Abstract—A network of social sensors estimate an unknown state of nature by performing Bayesian Social Learning, and myopically optimize individual reward functions. The decisions of the social sensors contain quantized information about the underlying state. How should a fusion center dynamically incentivize the social sensors for acquiring the information about the underlying state?

This paper presents four results. First, sufficient conditions on the model parameters are provided under which the optimal policy for the fusion center has a threshold structure. The optimal policy is determined in closed form, and is such that it switches between two exactly specified incentive policies at the threshold. Second, a sample path characterization of the optimal threshold policy is provided, i.e., the nature of the incentives (average trend over time) resulting from the fusion center employing the optimal threshold policy. It is shown that the optimal incentive sequence is a sub-martingale, i.e., the optimal incentives increase on average over time. Third, it is shown that it is possible for the fusion center to learn the true state asymptotically by employing a sub-optimal policy; in other words, controlled information fusion with social sensors can be consistent. Finally, uniform bounds on the average additional cost incurred by the fusion center for employing a sub-optimal policy are provided. This characterizes the trade-off between the cost of information acquisition and consistency, for the fusion center.

I. INTRODUCTION

A social sensor is an information processing system that differs from a physical sensor in the following ways:

i.) Social sensors influence the behavior of other sensors, whereas physical sensors typically do not affect other sensors.

ii.) Social sensors reveal quantized information (decisions) and have dynamics, whereas physical sensors are static with the dynamics modeled in the state equation. Social learning is the process by which social sensors are influenced by the behavior of other sensors in a multi-sensor network. The availability of online social media review platforms like Yelp, Expedia, Amazon etc, facilitates social learning; see [1], [2]. Social learning shares similarities with decentralized detection [3], [4] that falls within the class of team decision theory [5], [6]; but with key differences: Decentralized detection quantizes the observations, whereas social learning quantizes the Bayesian belief [7]. In decentralized detection the fusion policies are directly optimized, whereas in social learning the fusion rule is prescribed and is Bayesian.

Data fusion with physical sensors is a well studied problem. Information fusion with social sensors is challenging due to the fact that social learning leads to inefficiencies [7]–[9] like herds (sensors choose the same action irrespective of their private information) and informational cascades (information fusion results in no improvement in uncertainty). So having more social sensors need not always be advantageous (in terms of reduced mean square error between the state estimate and the true state).

Multi-sensor data fusion [10] on the other hand, refers to the problem of data acquisition, processing, and fusion of information, to provide a better estimate of the underlying state. A data fusion center gathers the information from the peripheral sensors (physical sensors) to make an informed decision regarding the desired parameter. Having more number of sensors leads to improvement in reliability, resolution, coverage, and confidence; see [10].

Traditionally, information fusion is open-loop; in this paper, we use feedback control to choose incentives to control how the sensors provide information. Hence we name the problem considered in this paper as controlled information fusion. The fusion is Bayesian and we are interested in designing the control laws for providing optimal incentives for social sensors that will result in accurate Bayesian estimates.

Information Fusion with Social Sensors

Controlled sensing refers to the stochastic control problem where a controller decides how the network of sensors should adapt so as to obtain improved measurements; or alternatively minimize a measurement cost. Some common applications include sensor scheduling [12], [13], measurement control [14], [15], and active hypothesis testing [16]–[18].

Controlled Sensing vs Controlled Information Fusion

Controlled sensing [11] refers to the stochastic control problem where a controller decides how the network of sensors should adapt so as to obtain improved measurements; or alternatively minimize a measurement cost. Some common applications include sensor scheduling [12], [13], measurement control [14], [15], and active hypothesis testing [16]–[18].

S. Bhatt (sh2376@cornell.edu) and V. Krishnamurthy (vikramk@cornell.edu) are with the Department of Electrical and Computer Engineering, Cornell Tech, Cornell University, New York, NY 10044.

1 A social (human) sensor provides information about its state (sentiment, social situation, quality of product) to a social network after interaction with other social sensors. In this paper, in line with a large body of literature, we adopt a more stylized definition: a social sensor performs social learning.
Many online social media review platforms like Amazon, Yelp, TripAdvisor, Airbnb etc, encourage sharing experiences of services, products, and vendors. Such review platforms offer the following benefits: (i) future customers are influenced by them, and (ii) the retailers can act on them to improve the quality of product or service. However, if such review platforms are to be a reliable source of information, the customers should leave an honest review. A review is honest if it reflects customer’s observations and experiences. How to dynamically incentivize the customers to encourage them to leave an honest review? (see Sec.V-A)

Main Results and Organization

In the context of controlled information fusion, this paper has 3 main results:

1. **Optimality of Threshold Incentive Policy**: Sec.III-B gives sufficient conditions on the model parameters under which the optimal incentive policy for the fusion center has a threshold structure (see Fig.1a). Indeed we will show that the optimal policy switches between two exactly specified incentive policies at the threshold, and hence is completely determined in closed form. Since the optimal policy is determined in closed form, the fusion center only needs to store the optimal incentive policy for choosing incentives to social sensors to reveal information to the fusion center. In general, the incentive policy under which the incentive policy is as in Fig.1a. Here $\mu^*(\pi) \in [0, 1]$ computed from the fusion of sensors’ decisions. $e_1$ and $e_2$ denote the indicator vectors. When $p = \mu^*(\pi) = 0$, the fusion center should not incentivize. The optimal policy is a choice between two exactly specified incentive policies, and hence is determined in closed form. Sec.V-A provides numerical examples (of more general cost functions) where the optimal incentive policy is a multi-threshold as in Fig.1b.

2. **Sub-martingale Property of Optimal Incentive Sequence**: While Sec.III-B establishes the structure of the optimal incentive policy, Sec.III-C establishes the sample path properties of the optimal incentive sequence. In particular, we show that the optimal incentive sequence is a sub-martingale; i.e., the incentives increase on average over time. This property is useful in assessing the reliability of the fusion center. In a related context, our result is similar to the super-martingale property of pricing policies in economics [22], [23]; which says that the optimal pricing policy for charging sensors (performing social learning) who purchase a product, is to start high, establish an elite customer base, and then decrease prices to increase profits.

3. **Consistency of Controlled Information Fusion**: Information fusion with social sensors is challenging due to the fact that social learning terminates after a finite horizon [11], [24] due to the formation of information cascades. Sec.IV shows how the fusion center can control the incentives and learn the true state asymptotically by employing a sub-optimal policy; in other words, how to control the incentives such that the information fusion with social sensors is consistent (convergence in probability). However, by employing a sub-optimal policy, the fusion center incurs additional cost. Therefore, uniform bounds on the average additional cost incurred by the fusion center for employing a sub-optimal policy are provided. These bounds characterize the trade-off between the cost of information acquisition and consistency, for the fusion center.

Sec.V presents numerical examples that provide additional insights on the main results.

II. **Social Learning Model and Fusion Center Objective**

We consider the setup illustrated in Fig.2. The fusion center controls the incentives given to the social sensors, and the social sensors share their decisions (quantized information on the underlying state) with the fusion center. Sec.II-A describes the controlled fusion social learning model that governs the manner in which the social sensors learn from each other, and how this behavior is influenced by the fusion center. Sec.II-B describes the insights on the main results.
formulates the objective of the fusion center that captures the trade-off between the cost of information acquisition from the social sensors versus the usefulness of the information measured.

A. Controlled Fusion Social Learning Model

Let \( k = 1, 2, \cdots \) denote the discrete time instants. It is assumed that each sensor decides once in a predetermined sequential order indexed by \( k \). Let \( x_k \in X = \{1, 2\} \) denote the state of nature. The state \( x \) is assumed to be a random variable chosen at \( k = 0 \). The fusion center dynamically chooses the incentives at each sensor \( k \) to estimate the realization of the random variable \( x \). Let the probability mass function of the state at time \( k = 1 \) be denoted as

\[
\pi_{k-1}(i) = P(x = i | a_1, \ldots, a_{k-1}).
\]

The state estimate is computed from the decisions of the social sensors \( a_1, \ldots, a_{k-1} \) and is termed as the public belief. Let the initial estimate be denoted as \( \pi_0 = (\pi_0(i), i \in X) \), where \( \pi_0(i) = P(x_0 = i) \). Let the belief space, i.e., the set of distributions \( \pi \) over the state be denoted as

\[
\Pi(2) = \{ \pi \in \mathbb{R}^2 : \pi(1) + \pi(2) = 1, 0 \leq \pi(i) \leq 1 \text{ for } i \in \{1, 2\} \}.
\]

Social Sensor Dynamics: A social sensor, unlike a physical sensor, has its own dynamics since it learns from previous actions of other social sensors. It receives an observation on the underlying state, computes an estimate (private belief) using the information revealed by other sensors (their decisions), and takes an action to myopically maximize a reward function. This action/decision is a quantization of the (private) belief, and is shared with the fusion center and other sensors.

1.) Social Sensor’s Private Observation: Each social sensor \( k \)’s obtains a noisy \( y_k \in Y = \{1, 2\} \) of the underlying state \( x \) with observation likelihood distribution:

\[
B_{ij} = P(y_k = j | x = i).
\]

The (discrete) observation likelihood distribution models the (limited) information gathering capabilities of the sensor.

2.) Social Learning and Private Belief update: Sensor \( k \) updates its private belief \( \pi_{k-1} \) by fusing observation \( y_k \) and the prior public belief \( \pi_{k-1} \), via the following classical Bayesian update

\[
\eta_{yk} = \frac{B_{yk} \pi_{k-1}}{\sum_{y} B_{yk} \pi_{k-1}},
\]

where \( B_{yk} \) denotes the diagonal matrix

\[
\begin{bmatrix}
P(y_k | x = 1) & 0 \\
0 & P(y_k | x = 2)
\end{bmatrix}
\]

and \( I \) denotes the 2-dimensional vector of ones.

3.) Social Sensor’s Action: Sensor \( k \) executes an action \( a_k \in \mathcal{A} = \{1, 2\} \) myopically to maximize the reward. Let \( r(x, y, a_k) \) denote the reward accrued if the sensor takes action \( a_k \) when the underlying state is \( x \) and the observation is \( y \).

For notational simplicity, we assume that all social sensors have the same reward function \( r(x, y, a) \) with

\[
r(x, y, a) = \delta_a p - \alpha_a I(a \neq x) - \beta_a I(a 
eq y) - \gamma_a.
\]

Here \( \delta_a, \alpha_a, \beta_a, \gamma_a \in [0, 1] \). The reward is inspired by the quasi-linear utility in [30]. Here, \( \delta_a \) is interpreted as the fraction of the monetary compensation \( p \) received, \( \alpha_a \) and \( \beta_a \) are the losses incurred for not taking appropriate actions, and \( \gamma_a \) is the cost incurred in information acquisition.

Let \( r(x, a) = \sum_{y=1}^{2} r(x, y, a) I_{xy} \). The sensor chooses an action \( a_k \) to maximize the reward:

\[
a_k = \arg \max_{a \in \mathcal{A}} r'(a \eta_{yk}, r_a = [r(1, a) r(2, a)]).
\]

Information Fusion cost: The fusion center minimizes the following cost of information fusion \( c(p_k) \), with

\[
c(p_k) = p_k - \Phi_s(k) I(a_k = y_k | \pi_{k-1}).
\]

Here models the trade-off between the cost of information acquisition (\( p_k \)) from the social sensors versus usefulness of the information \( I(a_k = y_k | \pi_{k-1}) \). The information from different sensors is allowed to be weighted differently using \( \Phi_s(k) \in (0, 1) \). For simplicity, we assume the weights to be same for all sensors; i.e \( \Phi_s(k) = \phi_s, \forall k \). Let us briefly discuss the cost [6]. We show in Sec. [17] that \( a = y \) corresponds to informative decisions. Since the sensors take

3The classical social learning model is sequential and each sensor acts once; see [7], [8], [23]. Repeated social learning over graphs is also well studied; see [25] and the references therein.

4Two states is common in machine learning using Crowdsourcing, for example, in image annotation, labelling etc; see [26]. More recently, in crowdsourced classification of autism spectrum disorder (ASD) and attention deficit hyperactivity disorder (ADHD); see [27].

5Each sensor being an expected reward maximizer is in line with being rational; see [24]. See [29], [29] for models considering a mix of rational and irrational sensors.

6Here informativeness is in the sense of Blackwell [11].
into account the actions or decisions of the preceding sensors, fusion of informative decisions leads to improved estimate of the parameter, and hence improves the usefulness of information (in terms of reduction in the uncertainty of the Bayesian state estimate) fused by the fusion center and the successive sensors.

**Public Belief Dynamics:** The fusion center shares sensor $k$’s decision with the multi-agent sensor network and the public belief (1) is updated (by the fusion center and subsequent sensors) according to the social learning Bayesian filter (see [11], [31]) as follows:

$$
\pi_k = T^\pi(\pi_{k-1}, a_k) = \frac{R_{a_k}^{\pi_{k-1}}\pi_{k-1}}{1\cdot R_{a_k}^{\pi_{k-1}}\pi_{k-1}}. \quad (7)
$$

Here, $R_{a_k}^{\pi_{k-1}} = \text{diag}(P(a_k|x = i, \pi_{k-1}), i \in \mathcal{X})$ is the decision or action likelihood matrix (compare with the observation likelihood matrix $B$ in (2)), where

$$
R_{a_k}^{\pi_{k-1}} = P(a_k|x = i, \pi_{k-1}) = \sum_{y \in \mathcal{Y}} P(a_k|y, \pi_{k-1})P(y|x = i), \quad (8)
$$

$$
P(a_k|y, \pi_{k-1}) = \begin{cases} 1 & \text{if } a_k = \arg \max_{a \in A} r_a'\eta_{yk}; \\ 0 & \text{otherwise.} \end{cases}
$$

Note that $\pi_k \in \Pi(2)$.

**Remark (Information Cascade).** Note that the (decision) likelihood probability (3) is an explicit function of the prior (public belief) $\pi_{k-1}$. This is unlike a standard Bayesian update (like (3)), where the likelihood is independent of the prior. This unusual update of the social learning filter leads to herding behavior: In (8), if the action becomes independent of the observation, $R_{a_k}^{\pi_{k-1}} = 1$ or 0. This in turn leads to information cascade, social learning stops as the public belief is frozen, as can be seen from (7). It can be shown that (Theorem 5.3.1, [11]) social learning stops in finite time.

**Information Fusion Incentive:** The fusion center incentivizes/compensates the social sensors for providing information about the underlying state. The fusion center dynamically adapts these incentives over time as the sensors perform social learning: each sensor will have a different expected reward. Let $F_k$ denote the history of past incentives and decisions $\{\pi_0, p_1, a_1, \ldots, p_{k-1}, a_k\}$ recorded by the fusion center and the network of social sensors. More technically,

$$
F_k := \sigma\text{-algebra generated by } (\pi_0, a_1, \ldots, a_k, p_1, \ldots, p_{k-1}). \quad (9)
$$

The fusion center chooses the incentive as $p_{k+1} = \mu(F_k)$, for the sensor $k + 1$ to provide information about its state via social learning. Here $\mu$ denotes the decision policy that associates the history $F_k$ with an incentive $p_{k+1}$. Since $F_k$ is increasing with time $k$ (filtration), to implement a controller, it is useful to obtain a sufficient statistic that does not grow in dimension. The public belief $\pi_k$ computed via the social learning filter (7) forms a sufficient statistic for $F_k$ and the incentive in is given as

$$
p_{k+1} = \mu(\pi_k) \in [0, 1]. \quad (10)
$$

**B. Controlled Information Fusion Objective**

Given the setup in Sec 11-A, the aim of the fusion center is estimate the state $x$ by minimizing the cost of information acquisition ($\rho$). As discussed in (7), the fusion center performs Bayesian fusion of the information revealed by the social sensors. For each initial distribution $\pi_0$, the following cost is associated for the fusion center:

$$
J_\mu(\pi_0) = E_\mu\{\sum_{k=1}^{\infty} \rho^k c_\mu(p_k)\mid \pi_0 = \pi\}. \quad (11)
$$

Here $p_k$ denotes the incentive, $\rho \in [0, 1)$ denotes an economic discount factor, $\mu$ denotes the decision policy for the fusion center that maps the public belief (1) to an incentive $p_k \in [0, 1]$, i.e., $p_k = \mu(\pi_{k-1})$, $c_\mu(p_k)$ denotes the cost of information fusion incurred at time $k$ by the fusion center by employing the decision policy $\mu$, and $E_\mu$ denotes the expectation conditioned on the policy $\mu$.

The policy $\mu$ can be restricted to the class of stationary (time invariant) policies for the infinite horizon discounted cost objective; see [11]. The fusion center aims to find the optimal stationary policy $\mu^*$ such that

$$
J_{\mu^*}(\pi_0) = \inf_{\mu \in \Pi} J_\mu(\pi_0) \quad (12)
$$

where $\mu$ denotes the class of stationary policies.

**Summary:** (11) is the optimization objective and (7) are the dynamics for the controlled information fusion problem considered in this paper. The model parameters are the sensors’ observation matrix $B$ in (2) and the reward $r_a$ in (4).

**C. Example: Social Media Review Platform**

We briefly motivate the above model and optimization objective in terms of online social media review platforms like Amazon or Airbnb. The information fusion objective is to estimate the product or service quality. The state $x \in \{1(\text{Bad quality}), 2(\text{Good quality})\}$, the observation $y \in \{1(\text{Bad experience}), 2(\text{Good experience})\}$, and the customers’ decision $a \in \{1(\text{Negative Review}), 2(\text{Positive Review})\}$. When the customer writes a good/ bad review when it has a good/ bad experience, the review is honest. The reward parameters $\delta_a\rho$ indicates the compensation in exchange for the review after tax. $\alpha_a$ models the cost incurred for making a decision not appropriate for the quality, $\beta_a$ models the cost incurred for acquiring information/ cost of writing the review. Here it is assumed that each customer leaves a review however, the nature of review depends on the optimization (5). The review platform shares the incentive value in addition to the reviews with the customers to avoid the perception of skewing reviews.  

---

7This can be achieved, for example, by providing free shipping and smaller handling time in case of an e-commerce social media review platform like Amazon.

8This is currently not provided by popular review sharing platforms. Disclosing incentives in the review context would ensure transparency, and would serve to reinforce the authenticity of the customer feedback.
Discussion

where $T$ is the optimal cost (value function) $V$ of the fusion center. If the estimate of $B$ is not incentivized, even for a finite action case, computing the optimal policies under which the optimal incentive policy has a threshold structure (as illustrated in Fig.1a) does not translate into practical solution methodologies, as the optimal cost $V(\pi)$ needs to be evaluated at each $\pi \in \Pi(2)$. (ii) The action (incentive) space for the information fusion center $p \in [0, 1]$ is a continuum. It is well known (11) that even for a finite action case, computing the optimal policies is a computationally intractable PSPACE hard problem.

B. Structure of the Optimal Incentive Policy

We wish to determine conditions under which the optimal incentive policy has the following intuitive threshold structure: don’t incentivize if the estimated $\pi < \pi^*$, and incentivize using an exactly specified incentive function otherwise. Some of the advantages of the threshold policy are: (i) To compute the threshold policy (as in Fig.12), one only needs to compute the single belief $\pi^*$; whereas a general policy (as in Fig.13) requires PSPACE hard dynamic programming recursion offline. (ii) To implement a controller with a threshold policy, one only needs to encode $\pi^*$ and the incentive function, so its practically useful.

Incentive Function: For future reference, we define the incentive function of the fusion center $\Delta(\eta_y) \in [0, 1]$ as

$$\Delta(\eta_y) = [l_1 - l_2] \frac{B_y \pi}{\Gamma B_y \pi} + l_3$$

where $\eta_y$ is the private belief update [3].

$$l_1 = \frac{\alpha_2 + \beta_2 B_{11} - \beta_1 B_{12}}{\delta_2 - \delta_1}, \quad l_2 = \frac{\alpha_1 - \beta_2 B_{21} + \beta_1 B_{22}}{\delta_2 - \delta_1}$$

$l_3 = \frac{\gamma_2}{\delta_2 - \delta_1}$, and $\alpha, \beta, \delta, \gamma$ are as in [3]. The incentive function (14) naturally arises by reformulating (13). The parameters in the incentive function are chosen such that $l_1 > 0$, $l_2 > 0$ and $l_3 > 0$. A sufficient condition is such that $\alpha_1 > \alpha_2 \geq \beta_1 > \beta_2$, $\delta_2 > \delta_1$, and $\gamma_2 > \gamma_1$.

Model Assumptions: We now give sufficient conditions under which the optimal incentive policy (13) has a threshold structure.

(A1) The observation distribution $B_{xy} = P(y|x)$ is TP2 (totally positive of order 2), i.e., the determinant of the matrix $B$ is non-negative.

(A2) The reward vector $r_a$ is supermodular, i.e., $r(1, 1) > r(2, 1)$ and $r(2, 2) > r(1, 2)$ for every $p \in [0, 1]$.

(A1) is an assumption on the underlying stochastic model, and enables the comparison of the posteriors. The observation distribution being TP2 (11) implies that in higher states, the probability of receiving higher observations is higher than in lower states.

(A2) is required for the problem to be non-trivial. If it does not hold and $r(i, 1) > r(i, 2)$ for $i = 1, 2$, then $a = 1$ always dominates $a = 2$; the sensors provide no useful information.

Main Result: (Optimality of Threshold Incentive Policy)

Theorem 1 below is our first main result. It provides a closed form expression for the optimal policy $\mu^*$ (through a threshold) of the controlled information fusion problem: the optimal policy has threshold structure (as illustrated in Fig.13). The choice over a continuum of actions is reduced to a choice between two exactly specified incentive policies.

Theorem 1. Under (A1) and (A2), the optimal incentive policy defined in (12) is given explicitly as:

$$\mu^*(\pi) = \begin{cases} 0 & \text{if } \pi(2) \in [0, \pi^*_2(2)) \\ \Delta(\eta_y=2) & \text{if } \pi(2) \in [\pi^*_2(2), 1]. \end{cases}$$

Here the threshold state $\pi^*_2(2) \in (0, 1)$ depends on the choice of $\phi_a \in (0, 1)$ defined in (6), and the parameters in the incentive function $\Delta(\eta_y=2)$ defined in (13).

The proof is given in the Appendix.

Discussion: According to Theorem 1, computing the optimal incentive policy is equivalent to finding the belief $\pi^*_2(2)$, below which it is optimal not to provide any incentive $p = 0$; and above which it is optimal to incentivize using $\Delta(\eta_y=2)$ at every belief, to minimize the cost (see Fig.14). Therefore, the controlled information fusion problem reduces to a finite dimensional optimization problem of finding a threshold state $\pi^*$. Theorem 1 provides a closed form expression for the optimal policy of the controlled information fusion problem: the choice over a continuum of actions is reduced to a choice between two exactly specified policies: $\mu(\pi) = 0$, $\forall \pi$ and $\mu(\pi) = \Delta(\eta_y=2)$, $\forall \pi$.

The practical usefulness of Theorem 1 stems from the following: (i) the search space of decision policies $\mu$ reduces from an infinite class of functions (over $\Pi(2)$) to those that switch once between the specified policies; (ii) at each instant (or belief) the fusion center only needs to decide between

9See also the proof of Theorem 2.
p = \Delta(\eta_{k-2}) \text{ and } p = 0; \text{ (iii) the region in the belief space } 
\Pi(2) \text{ where it is optimal to incentivize using } \Delta(\eta_{k-2}) \text{ is connected and convex (compare Fig.1a versus Fig.1b).}

C. Sub-martingale Property of Optimal Incentive Sequence

Theorem 1 characterized the structure of the optimal incentive policy for controlled information fusion. A natural question is: How does the actual sample path of the optimal incentive sequence behave? Theorem 2 below gives a complete sample path characterization of optimal incentive policy implemented by the fusion center. It is shown that when the fusion center aims to minimize the expected payout for gathering truthful information to reduce the uncertainty in the Bayesian state estimate, the incentive sequence is a sub-martingale, i.e., it increases on average over time.

**Theorem 2.** Consider the information fusion problem with optimal policy \( \mu^*(\pi) \) in (15). Under (A1), the optimal incentive sequence \( p_k = \mu^*(\pi_{k-1}) \) is a sub-martingale.

**Discussion:** Typically in stochastic control problems, it is difficult to characterize the optimal control sequence; one can only characterize the optimal control policy. Theorem 2 is interesting because we can characterize the optimal sequence of incentives as a sub-martingale. According to Theorem 2 the optimal incentive policy of the fusion center is such that the sample path of the incentive sequence displays an increasing trend, i.e., the incentives increase on average over time.

The usefulness of Theorem 2 stems from the following: (i) it gives a complete sample path characterization of the optimal incentive policy implemented by the fusion center; (ii) the sub-martingale property assures that the average incentives should always increase over time. This is useful in assessing the reliability of the fusion center.

**IV. Consistency of Controlled Information Fusion**

An elementary application of the martingale convergence theorem [12] shows that the social learning protocol [7] results in social sensors forming an information cascade; that is, after some time \( n^* \), all sensors choose the same action and social learning stops (see Theorem 5.3.1, (11)). Therefore, the true state can never be estimated using social learning, indeed, the belief will not converge to the true state asymptotically.

In this section, we show that by dynamically controlling the incentives over time, the fusion center can indeed learn the true state. However, this comes at the price of employing a suboptimal incentive policy. We further provide uniform bounds on the additional cost incurred for consistency [12].

**A. Controlled Information Fusion**

Fig. 3 shows the bi-directional interaction between the fusion center and the social sensor. The incentives chosen by the fusion center affects the reward function of the social sensors, and hence affects the decisions chosen. The decisions chosen in turn affect the estimate of the state (1) for the fusion center as in (7). Recall that social learning terminates after a finite horizon (see remark on Information cascade after (8)).

Theorem 3 below shows how to control the incentives to the social sensors to delay herding and information cascades, and hence estimate the state asymptotically. In particular, it is shown how the fusion center can control the incentives such that the fusion of Bayesian estimates is consistent. We will express the belief space \( \Pi(2) \) as a disjoint union of three connected regions to describe the sensors’ decision dynamics as a function of the incentive \( p \): a region \( P_1^p \) - where action \( a = 2 \) is optimal; a region \( P_3^p \) - where action \( a = 1 \) is optimal; and a region \( P_2^p \) - where action \( a = y \) is optimal. From (5), the decision of the social sensor depends on the private belief \( \eta_y \) and the reward \( r_a \) (defined in (4)). Therefore, define:

\[
\begin{align*}
\mathcal{P}_1^p &= \{ \pi \in \Pi(2) : (r_1 - r_2)\eta_{y=1} \leq 0 \} \\
\mathcal{P}_2^p &= \{ \pi \in \Pi(2) : (r_1 - r_2)\eta_{y=1} > 0 \land (r_1 - r_2)\eta_{y=2} \leq 0 \} \\
\mathcal{P}_3^p &= \{ \pi \in \Pi(2) : (r_1 - r_2)\eta_{y=2} > 0 \}
\end{align*}
\]

where \( r_a \) for \( a \in \{1,2\} \) are the social sensors’ rewards and \( \mathcal{P}^p \) models the explicit dependence of the width of the regions on the incentive parameter \( p \) through \( r_a, \eta_{y=1} \) and \( \eta_{y=2} \) denote the private belief updates after \( y = 1 \) and \( y = 2 \) respectively. The region \( \mathcal{P}_1^p \cap \mathcal{P}_3^p \) is the herding region and \( \mathcal{P}_2^p \) is the social learning region for any \( p \in [0,1] \).

**Theorem 3.** Under (A1) and (A2), the following relationship holds between the incentive \( p_k \) and the public belief \( \pi_{k+1} \):

\[
\begin{align*}
\pi_{k+1} \in \begin{cases}
\mathcal{P}_1^p & \text{iff } p_k \in \mathcal{P}_1^p \\
\mathcal{P}_2^p & \text{iff } p_k \in \mathcal{P}_2^p \\
\mathcal{P}_3^p & \text{iff } p_k \in \mathcal{P}_3^p 
\end{cases}
\end{align*}
\]

where the regions \( \mathcal{P}_i^p \) for \( i = 1,2,3 \) are defined in (16), and \( \Delta(\eta_y) \) is as in (14).

\[^{10}\text{See Appendix for definition.}\]

\[^{11}\text{Here average is over different iterations of the estimation process. For example, each round of labelling/classification in Crowdsourcing can be seen as one iteration.}\]

\[^{12}\text{Let the true state be } \theta = x. \text{ The pair } (\theta, \pi_k) \text{ is consistent, if } \pi_k \text{ converges to a point mass at } \theta \text{ in probability.}\]

\[^{13}\text{This is possible because of (A1) and (A2); see [7].}\]
Corollary 4. Let $p_k = \Delta(\eta_{yk}=2)$ for $k = 1, 2, \ldots$. The fusion of Bayesian estimates is consistent, i.e., the fusion center learns the true state asymptotically.

**Discussion:** We know that the fusion center can force the state estimates to be in the social learning region by choosing incentives in the range $p \in [\Delta(\eta_{y=2}), \Delta(\eta_{y=1})]$, see Fig. 4. From (16), Lemma 7 and Theorem 3 in the Appendix, the social sensors’ decision likelihood matrices $R^2_2$ (as in (7)) in regions $P^0_2$ and $P^1_2$, for any $p \in [0, 1]$ are

$$
\begin{bmatrix}
0 & 1 \\
1 & 0
\end{bmatrix}, \begin{bmatrix}
B_{11} & B_{12} \\
B_{21} & B_{22}
\end{bmatrix}, \text{ and } \begin{bmatrix}
1 & 0 \\
0 & 1
\end{bmatrix}
$$

respectively. In the herding region $P^0_1 \cup P^0_2$, the decision of the social sensor is independent of the public belief and the public belief (7) is frozen. In the social learning region $P^1_2$, the sensors take informative actions; i.e., each sensor acts according to its observation/valuation. In the social learning region, sensors take informative actions $\alpha = y$; or $R^2_S = B$. The observations are conditionally independent given the true state. Therefore, by suitably controlling the incentives, the fusion center fuses information that is i.i.d. on the true state. It is well known [33, 34] that fusion of Bayesian estimates is consistent (convergence in probability); i.e., for a point mass at the true state $\theta$ denoted as $q(\theta)$,

$$
\lim_{k \to \infty} P(|\pi_k - g(\theta)| > \epsilon) = 0 \quad \forall \epsilon > 0.
$$

In other words, the fusion center can learn the true state asymptotically by choosing the incentives as $p_k = \Delta(\eta_{yk}=2)$ for $k = 1, 2, \ldots$.

**B. Cost of consistency for the fusion center**

When the incentive policy is the optimal threshold policy (15), the fusion of Bayesian estimates computed from the social sensors’ decisions (7) is not consistent. This is because, the optimal incentive policy for the fusion center is such that below a certain threshold it is optimal to not incentivize (see Fig. 4). From Theorem 3 when the fusion center stops incentivizing $p = \mu^*(\pi) = 0$, the public belief is in the herding region $P^0_2$. In the herding region, social learning ceases and there is no improvement in uncertainty – mean square error between the state estimate and the true parameter remains at a fixed non-zero value. If, however, the fusion center chooses a sub-optimal policy (18), it will incur additional cost for the incentives; but the fusion of estimates computed from the social sensors’ decisions (7) will be consistent (Corollary 4). Theorem 5 below provides uniform bounds on the additional cost incurred by the fusion center for employing a sub-optimal incentive policy to reduce the uncertainty (mean square error) of the state estimate.

Consider the objective function for the fusion center:

$$
W_{\mu_c}(\pi) = \mathbb{E}_{\mu_c}\left\{ \sum_{k=1}^{\infty} \rho^k c_{\mu_c}(p_k) | \pi_0 = \pi \right\} 
$$

(17)

where $W_{\mu_c}(\pi)$ denotes the cost incurred by employing the sub-optimal policy (compare with (15))

$$
\mu_c(\pi) = \{ \Delta(\eta_{y=2}) \quad \forall \pi(2) \in [0, 1]\}. 
$$

(18)

**Theorem 5.** Let (A1) hold. The additional cost (on average) incurred by the fusion center for employing the sub-optimal policy $\mu_c(\pi)$ in (18) instead of the optimal policy $\mu^*(\pi)$ in (15) is bounded as:

$$
\sup_{\pi} |W_{\mu_c}(\pi) - J_{\mu^*(\pi)}| \leq 2 \left\{ 1 - \phi_s \right\} \frac{e^{2(\pi^*_s(1) - B_{21})^2}}{1 - \rho} - 1.
$$

(19)

where $B_{21} = \mathbb{P}(y = 1|x = 2)$, $\rho$ is the discount factor, $J_{\mu^*(\pi)}$ is the optimal cost (12), and $\pi^*_s$ denotes the optimal threshold in (15).

The proof is given in the Appendix.

**Discussion:** Theorem 5 characterizes the trade-off between consistency and cost of information acquisition. It says that when the fusion center employs a sub-optimal policy, the average additional cost incurred is bounded above by four parameters - the weight $\phi_s$ in the information fusion cost (6), discount factor $\rho$ that captures the degree of impatience of the fusion center, the optimal threshold $\pi^*_s$ in (15), and the
information gathering capability of the sensors \( B \) defined in \( 2 \).

The usefulness of Theorem 5 stems from the following: (i) It gives an upper bound on the additional discounted cost incurred when the fusion center chooses the incentives such that the fusion of Bayesian estimates computed as in \( 7 \) is consistent. (ii) It helps in choosing the weight \( \phi_s \) and the discount factor \( \rho \) for the fusion center.

**Summary:** Theorem 3 (together with Corollary 4) showed how the fusion center can employ a sub-optimal incentive policy such that information fusion with social sensors is consistent, and Theorem 5 gave an uniform bound on the average additional cost incurred for employing the sub-optimal policy.

**V. NUMERICAL RESULTS**

Sec V.A below illustrates a controlled information fusion with quadratic cost unlike \( 6 \). It is shown that a multi-threshold incentive policy is optimal for the fusion center. Sec V.B illustrates the sensitivity of the optimal threshold \( \pi^* \) to the parameters \( \phi_s \) (the weight in \( 6 \)) and \( \rho \) (discount factor in the objective \( 11 \)) that are chosen by the fusion center. Sec V.C illustrates the relation between the information gathering capabilities of the sensor (observation matrix \( B \) in \( 2 \)) and the average incentives provided by the fusion center.

Bellman’s equation \( 13 \) is solved by discretizing the state space \( \Pi(2) \). The optimal incentive policy and the optimal cost for the fusion center are computed by constructing a uniform grid of 1000 points for \( \pi(2) \in [0,1] \) and then implementing the policy and value iteration algorithm \( 11 \) for a duration of \( N = 100 \).

| \( \alpha_1 \) | 0.288 | \( \alpha_2 \) | 0.278 | \( \beta_1 \) | 0.11 |
| --- | --- | --- | --- | --- | --- |
| \( \beta_2 \) | 0.1 | \( \gamma_1 \) | 0.1 | \( \gamma_2 \) | 0.414 |

**TABLE I:** For \( \delta_1 = 0.3, \delta_2 = 0.95 \), the following parameters were obtained as a solution of \( \Delta(e_1) = 1 \) and \( \Delta(e_2) = 0 \) for the reward vector \( 4 \) parameters with the observation matrix \( B= \begin{pmatrix} 0.8 & 0.2 \\ 0.4 & 0.6 \end{pmatrix} \).

**A. Multi-threshold Incentive Policies**

This subsection illustrates numerically the nature of the optimal incentive policies for formulations of the information cost more general than \( 6 \), in particular we consider the entropy cost. We will see that the optimal incentive policy has a multi-threshold structure (as in Fig 1B).

**Expenditure & Entropy Cost for Information Fusion:** Suppose the fusion center aims to minimize the expenditure to receive truthful accounts of the information gathered by the social sensors in addition to minimizing the entropy of the state estimate, i.e.,

\[
e(c) = p + \psi_c(\pi)C_e - \phi_s Z(a = y|\pi)
\]

for \( \phi_s \in (0,1) \) denotes the scalar weight, \( p \) denotes the expenditure, \( \psi_c \) denotes the importance of the entropy cost, and \( C_e = - \sum_{i=1}^2 \pi(i) \log_2 \pi(i) \) for \( \pi(i) \in (0,1) \) and \( C_e \Delta 0 \) for \( \pi(i) = \{0,1\} \). Fig 7 shows the optimal cost and optimal policy for the fusion center when it considers entropy of the state estimate in addition to the expenditure in the information fusion cost \( 6 \). It can be seen that the optimal policy has a multi-threshold structure, and the optimal cost is discontinuous. A discontinuous cost implies a slight change in the initial conditions will lead to significantly different costs. Optimal policy being multi-threshold is unusual: it implies that if it is optimal to incentivize at a particular belief, it need not be optimal to do the same when the belief is larger.

**B. Sensitivity of Optimal Incentive Policy**

The following numerical results along with Theorem 5 provide a rationale for choosing the parameters: \( \phi_s \) – the weight in the information fusion cost \( 6 \) and \( \rho \) – the discount factor in the fusion center’s objective \( 11 \).

(i) **Usefulness of Information vs Incentivizing:** We illustrate the trade-off between usefulness of information and incentivizing in the information fusion cost \( 6 \), and see how it affects the threshold \( \pi^*_s \) in \( 15 \). Fig 5 shows the effect of increasing the weight \( \phi_s \) when the remaining parameters are the same. It can be seen that \( \pi^*_s \) is decreasing with \( \phi_s \). From Theorem 5 higher \( \phi_s \) implies that the additional cost for employing a sub-optimal policy is smaller; in other words, \( \pi^*_s(2) \) is smaller.

(ii) **Optimal cost vs Discount factor:** We illustrate the relation between total cost incurred by the fusion center for different discount factors \( \rho \) in the objective function \( 11 \). The discount factor models the degree of impatience of the fusion center, as the cost incurred at time \( k \) is \( \rho^k c(p_k) \). A smaller discount factor indicates that the fusion center pays more attention to the current costs than future costs. It is seen from Fig 6 that a higher discount factor leads to smaller (expected) costs for higher states. This indicates that it is beneficial for the fusion center to attach more importance to future costs as it should also take into account the benefit from sensors performing social learning.

**C. Sample Path of Optimal Incentives**

This subsection illustrates the sample path properties of the optimal incentive sequence over time (which was characterized in Theorem 2 to be a sub-martingale). Fig 8 shows the average incentives provided to the social sensors over time. The fusion center employs the optimal incentive policy \( 15 \) and fuses the information revealed by social sensors in a Bayesian way \( 7 \). Each sample path has a duration of \( N = 500 \), i.e, sequential information fusion from 500 social sensors. The figure shows the average over 100 independent such sample paths for three different observation likelihood matrices \( 2 \).

We consider the following observation likelihood matrices for illustrating the relation between the information gathering capabilities of the sensor \( 2 \) and the average incentives provided by the fusion center: \( B, B^2 \), and \( B^3 \). We know that \( B \) is more informative than \( B^2 \), which is in turn more informative than \( B^3 \), in the Blackwell sense \( 11 \) (see also Footnote 8).

**Parameters:** The parameters of the incentive function \( 14 \) for \( B^2 \) and \( B^3 \) are specified in Table 11.

In Fig 8 it can be seen that the range (or the slope) of the...
average incentives over the time horizon is highest for the case of observation matrix $B$ (compared to $B^2$ and $B^3$). This is intuitive, for example, when an online social media review platform like Amazon is soliciting honest reviews, gradual increase in the compensation when the quality looks promising (made possible by a more informative $B$) will lead to fair reviews and this in turn will increase the sales of the product or services in the future.

It can be seen from Fig.7 that the average incentives display an increasing trend. This is useful in assessing how reliable the fusion center is for providing incentives and estimating the state. If the average payout does not display an increasing trend, one could infer that either the reliability of the fusion center is compromised or the social sensors are playing the system.

Fig. 7: Multi-threshold incentive policy with the entropy cost. The regions in the belief space II(2) where it is optimal to not incentivize $\mu^*(\pi) = 0$ is no more connected and convex. Having a connected region in the belief space where it is optimal not to incentivize has implications on the confidence of the fusion center in implementing the incentive policy: once its optimal to incentivize at a certain belief, it need not be optimal to continue incentivizing when the belief is larger, i.e, when it is more certain about the estimate of the state. The optimal cost is discontinuous in Fig.7b, and this implies that a slight change in the initial conditions will lead to a significantly different cost.
offered in case of increasing trend. The zoomed in subfigure shows the increasing trend in case of observation matrix \( B \). The average incentives display an increasing trend. The zoomed in subfigure shows the increasing trend in case of observation matrix \( B \). It can be seen that average incentives offered in case of \( B^1 \) is higher than \( B^2 \) which in turn is higher than \( B \). This can be attributed to the higher value of \( \alpha \) in Table II.

![Incentives averaged over independent sample paths for the fusion center over time for observation matrices \( B, B^2 \) and \( B^3 \). The observation matrices are ordered in the decreasing order of informativeness (see Footnote 8). The parameters are specified in Tables I & II. The weight \( \phi_s = 0.4 \) in the information fusion cost (6) and the discount factor \( \rho = 0.6 \). It can be seen that the range (or the slope) of the average incentives over the time horizon is highest for the case of observation matrix \( B \). The average incentives display an increasing trend.](image)

TABLE II: The reward vector \( \mathbf{w} \) parameters for \( B^2 \) and \( B^3 \). For \( \delta_1 = 0.3, \delta_2 = 0.95 \), the following parameters were obtained as a solution of \( \Delta(e_1) = 1 \) and \( \Delta(e_2) = 0 \) for the reward vector \( \mathbf{w} \) parameters with observation matrix \( B \).

| Obs. matrix \( B^2 \) | \( \alpha_1 = 0.3132 \) | \( \alpha_2 = 0.3632 \) | \( \beta_1 = 0.11 \) |
|------------------------|------------------|------------------|------------------|
| Obs. matrix \( B^3 \)  | \( \beta_2 = 0.4 \) | \( \gamma_1 = 0.4 \) | \( \gamma_2 = 0.444 \) |

VI. CONCLUSION AND FUTURE WORK

Unlike data fusion involving physical sensors for tracking targets, this paper is motivated by information fusion with social sensors, which provide reviews on social media review platforms such as Amazon, Yelp, and Airbnb. Our main objective is to control the information fusion by dynamically providing incentives to the social sensors. We presented four main results. Theorem 1 showed that under reasonable conditions on the model parameters, the optimal incentive policy has a threshold structure. The optimal policy is determined in closed form, and is such that it switches once between two exactly specified incentive policies. Theorem 2 characterized the sample path property of the optimal incentive sequence that results from fusion center employing the optimal threshold policy. It was shown that the optimal incentive sequence is a sub-martingale. Theorem 3 showed how the fusion center can employ a sub-optimal policy and there by facilitate social learning indefinitely, to learn the true state asymptotically. In other words, it was shown how controlled information fusion with social sensors can be consistent. Finally, Theorem 5 provided uniform bounds on the average additional cost incurred, by employing a sub-optimal policy, for consistency.

While the formulation of the controlled information fusion problem applies to arbitrary finite state, observation and action spaces, our structural analysis of the optimal incentive policies are currently applicable only to the 2-state case. Our results for two states and observations, however, provide substantial insight into the nature of the complexity of controlled information fusion with social sensors, and highlight the means to derive structural results for the optimal policy in a multi-state case.

APPENDIX A

DEFINITIONS AND PRELIMINARIES:

**Definition 1.** First-Order Stochastic Dominance (FSD) \((\geq_s)\): Let \( \pi_1, \pi_2 \in \Pi(2) \) be any two belief state vectors. Then \( \pi_1 \geq_s \pi_2 \) if \( \sum_{i=j}^{2} \pi_1(i) \geq \sum_{i=j}^{2} \pi_2(i) \) for \( j \in \{1,2\} \).

Equivalently, \( \pi_2 \geq_s \pi_1 \) iff for all \( v \in \mathcal{V} \), \( v' \pi_2 \leq v' \pi_1 \), where \( \mathcal{V} \) denotes the space of 2-dimensional vectors \( v \), with non-increasing components, i.e, \( v_1 \geq v_2 \geq \ldots v_X \).

**Definition 2.** (Martingale [32]): Let \( \mathcal{F}_k \) denote the sigma algebra (as in [29]). A sequence \( \{X_k\} \) such that \( \mathbb{E}[|X_k|] < \infty \) is a martingale (with respect to \( \mathcal{F}_k \)) if

\[
\mathbb{E}[X_{k+1} | \mathcal{F}_k] = X_k, \text{ for all } k.
\]

If \( \mathbb{E}[X_{k+1} | \mathcal{F}_k] \geq X_k, \text{ for all } k, \) the sequence \( \{X_k\} \) is a sub-martingale.

**Definition 3.** ([32]) A sequence \( H_k \) is said to be a predictable sequence if \( H_k \in \mathcal{F}_{k-1} \).

In words, \( H_k \) may be predicted with certainty using the information available at time \( k-1 \).

**Lemma 6** ([11]). Under (A1), we have \( \sigma(\pi_1, a) \geq_s \sigma(\pi_2, a) \), where \( \sigma(\pi, a) = \mathbb{E}[B_{y_1} \pi | I' B_{y_2} \pi] \).

**Lemma 7** ([4]). The sensor decision likelihood probability matrix \( R^\pi \) in the social learning filter [4] is computed as

\[
R^\pi = BM^\pi \text{ where }
\]

\[
M^\pi_{y,a} = \mathbb{P}(a | y, \pi) = \mathbb{I}(r' \alpha B_{y \pi} > r' \alpha B_{y \pi}), \text{ where } \alpha = \mathcal{A}/a.
\]
Theorem 8 (7). Let (A1) and (A2) hold. The belief space \( \Pi(2) \) can be partitioned into at most 3 non-empty regions \( \mathcal{P}_1, \mathcal{P}_2, \mathcal{P}_3 \). On each of these regions, the sensor decision likelihood matrix \( R^\pi \) in (22) is a constant with respect to the belief state \( \pi \).

Theorem 9 (32). Let \( W_k \) be a sub-martingale. If \( H_k \geq 0 \) is predictable and each \( H_k \) is bounded, then \( (H, W)_k \) is a sub-martingale.

Theorem 9 corresponds to Theorem 5.2.5 in (32).

**APPENDIX B**

**Proofs**

**Proof of Theorem 1**

We first show that due to the structure of the social learning filter in (7), the choice of incentives reduces from a continuum to a finite number at every belief. Next, we show that the incentive function \( \Delta(\eta_y) \) is decreasing in \( \pi \) for any \( y \).

**Theorem 10.** Let \( \Delta(\eta_{y=1}) \) and \( \Delta(\eta_{y=2}) \) be two possible incentives at belief \( \pi \). Under (A1) and (A2), the Q function in (23) can be simplified as:

\[
Q(\pi, p) = \begin{cases} 
 p + \rho V(\pi) & \text{if } p \in [0, \Delta(\eta_{y=2})]; \\
 p - \phi_s + \rho E V(\pi) & \text{if } p \in [\Delta(\eta_{y=2}), \Delta(\eta_{y=1})]; \\
 p + \rho V(\pi) & \text{if } p \in [\Delta(\eta_{y=1}), 1].
\end{cases}
\]

and \( V(\pi) = \min Q(\pi, p) \). Here,

\[
E V(\pi) = I' B_1^{\pi=1} \pi \times V(\eta_{y=1}) + I' B_2^{\pi=2} \pi \times V(\eta_{y=2}).
\]

**Proof of Theorem 10**

From Lemma 3 and Theorem 8 we have

\[
R^\pi = \begin{bmatrix}
1 & 0 \\
1 & 0 \\
B_{11} & B_{12} \\
B_{21} & B_{22}
\end{bmatrix}
\begin{cases}
1 & \text{if } p \in [0, \Delta(\eta_{y=2})]; \\
1 & \text{if } p \in [\Delta(\eta_{y=2}), \Delta(\eta_{y=1})]; \\
0 & \text{if } p \in [\Delta(\eta_{y=1}), 1].
\end{cases}
\]

From (24), it is clear that the sensors’ decision

\[
a = \begin{cases}
1 & \text{if } p \in [0, \Delta(\eta_{y=2})]; \\
y & \text{if } p \in [\Delta(\eta_{y=2}), \Delta(\eta_{y=1})]; \\
2 & \text{if } p \in [\Delta(\eta_{y=1}), 1].
\end{cases}
\]

Therefore,

\[
\sum_{a \in A} V(T^\pi(\pi, a) \sigma(\pi, a) = \begin{cases}
V(\pi) & \text{if } p \in [0, \Delta(\eta_{y=2})]; \\
E V(\pi) & \text{if } p \in [\Delta(\eta_{y=2}), \Delta(\eta_{y=1})]; \\
V(\pi) & \text{if } p \in [\Delta(\eta_{y=1}), 1].
\end{cases}
\]

where \( E V(\pi) = I' B_1^{\pi=1} \pi \times V(\eta_{y=1}) + I' B_2^{\pi=2} \pi \times V(\eta_{y=2}). \)

The result follows.

Theorem 10 represents the Q function (13) over the range \([0, 1]\) into three regions. The following corollary highlights why such a partition is useful.

**Corollary 11.** At every public belief \( \pi \in \Pi(2) \), it is sufficient to choose one of the three incentives \( \{0, \Delta(\eta_{y=2}), \Delta(\eta_{y=1})\} \).

**Proof.** From Theorem 10 the instantaneous reward is a linear function in \( p \) and \( \argmin_{p \in [0, \Delta(\eta_{y=2})]} Q(\pi, p) = 0, \argmin_{p \in [\Delta(\eta_{y=2}), \Delta(\eta_{y=1})]} Q(\pi, p) = \Delta(\eta_{y=2}), \text{ and } \argmin_{p \in [\Delta(\eta_{y=1}), 1]} Q(\pi, p) = \Delta(\eta_{y=1}). \) These hold as for any value of \( p \) in each of the three regions, the corresponding continuation payoff is the same from Theorem 10.

**Lemma 12.** The incentive function \( \Delta(\eta_y) \) is decreasing in \( \pi \) for every \( y \).

**Proof.** The incentive function is given as (14), where \( l_1, l_2, l_3 > 0 \). With \( \pi = [1 - \pi(2), \pi(2)] \), differentiating w.r.t \( \pi(2) \),

\[
d(\Delta(\eta_y)) \frac{d\pi(2)}{d\pi(2)} = -(l_1 + l_2) B_{1y} B_{2y} < 0
\]

**Proof of Theorem 7**

From Corollary 11 the value function (13) is:

\[
V(\pi) = \min \{0, \Delta(\eta_{y=2}) - \phi_s + \rho E V(\pi), \Delta(\eta_{y=1}) + \rho V(\pi)\}.
\]

\[
\Rightarrow V(\pi) = \min \{0, \Delta(\eta_{y=2}) - \phi_s + \rho E V(\pi)\}
\]

as \( \Delta(\eta_{y=1}) \geq 0 \).

By using the value iteration algorithm 11 on (27), we have

\[
V_{n+1}(\pi) = \min \{0, \Delta(\eta_{y=2}) - \phi_s + \rho E V_n(\pi)\}
\]

with \( V_0(\pi) = 0 \) \( \forall \pi \).

From Lemma 12 the incentive function is decreasing. From the definition of First-Order Stochastic Dominance (21), and Lemma 6 we have \( E V_n(\pi) \) is decreasing in \( \pi \). Therefore, \( V_{n+1}(\pi) \) and hence \( V(\pi) \) is decreasing in \( \pi \).

Let \( V(0) \) and \( V(1) \) denote the values for \( \pi = \left[ \begin{array}{c} 1 \\ 0 \end{array} \right] \) and \( \pi = \left[ \begin{array}{c} 0 \\ 1 \end{array} \right] \). It is seen by substitution that \( E V(0) = V(0) \) and \( E V(1) = V(1) \).

By definition, we know that \( \Delta(\eta_y) \in [0, 1] \). Using Lemma 12 let \( \Delta(\xi_1) = 1 \) and \( \Delta(\xi_2) = 0 \). The value function for the fusion center is given by (27). We have the following:

1. For \( V(\pi) = \Delta(\eta_{y=2}) - \phi_s + \rho E V(\pi) \), \( V(0) = \frac{1 - \phi_s}{1 - \rho} > 0 \), and \( V(1) = \frac{1}{1 - \rho} < 0 \).
2. For \( V(\pi) = 0, V(0) = V(1) = 0 \).

The value function \( V(\pi) \) in (27) is decreasing with a positive value at \( e_1 \) and a negative value \( e_2 \), so must be zero at some point(s). Set \( \Sigma = \{\pi(2) | 0 = \Delta(\eta_{y=2}) - \phi_s + \rho E V(\pi)\} \). Since the value function \( V(\pi) \) is monotone in \( \pi \), the set \( \Sigma \) is convex.

Choosing \( \pi^*_2(2) = \{\pi(2) | \pi(2) > \pi(2) \forall \pi(2) \in \Sigma\} \), the result follows.

**Proof of Theorem 2**

We will first establish the property of the incentive function \( \Delta(\eta_y) \) as the optimal policy depends on it.

**Lemma 13.** Under (A1), \( \Delta(\eta_{y=1}) \) is concave in \( \pi \), and \( \Delta(\eta_{y=2}) \) is convex in \( \pi \).

**Proof of Lemma 13**

The incentive function \( \Delta(\eta_{y=2}) \) is given in (14). A differentiable function \( f : [0, 1] \rightarrow [0, 1] \) is convex if

\[
f(w_1) \geq f(w_2) + f'(w_2)(w_1 - w_2), \text{ for all } w_1, w_2 \in [0, 1].
\]
A function \( f \) is concave if \(-f\) is convex.

From (29) with \( u_1 = \pi_1(2) \) and \( u_2 = \pi_2(2) \), and using Lemma 3, it is verified that the function \( \Delta(\eta_{y=2}) \) is convex in \( \pi \).

Similarly, it can be shown that \( \Delta(\eta_{y=1}) \) is concave in \( \pi \).

**Proof of Theorem 2**

Consider the sub-optimal policy \( \hat{\mu}(\pi) \) given as

\[
\hat{\mu}(\pi) = \begin{cases} 
\Delta(\eta_{y=2}) - \epsilon & \text{if } \pi(2) \in [0, \pi_*(2)); \\
\Delta(\eta_{y=2}) & \text{if } \pi(2) \in [\pi_*(2), 1]. 
\end{cases}
\]

Here \( \epsilon > 0 \) and \( \pi_*(2) \in [0, 1] \). Let \( W_{k} = \hat{\mu}(\pi_{k-1}) \).

From Lemma 13, \( \Delta(\eta_{y=2}) \) is convex in \( \pi \). Let \( u^S(\pi_{k+1}) = \Delta(\eta_{y=2}) \) denote the price at time \( k + 1 \). So \( u^S(\pi) \) is convex in \( \pi \).

We know that the public belief \( \pi_k \) is a martingale, i.e., \( \mathbb{E}[\pi_{k+1} | F_k] = \pi_k \). For \( \epsilon \to 0 \),

\[
\mathbb{E}[W_{k+1} | F_k] = \mathbb{E}[u^S(\pi_{k+1}) | F_k] \geq u^S(\mathbb{E}[\pi_{k+1} | F_k]) \geq u^S(\pi_k) \geq W_k
\]

by Jensen’s inequality and martingale property of the public belief. Therefore \( W_{k} = \hat{\mu}(\pi_{k-1}) \) is a sub-martingale.

Consider a function \( \bar{\mu}(\pi) \) given by

\[
\bar{\mu}(\pi) = \begin{cases} 
0 & \text{if } \pi(2) \in [0, \pi_*(2)); \\
1 & \text{if } \pi(2) \in [\pi_*(2), 1]. 
\end{cases}
\]

Let \( H_k = \bar{\mu}(\pi_{k-1}) \). From Theorem 9, \( \mathbb{P}(H_k | \pi_k) \) is a sub-martingale. But \( (H.W)k = p_k \). Therefore, the optimal incentive sequence \( p_k = \mu^*(\pi_{k-1}) \) is a sub-martingale, \( \mathbb{E}[p_{k+1} | F_k] \geq p_k, \) i.e., it increases on average over time.

**Proof of Theorem 5**

We’ll prove that \( \pi \in P^p \) iff \( p \in [\Delta(\eta_{y=2}), \Delta(\eta_{y=1})] \). Other cases are proved similarly.

From (34), we can write

\[
\begin{align*}
0 = (1 - \gamma_1)- \beta_1 B_{12} & \quad (\delta_1 p - \alpha_1 - \beta_1 B_{22} - \gamma_1) \\
2 \Rightarrow & \quad (\delta_2 p - \alpha_2 - \beta_2 B_{21} - \gamma_2)
\end{align*}
\]

Let \( Z_k = \sum_{t=1}^{k} Z_t \) denote the binomial random variable. By Hoeffding’s Inequality, for \( \epsilon > 0 \),

\[
\mathbb{P}(Z_k > \mathbb{E}(Z) + \epsilon) \leq e^{-2k\epsilon^2}
\]

where \( \mathbb{E}(Z) = \mathbb{E}_{\mathbb{Y}}(Z) \).

Choosing \( \epsilon = \pi^*_s(1) - \mathbb{E}(Z) \), we have

\[
\begin{align*}
\mathbb{P}(Z_k > \pi^*_s(1)) & \leq e^{-2k(\pi^*_s(1) - \mathbb{E}(Z))^2} \\
\mathbb{P}(Z_k > \pi^*_s(1)) & \leq e^{-2k(\pi^*_s(1) - \pi^*_s(1))^2} \\
\mathbb{P}(\pi_k \in H) & = \mathbb{P}(Z_k > \pi^*_s(1))
\end{align*}
\]

Therefore, we have \( \mathbb{P}(\pi_k \in H) \leq e^{-2k(\pi^*_s(1) - \pi^*_s(1))^2} \).

**Proof of Theorem 5**

For any sub-optimal policy \( \mu_c \) and the corresponding cost \( W_{\mu_c}(\pi) \), it is clear that \( W_{\mu_c}(\pi) - J_{\mu^*_c}(\pi) \geq 0 \forall \pi \). Let \( I \) denote the indicator function. We have

\[
W_{\mu_c}(\pi) - J_{\mu^*_c}(\pi) = \mathbb{I}(\pi \in \mathbb{H}) \{ W_{\mu_c}(\pi) - J_{\mu^*_c}(\pi) \}
\]

\[
\Rightarrow \sup_{\pi} \{ W_{\mu_c}(\pi) - J_{\mu^*_c}(\pi) \} \leq \mathbb{I}(\pi \in \mathbb{H}) \{ W_{\mu_c}(\pi) - J_{\mu^*_c}(\pi) \}
\]

where \( \mathbb{H} \) is defined in (31). From Theorem 1, we know that \( J_{\mu^*_c}(\pi) = V(\pi) \) is monotone (non-increasing) in \( \pi \). Similar arguments can be used to establish that \( W_{\mu_c}(\pi) \) is monotone (non-increasing) in \( \pi \). Therefore, we have for (34)

\[
\sup_{\pi} \{ W_{\mu_c}(\pi) - J_{\mu^*_c}(\pi) \} \leq 2 \mathbb{I}(\pi \in \mathbb{H}) \{ W_{\mu_c}(\pi) - J_{\mu^*_c}(\pi) \}
\]
as $J_{\mu^*}(\pi) = 0 \forall \pi \in \mathcal{H}$ from (27) and Theorem 11. Let $c_{\mu}^*$ and $c_{\mu}^1$ denote the cost incurred by the fusion center employing a policy $\mu$ when it stops incentivizing and incentivizes respectively. We have for $\pi_0 = \pi$, the cost functions $J_{\mu^*}(\pi)$ and $W_{\mu^*}(\pi)$:

$$J_{\mu^*}(\pi) = E_{\mu^*}\left\{ \sum_{k=1}^{\infty} \rho^k \left( I(\pi_k \in \mathcal{H}) c_{\mu^*}(p_{\pi_k}) + I(\pi_k \notin \mathcal{H}) c_{\mu}^1(p_{\pi_k}) \right) \right\}$$

$$W_{\mu^*}(\pi) = E_{\mu^*}\left\{ \sum_{k=1}^{\infty} \rho^k \left( I(\pi_k \in \mathcal{H}) c_{\mu^*}(p_{\pi_k}) + I(\pi_k \notin \mathcal{H}) c_{\mu}^1(p_{\pi_k}) \right) \right\}$$

We have $c_{\mu^*}(p) = 0$, $c_{\mu^*}(p) = \Delta(\eta_{y=2}) - \phi_\pi$, $c_{\mu^1}(p) = \Delta(\eta_{y=2}) - \phi_\pi$, and $c_{\mu^1}(p) = \Delta(\eta_{y=2}) - \phi_\pi$. The set $\mathcal{H}$ defined in (33) is compact by definition. For the discount factor $\rho \in [0, 1)$ and bounded instantaneous costs, the average discounted cost is bounded (11). Therefore in (35),

$$\sup_{\pi} \{ I(\pi \in \mathcal{H}) W_{\mu^*}(\pi) \} = \max_{\pi} \{ I(\pi \in \mathcal{H}) W_{\mu^*}(\pi) \}$$

and $\bar{\pi} = \arg\max_{\pi} \{ I(\pi \in \mathcal{H}) W_{\mu^*}(\pi) \}$. We have for $\pi_0 = \bar{\pi}$,

$$\max_{\pi} \{ I(\pi \in \mathcal{H}) W_{\mu^*}(\pi) \} = E_{\mu^*}\left\{ \sum_{k=1}^{\infty} \rho^k \left( I(\pi_k \in \mathcal{H}) c_{\mu^*}(p_{\pi_k}) \right) \right\}$$

$$\leq E_{\mu^*}\left\{ \sum_{k=1}^{\infty} \rho^k I(\pi_k \in \mathcal{H}) \max_{\Delta(\eta_{y=2}) - \phi_\pi \in \mathcal{H}} s_{\mu^*}(p_{\pi_k}) \right\}$$

$$\leq (1 - \phi_\pi) \frac{\rho}{1 - \rho} \left\{ \sum_{k=1}^{\infty} I(\pi_k \in \mathcal{H}) \right\}$$

From Lemma 14, we have $P(\pi_k \in H) \leq e^{-2k(\pi^*_{\pi_k}) - 3 - B_{\pi_k})^2}$. Noting that under (A1), $B_{11} \geq B_{21}$, the result follows. □

REFERENCES

[1] E. Moretti, “Social learning and peer effects in consumption: Evidence from movie sales,” The Review of Economic Studies, vol. 78, no. 1, pp. 356–393, 2011.

[2] B. Irlach, C. Maglaras, and M. Scarsini, “Bayesian social learning with consumer reviews,” ACM SIGMETRICS Performance Evaluation Review, vol. 41, no. 4, pp. 28–28, 2014.

[3] J. N. Tsitsiklis et al., “Decentralized detection,” Advances in Statistical Signal Processing, vol. 2, no. 2, pp. 297–344, 1993.

[4] R. Viswanathan and P. K. Varshney, “Distributed detection with multiple sensors Part I. Fundamentals,” Proceedings of the IEEE, vol. 85, no. 1, pp. 54–63, 1997.

[5] J. N. Tsitsiklis and M. Athans, “On the complexity of decentralized decision making and detection problems,” IEEE Transactions on Automatic Control, vol. 30, no. 5, pp. 440–446, 1985.

[6] Y.-C. Ho, “Team decision theory and information structures,” Proceedings of the IEEE, vol. 68, no. 6, pp. 644–654, 1980.

[7] V. Krishnamurthy, “Quickest detection POMDPs with social learning: Interaction of local and global decision makers,” IEEE Transactions on Information Theory, vol. 58, no. 8, pp. 5563–5587, 2012.

[8] S. Bikchandani, D. Hirshleifer, and I. Welch, “A theory of fads, fashion, custom, and cultural change as informational cascades,” Journal of Political Economy, vol. 100, no. 5, pp. 992–1026, Oct., 1992.

[9] A. V. Banerjee, “A simple model of herd behavior,” The Quarterly Journal of Economics, vol. 107, no. 3, pp. 797–817, Aug., 1992.

[10] P. K. Varshney, Distributed detection and data fusion. Springer Science & Business Media, 2012.

[11] V. Krishnamurthy, Partially Observed Markov Decision Processes. Cambridge University Press, 2016.