This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2020.3046024, IEEE Transactions on Automatic Control

Controlled Sequential Information Fusion with Social Sensors

Sujay Bhatt Vikram Krishnamurthy, *Fellow, IEEE*

***Abstract*—A sequence of social sensors estimate an unknown parameter (modeled as a 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 information about the underlying state*?**

This paper presents five 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, it is shown that the optimal incentive sequence is a *submartingale*, 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. Fourth, 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. Finally, uniform bounds on the budget saved by employing policies that guarantee state estimation in finite time are provided.

***Index Terms*—social sensors, incentives, social learning, partially observed Markov decision process (POMDP), submartingale, threshold policies, uniform bounds, consistency.**

I. Introduction

A social (human) sensor provides information about its state (sentiment, quality of product) to a social network after interaction with other social sensors. It 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.

In this paper, in line with a large body of literature, we adopt a more stylized definition: a *social sensor performs social learning*. Social learning is an integral part of human behaviour and has been studied widely in economics, sociology (where the term groupthink is used), electrical engineering and computer science; see [1], [2]. It shares similarities with decentralized detection [3], [4] that falls within the class of team decision theory [5]; but with key differences: (i) Decentralized detection quantizes the observations, whereas social learning quantizes the Bayesian belief. (ii) In decentralized detection, the fusion policies are directly optimized, where as in social learning the fusion rule is prescribed and is Bayesian.

This paper considers a sequential decision making model of Bayesian social learning introduced in [6], [7], [8], where the social sensors learn from their predecessors. Each social sensor has a private signal on the underlying state and considers this in addition to the (bounded) information gathered by its predecessors. This interplay results in the well known inefficiencies such as the formation of herds (sensors choose the same action irrespective of their private information) and informational cascades (information fusion results in no improvement in uncertainty). In [9], some of the inefficiencies in the sequential social learning model is shown to arise due to the bounded nature of the information or beliefs used in decision making, and show that the true state is aggregated when the beliefs are unbounded. In this paper, similar to [10], we explore how such inefficiencies associated with social sensors can be controlled using an exogenous incentive by formulating controlled sequential information fusion as a *non-standard* Partially Observed Markov Decision Process (POMDP).

Information Fusion with Social Sensors

Information fusion with physical sensors is a well studied problem. In this paper, motivated by recent applications using online social media review platforms, we consider information fusion with social sensors. We consider the following problem: A sequence of social sensors estimate an unknown state of nature, and a fusion center aims to estimate the underlying state by incentivizing the social sensors. How should the fusion center dynamically incentivize the social sensors to acquire information about the underlying state? Equivalently, how can the fusion center optimize the trade-off between the cost of information acquisition from the social sensors versus the usefulness of the information in terms of reduction in uncertainty (mean-square error between the true state and the estimate) of the Bayesian state estimate. Similar problems with physical sensors were considered in [11], [12].

Multi-sensor data fusion [13] 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

S. Bhatt is currently at Baidu Research, Bellevue, WA, 98004 and V. Krishnamurthy is with the School of Electrical and Computer Engineering, Cornell University, Ithaca, New York, 14853.

E-mail: (sh2376@cornell.edu), (vikramk@cornell.edu).

This research was supported in part by the U. S. Army Research Office under grant W911NF-19-1-0365, Air Force Office of Scientific Research under grant FA9550-18-1-0007 and National Science Foundation under grant CCF-1714180.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2020.3046024, IEEE Transactions on Automatic Control

(a) Single Threshold Incentive Policy

(b) General (Multi-threshold) Incentive Policy

Fig. 1: Visual illustration of the main result of the paper, namely the optimal incentive policy for choosing incentives to social sensors to reveal information to the fusion center. In general, the incentive policy could be an arbitrary function of π as in Fig.1b; we provide conditions under which the incentive policy is as in Fig.1a. Here $\mu^*(\pi) \in [0,1]$ is the optimal incentive policy and π denotes the posterior (belief) 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.VII-A provides numerical examples (of more general cost functions) where the optimal incentive policy is multi-threshold as in Fig.1b.

decision regarding the desired parameter. Having more number of sensors leads to improvement in reliability, resolution, coverage, and confidence; see [13].

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.

In controlled sequential information fusion, the process of incentivization modifies the cost or reward function of the social sensors and hence directly affects the sensors' decisions (see Fig.2). The decisions are a quantization of the Bayesian estimate of the state, and hence controlling the incentives can shape the information that is subsequently fused.

Additional Related Literature

There are many works in Bayesian social learning, where the network structure is considered and social sensors repeatedly make decisions [14], [15], [16], [17]. In case of repeated decision making, inefficiencies like cascades and herds can be avoided, however, that comes at the cost of increased computational complexity for the individual social sensors.

The problem considered in this paper shares similarities with sequential hypothesis testing [18], however, with key differences: (i) Information from social sensors is correlated due to social learning. Information fusion with social learning thus leads to inefficiencies like herds and information cascades. So having more social sensors need not always be advantageous (in terms of reducing the mean square error between the state estimate and the true state). (ii) In general, information fusion with social sensors leads to multi-threshold policies (see Fig.1), unlike classical hypothesis testing. This has implications on the confidence of the information fusion center while announcing the true hypothesis – if it is optimal to announce the true hypothesis for a certain belief, it might not be optimal to make the announcement when the belief is larger (or more certain)!

Main Results and Organization

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

(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), when estimating a random variable. 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 threshold state π^* and the incentive function, so a threshold policy is practically useful.

(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*, when estimating a random variable. In particular, we show that the optimal incentive sequence is a sub-martingale; i.e, the incentives *increase* on average over time. The increase can be attributed to the fact that the senors polled for information at a later instant associate a higher value on average due to learning from their predecessors. 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 [19], [10]; 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. Sec.VI illustrates the difficulty of characterizing the structure of the optimal incentive sequence when the underlying state is changing according to a Markov chain. We provide conditions on the state transitions which guarantee that the optimal policy for estimating a random variable is near optimal for tracking the Markov chain.

(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 [1], [2] due to the formation of information cascades. We show that the inefficiencies in the sequential social learning model– herds and information cascades – can be controlled using the incentives (Corollary 4) and provide uniform bounds (Theorem 5 and Theorem 6) on the performance. Previous work [1], [20], [16] emphasized providing conditions on the observation distribution or action sampling distributions for

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2020.3046024, IEEE Transactions on Automatic Control

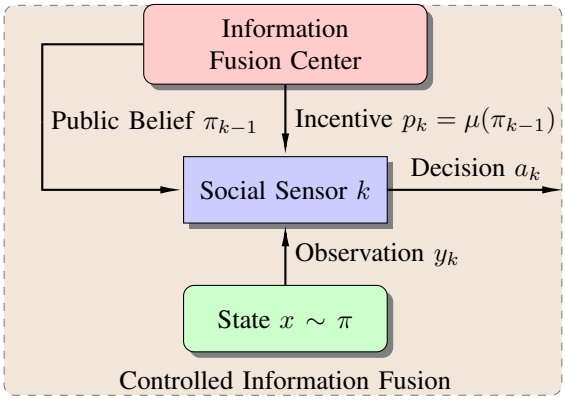


Fig. 2: A sequence of social sensors perform Bayesian social learning to estimate the underlying state x. Social sensor k chooses action a_k after myopically optimizing a reward function. The fusion center provides incentives $p_k \in [0,1]$ at each time k (or at each sensor k) and fuses the information gathered in a Bayesian way. Each incentive p_k is computed as a function μ of the posterior probability mass function (public belief) of the state π_{k-1} at time $k-1$. The public belief π_{k-1} is computed from the decisions of the first $k-1$ sensors. The decision a_k of social sensor k depends on the incentive p_k, the public belief π_{k-1}, and the private observation y_k of the state x.

the social sensors. 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. When it is sufficient to know the state with a degree of confidence, policies that guarantee state estimation in finite time are discussed. Uniform bounds on the budget saved as a result of estimating the state only upto a degree of confidence are provided.
Sec.VII presents numerical examples that provide additional insights on the main results. A discussion on extension to multiple actions and states is provided in Sec.VII-D.

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 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

In this subsection, we model: (i) dynamics of the social sensors; (ii) the information fusion cost for the fusion center that models the trade-off between incentives and the reduction in uncertainty in the state estimate. We also characterize the evolution of the posterior probability mass function of the state, and how the fusion center can make use of the available information to provide the incentives to the social sensors.

Let $k = 1, 2, \cdots$ denote discrete time. It is assumed that each social sensor is identical and acts once in a predetermined sequential order indexed by k. Let $x_0 (= x) \in \mathcal{X} = \{1, 2\}$ denote the state of nature, and is assumed to be a *random variable*[1] chosen at $k = 0$. Let the probability mass function of the state x at time $k-1$ be denoted as

$$\pi_{k-1}(i) = \mathbb{P}(x = i | a_1, \ldots, a_{k-1}). \tag{1}$$

The state estimate (1) 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 $\pi_0 = (\pi_0(i), i \in \mathcal{X})$, where $\pi_0(i) = \mathbb{P}(x = i)$. Let the belief space, i.e, the set of distributions π over the state be denoted as

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

Social Sensor Dynamics: A social sensor 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 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 obtains a noisy $y_k \in \mathcal{Y} = \{1, 2\}$ of the underlying state x with observation likelihood:

$$B_{ij} = \mathbb{P}(y_k = j | x = i). \tag{2}$$

The (discrete) observation likelihood models the (limited) information gathering capabilities of the sensor.
(2.) *Social Learning and Private Belief update*: Sensor k updates its private belief η_{y_k} by fusing observation y_k and the prior public belief π_{k-1}, via the following classical Bayesian update

$$\eta_{y_k} = \frac{B_{y_k} \pi_{k-1}}{\mathbf{1}' B_{y_k} \pi_{k-1}} \tag{3}$$

where B_{y_k} denotes the diagonal matrix with diagonal elements $[\mathbb{P}(y_k | x = 1), \mathbb{P}(y_k | x = 2)]$ and $\mathbf{1}'$ denotes the 2-dimensional row vector of ones.
(3.) *Social Sensor's Action*: Sensor k executes an action $a_k \in \mathcal{A} = \{1, 2\}$ myopically to maximize a reward function. Each sensor being an expected (and myopic) reward maximizer is rational [1]. This assumption implies that the social sensors have no altruistic concerns. The decision a_k of social sensor k is given by:

$$a_k = \arg \max_{a \in \mathcal{A}} r_a' \eta_{y_k}. \tag{4}$$

[1]Sec.VI discusses the estimation problem when the state is changing according to a Markov chain.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2020.3046024, IEEE Transactions on Automatic Control

Here $r_a = [r(1,a),\ r(2,a)]$, with r_a' denoting the transpose of the reward vector. We consider

$$r(1,a) = \delta_a p_k + \Gamma_{1a}, \quad r(2,a) = \delta_a p_k + \Gamma_{2a}, \quad \text{with}$$
$$\Gamma_{xa} = -\alpha_a \mathcal{I}(a \neq x) - \gamma_a. \tag{5}$$

Here $\delta_a \in [0,1]$, $\alpha_a, \gamma_a \in \mathbb{R}$ are the given parameters of the model and $\mathcal{I}$ denotes the indicator function. For an action $a \in \mathcal{A}$ of the social sensor, $\delta_a p$ indicates the effective incentive received by the social sensor; γ_a denotes the cost of taking the action; and α_a denotes the mis-representation or distortion weight [21]. Appendix C provides a detailed discussion of the reward function, including the case where the reward is an explicit function of the observation.

Tie-breaking rule: When $r_a'\eta_{y_k} = r_{\bar{a}}'\eta_{y_k}, \forall \bar{a} \in \mathcal{A}/\{a\}$, $a_k \sim \text{Uniform}(\mathcal{A})$, i.e, an action from the set $\mathcal{A}$ is chosen with probability $\frac{1}{|\mathcal{A}|}$, where $|\mathcal{A}|$ denotes the cardinality of set $\mathcal{A}$. The uniform sampling tie-breaking rule ensures that the public belief (1) is still a martingale. This is required in the proof of Theorem 2.

Public Belief Dynamics: The fusion center shares sensor k's decision with the social sensors and the public belief (1) is updated (by the fusion center and subsequent sensors) according to the social learning Bayesian filter (see [2], [22]) as follows:

$$\pi_k = T^{\pi}(\pi_{k-1}, a_k) = \frac{R_{a_k}^{\pi_{k-1}} \pi_{k-1}}{\mathbf{1}' R_{a_k}^{\pi_{k-1}} \pi_{k-1}}. \tag{6}$$

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

$$R_{ia}^{\pi} = \mathbb{P}(a_k | x = i, \pi_{k-1}) = \sum_{y \in \mathcal{Y}} \mathbb{P}(a_k | y, \pi_{k-1}) \mathbb{P}(y | x = i),$$
$$\mathbb{P}(a_k | y, \pi_{k-1}) = \begin{cases} 1 & \text{if } a_k = \arg\max_{a \in \mathcal{A}} r_a' \eta_{y_k}; \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

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

Remark (**Information Cascade**). Note that the (decision) likelihood (7) is an explicit function of the prior (public belief) π_{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 (7), if the action becomes independent of the observation, $R_{ia}^{\pi} = 1$ or 0. This in turn leads to information cascade, social learning stops as the public belief is frozen, as can be seen from (6). It can be shown that (Theorem 5.3.1, [2]) social learning stops in finite time.

Fusion Center Dynamics:

(1.) *Information Fusion cost:* The fusion center minimizes the following cost of information fusion $c(p_k)$, with

$$c(p_k) = p_k - \Phi_s(k) \mathcal{I}(a_k = y_k | \pi_{k-1}). \tag{8}$$

Here $\mathcal{I}$ denotes the indicator function. The cost function should model the trade-off between incentives and truthful information disclosure. Acting according to self valuations $(a = y)$ is in line with truthful information reporting in Peer Prediction literature; see [23]. We show in Sec.IV that $a = y$ corresponds to informative decisions. Here informativeness is in the sense of Blackwell [2]. One possible[2] cost function is (8). The information from different sensors is allowed to be weighed differently using $\Phi_s(k) \in (0,1)$. Here the subscript s is used to denote the cost when only social learning is considered (see Sec.VII-A for the case when entropy cost, in addition to the effect of social learning, is considered). For simplicity, we assume the weights to be same for all sensors; i.e $\Phi_s(k) = \phi_s,\ \forall\ k$. Appendix C provides a motivation of the information fusion cost using well studied models in economics [24], [1], [10].

(2.) *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 state estimate. Let $\mathcal{F}_k$ denote the history of past incentives and decisions $\{\pi_0, p_1, a_1, \cdots, p_{k-1}, a_k\}$ recorded by the fusion center and the social sensors. More technically, the sigma-algebra

$$\mathcal{F}_k := \Sigma(\pi_0, a_1, \ldots, a_k, p_1, \ldots, p_{k-1}). \tag{9}$$

The fusion center chooses the incentive as $p_{k+1} \in \mu_k(\mathcal{F}_k)$ for the sensor $k+1$ to provide information about its state via social learning. Here μ_k denotes a policy that associates the history $\mathcal{F}_k$ with an incentive p_{k+1}. Since $\mathcal{F}_k$ 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 π_k computed via the social learning filter (6) forms a sufficient statistic (see Sec.V for justification) for $\mathcal{F}_k$ and the incentive offered to social sensor $k+1$ is given as

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

The incentive is normalized to $[0,1]$ without loss of generality.

B. Controlled Information Fusion Objective

Given the setup in Sec.II-A, the aim of the fusion center is to estimate the state $x_0 (= x)$ by minimizing the cost of information acquisition (p). As discussed in (6), the fusion center performs Bayesian fusion of the information revealed by the social sensors.

Let $\bar{\boldsymbol{\mu}} = (\mu_0, \mu_1, \cdots)$ denote the sequence of policies employed by the fusion center at times $k = 0, 1, \cdots$. For each initial distribution π_0, the following cost is associated for the fusion center:

$$J_{\bar{\boldsymbol{\mu}}}(\pi) = \mathbb{E}_{\bar{\boldsymbol{\mu}}}\{\sum_{k=0}^{\infty} \rho^k c_{\mu_k}(p_k) | \pi_0 = \pi\}. \tag{11}$$

Here p_k denotes the incentive, $\rho \in [0,1)$ denotes an economic discount factor, μ_k denotes the decision policy (10) for the fusion center that maps the public belief π_k to an incentive $p_{k+1} \in [0,1]$, $c_{\mu_k}(p_k)$ denotes the cost of information fusion incurred at time k, and $\mathbb{E}_{\bar{\boldsymbol{\mu}}}$ denotes the expectation conditioned on the policy sequence $\bar{\boldsymbol{\mu}}$.

[2]In Sec.VII-A, we consider the information fusion cost that additionally has entropy of the state estimate.

The policy sequence $\bar{\boldsymbol{\mu}}$ can be restricted to the class of stationary (time invariant) policies $\boldsymbol{\mu} = (\mu, \mu, \cdots)$ for the infinite horizon discounted cost objective; see [2]. The fusion center aims to find the optimal stationary policy μ^* such that

$$J_{\mu^*}(\pi_0) = \inf_{\mu \in \boldsymbol{\mu}} J_\mu(\pi_0) \tag{12}$$

where $\boldsymbol{\mu}$ denotes the class of stationary policies.

Summary: (6) are the dynamics and (11) is the optimization objective 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 (5). The adaptive incentivizing problem is formulated as a non-standard (continuous actions) partially observed Markov decision process (POMDP) for the information fusion center to optimize the trade-off between the cost of information acquisition and consistency.

C. Example: Social Media Review Platform

We briefly motivate above set-up using an application on online social media review platforms like Amazon or Airbnb. Such review platforms offer the following benefits [25]: (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?

The state of a product $x \in \{1(\text{Bad}), 2(\text{Good})\}$ denotes quality, the observation $y \in \{1(\text{Bad}), 2(\text{Good})\}$ denotes experience, and the customers' decision $a \in \{1(\text{Neg. Review}), 2(\text{Pos. Review})\}$. When the customer writes a good/ bad review when it has a good/ bad experience, the review is honest. Here it is assumed that each customer leaves a review, however, the nature of review depends on the optimization (4). The information fusion objective is to estimate the product or service quality and, Amazon or AirBnb want to maximize the number of customers that report honest experiences. This informative feedback from the social sensors can be used by the retailers to improve the quality, and it will also benefit the future customers in that they are well informed before making a decision. In this sense, the objective (11) improves the overall welfare.

III. Structure of Optimal Incentive Policies

This section has three results. Sec.III-A formulates solving for the optimal incentive policy (12) as a stochastic dynamic programming problem. Sec.III-B provides sufficient conditions on the model parameters (B, r_a) under which the optimal incentive policy for the fusion center can be completely specified as a threshold policy. Sec.III-C provides a sample path characterization of the optimal incentive sequence when the fusion center employs the optimal policy.

A. Dynamic Programming Formulation

The optimal incentive policy μ^* in (12) and the corresponding optimal cost (value function) $V(\pi)$ satisfy the Bellman's stochastic dynamic programming equation [2]:

$$Q(\pi, p) = c(p) + \rho \sum_{a \in \mathcal{A}} V(T^\pi(\pi, a)) \sigma(\pi, a),$$

$$V(\pi) = \min_{p \in [0,1]} Q(\pi, p), \quad J_{\mu^*}(\pi_0) = V(\pi_0), \quad \text{and} \tag{13}$$

$$\mu^*(\pi) = \arg\min_{p \in [0,1]} Q(\pi, p). \tag{14}$$

where $T^\pi(\pi, a)$ is defined in (6) and $\sigma(\pi, a) = \mathbf{1}' R_a^\pi \pi$, and $c(p)$ is the information fusion cost defined in (8).

Discussion: Even though Bellman's equation (13) specifies the optimal policy, it has two problems:

(i) The state (belief) space $\Pi(2)$ is an uncountable set. Hence the dynamic programming equation (13) 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 [2] 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 estimate $\pi < \pi^*$, 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.1a), one only needs to compute the single belief π^*; whereas a general policy (as in Fig.1b) requires PSPACE hard dynamic programming recursion offline. (ii) To implement a controller with a threshold policy, one only needs to encode π^* 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

$$\Delta(\eta_y) = [l_1 \ -l_2] \frac{B_y \pi}{\mathbf{1}' B_y \pi} + l_3 \tag{15}$$

where η_y is the private belief update (3) with π_k replaced by π,

$$l_1 = \frac{\alpha_2}{\delta_2 - \delta_1}, \ l_2 = \frac{\alpha_1}{\delta_2 - \delta_1}, \ l_3 = \frac{\gamma_2 - \gamma_1}{\delta_2 - \delta_1}.$$

The incentive function (15) naturally arises by reformulating (4). A set of parameters in the incentive function that ensure $\Delta(\eta_y) \in [0, 1]$ are $l_1 > 0$, $l_2 > 0$ and $l_3 > 0$. A *sufficient condition* is that $\alpha_1 > \alpha_2$, $\delta_2 > \delta_1$ and $\gamma_2 > \gamma_1$. For other forms of reward functions (see Appendix C), the expression for $\Delta(\eta_y) \in [0, 1]$ and the conditions on the model parameters are suitably derived.

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

(A1) The observation distribution $B_{xy} = \mathbb{P}(y|x)$ 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 [2] 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. (see Sec.VII-D for assumptions in non-binary environments)

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 $\mu^*(\pi)$ of the controlled information fusion problem: the optimal policy has threshold structure (as illustrated in Fig.1a). The choice over a continuum of actions is reduced to a choice between two exactly specified incentive policies. The optimal policy is not unique. There exists a version of the optimal policy having the structure as in Theorem 1.

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

$$\mu^*(\pi) = \begin{cases} 0 & \text{if } \pi(2) \in [0, \pi_s^*(2)); \\ \Delta(\eta_{y=2}) & \text{if } \pi(2) \in [\pi_s^*(2), 1]. \end{cases} \tag{16}$$

Here the threshold state $\pi_s^*(2) \in (0,1)$ depends on the choice of $\phi_s \in (0,1)$ defined in (8), and the parameters in the incentive function $\Delta(\eta_{y=2})$ defined in (15).

The auxiliary results required for the proof are provided in the Appendix. These results show that due to the structure of the social learning filter in (6), the choice of incentives reduces from a continuum $[0,1]$ to a finite number at every belief. Also, the incentive function $\Delta(\eta_y)$ is decreasing in π for any y.

Proof. From Lemma 1 (see Appendix B), the value function (13) can be expressed as:

$$\begin{aligned} V(\pi) &= \min\{\rho V(\pi), \Delta(\eta_{y=2}) - \phi_s + \rho \mathbb{E} V(\pi), \\ &\qquad \Delta(\eta_{y=1}) + \rho V(\pi)\}. \\ \Rightarrow V(\pi) &= \min\{0, \Delta(\eta_{y=2}) - \phi_s + \rho \mathbb{E} V(\pi)\}, \end{aligned} \tag{17}$$

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

By using the value iteration algorithm [2] on (17), we have

$$V_{n+1}(\pi) = \min\{0, \Delta(\eta_{y=2}) - \phi_s + \rho \mathbb{E} V_n(\pi)\} \tag{18}$$

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

From Lemma 2 (see Appendix B), the incentive function is decreasing. From the definition of First-Order Stochastic Dominance (37), and Proposition 1, we have $\mathbb{E}V_n(\pi)$ is decreasing in π. Therefore, $V_{n+1}(\pi)$ and hence $V(\pi)$ is decreasing in π. Let $V(0)$ and $V(1)$ denote the values for $\pi = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$. It is seen by substitution that $\mathbb{E}V(0) = V(0)$ and $\mathbb{E}V(1) = V(1)$. By definition, we know that $\Delta(\eta_y) \in [0,1]$. Using[3] Lemma 2, let $\Delta(e_1) > \phi_s$ and $\Delta(e_2) < \phi_s$. The value function for the fusion center is given by (17). We have the following:

1) For $V(\pi) = \Delta(\eta_{y=2}) - \phi_s + \rho \mathbb{E} V(\pi)$, $V(0) = \frac{\Delta(e_1) - \phi_s}{(1-\rho)} > 0$, and $V(1) = \frac{\Delta(e_2) - \phi_s}{(1-\rho)} < 0$.
2) For $V(\pi) = 0$, $V(0) = V(1) = 0$.

The value function $V(\pi)$ in (17) is decreasing with a positive value at e_1 and a negative value e_2, so must be zero at some point(s). Let $\Sigma = \{\pi(2) | 0 = \Delta(\eta_{y=2}) - \phi_s + \rho \mathbb{E} V(\pi)\}$. Since the value function $V(\pi)$ is monotone in π, the set Σ is convex. Choosing $\pi_s^*(2) = \{\hat{\pi}(2) | \hat{\pi}(2) > \pi(2) \ \forall \ \pi(2) \in \Sigma\}$, the result follows. □

According to Theorem 1, computing the optimal incentive policy is equivalent to finding the belief $\pi_s^*(2)$, 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.1a). Therefore, the controlled information fusion problem reduces to a finite dimensional optimization problem of finding a threshold state π^*. 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 μ 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 $p = \Delta(\eta_{y=2})$ and $p = 0$; (iii) the region in the belief space $\Pi(2)$ where it is optimal to incentivize using $\Delta(\eta_{y=2})$ 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 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[4] over time.

Theorem 2. *Consider the information fusion problem with optimal policy $\mu^*(\pi)$ in (16). Under (A1), the optimal incentive sequence $p_k = \mu^*(\pi_{k-1})$ is a sub-martingale.*

Proof. 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 3 (see Appendix B), $\Delta(\eta_{y=2})$ is convex in π.

[3]Note that after normalization $\Delta(e_1) = 1$ and $\Delta(e_2) = 0$.

[4]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.

Let $u^S(\pi_{k+1}) = \Delta(\eta_{y_k=2})$ denote the price at time $k+1$. So $u^S(\pi)$ is convex in π.
We know that the public belief π_k is a martingale ([1]), i.e, $\mathbb{E}[\pi_{k+1}|\mathcal{F}_k] = \pi_k$. For $\epsilon \to 0$,

$$\begin{aligned}\mathbb{E}[W_{k+1}|\mathcal{F}_k] = \mathbb{E}[u^S(\pi_{k+1})|\mathcal{F}_k] &\geq u^S(\mathbb{E}[\pi_{k+1}|\mathcal{F}_k]) \\ &\geq u^S(\pi_k) \geq W_k\end{aligned}$$

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 Proposition 4 (see Appendix A), $(H.W)_k$ 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, $\mathbb{E}[p_{k+1}|\mathcal{F}_k] \geq p_k$, i.e, it increases on average over time. □

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 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.
The increase in incentives over time can be attributed to the fact that the senors polled for information at a later instant have more accurate estimate of the state due to learning from predecessors, and hence require higher compensation to reveal the same.

IV. Consistency of Controlled Information Fusion

An elementary application of the martingale convergence theorem [26] shows that the social learning protocol (6) 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, [2]). 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 sub-optimal incentive policy. We further provide uniform bounds on the additional cost incurred for consistency[5]. When it is sufficient to know the state with a degree of confidence, policies that guarantee state estimation in finite time are discussed. We also provide uniform bounds on the budget saved as a result of estimating the state only upto a degree of confidence.

[5]Let the true state be $x = \theta$. The pair (θ, π_k) is consistent, if π_k converges to a point mass at θ in probability.

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 (6). Recall that social learning terminates after a finite horizon (see remark on Information cascade after (7)). 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.

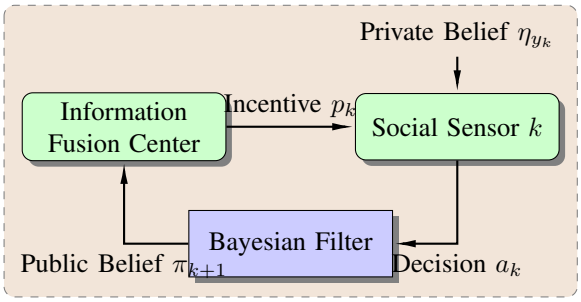


Fig. 3: Bi-directional interaction between the information fusion center and the social sensor. The fusion center provides an incentive p_k to the social sensor, which has a private belief η_{y_k} after observation y_k. The social sensor takes a decision a_k and this quantized information on the underlying state is used to update the public belief π_{k+1} using a social learning Bayesian filter (6). The incentive p_k at time k directly modifies the reward function of the social sensor, and hence affects the state estimate π_{k+1} at time $k+1$.

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 $\mathcal{P}_1^p$ - where action $a = 2$ is optimal; a region $\mathcal{P}_3^p$ - where action $a = 1$ is optimal; a region $\mathcal{P}_2^p$ - where action $a = y$ is optimal. This partition is possible because of (A1) and (A2); see [27]. From (4), the decision of the social sensor depends on the private belief η_y and the reward r_a (defined in (5)). Therefore, define:

$$\begin{aligned}\mathcal{P}_1^p &= \{\pi \in \Pi(2) : (r_1 - r_2)'\eta_{y=1} \leq 0\} \\ \mathcal{P}_2^p &= \{\pi \in \Pi(2) : (r_1 - r_2)'\eta_{y=1} > 0 \cap (r_1 - r_2)'\eta_{y=2} \leq 0\} \\ \mathcal{P}_3^p &= \{\pi \in \Pi(2) : (r_1 - r_2)'\eta_{y=2} > 0\} \end{aligned} \quad (19)$$

where r_a for $a = \{1, 2\}$ 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 $\mathcal{P}_1^p \cup \mathcal{P}_3^p$ is the *herding* region and $\mathcal{P}_2^p$ is the *social learning* region for any $p \in [0, 1]$.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2020.3046024, IEEE Transactions on Automatic Control

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

$$\pi_{k+1} \in \begin{cases} \mathcal{P}_3^p & \text{iff } p_k \in [0, \Delta(\eta_{y_k=2})); \\ \mathcal{P}_2^p & \text{iff } p_k \in [\Delta(\eta_{y_k=2}), \Delta(\eta_{y_k=1})); \\ \mathcal{P}_1^p & \text{iff } p_k \in [\Delta(\eta_{y_k=1}), 1]. \end{cases}$$

where the regions $\mathcal{P}_i^p$ for $i = 1, 2, 3$ are defined in (19), and $\Delta(\eta_y)$ is as in (15).

Proof. We'll prove that $\pi \in \mathcal{P}_2^p$ iff $p \in [\Delta(\eta_{y=2}), \Delta(\eta_{y=1}))$. Other cases are proved similarly. We can write

$$r_1 = [(\delta_1 p - \gamma_1) \quad (\delta_1 p - \alpha_1 - \gamma_1)], \tag{20}$$
$$r_2 = [(\delta_2 p - \alpha_2 - \gamma_2) \quad (\delta_2 p - \gamma_2)]. \tag{21}$$

By definition,

$$\mathcal{P}_2^p = \{\pi \in \Pi(2) : (r_1 - r_2)' \eta_{y=1} > 0 \cap (r_1 - r_2)' \eta_{y=2} \leq 0\}.$$

We have,

$$(r_1 - r_2)' \eta_{y=1} > 0 \Leftrightarrow p < \frac{1}{\delta_2 - \delta_1} \Big[[\alpha_2 \ -\alpha_1] \eta_{y=1} + (\gamma_2 - \gamma_1) \Big] = \Delta(\eta_{y=1}).$$
$$(r_1 - r_2)' \eta_{y=2} \leq 0 \Leftrightarrow p \geq \frac{1}{\delta_2 - \delta_1} \Big[[\alpha_2 \ -\alpha_1] \eta_{y=2} + (\gamma_2 - \gamma_1) \Big] = \Delta(\eta_{y=2}).$$

□

According to Theorem 3, relation between the incentive p_k at time k and the state estimate (public belief π_k) at the next instant $k+1$ is such that, when p_k belongs to the intervals defined by the private beliefs (in the incentive function $\Delta(\eta_{y_k})$), the widths of the herding and social learning regions change (see Fig.4) so that the public belief (π_{k+1}) belongs to the desired $\mathcal{P}_i^p$. Fig.4 shows the variation of the width of the regions with respect to the incentive parameter p. Theorem 3 characterizes the sensitivity of the regions $\mathcal{P}_1^p, \mathcal{P}_2^p, \mathcal{P}_3^p$ with respect to the incentive $p \in [0, 1]$, and Corollary 1 below shows how to stop the information cascade so that social learning can proceed indefinitely so that the state estimate converges to the true state.

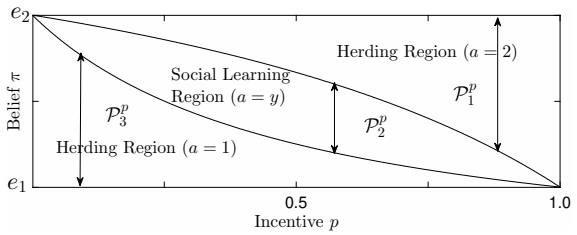


Fig. 4: Herding ($\mathcal{P}_1^p \cup \mathcal{P}_3^p$) and social learning ($\mathcal{P}_2^p$) regions with respect to the incentive parameter p. It is seen that when the incentives are small (close to 0), the sensors herd on low quality actions ($a = 1$); and when the incentives are high (close to 1), the sensors herd on high quality actions ($a = 2$); however, only the actions in the social learning region are informative or reflect the sensors' true valuation.

Corollary 1. *Let $p_k = \Delta(\eta_{y_k=2})$ 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 (19), Proposition 2 and Proposition 3 in the Appendix, the social sensors' decision likelihood matrices R_a^π (as in (6)) in regions $\mathcal{P}_1^p, \mathcal{P}_2^p$, and $\mathcal{P}_3^p$ for any $p \in [0, 1]$ are

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

respectively. In the herding region $\mathcal{P}_1^p \cup \mathcal{P}_3^p$, the decision of the social sensor is independent of the public belief and the public belief (6) is frozen. In the social learning region $\mathcal{P}_2^p$, the sensors take informative actions; i.e, each sensor acts according to its observation/valuation. Informativeness is in the sense of Blackwell; see [2]. For any two observation matrices $\mathcal{O}_1$ and $\mathcal{O}_2$, $\mathcal{O}_1$ is more informative than $\mathcal{O}_2$ in the Blackwell sense ($\mathcal{O}_1 \succ_B \mathcal{O}_2$) if $\mathcal{O}_2 = \mathcal{O}_1 \Gamma$, for any stochastic matrix Γ. When the sensors act according to their observations, $\pi \in \mathcal{P}_2^p$, and the decision likelihood matrix in (6) $R_S^\pi = B$; and when the sensors don't act according to the observations (they herd), $\pi \in \mathcal{P}_3^p$, the decision likelihood matrix $R_H^\pi = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$. We have for $\Gamma = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$, $R_H^\pi = R_S^\pi \Gamma \Rightarrow R_S^\pi \succ_B R_H^\pi$.

In the social learning region, sensors take informative actions $a = y$; or $R_a^\pi = B$. 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 [28], [29] that fusion of Bayesian estimates is consistent (convergence in probability); i.e, for a point mass at the true state θ denoted as $g(\theta)$, $\lim_{k\to\infty} \mathbb{P}(|\pi_k - g(\theta)| > \epsilon) = 0 \ \forall \ \epsilon > 0$. In other words, the fusion center *can* learn the true state asymptotically by choosing the incentives as $p_k = \Delta(\eta_{y_k=2})$ for $k = 1, 2, \ldots$.

B. Cost of consistency for the fusion center

When the incentive policy is the optimal threshold policy (16), the fusion of Bayesian estimates computed from the social sensors' decisions (6) 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.1a). From Theorem 3, when the fusion center stops incentivizing $p = \mu^*(\pi) = 0$, the public belief is in the herding region $\mathcal{P}_3^p$. 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 (23), it will incur additional cost for the incentives; but the fusion of estimates computed from the social sensors' decisions (6) will be consistent (Corollary 1). Theorem 4 below provides uniform bounds on the additional cost incurred by the fusion center for employing a sub-optimal incentive policy that results in consistent information fusion. Consider the objective function for the fusion center:

$$W_{\mu_c}(\pi) = \mathbb{E}_{\mu_c}\{\sum_{k=0}^{\infty} \rho^k c_{\mu_c}(p_k) | \pi_0 = \pi\} \tag{22}$$

where $W_{\mu_c}(\pi)$ denotes the cost incurred by employing the sub-optimal policy (compare with (16))

$$\mu_c(\pi) = \{\Delta(\eta_{y=2}) \ \forall \ \pi(2) \in [0,1]\}. \tag{23}$$

Theorem 4. *Let (A1) hold. The additional cost (on average) incurred by the fusion center for employing the sub-optimal policy $\mu_c(\pi)$ in (23) instead of the optimal policy $\mu^*(\pi)$ in (16) is bounded as:*

$$\sup_{\pi} |W_{\mu_c}(\pi) - J_{\mu^*}(\pi)| \leq 2\frac{(1-\phi_s)}{1-\rho} \tag{24}$$

where $J_{\mu^*}(\pi)$ is the optimal cost (12).

Proof. Define the following region in the belief space $\Pi(2)$:

$$\mathcal{H} = \{\pi | \pi(2) \leq \pi^*(2)\}. \tag{25}$$

Here $\mathcal{H}$ denotes the region where the optimal policy in (16) is such that $\mu^*(\pi) = 0$. For any sub-optimal policy μ_c and the corresponding cost $W_{\mu_c}(\pi)$, it is clear that $W_{\mu_c}(\pi) - J_{\mu^*}(\pi) \geq 0 \ \forall \ \pi$. Also, $W_{\mu_c}(e_2) = J_{\mu^*}(e_2)$. Let $\mathcal{I}$ denote the indicator function. We have

$$\begin{aligned} W_{\mu_c}(\pi) - J_{\mu^*}(\pi) &= \mathcal{I}(\pi \in \mathcal{H})\{W_{\mu_c}(\pi) - J_{\mu^*}(\pi)\} \\ &\quad + \mathcal{I}(\pi \notin \mathcal{H})\{W_{\mu_c}(\pi) - J_{\mu^*}(\pi)\} \\ \Rightarrow \sup_{\pi} |W_{\mu_c}(\pi) - J_{\mu^*}(\pi)| &\leq \Big\{ \sup_{\pi} \mathcal{I}(\pi \in \mathcal{H})\{W_{\mu_c}(\pi) - J_{\mu^*}(\pi)\} \\ &\quad + \sup_{\pi} \mathcal{I}(\pi \notin \mathcal{H})\{W_{\mu_c}(\pi) - J_{\mu^*}(\pi)\} \Big\}. \end{aligned}$$

where $\mathcal{H}$ is defined in (25). From Theorem 1, we know that $J_{\mu^*}(\pi) = V(\pi)$ is monotone (non-increasing) in π. Similar arguments can be used to establish that $W_{\mu_c}(\pi)$ is monotone (non-increasing) in π. Therefore, we have for

$$\sup_{\pi} |W_{\mu_c}(\pi) - J_{\mu^*}(\pi)| \leq 2\Big\{ \sup_{\pi} \mathcal{I}(\pi \in \mathcal{H}) W_{\mu_c}(\pi) \Big\} \tag{26}$$

as $J_{\mu^*}(\pi) = 0 \ \forall \ \pi \in \mathcal{H}$ from (17) and Theorem 1. Equation (26) follows from

$$\mathbb{E}_{\mu_c}\Big\{ \sum_{k=0}^{\infty} \rho^k \{c_{\mu_c}(p_k)\} \Big\} > \mathbb{E}_{\mu_c}\Big\{ \sum_{k=0}^{\infty} \rho^k \{\mathcal{I}(\pi_k \in \mathcal{H}) c_{\mu_c}(p_k)\} \Big\}.$$

The set $\mathcal{H}$ defined in (25) is compact by definition. For the discount factor $\rho \in [0,1)$ and bounded instantaneous costs, the cumulative discounted cost is bounded [2]. Therefore in (26),

$$\sup_{\pi}\{\mathcal{I}(\pi \in \mathcal{H}) W_{\mu_c}(\pi)\} = \max_{\pi}\{\mathcal{I}(\pi \in \mathcal{H}) W_{\mu_c}(\pi)\}$$

and $\tilde{\pi} = \text{argmax}_{\pi}\{\mathcal{I}(\pi \in \mathcal{H}) W_{\mu_c}(\pi)\}$. We have for $\pi_0 = \tilde{\pi}$,

$$\begin{aligned} \max_{\pi}\{\mathcal{I}(\pi \in \mathcal{H}) W_{\mu_c}(\pi)\} &= \mathbb{E}_{\mu_c}\Big\{ \sum_{k=0}^{\infty} \rho^k \{c_{\mu_c}(p_k)\} \Big| \pi_0 = \tilde{\pi} \Big\} \\ &\leq \mathbb{E}_{\mu_c}\Big\{ \sum_{k=0}^{\infty} \rho^k \max_{\Delta(\eta_{y=2}):\pi \in \mathcal{H}} c_{\mu_c}(p_k) \Big\} \\ &= (1-\phi_s)\mathbb{E}\Big\{ \sum_{k=0}^{\infty} \rho^k \Big\} = \frac{(1-\phi_s)}{1-\rho}. \end{aligned}$$

□

$W_{\mu_c}(\pi)$ and $J_{\mu^*}(\pi)$ are decreasing in π, and can be established using similar arguments as in Theorem 1 and $W_{\mu_c}(\pi) - J_{\mu^*}(\pi) \geq 0 \ \forall \ \pi$. Theorem 4 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 the weight ϕ_s in the information fusion cost (8), discount factor ρ that captures the degree of impatience of the fusion center.

The usefulness of Theorem 4 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 (6) is consistent. (ii) It helps in choosing the weight ϕ_s and the discount factor ρ for the fusion center.

C. Finite time bounds for the fusion center

In Sec.IV-B, it was shown that by employing a sub-optimal policy the fusion center can estimate the true state asymptotically. However, it is often enough to know the state with a degree of confidence. In this section, we obtain uniform bounds on the budget saved by estimating the state upto a degree of confidence.

The degree of confidence characterizes regions in the belief space $\Pi(2)$ that can be used to estimate the states. For a degree of confidence $\vartheta \in (0,1)$, any belief in the confidence region $\pi(2) \in [0, \vartheta]$ is identified with state $x = 1$, and any belief in the confidence region $\pi(2) \in [1-\vartheta, 1]$ is identified with state $x = 2$. For example, when the public belief (posterior) is such that $\pi(2) \in [0.9, 1]$, then the fusion center is (atleast) 90% confident that the state $x = 2$, and if $\vartheta < 0.1$, the state is estimated as $x = 2$. For a degree of confidence $\vartheta \in (0,1)$, consider using the following policy

$$\mu_f(\pi) = \begin{cases} 0 & \text{if } \pi(2) \in [0, \vartheta]; \\ \Delta(\eta_{y_k}) & \text{if } \pi(2) \in (\vartheta, 1-\vartheta); \\ 0 & \text{if } \pi(2) \in [1-\vartheta, 1]. \end{cases} \tag{27}$$

It can be shown using martingale convergence theorem [26] that when using the policy (27), the public belief hits one of the two confidence regions in finite time. The arguments are similar to those used to establish information cascades occur in finite time in [6] and in [2][Theorem 5.3.1].

The following theorem provides a bound on the budget saved by employing the policy in (27) instead of the policy (23). Let $\pi_\vartheta = \begin{bmatrix} \vartheta \\ 1-\vartheta \end{bmatrix}$ and $\eta_\vartheta = \frac{B_{y=2}\pi_\vartheta}{\mathbf{1}' B_{y=2}\pi_\vartheta}$.

Theorem 5. *Let (A1) hold. For a degree of confidence ϑ, the budget saved by the fusion center by employing the policy $\mu_f(\pi)$ in (27) instead of the policy $\mu_c(\pi)$ in (23) is bounded as:*

$$\sup_{\pi} |W_{\mu_c}(\pi) - W_{\mu_f}(\pi)| \leq 2\frac{(1-\phi_s)}{1-\rho} + \frac{|\Delta(\eta_\vartheta) - \phi_s|}{1-\rho}. \tag{28}$$

where ρ is the discount factor.

Discussion: The proof follows using arguments similar to Theorem 5 in the paper. Theorem 5 provides an uniform bound on the budget saved by employing the policy $\mu_f(\pi)$ in (27) instead of $\mu_c(\pi)$ in (23). A bound on the budget saved with respect to the optimal policy $\mu^*(\pi)$ can be obtained from

Theorem 5 and Theorem 4 using the triangle inequality of the norm,

$$|J_{\mu^*}(\pi)-W_{\mu_f}(\pi)| \leq |W_{\mu_c}(\pi)-J_{\mu^*}(\pi)|+|W_{\mu_c}(\pi)-W_{\mu_f}(\pi)|.$$

In Theorem 5, the fact (Lemma 2) that $\Delta(\eta_{y=2})$ is decreasing in π, and $|\varepsilon| \geq \varepsilon$ is utilized in deriving the bounds.

V. Strategic behaviour in Social Sensors

The information fusion center polls the social sensors in a pre-determined order and they decide what information to reveal, i.e, it was assumed that the sensors do not hide their signals and are not strategic. However, the rewards can be suitably designed so that the sensors reveal information when polled. In this section, we show how to design the reward functions to prevent the social sensors from being strategic. This implies that the social sensors have no forward-looking tendencies and reward function of the social sensors has no externalities, and the public belief (1) forms a sufficient statistic for the history of past actions and incentives.

Under an additional minor restriction on the reward parameters, it is shown below that the social sensors have no incentive to delay or hide their signals. This restriction is independent of the actual form of the rewards when the rewards in both states are non-zero.

A. Social sensors do not display contrarian behavior

The optimal policy for the fusion center dictates that it either incentivize or not incentivize, see Theorem 1. When the fusion center is offering incentives ($\Delta(\eta_{y=2})$), from Theorem 3, it is seen that it is optimal for the social sensors to act according to their observations. As the social sensors are assumed to be Bayes rational, they have no incentive to deviate. When the fusion center is not incentivizing, the sensors always herd.

B. Social sensors are not strategic

Let $\mathcal{R}_H$ and $\mathcal{R}_S$ denote the regions where the fusion center does not incentivize ($\mu(\pi) = 0$) and incentivizes ($\mu(\pi) = \Delta(\eta_{y=2})$) respectively. A social sensor deciding at time k considers the following scenarios:

a.) $\pi_k \in \mathcal{R}_S$ and $\pi_{k+1} \in \mathcal{R}_H$. In other words if the sensor delays revealing information and the belief update after the next $(k+1)$ sensors' decision belongs to the region where there is no incentivization.
$p_{k+1} = 0$, so the social sensor k would be better off revealing at time k.

b.) $\pi_k,\ \pi_{k+1} \in \mathcal{R}_S$ and $\pi_{k+1}(2) < \pi_k(2)$.
Consider the rewards for the social sensor from (5),

$$\begin{aligned} r_1 &= [\delta_1 p + \Gamma_{11} \quad \delta_1 p + \Gamma_{21}] \\ r_2 &= [\delta_2 p + \Gamma_{12} \quad \delta_2 p + \Gamma_{22}] \end{aligned} \tag{29}$$

Assume $\Gamma_{ij} > 0$ for all i, j without loss of generality. Note that the reward vector r_a is also required to be supermodular for any p , so $\Gamma_{11} > \Gamma_{21}$ and $\Gamma_{22} > \Gamma_{12}$. Let $T(\pi, y_k) = \frac{B_{y_k}\pi}{\mathbf{1}' B_{y_k}\pi}$ denote the private belief of sensor k. There are two possible observations for the social sensor k, $y_k = 1, 2$. We will establish the result for $y_k = 1$, and the result follows immediately for $y_k = 2$.
Let $\bar{r}_a = [\delta_a p_{k+1} + \Gamma_{1a} \quad \delta_a p_{k+1} + \Gamma_{2a}]$.

Theorem 6. *Let the observation of sensor k be $y_k = 1$. There is a discount factor $D \in (0, 1]$ such that $r_1' T(\pi_k, y_k = 1) \geq D\ \bar{r}_1' T(\pi_{k+1}, y_k = 1)$.*

Proof. From the definition of First-order stochastic dominance and TP2 on B, we have the following[6]

$$\begin{aligned} & r_1' T(\pi_k, y_k = 1) \leq \bar{r}_1' T(\pi_{k+1}, y_k = 1) \\ \therefore\ & r_1' T(\pi_k, y_k = 1) > D\ \bar{r}_1' T(\pi_{k+1}, y_k = 1), \\ & \text{where } D = \frac{r_1' T(\pi_k, y_k = 1)}{\bar{r}_1' T(\pi_{k+1}, y_k = 1)} - \epsilon, \text{for } \epsilon > 0. \end{aligned}$$

Considering the largest possible deviation[7] $\pi_k(2) = 1$ and $\pi_{k+1}(2) = 0$, it is easily seen that the smallest value for $D = \frac{\Gamma_{21}}{\delta_1 + \Gamma_{11}} - \epsilon < 1$. □

Discussion: The social sensors are not more forward looking than D from Theorem 6. By suitably choosing the reward parameters, we can obtain $D = 1$. This implies that the social sensors have no incentive to deviate when $y_k = 1$.

c.) $\pi_k,\ \pi_{k+1} \in \mathcal{R}_S$ and $\pi_{k+1}(2) > \pi_k(2)$. By using similar arguments as in Theorem 6, we obtain the discount factor $D = \frac{\Gamma_{12}}{\delta_2 + \Gamma_{22}} - \epsilon < 1$. By suitably choosing the reward parameters, we can obtain $D = 1$. This implies that the social sensors have no incentive to deviate when $y_k = 2$. Also, the result follows immediately for $y_k = 1$.

It was shown that when the reward parameters are chosen so that $D = 1$, myopically maximizing the expected reward is a Markov perfect equilibrium.

VI. Controlled Information Fusion with Dynamic States

So far, we considered the problem of incentivized information fusion for estimating the random variable $x \in \mathcal{X}$. In this section, we consider the information fusion to estimate the state of a Markov chain x_k for $k = 0, 1, 2, \cdots$ with social sensors. The dynamic states might correspond to, for example, a change in the product/ service quality on AirBnb or Amazon.

Let the state x_k evolve as a Markov chain on the space $\mathcal{X}$ with a transition probability matrix P and an initial distribution π_0 in (1). Below we briefly highlight the changes in the social learning model in Sec.II-A for the case of dynamic states.
The private belief update in (3) for the social sensors taking the possible state change into account is given as

$$\eta_{y_k} = \frac{B_{y_k} P' \pi_{k-1}}{\mathbf{1}' B_{y_k} P' \pi_{k-1}} \tag{30}$$

The public belief update in (6) taking the possible state change into account is given as

$$\pi_k = T^{\pi}(\pi_{k-1}, a_k) = \frac{R_{a_k}^{\pi_{k-1}} P' \pi_{k-1}}{\mathbf{1}' R_{a_k}^{\pi_{k-1}} P' \pi_{k-1}}. \tag{31}$$

[6]Note that for any $a, b > 0$, $a < b \Rightarrow a > (\frac{a}{b} - \epsilon)b$ for any $\epsilon > 0$.

[7]Note that $\pi_k,\ \pi_{k+1} \in \mathcal{R}_S$. Clearly, this is included in $\pi_k(2), \pi_{k+1}(2) \in [0, 1]$.

The optimal incentive policy in case of a random variable $\mu^*(\pi)$ in Theorem 1 is near optimal for the case of dynamic states, when transitions out of the current state is allowed only with a small probability. This is shown in Theorem 7 below.

Let $\mu^*(\pi)$ denote the optimal policy for estimating/ localizing the random variable ($P = I$); and $\mu^*_\epsilon(\pi)$ denote the optimal policy for estimating/ tracking the state of a Markov chain with $P = \begin{bmatrix} 1-\epsilon_1 & \epsilon_1 \\ \epsilon_2 & 1-\epsilon_2 \end{bmatrix}$, where $\epsilon_1, \epsilon_2 > 0$.

Theorem 7. *Let $\rho \in [0,1)$ denote the economic discount factor. Let $V_{\mu^*}(\pi)$ and $V_{\mu^*_\epsilon}(\pi)$ denote the optimal costs incurred by employing the optimal policy $\mu^*(\pi)$ and $\mu^*_\epsilon(\pi)$ respectively. The following holds:*

$$V_{\mu^*}(\pi) - V_{\mu^*_\epsilon}(\pi) \leq \frac{2\rho(1-\phi_s)(\epsilon_1+\epsilon_2)}{(1-\rho)^2} \quad \times \max\{|B_{21}-B_{11}|, |B_{22}-B_{12}|\}. \quad (32)$$

Discussion: The proof follows from [30][Theorem 2]. Theorem 7 says that the policy $\mu^*(\pi)$ incurs a total cost $V_{\mu^*}(\pi)$ that is within $O(\epsilon_1+\epsilon_2)$ of the total cost $V_{\mu^*_\epsilon}(\pi)$. When $\epsilon_1, \epsilon_2 << 1$, the policy $\mu^*(\pi)$ for the state localization problem ($P = I$) is near optimal for the state tracking problem ($P \neq I$).

Characterizing the nature of the optimal incentive sequence (as in Sec.III) in case of a random variable relied on the crucial fact that the belief is a martingale unconditional on the state. However, when the states are changing, the public belief (31) is not a martingale (see [1]). This implies that, even though, $\mu^*(\pi)$ is near optimal, the incentive sequence that results from the fusion center employing $\mu^*(\pi)$ need not show an increasing trend on average.

VII. Numerical Results

Sec.VII-A below illustrates controlled information fusion with quadratic cost unlike (8). It is shown that a multi-threshold incentive policy is optimal for the fusion center. Sec.VII-B illustrates the sensitivity of the optimal threshold (16) to the parameters ϕ_s (the weight in (8)) and ρ (discount factor in the objective (11)) that are chosen by the fusion center. Sec.VII-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. Sec.VII-D discusses the formulation and a numerical simulation for the controlled information fusion in non-binary environments.

Bellman's equation (13) is solved by discretizing the belief 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 [2] for a duration of $N = 100$.

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 (8), in particular we consider the *entropy* cost. We will see that the optimal incentive policy has a multi-threshold structure (as in Fig.1b).

$\alpha_1 = 0.288$	$\alpha_2 = 0.278$	$\beta_1 = 0.11$
$\beta_2 = 0.1$	$\gamma_1 = 0.1$	$\gamma_2 = 0.414$

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 (44) parameters with the observation matrix B=$\begin{bmatrix} 0.8 & 0.2 \\ 0.4 & 0.6 \end{bmatrix}$.

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,

$$c(p) = p + \psi_e(\pi)C_e(\pi) - \phi_e \mathcal{I}(a = y|\pi) \quad (33)$$

where $\phi_e \in (0,1)$ denotes the scalar weight, p denotes the expenditure, ψ_e denotes the importance of the entropy cost, and $C_e(\pi) = -\sum_{i=1}^2 \pi(i)\log_2\pi(i)$ for $\pi(i) \in (0,1)$ and $C_e(\pi) \stackrel{\Delta}{=} 0$ 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 (8). 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 4 provide a rationale for choosing the parameters: ϕ_s – the weight in the information fusion cost (8) and ρ – 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 (8), and see how it affects the threshold π^*_s in (16). Fig.5 shows the affect of increasing the weight ϕ_s when the remaining parameters are the same. It can be seen that π^*_s is decreasing with ϕ_s. From Theorem 4, higher ϕ_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 ρ 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.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2020.3046024, IEEE Transactions on Automatic Control

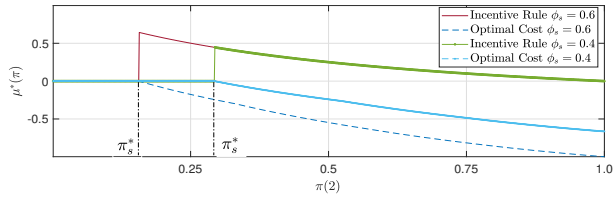


Fig. 5: Usefulness of information vs Incentivizing trade-off for the fusion center. It can be seen that $\pi^*(2)$ is decreasing with ϕ_s – a higher weight will necessitate incentivizing sooner. According to Theorem 4, higher ϕ_s implies that the additional cost for employing a sub-optimal policy is smaller; in other words, π_s^* is smaller. The parameters of the incentive function (15) are given in Table I and the discount factor $\rho = 0.4$. Here ϕ_s denotes the weight in the information fusion cost (8).

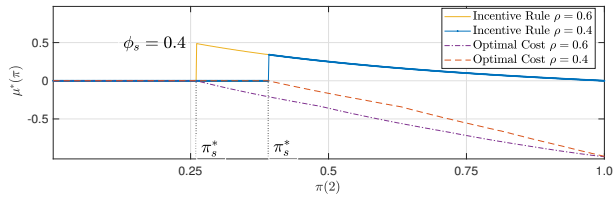
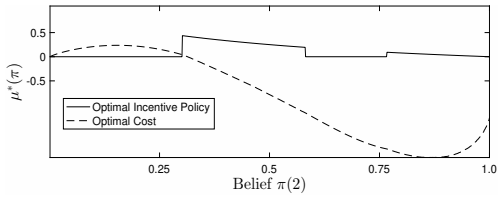
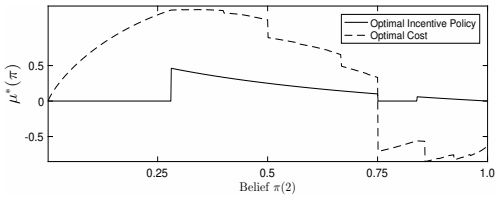


Fig. 6: Optimal cost vs Discount factor. It is seen that a higher discount factor leads to smaller (expected) costs for higher states. This indicates that it is beneficial for the fusion center to attach more importance to future costs as it should also take into account the benefit from sensors performing social learning. The parameters of the incentive function (15) are specified in Table I and the weight $\phi_s = 0.4$. Here ρ denotes the discount factor in the objective (11) and ϕ_s denotes the weight in the information fusion cost (8).



(a) The parameters are in Table I with $\phi_e = 0.25$, discount factor $\rho = 0.8$, and $\psi_e(\pi) = 0.1 - \pi^2(2)$. Here $\psi_e(\pi)$ captures the requirement of higher weight when the belief is smaller.



(b) Discontinuous optimal cost. The parameters are in Table I with $\phi_e = 0.4$, discount factor $\rho = 0.6$, and $\psi_e(\pi) = 0.6 \times \mathcal{I}(\pi(2) < 0.75) - 0.35 \times \mathcal{I}(\pi(2) > 0.75)$. Here $\psi_e(\pi)$ captures the requirement of higher weight when the belief is smaller.

Fig. 7: Multi-threshold incentive policy with entropy cost. The regions in the belief space $\Pi(2)$ 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 it is 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.

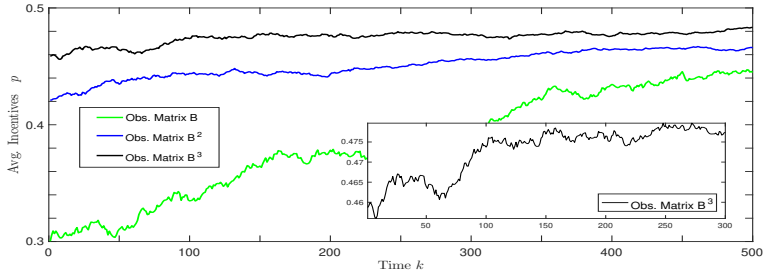


Fig. 8: The figure shows the incentives averaged over independent sample paths for the fusion center over time for observation matrices B, B^2 and B^3. The observation matrices are ordered in the decreasing order of informativeness (see Footnote 8). The parameters are specified in Tables I & II. The weight $\phi_s = 0.4$ in the information fusion cost (8) 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. The zoomed in subfigure shows the increasing trend in case of observation matrix B^3. It can be seen that average incentives offered in case of B^3 is higher than B^2 which in turn is higher than B.

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 (16) and fuses the information revealed by social sensors in a Bayesian way (6).

Obs. matrix B^2	$\alpha_1 = 0.3132$	$\alpha_2 = 0.3032$	$\beta_1 = 0.11$
	$\beta_2 = 0.1$	$\gamma_1 = 0.1$	$\gamma_2 = 0.414$
Obs. matrix B^3	$\alpha_1 = 0.3233$	$\alpha_2 = 0.3133$	$\beta_1 = 0.11$
	$\beta_2 = 0.1$	$\gamma_1 = 0.1$	$\gamma_2 = 0.414$

TABLE II: The reward vector (44) 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 (44) parameters with observation matrix B= $\begin{bmatrix} 0.8 & 0.2 \\ 0.4 & 0.6 \end{bmatrix}$.

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 [2].

Parameters: The parameters of the incentive function (15) using the resolution dependent reward (44) for B^2 and B^3 are specified in Table II. 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). It can be seen from Fig.8 that the average incentives display an increasing trend.

D. Controlled Information Fusion in non-binary environments

In this section, we briefly discuss the formulation for multiple states. Partial results on social learning with multiple states and 2 actions appears in [27]. In the controlled fusion problem considered in this paper, the social sensors reveal the observation to the fusion center. This requires that the cardinality of $\mathcal{A}$ and $\mathcal{Y}$ be equal. Due to the complexity of analyzing the structural results for the optimal policy in case of multiple actions and states, we only describe the formulation and illustrate the incentive policy using a numerical simulation for a $\mathcal{X} = \mathcal{A} = \mathcal{Y} = \{1, 2, 3\}$. When $|\mathcal{X}| = 3$, the public belief is in the belief space

$$\Pi(3) \stackrel{\Delta}{=} \{\pi \in \mathbb{R}^2 : \sum_i \pi(i) = 1, 0 \le \pi(i) \le 1 \text{ for } i \in \{1, 2, 3\}\}.$$

The number of regions in the space $\Pi(3)$ that need be considered for analyzing the structural results of the optimal incentive policy are 5 (see (35) below) as opposed to 3 in (19).

Model Assumptions:

(A'1) The observation distribution $B_{xy} = \mathbb{P}(y|x)$ is TP2 (totally positive of order 2), i.e, all second order minors of matrix B are non-negative.

(A'2) The reward vector r_a is supermodular, i.e, $r_{a+1} - r_a$ is an increasing vector for $a = \{1, 2\}$ and every $p \in [0, 1]$.

The social sensors' decision $a(\pi, y) = \arg\max r_a' \eta_y$ is increasing in π and y under (A'1) and (A'2); see [2]. This can be used to establish the single crossing condition,

$$\{\pi \in \Pi(3) : (r_a - r_{a+1})' \eta_y \le 0\} \subseteq \{\pi \in \Pi(3) : (r_a - r_{a+1})' \eta_{y+1} \le 0\}. \quad (34)$$

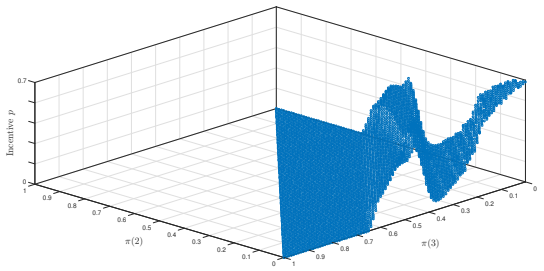


Fig. 9: Optimal Incentive Policy for social learning weight $\phi_s = 0.6$, $\rho = 0.8$, $\beta_1 = 0.6771, \beta_2 = 0.5465, \beta_3 = 0.7113$, $\delta_1 = 0.3, \delta_2 = 0.4, \delta_3 = 0.5$, $\gamma_1 = 0.5, \gamma_2 = 0.3, \gamma_3 = 0.2$, and $\alpha_a = 0$ for all $a \in \{1, 2, 3\}$. The belief space $\Pi(3)$ was discretized into a grid of 5151 belief points using Freudenthal triangulation [2]. The incentive $p \in [0, 1]$ was discretized into 50 values. Note that the optimal policy is a non-convex, non-monotone function of the belief.

We now can define the following regions in the belief simplex $\Pi(3)$ (compare with (19)):

$$\begin{aligned}
\mathcal{P}_1^p &= \{\pi \in \Pi(3) : (r_1 - r_3)' \eta_{y=1} \cap (r_2 - r_3)' \eta_{y=1} \le 0\},\\
\mathcal{P}_2^p &= \{\pi \in \Pi(3) : (r_1 - r_3)' \eta_{y=2} \le 0 \cap (r_2 - r_3)' \eta_{y=2} \le 0\\
&\quad \cap \ (r_1 - r_2)' \eta_{y=1} \le 0\},\\
\mathcal{P}_3^p &= \{\pi \in \Pi(3) : (r_1 - r_3)' \eta_{y=3} \le 0 \cap (r_2 - r_3)' \eta_{y=3} \le 0\\
&\quad \cap (r_1 - r_2)' \eta_{y=2} \le 0 \cap (r_2 - r_3)' \eta_{y=2} > 0\\
&\quad \cap \ (r_1 - r_2)' \eta_{y=1} > 0 \cap \{(r_1 - r_3)' \eta_{y=1} > 0\},\\
\mathcal{P}_4^p &= \{\pi \in \Pi(3) : (r_1 - r_2)' \eta_{y=3} \le 0 \cap (r_2 - r_3)' \eta_{y=3} > 0\\
&\quad \cap \ (r_1 - r_2)' \eta_{y=2} > 0 \cap (r_1 - r_3)' \eta_{y=2} > 0\},\\
\mathcal{P}_5^p &= \{\pi \in \Pi(3) : (r_1 - r_2)' \eta_{y=3} > 0 \cap (r_1 - r_3)' \eta_{y=3} > 0\}.
\end{aligned} \quad (35)$$

The value function for the fusion center is given by:

$$\begin{aligned}
V(\pi) &= \min\{c(p) + \rho \sum_a \sum_{j=1}^5 V(T^j(\pi, a)) \sigma(\pi, a) \mathcal{I}(\pi \in \mathcal{P}_j^p)\},\\
V(\pi) &= \min_{p \in [0,1]} \Big\{ p - \phi_s \mathcal{I}(\pi \in \mathcal{P}_3^p) +\\
&\qquad \rho \sum_a \sum_{j=1}^5 V(T^j(\pi, a)) \sigma(\pi, a) \mathcal{I}(\pi \in \mathcal{P}_j^p) \Big\}.
\end{aligned} \quad (36)$$

Here $T^j(\pi, a) = \frac{R_a^j \pi}{\mathbf{1}' R_a^j \pi}$, with $R^j = BM^j$ for $j = 1, 2, \cdots, 5$. Fig.9 shows the optimal incentive policy for a 3 state, observation, and action model. Lemma 2 in the paper can be used to find the matrices M^j for $j = 1, 2, \cdots, 5$. The observation distribution for the controlled fusion problem for 3 states and actions is chosen as:

$$B = \begin{bmatrix} 0.7479 & 0.1986 & 0.0536 \\ 0.6023 & 0.2543 & 0.1434 \\ 0.2785 & 0.2459 & 0.4756 \end{bmatrix}.$$

The value iteration algorithm based on (36) was run for a horizon $N = 100$.

Remark (**Approximation Methods**). Using the structural results in this paper, stochastic approximation algorithms can be modified to search for optimal policies restricted to the class of policies having the threshold structure; see [2][Section 12.4.2].

VIII. 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 five 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 thereby 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. Theorem 4 provided uniform bounds on the average additional cost incurred, by employing a sub-optimal policy, for consistency. Theorem 5 provided uniform bounds on the budget saved by employing a policy that estimates the state with a degree of confidence, instead of the optimal policy. Finally, Theorem 7 established that the optimal policy for estimating a random variable is near optimal for tracking a changing state, when out-of-state transition probabilities are small.

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. We briefly discussed the formulation for the case of 3 states, observations and actions, and highlighted the difficulty in deriving structural results in non-binary environments.

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

$$\sum_{i=j}^{2} \pi_1(i) \geq \sum_{i=j}^{2} \pi_2(i) \text{ for } j \in \{1,2\}. \tag{37}$$

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, $v_1 \geq v_2 \geq \dots v_X$.

Definition 2. (Martingale [26]): Let $\mathcal{F}_k$ denote the sigma algebra (as in (9)). 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 $\mathbb{E}[X_{k+1}|\mathcal{F}_k] \geq X_k$, for all k., the sequence $\{X_k\}$ is a *sub-martingale*.

Definition 3. ([26]) 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 $k-1$.

Proposition 1 ([2])**.** *Under (A1), we have* $\sigma(\pi_1, a) \geq_s \sigma(\pi_2, a)$, *where* $\sigma(\pi, a) = \begin{bmatrix} \boldsymbol{1}' B^{\pi}_{y=1}\pi \\ \boldsymbol{1}' B^{\pi}_{y=2}\pi \end{bmatrix}$.

Proposition 2 ([27])**.** *The sensor decision likelihood matrix* R^π *in the social learning filter (6) is computed as*

$$R^\pi = BM^\pi \text{ where } M^\pi_{y,a} = \mathbb{P}(a|y,\pi) = \mathcal{I}(r'_a B_y \pi > r'_{\bar{a}} B_y \pi), \quad \text{with } \bar{a} = \mathcal{A}/a. \tag{38}$$

Proposition 3 ([27])**.** *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^π *in (38) is a constant with respect to the belief state* π.

Proposition 4 ([26])**.** *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.*

Proposition 4 appears in [26][Theorem 5.2.5].

Appendix B
Auxiliary Results

Lemma 1. *Let* $\Delta(\eta_{y=1})$ *and* $\Delta(\eta_{y=2})$ *be two possible incentives at belief* π. *Under (A1) and (A2), the Q function in (13) can be simplified as:*

$$Q(\pi,p) = \begin{cases} p + \rho V(\pi) & \text{if } p \in [0, \Delta(\eta_{y=2})); \\ p - \phi_s + \rho \mathbb{E}V(\pi) & \text{if } p \in [\Delta(\eta_{y=2}), \Delta(\eta_{y=1})); \\ p + \rho V(\pi) & \text{if } p \in [\Delta(\eta_{y=1}), 1]. \end{cases} \tag{39}$$

and $V(\pi) = \min Q(\pi, p)$. *Here,* $\mathbb{E}V(\pi) = \boldsymbol{1}' B^{\pi}_{y=1}\pi \times V(\eta_{y=1}) + \boldsymbol{1}' B^{\pi}_{y=2}\pi \times V(\eta_{y=2})$.

Proof. From Proposition 2 and Proposition 3, we have

$$R^\pi = \begin{cases} \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} & \text{if } p \in [0, \Delta(\eta_{y=2})); \\ \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} & \text{if } p \in [\Delta(\eta_{y=2}), \Delta(\eta_{y=1})); \\ \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix} & \text{if } p \in [\Delta(\eta_{y=1}), 1]. \end{cases} \tag{40}$$

From (40), it is clear that the sensors' decision

$$a = \begin{cases} 1 & \text{if } p \in [0, \Delta(\eta_{y=2})); \\ y & \text{if } p \in [\Delta(\eta_{y=2}), \Delta(\eta_{y=1})); \\ 2 & \text{if } p \in [\Delta(\eta_{y=1}), 1]. \end{cases} \tag{41}$$

Therefore,

$$\sum_{a \in \mathcal{A}} V(T^\pi(\pi, a))\sigma(\pi, a) = \begin{cases} V(\pi) & \text{if } p \in [0, \Delta(\eta_{y=2})); \\ \mathbb{E}V(\pi) & \text{if } p \in [\Delta(\eta_{y=2}), \Delta(\eta_{y=1})); \\ V(\pi) & \text{if } p \in [\Delta(\eta_{y=1}), 1]. \end{cases} \tag{42}$$

where $\mathbb{E}V(\pi) = \mathbf{1}' B^{\pi}_{y=1}\pi \times V(\eta_{y=1}) + \mathbf{1}' B^{\pi}_{y=2}\pi \times V(\eta_{y=2})$. □

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

Corollary 2. *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 Lemma 1, the instantaneous reward is a linear function in p and

$$\begin{aligned}\underset{p\in[0,\Delta(\eta_{y=2}))}{\operatorname{argmin}} \quad & Q(\pi,p)=0,\\ \underset{p\in[\Delta(\eta_{y=2}),\Delta(\eta_{y=1}))}{\operatorname{argmin}} \quad & Q(\pi,p)=\Delta(\eta_{y=2}),\\ \underset{p\in[\Delta(\eta_{y=1}),1]}{\operatorname{argmin}} \quad & Q(\pi,p)=\Delta(\eta_{y=1}).\end{aligned}$$

These hold as for any value of p in each of the three regions, the corresponding continuation payoff is the same from Lemma 1. □

Lemma 2. *The incentive function $\Delta(\eta_y)$ is decreasing in π for every y.*

Proof. The incentive function is given as (15), where $l_1, l_2, l_3 > 0$. With $\pi = [1-\pi(2),\ \pi(2)]'$, differentiating w.r.t $\pi(2)$,

$$\frac{d(\Delta(\eta_y))}{d\pi(2)} = -(l_1+l_2)B_{1y}B_{2y} < 0. \qquad \square$$

Lemma 3. *Under (A1), $\Delta(\eta_{y=1})$ is concave in π, and $\Delta(\eta_{y=2})$ is convex in π.*

Proof. The incentive function $\Delta(\eta_{y=2})$ is given in (15). A differentiable function $f:[0,1]\to[0,1]$ is convex if

$$f(w_1) \geq f(w_2) + f'(w_2)(w_1-w_2), \text{for all } w_1, w_2 \in [0,1]. \tag{43}$$

From (43) with $w_1 = \pi_1(2)$ and $w_2 = \pi_2(2)$, and using Proposition 1, it is verified that the function $\Delta(\eta_{y=2})$ is convex in π. Similarly, it can be shown that $\Delta(\eta_{y=1})$ is concave in π. □

APPENDIX C
DISCUSSION OF REWARD FUNCTIONS

A. Social Sensor's Reward Function

The nature of the results, specifically, the structural results (Theorem 1 and Theorem 3 in Sec.III); characterization of optimal incentive sequence (Theorem 2 in Sec.III); and the uniform bounds (Theorem 5 and Theorem 6 in Sec.IV); is unaffected by the choice of the form of reward functions below.

a.) (Resolution dependent reward): This form of reward function can be used to explicitly capture the effect or the influence of the observation distribution (resolution) matrix B of the social sensors on the actions. Let $r(x,y,a)$ denote the reward accrued if the sensor takes action a when the underlying state is x and the observation is y. The reward function is given as:

$$r(x,a) = \sum_y r(x,y,a)B_{xy}. \tag{44}$$

Here $r(x,y,a) = \delta_a p - \alpha_a \mathcal{I}(a\neq x) - \beta_a \mathcal{I}(a\neq y) - \gamma_a$, $\delta_a \in [0,1]$, $\alpha_a, \beta_a, \gamma_a \in \mathbb{R}$ are the given parameters of the model and $\mathcal{I}$ denotes the indicator function. For an action $a \in \mathcal{A}$ of the social sensor, $\delta_a p$ the effective incentive received (see discussion below) by the social sensor; γ_a denotes the cost of taking the action; α_a and β_a denote the mis-representation or distortion weights.

b.) (Resolution independent reward): This form of the reward function is not explicitly dependent on the resolution of the social sensors, i.e,

$$r(x,a) = \delta_a p - \alpha_a \mathcal{I}(a\neq x) - \gamma_a. \tag{45}$$

c.) (Realization dependent reward): This form of reward function explicitly depends on the private observation or realization y_k for the social sensor k, i.e,

$$r(x,y_k,a) = \delta_a p - \alpha_a \mathcal{I}(a\neq x) - \beta_a \mathcal{I}(a\neq y_k) - \gamma_a. \tag{46}$$

d.) (General state-action reward): This form of reward function models a general state-action reward function, i.e,

$$r(x,a) = \delta_a p + \Gamma^y_{xa} \tag{47}$$

The parameter Γ^y_{xa} is any function of the resolution, realization, state, and action.

Motivation: The social sensor k receives a noisy observation y_k of the state x. The term $\beta_a I(a\neq y_k)$ models the distortion cost [21] induced by the sensor's realization in equation (46). For social sensor k, $I(a_k \neq y_k)$ is the binary distance function [21] of the distortion or mis-representation of the received information y_k as a_k. In case of (44), the term $\beta_a I(a\neq y)$ captures the inherent distortion that can result from the sensor's observation matrix B.

The information fusion center offers a single incentive $p_k \in [0,1]$ to the social sensor k by using the information from the actions of the previous social sensors contained in the public belief π_{k-1} (see (10)). The weight δ_a helps to model asymmetric incentives for the different actions of the social sensor, and determines the effective incentive received by the social sensor for choosing different actions. The asymmetry is required to derive a feedback (public belief dependent) policy for the information fusion center to choose the future price. Symmetry ($\delta_{a=2} = \delta_{a=1}$) results in open loop or static prices (as the dependency cancels out) for the information fusion center. Since we are interested in dynamically changing the incentives to incorporate learning from the previous social sensors, we choose $\delta_{a=2} \neq \delta_{a=1}$.

B. Information Fusion Cost

The cost function for the fusion center is motivated by the revenue maximization problem with social learning literature [19], [24], [1], [10]:

$$\sum_{k=0}^{\infty} \rho^k (p_k - c)\mathcal{I}(a_k = \text{buy}). \tag{48}$$

Here (48) is the objective function of a monopoly that dynamically charges a price p_k for a product that costs c to manufacture, to a social sensor k that learns about the

underlying value (state) of the product from the decisions of other social sensors. The monopoly's objective is to maximize the revenue collected. The price p_k is selected (using the optimal pricing policy) so as to influence or elicit the desired behavior (buy or not buy) from the social sensors.

A modification of (48) motivated by controlled information fusion applications in the presence of social learning is given by (8) and (11). Here p_k is the incentive offered by the fusion center and $\Phi_s(k) \in (0,1)$ is the weight attached to the usefulness of the information acquired from sensor k. The objective of the information fusion is to maximize the number of sensors that act according to their observations, and estimate the underlying state. Since the sensors take 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.

REFERENCES

[1] C. Chamley, *Rational Herds: Economic Models of Social Learning*. Cambridge University Press, 2004.

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

[3] J. 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.

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

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

[6] S. Bikhchandani, 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.

[7] I. Welch, "Sequential sales, learning, and cascades," *Journal of Finance*, vol. 47, no. 2, pp. 695–732, June, 1992.

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

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

[10] S. Bose, G. Orosel, M. Ottaviani, and L. Vesterlund, "Monopoly pricing in the binary herding model," *Economic Theory*, vol. 37, no. 2, pp. 203–241, 2008.

[11] G. M. Lipsa and N. C. Martins, "Remote state estimation with communication costs for first-order LTI systems," *IEEE Transactions on Automatic Control*, vol. 56, no. 9, pp. 2013–2025, 2011.

[12] J. Arabneydi and A. G. Aghdam, "Data Collection versus Data Estimation: A Fundamental Trade-off in Dynamic Networks," *IEEE Transactions on Network Science and Engineering*, 2020.

[13] P. K. Varshney, *Distributed Detection and Data Fusion*. Springer Science & Business Media, 2012.

[14] A. Orléan, "Bayesian interactions and collective dynamics of opinion: Herd behavior and mimetic contagion," *Journal of Economic Behavior & Organization*, vol. 28, no. 2, pp. 257–274, 1995.

[15] D. Gale and S. Kariv, "Bayesian learning in social networks," *Games and Economic Behavior*, vol. 45, no. 2, pp. 329–346, 2003.

[16] D. Acemoglu, M. A. Dahleh, I. Lobel, and A. Ozdaglar, "Bayesian learning in social networks," *The Review of Economic Studies*, vol. 78, no. 4, pp. 1201–1236, 2011.

[17] H. Salami, B. Ying, and A. H. Sayed, "Social learning over weakly connected graphs," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 3, no. 2, pp. 222–238, 2017.

[18] S. Nitinawarat and V. V. Veeravalli, "Controlled sensing for sequential multihypothesis testing with controlled markovian observations and nonuniform control cost," *Sequential Analysis*, vol. 34, no. 1, pp. 1–24, 2015.

[19] S. Bose, G. Orosel, M. Ottaviani, and L. Vesterlund, "Dynamic monopoly pricing and herding," *The RAND Journal of Economics*, vol. 37, no. 4, pp. 910–928, 2006.

[20] D. Acemoglu, K. Bimpikis, and A. Ozdaglar, "Dynamics of information exchange in endogenous social networks," *Theoretical Economics*, vol. 9, no. 1, pp. 41–97, 2014.

[21] S. Obraztsova, O. Lev, E. Markakis, Z. Rabinovich, and J. S. Rosenschein, "Distant truth: Bias under vote distortion costs," in *Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems*, pp. 885–892, International Foundation for Autonomous Agents and Multiagent Systems, 2017.

[22] V. Krishnamurthy and S. Bhatt, "Sequential detection of market shocks with risk-averse CVaR social sensors," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 6, pp. 1061–1072, 2016.

[23] N. Miller, P. Resnick, and R. Zeckhauser, "Eliciting informative feedback: The peer-prediction method," *Management Science*, vol. 51, no. 9, pp. 1359–1373, 2005.

[24] M. Ottaviani, *Social learning in markets*. PhD thesis, Massachusetts Institute of Technology, 1996.

[25] D. Acemoglu, A. Makhdoumi, A. Malekian, and A. Ozdaglar, "Fast and slow learning from reviews," tech. rep., National Bureau of Economic Research, 2017.

[26] R. Durrett, *Probability: Theory and Examples*. Cambridge University Press, 2010.

[27] 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.

[28] P. Diaconis and D. Freedman, "On the consistency of Bayes estimates," *The Annals of Statistics*, pp. 1–26, 1986.

[29] A. W. Van der Vaart, *Asymptotic Statistics*. Cambridge University Press, 1998.

[30] S. Ross, M. Izadi, M. Mercer, and D. Buckeridge, "Sensitivity analysis of POMDP value functions," in *International Conference on Machine Learning and Applications, 2009. ICMLA'09.*, pp. 317–323, IEEE, 2009.

Sujay Bhatt received the Ph.D. degree in Electrical & Computer Engineering from Cornell University in 2019. He previously received the M. Tech degree in Electrical Engineering from Indian Institute of Technology, Bombay in 2014. His research interests include stochastic control, reinforcement learning, multi-armed bandits, statistical inference, and game theory.

Vikram Krishnamurthy (F'05) received the Ph.D. degree from the Australian National University in 1992. He is currently a professor in the School of Electrical & Computer Engineering, Cornell University. From 2002-2016 he was a Professor and Canada Research Chair at the University of British Columbia, Canada. His research interests include statistical signal processing and stochastic control in social networks and adaptive sensing. He served as Distinguished Lecturer for the IEEE Signal Processing Society and Editor-in-Chief of the IEEE Journal on Selected Topics in Signal Processing. In 2013, he was awarded an Honorary Doctorate from KTH (Royal Institute of Technology), Sweden. He is author of the books *Partially Observed Markov Decision Processes* and *Dynamics of Engineered Artificial Membranes and Biosensors* published by Cambridge University Press in 2016 and 2018, respectively.