly JOurnal of Economics, vol. 107, pp. 797-817, 1992.
[4] L. Smith, P. Sorensen, *Pathological Outcomes of Observational Learning*, Econometrica, vol. 68, pp. 371-398, 2000.
[5] D. Sgroi, *Optimizing Information in the Herd: Guinea Pigs, Profits, and Welfare*, Games and Economic Behaviour, vol. 39, pp. 137-166, 2002.
[6] D. Acemoglu, M. Dahleh, I. Lobel, and A. Ozdaglar, *Bayesian learning in social networks*, Review of Economic Studies, vol. 78, pp. 1201-1236, 2011.