SELFISH LEARNING: LEVERAGING THE GREED IN SOCIAL LEARNING

Ravi Kiran Raman

University of Illinois, Urbana-Champaign

Srilakshmi Pattabiraman

University of Texas, Austin

ABSTRACT

We introduce a sequential Bayesian binary hypothesis testing problem under social learning, termed *selfish learning*, where agents work to maximize their individual rewards. In particular, each agent receives a private signal and is aware of decisions made by earlier-acting agents. Beside inferring the underlying hypothesis, agents also decide whether to stop and declare, or pass the inference to the next agent. The employer rewards only correct responses and the reward per worker decreases with the number of employees used for decision making. We characterize decision regions of agents in the infinite and finite horizon. In particular, we show that the decision boundaries in the infinite horizon are the solutions to a Markov Decision Process with discounted costs, and can be solved using value iteration. In the finite horizon, we show that team performance is enhanced upon appropriate incentivization when compared to sequential social learning.

***Index Terms*—** social learning, selfish agents, MDP, incentive mechanism, time-constrained decision making

1. INTRODUCTION

Multi-agent systems in decision making tasks often benefit from reinforcing private information about the underlying hypothesis with public information from other agents in them. Sequential social learning is one such multi-agent system wherein a set of workers, in some predetermined order, perform local inference regarding the underlying hypothesis using a private signal, their individual beliefs, and the decisions made by earlier-acting agents. The last-acting agent declares the collective decision after taking into consideration the decisions made by earlier-acting agents.

Propagating local inferences provides additional information regarding the underlying problem thereby increasing the accuracy of inference. Thus, it is typically beneficial to use the social learning framework with rational human workers.

Such social learning systems have been extensively studied [1–3]. Also referred to as observational learning, the notions of "herding" and conformism have in particular been explored in detail in the social learning construct [4–6]. Benefits of including side information beside local decisions have been considered in the context of information cascades [7].

This work was supported in part by NSF Grant CCF-1717530.

Sequential social learning has also been generalized to social networks where agents learn from neighbors in [8]. This model has also been explored in fair generality in [9]. In addition, the social learning setup has also been studied under probability reweighting [10], quantized priors [11], and distributed detection with sensor fusion [12].

Social learning has been explored for quickest change detection through a POMDP formulation in [13]. Notwithstanding [13], prior work has predominantly focused on static systems with a set of agents acting in a predetermined order.

Expected utility theory [14, 15] argues that rational agents behave in a manner that maximizes their expected utility. In the social learning setting, this translates to workers making decisions that maximize their individual reward. Thus, if the employer pays a non-increasing reward per worker as a function of the number of employees used in decision making, the employees are incentivized to stop early when they are confident of their decisions. We call this behavior *selfish learning*. In this work, we consider independent private observations with additive Gaussian noise and characterize the optimal decision regions for each selfish worker.

In the infinite horizon (i.e., infinite set of workers), we show that the conditional expected rewards constitute a Markov decision process (MDP) with discounted costs. Solving the Bellman equations using value iteration on a quantized state space, we show that optimal decision boundaries can be approximated through pre-employment training.

In the finite horizon (i.e., finite number of employees), we show that for appropriate incentivization, the selfish learning achieves higher accuracy and lesser time for detection when compared to sequential social learning.

2. PROBLEM DEFINITION

Consider sequential binary hypothesis testing using a set of workers. The n-th worker receives a private signal Y_n, depending on the true hypothesis $H^* \in \{0, 1\}$, and decisions of the earlier-acting $n-1$ workers (termed public signals).

We first introduce the notations. Let $\mathbb{P}_h[\cdot] = \mathbb{P}[\cdot|H^* = h]$, $x^{(n)} = \{x_1, \ldots, x_n\}$, $[n] = \{1, \ldots, n\}$, and $f_h(\cdot)$ be the probability density function of private signal, given $H^* = h$.

Consider the task of detecting presence of a signal, in the presence of additive Gaussian noise. Let the prior probability be $p_0 = \mathbb{P}[H^* = 0]$. Then, the private signal of worker n

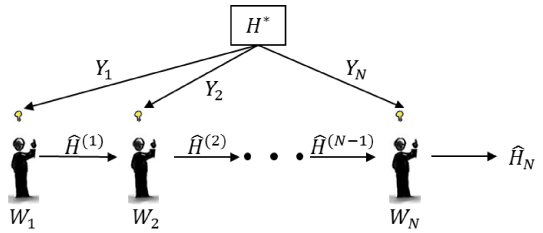

Fig. 1. Selfish learning model. Each worker has the choice to stop and declare an answer based on private & public signals.

is $Y_n = H^* + Z_n$, where $Z_n \sim \mathcal{N}(0,1)$. The results can be extended to arbitrary noise variance by the scaling signals. Worker n infers the hypothesis as $\hat{H}_n \in \{0,1\}$.

In addition, each worker can choose to stop and declare when they are confident of their inference. Let $\hat{S}_n = 1$ if worker n chooses to stop and declare, 0 otherwise. Let this stopping time be given by the random variable N.

We consider an employer who rewards quicker, accurate decisions by paying more if the system uses fewer workers to infer, and if the decision is correct. Thus, if the final decision is taken by the n-th worker, then, the reward for worker i is

$$R_i = g_n \mathbf{1}\left\{N = n, \hat{H}_n = H^*, i \le n\right\}. \tag{1}$$

We will assume that g_n is non-increasing with n.

Here we assume that workers are selfish and seek to maximize their expected reward conditional on the private and public signals, i.e., worker n decides

$$(\hat{H}_n, \hat{S}_n) = \arg\max_{(H,S)} \mathbb{E}\left[g_N \mathbf{1}\left\{\hat{H}_N = H^*\right\} \middle| \hat{H}^{(n)}, \hat{S}^{(n)}\right], \tag{2}$$

where $\hat{H}^n = \hat{H}^{(n-1)} \cup H$ and $\hat{S}^{(n)} = \hat{S}^{(n-1)} \cup S$. The selfish learning system is depicted in Fig. 1.

3. DECISION BOUNDARIES

Given the set of public signals, each worker makes the decision according to the strength of the private signal. For additive Gaussian noise model the decision regions are defined by private signal thresholds, determined by the public signals.

When the private signal is strong, the worker stops and declares, and continues otherwise. Thus, decision boundaries of the n-th worker are as shown in Fig. 2, where $\alpha_n, \beta_n, \gamma_n$ are obtained according to the public signals. Thus, the decision of the n-th worker upon receiving Y_n is

$$\hat{H}_n = \mathbf{1}\{Y_n > \gamma_n\}, \text{ and } \hat{S}_n = \mathbf{1}\{Y_n \notin (\alpha_n, \beta_n)\}. \tag{3}$$

Let the posterior probability for the n-th worker of the hypothesis being 0 given the public signals be q_n, i.e.,

$$q_n = \mathbb{P}\left[H^* = 0 \middle| \hat{H}^{(n-1)} = h^{(n-1)}, \hat{S}^{(n-1)} = 0\right]. \tag{4}$$

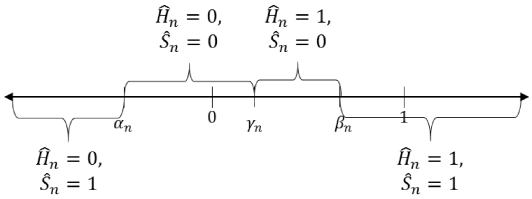

Fig. 2. Decision boundaries of n-th worker, given public signals. The worker stops only if the private signals are strong.

For ease, whenever evident from context, we will use $\hat{H}^{(n)}, \hat{H}_n$ to depict the events $\{\hat{H}^{(n)} = h^{(n)}\}, \{\hat{H}_n = h_n\}$. Then, from Bayes' rule, the posterior evolves as follows:

$$\frac{q_{n+1}}{1-q_{n+1}} = \frac{q_n}{1-q_n} \frac{\mathbb{P}_0\left[\hat{H}_n \middle| \hat{H}^{(n-1)}\right]}{\mathbb{P}_1\left[\hat{H}_n \middle| \hat{H}^{(n-1)}\right]}. \tag{5}$$

Threshold γ_n is used to infer the hypothesis accurately and is not affected by the reward. Further, as we have a social learning system incentivized to be correct, worker n uses the posterior probability q_n to compute the maximum a posteriori (MAP) estimate. Thus $\gamma_n = \frac{1}{2} + \log\left(\frac{q_n}{1-q_n}\right)$.

The stopping thresholds are defined by the private signal $Y_n = y_n$ for which the expected reward obtained upon declaring matches that obtained upon passing.

First, let us compute the expected reward if worker n stops and declares 1, i.e., $y_n \ge \gamma_n$. Then, from independence of private signals and Bayes rule, expected reward is given by

$$\begin{aligned}\mathbb{E}\left[R_N \middle| \begin{pmatrix}\hat{H}^{(n-1)}, Y_n, \\ N = n\end{pmatrix}\right] &= g_n \mathbb{P}\left[H^* = 1 \middle| \hat{H}^{(n-1)}, Y_n\right] \\ &= g_n \frac{(1-q_n) f_1(y_n)}{q_n f_0(y_n) + (1-q_n) f_1(y_n)}.\end{aligned} \tag{6}$$

Similarly, when worker n stops and declares $\hat{H}_n = 0$,

$$\mathbb{E}\left[R_N \middle| \begin{pmatrix}\hat{H}^{(n-1)}, Y_n, \\ N = n\end{pmatrix}\right] = g_n \frac{q_n f_0(y_n)}{q_n f_0(y_n) + (1-q_n) f_1(y_n)}.$$

We now compute expected reward when the worker passes. Assume that $y_n \in [\gamma_n, \beta_n]$, i.e., $\hat{H}_n = 1, \hat{S}_n = 0$. Then, expected reward is

$$\begin{aligned}\mathbb{E}\left[R_N \middle| \begin{pmatrix}\hat{H}^{(n-1)}, Y_n, \\ N > n\end{pmatrix}\right] &= \tilde{q}_n \mathbb{E}_0\left[R_N | \hat{H}^{(n)}, N > n\right] \\ &\quad + (1-\tilde{q}_n)\mathbb{E}_1\left[R_N | \hat{H}^{(n)}, N > n\right],\end{aligned}$$

where $\hat{H}_n = \mathbf{1}\{y_n \in (\gamma_n, \beta_n)\}$, and $\frac{\tilde{q}_n}{1-\tilde{q}_n} = \frac{q_n}{1-q_n}\frac{f_0(y_n)}{f_1(y_n)}$.

Setting the expected rewards upon passing and that upon stopping equal at $Y_n \in \{\alpha_n, \beta_n\}$, we get

$$\alpha_n = \gamma_n - \left[\log\left[\frac{\mathbb{E}_1\left[R_N|\hat{H}^{(n)}, N>n\right]}{g(n)-\mathbb{E}_0\left[R_N|\hat{H}^{(n)}, N>n\right]}\right]\right]^+, \tag{7}$$

$$\beta_n = \gamma_n + \left[\log\left[\frac{\mathbb{E}_0\left[R_N|\hat{H}^{(n)}, N>n\right]}{g(n)-\mathbb{E}_1\left[R_N|\hat{H}^{(n)}, N>n\right]}\right]\right]^+, \tag{8}$$

where, $x^+ = \max\{0, x\}$. Eqns. (7) and (8) relate the incentives to decision boundaries. We now consider infinite and finite horizons of the problem.

4. LEARNING IN THE INFINITE HORIZON

We now study infinite horizon selfish learning. We prove the reduction to a discounted cost Markov Decision Process (MDP). This results in a dynamic programming formulation which can be solved to obtain optimal decision boundaries.

Let the expected reward under an incentive profile $\{g_n\}_{n\in\mathbb{N}}$ and prior p, conditional on the true hypothesis $H^* = h$ be $\mathcal{R}_h(g,p) = \mathbb{E}[R_N|H^*=h; g, p]$. Let $\tilde{g}^{(n)}_{n'-n} = \frac{g_{n'}}{g_n}$, for all $n' > n$. Then, from (7) and (8),

$$\alpha_n = \gamma_n - \left[\log\left[\frac{\mathcal{R}_1\left(\tilde{g}^{(n)}, q^{(0)}_{n+1}\right)}{1-\mathcal{R}_0\left(\tilde{g}^{(n)}, q^{(0)}_{n+1}\right)}\right]\right]^+, \tag{9}$$

$$\beta_n = \gamma_n + \left[\log\left[\frac{\mathcal{R}_0\left(\tilde{g}^{(n)}, q^{(1)}_{n+1}\right)}{1-\mathcal{R}_1\left(\tilde{g}^{(n)}, q^{(1)}_{n+1}\right)}\right]\right]^+, \tag{10}$$

where $q^{(h)}_{n+1}$ is the posterior given $\hat{H}_n = h$, for $h \in \{0,1\}$. That is, the optimal decision boundaries for worker n depend on the conditional expected rewards of the first worker, if the problem had a prior of $q^{(h)}_{n+1}$ and an incentive profile of $\tilde{g}^{(n)}$.

Consider $g_n = \rho^{n-1}$, for some $\rho \in (\frac{1}{2}, 1)$. Then, $\tilde{g}^{(n)}_m = \rho g_m$. From (9) and (10) optimal stopping boundaries are only a function of posteriors for a fixed ρ.

We suppress the payoff structure now for convenience. Let $u_n = (\alpha_n, \beta_n)^T \in \mathbb{R}^2$ be the optimal decision boundaries. The conditional expected reward for $h \in \{0,1\}$, starting from a posterior q_n, satisfies

$$\mathcal{R}_h(q_n) = C^{(h)}(q_n, u_n) + \rho\mathbb{P}_h[Y_n \in (\alpha_n, \gamma_n)]\,\mathcal{R}_h\left(q^{(0)}_{n+1}\right) + \rho\mathbb{P}_h[Y_n \in (\gamma_n, \beta_n)]\,\mathcal{R}_h\left(q^{(1)}_{n+1}\right), \tag{11}$$

where $C^{(h)} : [0,1]\times\mathbb{R}^2 \to [0,1]$, is defined as

$$C^{(h)}(q_n, u_n) = \begin{cases} \mathbb{P}_0[Y_n \le \alpha_n] & \text{, if } h = 0 \\ \mathbb{P}_1[Y_n \ge \beta_n] & \text{, if } h = 1. \end{cases}$$

For each $h \in \{0,1\}$, let $\mathrm{J}^{(h)}(q) = \mathcal{R}_h(q)$. Further, for any $q, q' \in [0,1]$, and $u = (\alpha,\beta)^T$, let

$$P^{(0)}_{q,q'}(u) = \begin{cases} \mathbb{P}_0[Y\in(\alpha,\gamma)] & \text{, if } \frac{q'}{1-q'} = \frac{q}{1-q}\frac{\mathbb{P}_0[Y\in(\alpha,\gamma)]}{\mathbb{P}_1[Y\in(\alpha,\gamma)]}, \\ \mathbb{P}_0[Y\in(\gamma,\beta)] & \text{, if } \frac{q'}{1-q'} = \frac{q}{1-q}\frac{\mathbb{P}_0[Y\in(\gamma,\beta)]}{\mathbb{P}_1[Y\in(\gamma,\beta)]}, \\ \mathbb{P}_0[Y\notin(\alpha,\beta)] & \text{, if } q' = \phi \end{cases},$$

where ϕ is a proxy state corresponding to stopping. We fix the reward of this state as 0. Analogously define $P^{(1)}_{q,q'}(u)$.

Then, $\mathcal{R}_h$ are solutions to the Bellman equations

$$J^{(h)}(p) = \max_{u\in\mathbb{R}^2} C^{(h)}(p,u) + \rho\mathbb{E}_{q\sim P^{(h)}_{p,q}(u)}\left[J^{(h)}(q)\right]. \tag{12}$$

Let $T^{(h)}(\cdot)$ be the operator from (12), such that $J^{(h)} = T^{(h)}(J^{(h)})$. We suppress hypothesis h, owing to symmetry.

Theorem 1. *Operator T is a contraction under $\|\cdot\|_\infty$ norm.*

The proof follows similar to standard proofs of contraction of cost operator in reinforcement learning using monotonicty and discounted increase of the $T(\cdot)$ operator. From Banach's fixed point theorem [16], we get the following.

Corollary 2. *There exists a unique solution J^* to $J = T(J)$.*

Thus value iteration can be used to determine J^*.

Lemma 3. *For any $\hat{J}_0$, if $\hat{J}_{n+1} = T(\hat{J}_n)$, for all $n \ge 0$, then*

$$\|\hat{J}_n - J^*\|_\infty \le \frac{\rho^n}{1-\rho}\|\hat{J}_1 - \hat{J}_0\|_\infty. \tag{13}$$

The result follows from Thm. 1 and Cor. 2. However, J is a function on $[0,1]$ and so it is expensive to store and update the value iterations. It is also computationally expensive to compute $T(J)$ as it requires solving (9) and (10) in $\mathbb{R}^2$.

Note that $J^{(h)}(p) \le 1$ for all p, and is continuous in p. Hence we can approximate the conditional cost by quantizing the state-action space. We quantize the space of priors, $[0,1]$, into K levels, $\mathcal{P}_K = \{p_1, \ldots, p_K\}$. Similarly, quantize the action space as $\mathcal{U}_{K'} = \{u^{(1)}, \ldots, u^{(K')}\}$, where $u^{(i)} \in \mathbb{R}^2$.

Let the fixed point for the reduced version of the problem be $\tilde{J}^{(K,h)}$, state-transition probability be $\tilde{P}^{(h)}_{q,q'}(u)$ where the states and actions are appropriately replaced by their quantized equivalents, and the corresponding operator be $\tilde{T}^{(K,h)}$.

To perform value iteration, choose an initial cost vector $\hat{J}^{(K,h)}_0 \in \mathbb{R}^K$, and iteratively apply $\hat{J}^{(K,h)}_{t+1} = \tilde{T}^{(K,h)}(\hat{H}^{(K,h)}_t)$ using (9), and (10). Let the converged estimate be $\hat{J}^{(K,h)}_T$. Finally, approximate $J^{(h)}$ by choosing sufficiently large K, K' and interpolating the estimate $\hat{J}^{(K,h)}_T$.

Such a reduction to Bellman equations indicates that by training workers prior to employment, we can obtain optimal performance of selfish workers under geometric payoff in the infinite horizon [16]. In particular, owing to limited computational resources, humans tend to perceive quantized estimates of prior and posterior probabilities [17]. Thus, it suffices to estimate the approximate versions from the value iteration.

Further guarantees on convergence and the feasibility of training with humans are beyond the scope of this paper.

5. LEARNING UNDER FINITE HORIZON

We now consider finite horizon selfish learning where the employer can use a maximum of M workers. Thus, the incentive is now equivalently given by a vector $g \in \mathbb{R}^M$.

The system always declares a final decision. Thus, $\alpha_M = \gamma_M = \beta_M$, for any sequence of prior decisions $\hat{H}^{(M-1)}$.

We study the relationship between incentives, accuracy, and stopping time. Let $\beta_n = -\alpha_n = \infty$, if $N < n$. Then,

$$\mathbb{P}\left[\hat{H}_N = H^*\right] = p_0 \sum_{n=1}^{M} \mathbb{E}_0\left[\Phi(\alpha_n)\right] + \bar{p}_0 \sum_{n=1}^{M} \mathbb{E}_1\left[Q(\beta_n - 1)\right],$$

where $\bar{p}_0 = 1 - p_0$ and $\Phi(x) = 1 - Q(x)$. Similarly,

$$\mathbb{P}\left[N = n\right] = p_0 \mathbb{E}_0\left[\Phi(\alpha_n) + Q(\beta_n)\right] + \bar{p}_0 \mathbb{E}_1\left[\Phi(\alpha_n - 1) + Q(\beta_n - 1)\right].$$

Finally, expected cost to employer is $\mathbb{E}\left[\text{Cost}\right] = \mathbb{E}\left[N g_N\right]$.

Theorem 4. *For any $M \geq 2$, if $g_n = 1$ for all $n \in [M]$, then*

$$\mathbb{P}\left[\hat{H}_N = H^*\right] \geq \mathbb{P}\left[\hat{H}_M = H^*\right], \text{ and } \mathbb{E}\left[N\right] < M. \quad (14)$$

The proof follows from the observation that a worker stops if and only if the probability of success exceeds that from passing. Further, for the Gaussian model there is a non-zero probability of stopping before M.

Thus, by including the stopping option, we get quicker and more accurate detection compared to social learning.

5.1. Two Worker System

Consider a system with two workers, $M = 2$. Here, worker 1 can stop and declare, or pass to worker 2, who behaves as a MAP decoder, using the posterior probability.

Now, $\alpha_2^{(h)} = \gamma_2^{(h)} = \beta_2^{(h)} = \frac{1}{2} + \log\left(\frac{q_2^{(h)}}{1-q_2^{(h)}}\right)$, for any $\hat{H}_1 = h \in \{0, 1\}$. Here, according to (5),

$$\frac{q_2^{(h)}}{1 - q_2^{(h)}} = \frac{p_0}{(1-p_0)} \frac{(Q(x_h) - Q(\gamma_1))}{(Q(x_h - 1) - Q(\gamma_1 - 1))}, \quad (15)$$

where $x_0 = \alpha_1, x_1 = \beta_1$. Further, $\gamma_1 = \frac{1}{2} + \log\left(\frac{p_0}{1-p_0}\right)$, and α_1, β_1 are obtained from (7) and (8).

The cost-accuracy tradeoff for various priors is shown in Fig. 3. Note that the inference error decreases with increasing budget, and that the expected cost is strictly less than 2.

We now compare selfish learning with sequential social learning, for various incentives. Let $g_1 = 1$. When g_2 is low, the first worker benefits more from stopping and declaring and so declares more often which results in lesser accuracy.

As g_2 increases, the first worker gains more from passing weak signals to the second worker and hence accuracy improves and is maximized for $g_2 = 1$ (cost also increases).

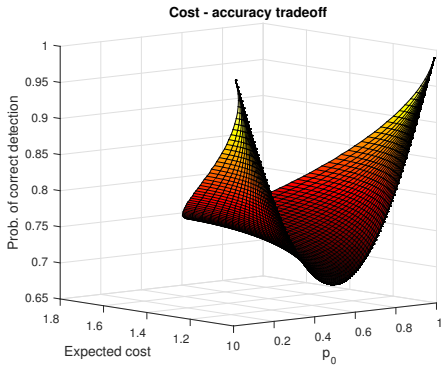

Fig. 3. Accuracy-cost tradeoff for various priors. For a given prior, accuracy can be improved by allocating larger budgets.

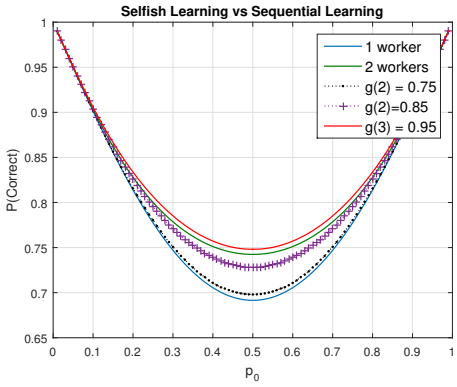

Fig. 4. Comparison of sequential and selfish learning for varying payments.

The performance of the selfish learner for various cost profiles, as compared with sequential learning is shown in Fig. 4. As expected, for any $g_2 > \frac{1}{2}$, selfish learning performs better than using just one worker.

Further, there exists $\theta < 1$ such that for $g_2 \in [\theta, 1]$, selfish learning outperforms social learning with 2 workers. Thus, through an appropriate choice of incentive, the employer can leverage worker greed to improve accuracy.

6. CONCLUSION

In this work we introduce *selfish learning* that considers the sequential social learning with agents who aim to maximize individual expected reward. We characterize decision regions for each agent under an additive Gaussian noise model.

In the infinite horizon we prove a novel reduction to an MDP with discounted costs. Through value iteration on a quantized state-action space, we approximate the decision regions. In the finite horizon, we showed that appropriate incentivization reduces the expected cost and time for decision making, and achieves better accuracy than social learning.

7. REFERENCES

[1] Glenn Ellison and Drew Fudenberg, "Rules of thumb for social learning," *J. Polit. Econ.*, vol. 101, no. 4, pp. 612–643, Aug. 1993.

[2] Lones Smith and Peter Sørensen, "Pathological outcomes of observational learning," *Econometrica*, vol. 68, no. 2, pp. 371–398, Mar. 2000.

[3] Abhijit Banerjee and Drew Fudenberg, "Word-of-mouth learning," *Games Econ. Behav.*, vol. 46, no. 1, pp. 1 – 22, Jan. 2004.

[4] Abhijit V. Banerjee, "A simple model of herd behavior," *Quart. J. Econ.*, vol. 107, no. 3, pp. 797–817, Aug. 1992.

[5] Sushil Bikhchandani, David Hirshleifer, and Ivo Welch, "Learning from the behavior of others: Conformity, fads, and informational cascades," *J. Econ. Perspect.*, vol. 12, no. 3, pp. 151–170, 1998.

[6] Venkatesh Bala and Sanjeev Goyal, "Conformism and diversity under social learning," *Econ. Theor.*, vol. 17, no. 1, pp. 101–120, Jan. 2001.

[7] T. N. Le, V. G. Subramanian, and R. A. Berry, "Quantifying the utility of imperfect reviews in stopping information cascades," Dec. 2016, pp. 6990–6995.

[8] Douglas Gale and Shachar Kariv, "Bayesian learning in social networks," *Games Econ. Behav.*, vol. 45, no. 2, pp. 329–346, Nov. 2003.

[9] Daron Acemoglu, Munther A. Dahleh, Ilan Lobel, and Asuman Ozdaglar, "Bayesian learning in social networks," *Rev. Econ. Stud.*, vol. 78, no. 4, pp. 1201–1236, Oct. 2011.

[10] Joong Bum Rhim and Vivek K Goyal, "Social teaching: Being informative vs. being right in sequential decision making," in *Proc. 2013 IEEE Int. Symp. Inf. Theory*, July 2013, pp. 2602–2606.

[11] Joong Bum Rhim, Lav R. Varshney, and Vivek K Goyal, "Quantization of prior probabilities for collaborative distributed hypothesis testing," *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4537–4550, Sept. 2012.

[12] Joong Bum Rhim and Vivek K Goyal, "Distributed hypothesis testing with social learning and symmetric fusion," *IEEE Trans. Signal Process.*, vol. 62, no. 23, pp. 6298–6308, Dec. 2014.

[13] V. Krishnamurthy, "Quickest detection pomdps with social learning: Interaction of local and global decision makers," *IEEE Trans. Inf. Theory*, vol. 58, no. 8, pp. 5563–5587, Aug 2012.

[14] Peter C Fishburn, *Nonlinear preference and utility theory*, vol. 5, Johns Hopkins University Press Baltimore, 1988.

[15] John Quiggin, "A theory of anticipated utility," *J. Econ. Behav. Org.*, vol. 3, no. 4, pp. 323 – 343, 1982.

[16] Dimitri P. Bertsekas, *Dynamic Programming and Optimal Control*, vol. 1, Athena Scientific, Belmont, MA, 3 edition, 2005.

[17] Kush R. Varshney and Lav R. Varshney, "Quantization of prior probabilities for hypothesis testing," *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 4553–4562, Oct. 2008.