

# Robust Decentralized Detection and Social Learning in Tandem Networks

Jack Ho, *Student Member, IEEE*, Wee Peng Tay, *Senior Member, IEEE*, Tony Q. S. Quek, *Senior Member, IEEE*, and Edwin K. P. Chong, *Fellow, IEEE*

**Abstract**—We study a tandem of agents who make decisions about an underlying binary hypothesis, where the distribution of the agent observations under each hypothesis comes from an uncertainty class defined by a 2-alternating capacity. We investigate both decentralized detection rules, where agents collaborate to minimize the error probability of the final agent, and social learning rules, where each agent minimizes its own local minimax error probability. We then extend our results to the infinite tandem network, and derive necessary and sufficient conditions on the uncertainty classes for the minimax error probability to converge to zero when agents know their positions in the tandem. On the other hand, when agents do not know their positions in the network, we study the cases where agents collaborate to minimize the asymptotic minimax error probability, and where agents seek to minimize their worst-case minimax error probability (over all possible positions in the tandem). We show that asymptotic learning of the true hypothesis is no longer possible in these cases, and derive characterizations for the minimax error performance.

**Index Terms**—Social learning, decentralized detection, tandem networks, robust hypothesis testing

## I. INTRODUCTION

In this paper, we formulate and study the robust social learning problem in a tandem network. A tandem network consists of agents connected in a serial fashion, where each agent receives an opinion about a binary hypothesis from a previous agent, and makes a decision about a binary hypothesis based on the previous agent's opinion and its own observation. Despite the simple structure of the tandem network, studying it can lead to insights about more complicated network structures such as those in social networks or Internet of Things (IoT) networks. The tandem network approximates a single information flow in a network, and it and its variants have been widely studied in [1]–[10].

Copyright (c) 2015 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

Part of this paper was presented at the IEEE International Conference on Acoustics, Speech and Signal Processing, May 2013. This research is supported in part by the Singapore Ministry of Education Academic Research Fund Tier 2 grants MOE2013-T2-2-006 and MOE2014-T2-1-028. This research is also supported by in part by the A\*STAR SERC under Grant 1224104048, and the MOE ARF Tier 2 grant MOE2014-T2-2-002. J. Ho and W.P. Tay are with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore. E-mail: JHO008@e.ntu.edu.sg, wptay@ntu.edu.sg. T.Q.S. Quek is with the Information Systems Technology and Design Pillar, Singapore University of Technology and Design, Singapore. E-mail: tonyquek@sutd.edu.sg. Edwin K.P. Chong is with the Department of Electrical and Computer Engineering, Colorado State University, USA. E-mail: edwin.chong@colostate.edu.

In our model, each agent's decision is based on a local error criterion, which it selfishly tries to optimize. This behavior is present in social networks, where users are mainly concerned with spreading only locally accurate information. In this paper, we call this *social learning* [9]–[17]. One such application of social learning is in the case of participatory sensing, where inference about a phenomenon of interest is made through the help of agents in the network [11], [18], [19]. For example, this can occur when users send a picture of litter in a park to a social sensing platform [20], [21] or report congested road conditions. [22].

On the other hand, if the agents' decision rules are designed to minimize the error criterion of the last agent in the network, or the asymptotic error probability in the case of an infinite network, this is known as *decentralized detection* [8], [23], [24]. One major application of decentralized detection is in sensor networks with a fusion center [25]–[27]. If the fusion center is able to relay information to the other agents, it will be able to select a set of globally optimal decision rules for every agent. However, many practical networks, such as social networks, do not have a fusion center. Furthermore, even for networks with fusion centers, the fusion center may not be able to easily communicate with the other agents. This is true of the participatory sensing examples above.

In the above examples, it is assumed that each agent knows the distribution of its private observation, and that of its predecessor, as well as its position in the network. However, in a real-life network, this is generally not the case. In this paper, we investigate what happens when one or both of these assumptions do not hold.

### A. Related Work

Binary hypothesis testing in a tandem network model is studied in [3], [5], which shows that learning the true hypothesis asymptotically is possible with unbounded likelihood ratios, and not possible with bounded likelihood ratios when agents transmit only 1-bit messages. Decentralized detection policies for tandem networks are also considered in [6], and conditions for the error probability approaching zero as the number of agents grows are derived. This is a network where each agent after the first receives exactly one decision from its predecessor. The authors also study a sub-optimal scheme where each sensor “selfishly” tries to minimize its own error, as opposed to the error of the root agent. The reference [8] shows that the rate of error decay is at most sub-exponential.

Feedforward networks, in which an agent obtains information from a subset of previous agents not necessarily just

the immediate predecessor, have been studied in [9], [10], [17]. In [10], agents are able to access the decisions of their  $K$  most recent predecessors. It is demonstrated that almost sure learning is impossible for any value of  $K$ , and learning in probability is possible for  $K \geq 2$ . A new model where forward looking agents try to maximize the discounted sum of the probability of a right decision is also considered. The reference [9] studies the decentralized detection problem in a game theoretic setting, and examines the effect of obtaining information from different sets of previous agents on the rate of error decay. The reference [28] examines the asymptotic error rate of feedforward topologies under two types of broadcast errors, namely erasure and random flipping.

All the above works assume that agent's observations are drawn from *known* distributions under each hypothesis. This assumption may not hold in practical networks like IoT networks, in which sensors' observation distributions may change over time, or in social networks, in which agents' observations may be affected by the agents' mood at a particular time. The robust detection framework was first proposed by [29] for a single agent to model the case where the observation distributions are not known exactly. A survey of results in this area can be found in [30]. The underlying probability distributions governing the agent observations are assumed to belong to different uncertainty classes under different hypotheses, and it is shown that under a minimax error criterion, the optimal decision rule for the agent is a likelihood ratio test based on the pair of least favorable distributions (LFDs). Subsequently, the work [31] investigates robust detection in a finite parallel configuration, with and without a fusion center. The problem of robust sequential detection is studied in [32]. Robust social learning however has not been addressed in these works. In addition, robust detection and learning have not been studied for the tandem network.

## B. Our Contributions

In this paper, we consider robust binary hypothesis detection and social learning in a tandem network in which the observation models of agents under each hypothesis are uncertainty classes of probability distributions defined by 2-alternating capacities [33]. Our main goals are to obtain the optimal agent policies under a minimax error criterion, under both decentralized detection and social learning frameworks, and characterize the asymptotic minimax error probabilities under various scenarios. The 2-alternating capacities are used because they allow us to define general uncertainty classes, which include the  $\epsilon$ -contamination uncertainty class [29]. Our main contributions are the following:

- 1) For the tandem network, we show that the solutions to the robust decentralized detection and social learning problems are equivalent to the respective solutions of the corresponding classical hypothesis testing problem where all the private observations are distributed according to the LFDs (Theorem 2). Our proof can be extended to general tree topologies, which generalizes a result in [31], where a parallel topology is considered.

- 2) We show that when the uncertainty classes for all agent observations are the same, and agents know their positions in the tandem, asymptotically learning the true hypothesis under both decentralized detection and social learning frameworks is not possible unless the log-likelihood distributions between the LFDs is almost surely unbounded (Theorem 4 and Theorem 5).
- 3) When agents know their positions in the network, we show that asymptotically learning the true hypothesis under social learning is achievable only if there are two subsequences of agents such that the log-likelihood ratio between the LFDs is almost surely unbounded from above in one of the subsequences and almost surely unbounded from below in the other. If in addition, the uncertainty classes converge to singletons, then asymptotic learning of the true hypothesis is guaranteed. (Theorem 6).
- 4) When agents do not know their positions in the tandem, we show that it is not possible to asymptotically learn the true hypothesis. We investigate the cases where agents collaborate to minimize the asymptotic minimax error probability, and where agents seek to minimize their worst-case minimax error probability (over all possible positions in the tandem), and characterize the minimax error performance in these approaches (Theorems 7 and 8).

In this paper, we consider only robust decentralized detection and social learning in tandem networks, which are very simple in structure. Social networks and IoT networks are much more complex in practice. Therefore, our results are limited, and can only be applied *heuristically* to more practical networks. Our analysis forms the foundation for studying more complex networks like trees and general loopy graphs, and provides insights into designing optimal decision rules for such networks. For example, although using likelihood ratio tests based on LFDs at each agent is not known to be optimal for loopy graphs, we expect this to produce reasonable results in practice. Addressing the performance of social learning in more complex networks is part of future research.

The rest of this paper is organized as follows. In Section II, we introduce the robust decentralized detection and social learning problem in a tandem network. In Section III, we provide a characterization for the agents' optimal decision rules in both decentralized detection and social learning in a tandem network. We then derive necessary and sufficient conditions for asymptotically learning the true hypothesis under various simplifications in Section IV. We also study the case where agents do not know their positions in the tandem in this section. In Section V, we illustrate some of our results using a numerical example. Lastly, we conclude in Section VI.

## II. PROBLEM FORMULATION

We consider a tandem network consisting of  $N$  agents, with agent 1 being the first agent and  $N$  being the last (see Figure 1). Consider a binary hypothesis testing problem in which the true hypothesis  $H$  is  $H_i$  with prior probability  $\pi_i \in (0, 1)$ , for  $i = 0, 1$ . For  $k = 1, \dots, N$ , and conditioned on  $H = H_i$ , each agent  $k$  in the network makes an observation  $Y_k$ , with

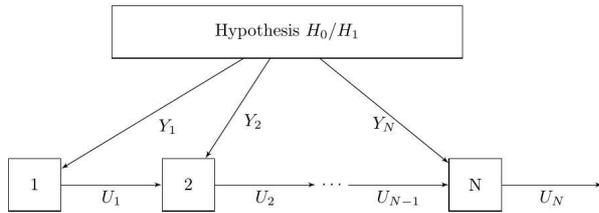


Fig. 1. Hypothesis testing in a tandem network.

distribution  $P_{i,k}$ . In various applications like inference in a social network, the distribution  $P_{i,k}$  may be unknown a priori. We therefore assume that  $P_{i,k}$  belongs to an uncertainty class  $\mathcal{P}_{i,k}$  of distributions, and is unknown to the agents. We assume that  $\mathcal{P}_{0,k}$  and  $\mathcal{P}_{1,k}$  are disjoint and that all distributions in  $\mathcal{P}_{0,k}$  and  $\mathcal{P}_{1,k}$  are absolutely continuous with respect to one another. Furthermore, we assume that conditioned on the true hypothesis, the observations of each agent are independent from one another. In the following, we first describe the decision making process of the agents under the decentralized detection and social learning formulations, and then the uncertainty class models we adopt in this paper.

#### A. Minimax Decentralized Detection and Social Learning

Each agent  $k$  makes a decision  $U_k = \phi_k(Y_k, U_{k-1}) \in \{0, 1\}$  about the hypothesis  $H$ , where  $\phi_k$  is an agent decision rule whose decision  $i$  corresponds to deciding in favor of  $H_i$ , and  $U_0 \equiv 0$ . For  $i = 0, 1$ , let  $P_i^{(N)} = P_{i,1} \times P_{i,2} \times \dots \times P_{i,N}$ . Similarly, define  $\mathcal{P}_i^{(N)} = \mathcal{P}_{i,1} \times \mathcal{P}_{i,2} \times \dots \times \mathcal{P}_{i,N}$ . In the decentralized detection problem, our aim is to find a sequence of decision rules  $\phi^{(N)} = (\phi_1, \phi_2, \dots, \phi_N)$  to minimize the maximum probability of error given by

$$P_N^{\text{DD}}(\phi^{(N)}) = \sup_{(P_0^{(N)}, P_1^{(N)}) \in \mathcal{P}_0^{(N)} \times \mathcal{P}_1^{(N)}} P_{e,N}(\phi^{(N)}, P_0^{(N)}, P_1^{(N)}), \quad (1)$$

where

$$P_{e,N}(\phi^{(N)}, P_0^{(N)}, P_1^{(N)}) = \pi_0 P_{F,N}(\phi^{(N)}, P_0^{(N)}) + \pi_1 P_{M,N}(\phi^{(N)}, P_1^{(N)}). \quad (2)$$

In (2),  $P_{F,N}$  and  $P_{M,N}$  are the false alarm and missed detection probabilities of agent  $N$  respectively, given the decision rules  $\phi^{(N)}$  and the agents' observation distributions  $P_i^{(N)}$ ,  $i = 0, 1$ . When there is only one distribution in each of the uncertainty classes, we recover the classical Bayesian hypothesis testing problem.

In the social learning problem, the first agent chooses its decision rule  $\phi_1$  to minimize

$$\sup_{(P_{0,1}, P_{1,1}) \in \mathcal{P}_{0,1} \times \mathcal{P}_{1,1}} P_{e,1}(\phi_1, P_{0,1}, P_{1,1}).$$

Each other agent  $k$ , given the decision rules  $\phi^{(k-1)}$  of the previous agents  $1, \dots, k-1$ , is able to derive the false alarm

and miss detection probabilities of agent  $k-1$ . It then seeks to find  $\phi_k$  to minimize

$$P_k^{\text{SL}}(\phi_k | \phi^{(k-1)}) = \sup_{(P_0^{(k)}, P_1^{(k)}) \in \mathcal{P}_0^{(k)} \times \mathcal{P}_1^{(k)}} P_{e,k}(\phi^{(k)}, P_0^{(k)}, P_1^{(k)}). \quad (3)$$

In contrast to the decentralized detection problem in (1), each agent *myopically* seeks to minimize its local maximum probability of error.

#### B. Uncertainty Classes

To model the uncertainty classes  $\mathcal{P}_{i,k}$ , for  $i = 0, 1$ , let  $(\mathcal{Y}, \mathcal{A})$  be a common measurable space for the observation of each agent  $k$  in the network, where  $\mathcal{Y}$  is a Polish metric space. We use the concept of 2-alternating capacities introduced by [33] to characterize the uncertainty classes for which LFDs exist for minimax tests between two hypotheses. Due to the technicalities involved and space constraint, we refer the reader to [33] for the mathematical definition of a 2-alternating capacity. For this paper, it suffices to note that several natural uncertainty classes like the  $\epsilon$ -contamination uncertainty class, as shown below, are special cases of the uncertainty classes defined by 2-alternating capacities.

For each hypothesis  $H_i$  and agent  $k$ , let  $v_{i,k}$  be a 2-alternating capacity. Define the uncertainty class associated with  $H_i$  at agent  $k$  to be

$$\mathcal{P}_{i,k} = \{Q \in \mathcal{R} \mid Q(A) \leq v_{i,k}(A), \forall A \in \mathcal{A}\}, \quad (4)$$

where  $\mathcal{R}$  is the set of all probability measures on  $(\mathcal{Y}, \mathcal{A})$ . An example of an uncertainty class is the  $\epsilon$ -contamination uncertainty class [33]. For each hypothesis  $H_i$ , let  $P_i^* \in \mathcal{R}$  be a nominal probability distribution. A natural uncertainty class is a set of distributions that are within a certain "distance" (in some sense) of the nominal distribution  $P_i^*$ . Let  $v_{i,k} = (1 - \epsilon_{i,k})P_i^* + \epsilon_{i,k}R$ , where  $\epsilon_{i,k} \in [0, 1]$  is called the contamination value. It is easy to see that (4) reduces to

$$\mathcal{P}_{i,k} = \{Q : Q = (1 - \epsilon_{i,k})P_i^* + \epsilon_{i,k}R, R \in \mathcal{R}\}, \quad (5)$$

which we call an  $\epsilon$ -contamination uncertainty class. Another example of an uncertainty class of distributions within a "distance" to a given nominal distribution  $P_i^*$  is the total variation uncertainty class [33]. In this case, we let  $v_{i,k} = \min\{P_i^* + \epsilon_{i,k}, 1\}$ , and

$$\mathcal{P}_{i,k} = \left\{ Q : \sup_{A \in \mathcal{A}} |Q(A) - P_i^*(A)| \leq \epsilon_{i,k} \right\},$$

is the class of distributions within total variation distance  $\epsilon_{i,k}$  of  $P_i^*$ . Our subsequent results hold for all uncertainty classes defined by 2-alternating capacities under various technical conditions. However, for illustration purposes, we will use  $\epsilon$ -contamination uncertainty classes throughout this paper as examples to interpret our theorems.

For each agent  $k$ , let  $p_{i,k}$  be the density (with respect to some common measure) of  $P_{i,k}$ . The least favorable distributions (LFDs) for two given uncertainty classes  $\mathcal{P}_{0,k}$  and  $\mathcal{P}_{1,k}$

is defined in [33] to be the pair of distributions  $(Q_{0,k}, Q_{1,k})$  with densities  $(q_{0,k}, q_{1,k})$  such that

$$P_{0,k}(l_q(Y_k) > t) \leq Q_{0,k}(l^*(Y_k) > t)$$

and

$$Q_{1,k}(l_q(Y_k) > t) \leq P_{1,k}(l^*(Y_k) > t),$$

for all  $t \geq 0$ , where  $l^*(Y_k)$  is the likelihood ratio  $q_{1,k}(Y_k)/q_{0,k}(Y_k)$ .

For the example of the  $\epsilon$ -contamination uncertainty classes in (5), we have

$$q_{0,k}(y) = \begin{cases} (1 - \epsilon_{0,k})p_0^*(y) & \text{for } p_1^*(y)/p_0^*(y) < c' \\ (1 - \epsilon_{0,k})p_1^*(y)/c'' & \text{for } p_1^*(y)/p_0^*(y) \geq c' \end{cases}$$

$$q_{1,k}(y) = \begin{cases} (1 - \epsilon_{1,k})p_1^*(y) & \text{for } p_1^*(y)/p_0^*(y) > c' \\ c'(1 - \epsilon_{1,k})p_0^*(y) & \text{for } p_1^*(y)/p_0^*(y) \leq c', \end{cases}$$

where  $0 \leq c' < c'' \leq \infty$  are determined such that  $q_{0,k}$  and  $q_{1,k}$  are probability densities. Note that  $c' = 0$  if and only if  $\epsilon_{1,k} = 0$ , and  $c'' = \infty$  if and only if  $\epsilon_{0,k} = 0$ . Let  $b_k = (1 - \epsilon_{1,k})/(1 - \epsilon_{0,k})$ . We then have

$$\frac{q_{1,k}(y)}{q_{0,k}(y)} = \begin{cases} b_k c' & \text{for } p_1^*(y)/p_0^*(y) \leq c' \\ b_k \cdot \frac{p_1^*(y)}{p_0^*(y)} & \text{for } c' < \frac{p_1^*(y)}{p_0^*(y)} < c'' \\ b_k c'' & \text{for } p_1^*(y)/p_0^*(y) \geq c''. \end{cases} \quad (6)$$

In many of our theorems, we require the presence of LFDs between the uncertainty classes. The following theorem which proves the existence of LFDs for uncertainty classes derived from 2-alternating capacities was shown in [33].

**Theorem 1.** *Let  $v_{0,k}$  and  $v_{1,k}$  be 2-alternating capacities. Then for  $i = 0, 1$ , there exist probability measures  $Q_{i,k} \in \mathcal{P}_{i,k}$  such that for all  $t > 0$ ,*

$$Q_{0,k}(l^*(Y_k) \geq t) = v_{0,k}(l^*(Y_k) \geq t)$$

$$Q_{1,k}(l^*(Y_k) < t) = v_{1,k}(l^*(Y_k) < t).$$

In the rest of this paper, for any random variable  $Y$  with distributions drawn from a given pair of uncertainty classes, we let  $l^*(Y)$  be the likelihood ratio  $q_1(Y)/q_0(Y)$ , where  $q_0$  and  $q_1$  are the respective densities of the LFDs of the given uncertainty classes. In addition, we use  $l^*(Y = y)$  to denote the realization of  $l^*(Y)$  when  $Y = y$ . A sequence  $x_1, x_2, \dots, x_n$  is denoted as  $(x_i)_{i=1}^n$ .

### III. ROBUST LEARNING IN A TANDEM NETWORK

When there is only a single agent, the minimax error  $\inf_{\phi} P_1^{\text{SL}}(\phi)$  is achieved by setting  $\phi$  to be a likelihood ratio test using the LFDs  $(Q_{0,1}, Q_{1,1})$  [33]. A similar result is proven in [31] for a parallel network configuration. In this section, we show the same result for the tandem network. We do this by first showing that for any sequence of agent decision rules that consists of likelihood ratio tests between LFDs, the error probabilities (1) and (3) are maximized when all the private agent observations are drawn from the corresponding LFDs. We then show that when agents' private observations are drawn from the respective LFDs, then likelihood ratio tests using LFDs minimize the error probabilities.

We can now present our first result. For all  $k \geq 1$ , let  $(Q_{0,k}, Q_{1,k})$  be the LFDs for  $(\mathcal{P}_{0,k}, \mathcal{P}_{1,k})$ , and  $Q_i^{(N)} = Q_{i,1} \times Q_{i,2} \times \dots \times Q_{i,N}$  for  $i = 0, 1$ .

**Theorem 2.** *Let  $\phi^{(N)}$  be any sequence of monotone likelihood ratio tests based on the LFDs  $(Q_0^{(N)}, Q_1^{(N)})$  for a tandem network. Then for all  $(P_0^{(N)}, P_1^{(N)}) \in \mathcal{P}_0^{(N)} \times \mathcal{P}_1^{(N)}$ , we have*

$$P_{F,N}(\phi^{(N)}, Q_0^{(N)}) \geq P_{F,N}(\phi^{(N)}, P_0^{(N)}), \quad (7)$$

and

$$P_{M,N}(\phi^{(N)}, Q_1^{(N)}) \geq P_{M,N}(\phi^{(N)}, P_1^{(N)}). \quad (8)$$

*Proof:* We proceed by mathematical induction on  $N$ . From the definition of the LFDs, the theorem holds for  $N = 1$ . We now assume that it holds for  $N < i$ . We also make use of the following two lemmas, the first of which is proved in Appendix A-A, while the second is shown in [31].

**Lemma 1.** *For any  $N$ ,  $Q_1^{(N)}(U_N = 1) \geq Q_0^{(N)}(U_N = 1)$  and  $Q_1^{(N)}(U_N = 0) \leq Q_0^{(N)}(U_N = 0)$ .*

**Lemma 2.** *Let  $Z_1$  and  $Z_2$  be non-negative, independent random variables. If for  $k = 1, 2$ , we have*

$$F(Z_k > t) \geq G(Z_k > t), \quad \forall t \geq 0, \quad (9)$$

then

$$F(Z_1 Z_2 > t) \geq G(Z_1 Z_2 > t), \quad \forall t \geq 0.$$

If (9) holds, we say that  $Z_k$  is stochastically larger under  $F$  than  $G$ . Since the observation of agent  $k$  is independent from the decision it receives, for  $i = 0, 1$  we have

$$Q_i^{(k)}(l^*(Y_k) > t) = Q_{i,k}(l^*(Y_k) > t).$$

From the definition of the LFDs,  $l^*(Y_k)$  is stochastically larger under  $Q_0^{(k)}$  than under any other distribution  $P_0^{(k)} \in \mathcal{P}_0^k$ . To show the same for  $l^*(U_{k-1})$ , we obtain from Lemma 1 that

$$l^*(U_{k-1} = 1) = \frac{Q_1^{(k-1)}(U_{k-1} = 1)}{Q_0^{(k-1)}(U_{k-1} = 1)} \geq 1 \geq \frac{1 - Q_1^{(k-1)}(U_{k-1} = 1)}{1 - Q_0^{(k-1)}(U_{k-1} = 1)} \geq l^*(U_{k-1} = 0).$$

For any  $l^*(U_{k-1} = 0) < t < l^*(U_{k-1} = 1)$ ,

$$Q_0^{(k)}(l^*(U_{k-1}) > t) = Q_0^{(k)}(U_{k-1} = 1) \geq P_0^{(k)}(U_{k-1} = 1) = P_0^{(k)}(l^*(U_{k-1}) > t),$$

where the inequality follows from the induction hypothesis. Note that  $l^*(U_{k-1})$  only takes the two values  $l^*(U_{k-1} = 0)$  and  $l^*(U_{k-1} = 1)$ , so for any  $t$  not in between these values, the equality  $Q_0^{(k)}(l^*(U_{k-1}) > t) = P_0^{(k)}(l^*(U_{k-1}) > t)$  is trivially true. Therefore,  $l^*(U_{k-1})$  is stochastically larger under  $Q_0^{(k)}$ .

From Lemma 2, the product of  $l^*(Y_k)$  and  $l^*(U_{k-1})$  is stochastically larger under  $Q_0^{(k)}$  than under any other distribution  $P_0^{(k)}$  as well. Therefore, we have

$$\begin{aligned} Q_0^{(k)}(U_k = 1) &= Q_0^{(k)}(l^*(U_{k-1}, Y_k) > t_k) \\ &\geq P_0^{(k)}(l^*(U_{k-1}, Y_k) > t_k) \\ &= P_0(U_k = 1). \end{aligned}$$

The proof for the missed detection probability inequality (8) is similar, and the induction is complete. The theorem is now proved.  $\blacksquare$

**Theorem 3.** Let  $\phi_*^{(N)}$  be an optimal sequence of decision rules such that

$$\phi_*^{(N)} = \arg \min_{\phi^{(N)}} P_{e,N}(\phi^{(N)}, Q_0^{(N)}, Q_1^{(N)}).$$

Then,  $\phi_*^{(N)}$  minimizes  $P_N^{DD}(\cdot)$  in (1). Similarly, for each  $k \geq 1$ , define  $\psi_{*,k}$  recursively as

$$\psi_{*,k} = \arg \min_{\psi_k} P_{e,k}((\psi_*^{(k-1)}, \psi_k), Q_0^{(k)}, Q_1^{(k)}),$$

where  $\psi_{*,0}$  is ignored. Then,  $\psi_{*,k}$  minimizes  $P_k^{SL}(\cdot | \psi_*^{(k-1)})$  in (3) for all  $k \geq 1$ .

*Proof:* In [23], it was shown that  $\phi_*^{(N)}$  is a sequence of likelihood ratio tests based on  $(Q_0^{(N)}, Q_1^{(N)})$ . Hence, for any sequence of decision rules  $\phi^{(N)}$ , we have from Theorem 2,

$$\begin{aligned} &\sup_{(P_0^{(N)}, P_1^{(N)}) \in \mathcal{P}_0^{(N)} \times \mathcal{P}_1^{(N)}} P_{e,N}(\phi_*^{(N)}, P_0^{(N)}, P_1^{(N)}) \\ &= P_{e,N}(\phi_*^{(N)}, Q_0^{(N)}, Q_1^{(N)}) \\ &\leq P_{e,N}(\phi^{(N)}, Q_0^{(N)}, Q_1^{(N)}) \\ &\leq \sup_{(P_0^{(N)}, P_1^{(N)}) \in \mathcal{P}_0^{(N)} \times \mathcal{P}_1^{(N)}} P_{e,N}(\phi^{(N)}, P_0^{(N)}, P_1^{(N)}) \end{aligned} \quad (10)$$

where the inequality in (10) follows from the theorem assumption. Therefore, the minimax error in the decentralized detection problem is equal to the minimum error when all the distributions of the private observations are equal to the LFDs. A similar argument holds for the social learning problem in the second part of this theorem. The proof is now complete.  $\blacksquare$

We remark that Theorem 3 can be extended to general tree topologies where the last agent is at the root of the tree, and agent decisions are passed towards the root. This is because in such a topology, the decisions received by each agent are mutually independent. Hence,

$$l^*(Y_k, U_{k_1}, \dots, U_{k_m}) = l^*(Y_k) \prod_{j=1}^m l^*(U_{k_j}),$$

where  $k_1, \dots, k_m$  are the agents that agent  $k$  is receiving decisions from. The rest of the proof then proceeds by showing that Theorem 2 holds for all agents in the tree topology by induction on the distance from the root agent. We can then use this result to prove a variant of Theorem 3 for general tree topologies.

#### IV. ASYMPTOTIC DETECTION AND SOCIAL LEARNING

Theorem 3 shows that there is no loss in optimality in both the decentralized detection and social learning problems if agents in a tandem network are restricted to monotone likelihood ratio tests based on the LFDs. It however does not tell us the smallest minimax error probability that is achievable at each agent. In this section, we study the minimax error probability in long tandems under various technical assumptions in order to simplify the problem. In particular, we investigate the conditions under which the minimax error probability converges to zero as the number of agents increases.<sup>1</sup> We first consider the case where every agent has identical uncertainty classes, and provide necessary and sufficient conditions for the minimax error probability to approach zero under both decentralized detection and social learning. We will then proceed to analyze social learning in long tandems where the uncertainty classes can differ. Finally, we study the achievable asymptotic minimax error probability when agents do not know their own positions in the tandem.

##### A. Identical Uncertainty Classes

In this subsection, we make the following assumption that the uncertainty classes of every agent are identical.

**Assumption 1.** For all  $k \geq 1$ ,  $\mathcal{P}_{0,k} = \mathcal{P}_0$ ,  $\mathcal{P}_{1,k} = \mathcal{P}_1$ , and the LFDs of each agent's uncertainty classes are  $(Q_0, Q_1)$ .

The following two results give necessary and sufficient conditions for the minimax error probability to approach zero under the decentralized detection and social learning frameworks, respectively.

**Theorem 4** (Decentralized detection). *Suppose that Assumption 1 holds, and that the decision rules for every agent are chosen so as to minimize  $P_N^{DD}(\phi^{(N)})$  in (1). Then  $P_N^{DD}(\phi^{(N)}) \rightarrow 0$  as  $N \rightarrow \infty$  if and only if the log-likelihood ratio of  $Q_0$  versus  $Q_1$  is almost surely unbounded from either above or below.*

*Proof:* See Appendix A-B.  $\blacksquare$

**Theorem 5** (Social learning). *Suppose that Assumption 1 holds, and that the decision rule for each agent is chosen sequentially so as to minimize  $P_k^{SL}(\phi_k | \phi^{(k-1)})$  in (3). Then  $P_k^{SL}(\phi_k | \phi^{(k-1)}) \rightarrow 0$  as  $k \rightarrow \infty$  if and only if the log-likelihood ratio of  $Q_0$  versus  $Q_1$  is almost surely unbounded from both above and below.*

*Proof:* It was demonstrated in Proposition 3 of [6] that using social learning decision rules, the error probability in a tandem network where all agents have the same observation distributions will converge to zero if and only if the log-likelihood ratios between the two probability distributions is unbounded from both above and below. The sufficiency of the given condition then follows immediately. To show that it is also necessary, observe that if the log-likelihood ratio of  $Q_1$  versus  $Q_0$  is bounded from either above or below respectively, and Theorem 3 and Proposition 3 of [6] implies

<sup>1</sup>In the case where the agents' uncertainty classes are singletons, [9] calls this *asymptotic learning*.

that  $P_N^{\text{DD}}(\phi^{(N)})$  is bounded away from zero. The proof is now complete. ■

For the  $\epsilon$ -contamination uncertainty classes defined in (5) with  $\epsilon_{i,k} = \epsilon_i$  for  $i = 0, 1$  and all  $k \geq 1$ , Theorem 4 implies that learning can only occur if at least one of the following is true:

- 1)  $\epsilon_0 = 0$  and the log-likelihood ratio of  $P_1^*$  versus  $P_0^*$  is almost surely unbounded from above,
- 2)  $\epsilon_1 = 0$  and the log-likelihood ratio of  $P_1^*$  versus  $P_0^*$  is almost surely unbounded from below.

Similarly, Theorem 5 implies that learning can only occur if  $\epsilon_0 = \epsilon_1 = 0$  and the log-likelihood ratio of  $P_1^*$  versus  $P_0^*$  is almost surely unbounded.

### B. Varying uncertainty classes for social learning

In this subsection, we relax the assumption of identical uncertainty classes for all agents in the previous subsection, and study the effect of varying uncertainty classes on the asymptotic error probability in the social learning framework. We make the following assumptions.

**Assumption 2.** *Each agent  $k \geq 1$  knows its own uncertainty class, as well as those of its predecessors, and its position in the tandem network.*

We will show that under these assumptions, learning the true hypothesis can occur only if there exist infinite subsequences  $(k_n)_{n \geq 1}$  and  $(j_n)_{n \geq 1}$  (which may be distinct) such that  $\log(l^*(Y_{k_n}))$  is almost surely unbounded from above and  $\log(l^*(Y_{j_n}))$  is almost surely unbounded from below. Furthermore, we will show that learning will occur if in addition, both uncertainty classes  $\mathcal{P}_{0,k_n}$  and  $\mathcal{P}_{1,j_n}$  converge to singletons.

We first observe that under the social learning framework, agents minimize their local maximum error probability. This implies that the minimax error probability of each agent  $k$  is non-increasing in  $k$ . This is because any agent can simply pass on the decision of the previous agent if no other decision rule leads to a decrease in minimax error probability.

For ease of notation, we let  $Q_{F,k} = P_{F,k}(\phi^{(k)}, Q_0^{(k)})$  and  $Q_{M,k} = P_{M,k}(\phi^{(k)}, Q_0^{(k)})$  be the LFD false alarm and miss detection probabilities of agent  $k$  respectively, where  $\phi^{(k)}$  is the sequence of optimal social learning decision rules that minimizes  $P_k^{\text{SL}}(\phi_k | \phi^{(k-1)})$  in (3). From [6], it can be shown that the decision rule used by agent  $k$  is of the form

$$u_k = \begin{cases} 0 & \text{if } l^*(Y_k) < \frac{\pi_0 Q_{F,k-1}}{\pi_1(1-Q_{M,k-1})} \\ 1 & \text{if } l^*(Y_k) \geq \frac{\pi_0(1-Q_{F,k-1})}{\pi_1 Q_{M,k-1}} \\ u_{k-1} & \text{if } \frac{\pi_0 Q_{F,k-1}}{\pi_1(1-Q_{M,k-1})} \leq l^*(Y_k) < \frac{\pi_0(1-Q_{F,k-1})}{\pi_1 Q_{M,k-1}} \end{cases} \quad (11)$$

We now have the following two theorems.

**Theorem 6.** *Suppose that Assumption 2 holds, and that each agent  $k$  in a tandem network chooses decision rule  $\phi_k$  to minimize  $P_k^{\text{SL}}(\phi_k | \phi^{(k-1)})$  in (3). Then  $P_k^{\text{SL}}(\phi_k | \phi^{(k-1)}) \rightarrow 0$  as  $k \rightarrow \infty$  only if there exist infinite subsequences  $(k_n)_{n \geq 1}$  and  $(j_n)_{n \geq 1}$  (which may be distinct) such that  $\log(l^*(Y_{k_n}))$*

*is almost surely unbounded from above and  $\log(l^*(Y_{j_n}))$  is almost surely unbounded from below. In addition, if  $\mathcal{P}_{0,k_n}$  and  $\mathcal{P}_{1,j_n}$  each converges to a singleton as  $n \rightarrow \infty$ , then  $P_k^{\text{SL}}(\phi_k | \phi^{(k-1)}) \rightarrow 0$  as  $k \rightarrow \infty$ .*

*Proof:* See Appendix A-C. ■

Theorem 6 provides separate necessary and sufficient conditions for asymptotic social learning to occur. For asymptotic learning to occur, one would expect that the observations of the agents become more and more informative as  $k$  increases. This is why the log-likelihood ratio of the LFDs must be almost surely unbounded in the two subsequences in Theorem 6. However, without any other assumptions on the uncertainty classes, it is possible to construct uncertainty classes for each agent such that the LFD log-likelihood ratio becomes unbounded with  $k$ , but the observation becomes less and less informative. If we add the assumption that the uncertainty classes converge to singletons, then the above construction is not possible, and asymptotic learning is guaranteed.

For the  $\epsilon$ -contamination uncertainty classes in (5), the above necessary and sufficient conditions are identical.

**Corollary 1.** *If the uncertainty classes in Theorem 6 are  $\epsilon$ -contamination uncertainty classes as defined in (5), then  $P_k^{\text{SL}}(\phi_k | \phi^{(k-1)}) \rightarrow 0$  as  $k \rightarrow \infty$  if and only if the the log-likelihood ratio of  $P_0^*$  and  $P_1^*$  is almost surely unbounded from both above and below, and there exist infinite subsequences  $(k_n)_{n \geq 1}$  and  $(j_n)_{n \geq 1}$  such that  $\epsilon_{0,k_n} \rightarrow 0$  and  $\epsilon_{1,j_n} \rightarrow 0$  as  $n \rightarrow \infty$ .*

*Proof:* From (6), a subsequence of agents with  $\log l^*(Y_{j_n})$  almost surely unbounded from below would require the log-likelihood ratio of  $P_1^*$  versus  $P_0^*$  being unbounded, as well as  $c' \rightarrow 0$ , which is equivalent to  $\epsilon_{1,j_n} \rightarrow 0$ . Similarly, we also need another subsequence where  $\epsilon_{0,k_n} \rightarrow 0$ . Since this implies that both the uncertainty classes converge to the distributions  $P_1^*$  and  $P_0^*$  respectively, this is also a sufficient condition. ■

We observe that agents in a tandem network in the social learning framework have total error probabilities at least that of agents adopting decentralized detection rules. Hence, Theorem 6 also provides a sufficient condition for the minimax error probability (1) under a decentralized detection framework to converge to zero.

### C. Unknown agent positions

In a social network, users have to make their decisions not knowing how many hops information has been propagated from a source node. We model this in a tandem network by assuming that each agent has no knowledge of its position in the network. We make the following assumption, in addition to Assumption 1, in this subsection.

**Assumption 3.** *Every agent  $k > 1$  uses the same decision rule.*

Except for agent 1 (which knows its position in the tandem because it does not receive any preceding messages), every other agent in the tandem does not know its own position, and has access to exactly the same information when it comes to choosing a decision rule, which is based solely on the

uncertainty class of the agent. Therefore, it is natural to make Assumption 3.

Because of Assumption 3, any sequence of  $k$  agent decision rules has the form  $\phi^{(k)} = (\phi_1, \phi^{k-1})$ , where  $\phi_1$  is the decision rule used by the first agent, and  $\phi$  is the decision rule used by every other agent with  $\phi^{k-1} = (\phi, \dots, \phi)$  consisting of  $k-1$  copies of  $\phi$ . For simplicity, and by abusing notation, we replace  $\phi^{(k)}$  in our notations by  $(\phi_1, \phi)$  in the sequel. In the following, we consider two different scenarios.

1) *Minimizing asymptotic error*: We consider the case where agents are collaborating to minimize the asymptotic error. This might occur when there is a chain of agents trying to relay some information to a fusion center, but each agent is unsure of how many other agents there are between itself and the fusion center. For a given decision rule  $\phi$ , the asymptotic maximum error probability is given by

$$P_\infty^{\text{DD}}(\phi) = \lim_{k \rightarrow \infty} P_k^{\text{DD}}(\phi_1, \phi), \quad (12)$$

where we have implicitly assumed that  $P_\infty^{\text{DD}}$  does not depend on  $\phi_1$ . This assumption is valid, as shown in the next theorem, which also provides a characterization for the optimal  $\phi$  that achieves the asymptotic minimax error probability, defined as

$$\inf_{\phi} P_\infty^{\text{DD}}(\phi).$$

In the previous sections, we had no need to consider randomized decision rules. This is because under the decentralized detection framework, the final error probability under a randomized sequence of decision rules is no less than the minimum final error probability of each of the respective deterministic sequences of decision rules. Similarly, under the social learning framework, the error probability of an agent using a randomized decision rule is no less than the minimum error probability under each of the deterministic rules. However, this property does not hold for the asymptotic maximum error probability. A randomized version of two deterministic decision rules may yield a lower asymptotic maximum error probability than either of the two deterministic rules. Thus, we introduce the randomized likelihood ratio test:

$$u_k = \begin{cases} 0 & \text{if } l^*(Y_k) < t_1 \\ A & \text{if } l^*(Y_k) = t_1 \\ u_{k-1} & \text{if } t_1 < l^*(Y_k) < t_0 \\ B & \text{if } l^*(Y_k) = t_0 \\ 1 & \text{if } l^*(Y_k) > t_0, \end{cases} \quad (13)$$

where  $A$  is equal to 0 with probability  $p$  and equal to  $u_{k-1}$  with probability  $1-p$ ,  $B$  is equal to  $u_{k-1}$  with probability  $q$  and equal to 1 with probability  $1-q$ , and  $t_1$ ,  $t_0$ ,  $p$  and  $q$  are constants to be determined.

We now show how to obtain the minimax asymptotic error probability. To do so, we start with the following two lemmas.

**Lemma 3.** *Suppose that Assumptions 1 and 3 hold. Then,*

$$\begin{aligned} & \lim_{k \rightarrow \infty} P_{e,k}((\phi_1, \phi), Q_0^k, Q_1^k) \\ &= \frac{\pi_0 Q_0(\phi(Y_1, 0) = 1)}{Q_0(\phi(Y_1, 0) = 1) + Q_0(\phi(Y_1, 1) = 0)} \\ &+ \frac{\pi_1 Q_1(\phi(Y_1, 1) = 0)}{Q_1(\phi(Y_1, 1) = 0) + Q_1(\phi(Y_1, 0) = 1)}. \end{aligned}$$

*Proof:* We have the two following recurrence relations:

$$\begin{aligned} & P_{F,k}((\phi_1, \phi), Q_0^k) \\ &= Q_0^k(U_k = 1) \\ &= Q_0^{k-1}(U_{k-1} = 0) \cdot Q_0(U_k = 1 | U_{k-1} = 0) \\ &\quad + Q_0^{k-1}(U_{k-1} = 1) \cdot Q_0(U_k = 1 | U_{k-1} = 1) \\ &= (1 - Q_0^{k-1}(U_{k-1} = 1)) \cdot Q_0(\phi(Y_k, 0) = 1) \\ &\quad + Q_0^{k-1}(U_{k-1} = 1) \cdot Q_0(\phi(Y_k, 1) = 1) \\ &= P_{F,k-1}((\phi_1, \phi), Q_0^{k-1}) \\ &\quad \cdot [Q_0(\phi(Y_1, 1) = 1) - Q_0(\phi(Y_1, 0) = 1)] \\ &\quad + Q_0(\phi(Y_1, 0) = 1), \end{aligned} \quad (14)$$

and

$$\begin{aligned} & P_{M,k}((\phi_1, \phi), Q_0^k) \\ &= Q_1^k(U_k = 0) \\ &= Q_1^{k-1}(U_{k-1} = 0) \cdot Q_1(U_k = 0 | U_{k-1} = 0) \\ &\quad + Q_1^{k-1}(U_{k-1} = 1) \cdot Q_1(U_k = 0 | U_{k-1} = 1) \\ &= Q_1^{k-1}(U_{k-1} = 0) \cdot Q_1(\phi(Y_k, 0) = 0) \\ &\quad + (1 - Q_1^{k-1}(U_{k-1} = 0)) \cdot Q_0(\phi(Y_k, 1) = 0) \\ &= P_{M,k-1}((\phi_1, \phi), Q_1^{k-1}) \\ &\quad \cdot [Q_1(\phi(Y_1, 1) = 1) - Q_1(\phi(Y_1, 0) = 1)] \\ &\quad + Q_1(\phi(Y_1, 0) = 0), \end{aligned} \quad (15)$$

The first recurrence relation converges linearly to

$$\frac{Q_0(\phi(Y_1, 0) = 1)}{Q_0(\phi(Y_1, 0) = 1) + Q_0(\phi(Y_1, 1) = 0)}, \quad (16)$$

and the second recurrence relation converges linearly to

$$\frac{\pi_1 Q_1(\phi(Y_1, 1) = 0)}{Q_1(\phi(Y_1, 1) = 0) + Q_1(\phi(Y_1, 0) = 1)}. \quad (17)$$

Hence, the proof is complete.  $\blacksquare$

Next, we show that if the observation distributions are all drawn from the LFDs, then the decision rule that minimizes that asymptotic error probability is a randomized likelihood ratio test between  $Q_0$  and  $Q_1$ .

**Lemma 4.** *Suppose that Assumptions 1 and 3 hold. For any  $\phi_1$ , let*

$$\phi_* = \arg \min_{\phi} \lim_{k \rightarrow \infty} P_{e,k}((\phi_1, \phi), Q_0^k, Q_1^k).$$

*Then, there is no loss in optimality if  $\phi_*$  is restricted to be a randomized likelihood ratio test between  $Q_0$  and  $Q_1$ .*

*Proof:* Assume that  $\phi_*$  is not a randomized likelihood ratio test between  $Q_0$  and  $Q_1$ . For any agent  $k \geq 2$ , consider a randomized likelihood ratio test  $\phi'$  in the form presented in (13), where  $t_0, t_1, p, q$  are chosen such that

$$\begin{aligned} Q_0(\phi'(Y_k, 0) = 1) &= Q_0(l^*(Y_k) \geq t_0) \\ &= Q_0(\phi_*(Y_k, 0) = 1) \\ Q_1(\phi'(Y_k, 1) = 0) &= Q_1(l^*(Y_k) < t_1) \\ &= Q_1(\phi_*(Y_k, 1) = 0). \end{aligned}$$

From the Neyman-Pearson lemma, we then have

$$\begin{aligned} Q_1(\phi'(Y_k, 0) = 0) &= Q_1(l^*(Y_k) < t_0) \\ &\leq Q_1(\phi_*(Y_k, 0) = 0) \end{aligned}$$

$$\begin{aligned} Q_0(\phi'(Y_k, 1) = 1) &= Q_0(l^*(Y_k) < t_1) \\ &\leq Q_0(\phi_*(Y_k, 1) = 1). \end{aligned}$$

Hence, from (14) and (15), it is clear that for any  $\phi_1$  and any  $k \geq 1$ , we have

$$P_{F,k}((\phi_1, \phi_*), Q_0^k, Q_1^k) \geq P_{F,k}((\phi_1, \phi'), Q_0^k, Q_1^k),$$

and

$$P_{M,k}((\phi_1, \phi_*), Q_0^k, Q_1^k) \geq P_{M,k}((\phi_1, \phi'), Q_0^k, Q_1^k),$$

and hence

$$P_{e,k}((\phi_1, \phi_*), Q_0^k, Q_1^k) \geq P_{e,k}((\phi_1, \phi'), Q_0^k, Q_1^k).$$

By choosing suitable values of  $t_0, t_1, p, q$ , we can set  $Q_0(\phi(Y_k, 0) = 1)$  and  $Q_0(\phi(Y_k, 1) = 1)$  to any value between 0 and 1. Similarly, we can also set  $Q_1(\phi(Y_k, 1) = 0)$  and  $Q_1(\phi(Y_k, 0) = 0)$  to any value between 0 and 1. Hence,

$$\min_{\phi} \lim_{k \rightarrow \infty} P_{e,k}((\phi_1, \phi), Q_0^k, Q_1^k)$$

is attainable, and the decision rule used to attain it can be assumed to be in the form of a randomized likelihood ratio test.  $\blacksquare$

The proof of Lemma 4 shows that the asymptotic minimax error probability is attainable even when the likelihood ratio of  $Y_k$  is not continuous. To avoid cumbersome notation, for the rest of the paper, we assume that the likelihood ratio of  $Y_k$  is continuous. It is easy to extend our results if this is not the case.

We can now prove the following theorem, which provides an expression for the minimax asymptotic error.

**Theorem 7.** *Suppose that Assumptions 1 and 3 hold, and let  $\phi_*$  be the randomized likelihood ratio test such that*

$$\phi_* = \arg \min_{\phi} \lim_{k \rightarrow \infty} P_{e,k}((\phi_1, \phi), Q_0^k, Q_1^k).$$

Then,

$$\inf_{\phi} P_{\infty}^{DD}(\phi) = \lim_{k \rightarrow \infty} P_{e,k}((\phi_1, \phi_*), Q_0^k, Q_1^k).$$

*Proof:* From Theorem 2, we have

$$\begin{aligned} \lim_{k \rightarrow \infty} P_{e,k}((\phi_1, \phi_*), Q_0^k, Q_1^k) \\ &= \lim_{k \rightarrow \infty} \sup_{(P_0^{(k)}, P_1^{(k)}) \in \mathcal{P}_0^k \times \mathcal{P}_1^k} P_{e,k}((\phi_1, \phi_*), P_0^{(k)}, P_1^{(k)}) \\ &= P_{\infty}^{DD}(\phi_*). \end{aligned}$$

By the theorem assumption, for any decision rule  $\phi$  we have

$$\lim_{k \rightarrow \infty} P_{e,k}((\phi_1, \phi_*), Q_0^k, Q_1^k) \leq \lim_{k \rightarrow \infty} P_{e,k}((\phi_1, \phi), Q_0^k, Q_1^k).$$

Hence, for any decision rule  $\phi$ , we have

$$\begin{aligned} P_{\infty}^{DD}(\phi) &= \lim_{k \rightarrow \infty} \sup_{P_0^{(k)} \in \mathcal{P}_0^k, P_1^{(k)} \in \mathcal{P}_1^k} P_{e,k}((\phi_1, \phi), P_0^{(k)}, P_1^{(k)}) \\ &\geq \lim_{k \rightarrow \infty} P_{e,k}((\phi_1, \phi), Q_0^k, Q_1^k) \\ &\geq \lim_{k \rightarrow \infty} P_{e,k}((\phi_1, \phi_*), Q_0^k, Q_1^k) \\ &= P_{\infty}^{DD}(\phi_*). \end{aligned}$$

The proof is now complete.  $\blacksquare$

Theorem 7 states that each agent should find the decision rule to optimize the asymptotic minimax error as if its observations were distributed according to the LFDs of the uncertainty class. This is consistent with our results when agents do know their positions (Theorem 2). However, in general, the exact threshold values for  $\phi_*$  are difficult to find analytically, but can be computed using numerical methods. Together with Lemma 3, since the uncertainty classes  $\mathcal{P}_{0,k}$  and  $\mathcal{P}_{1,k}$  are disjoint for all  $k$ , Theorem 7 shows that asymptotically learning the true hypothesis is impossible when agents do not know their own positions and also provides an expression for the asymptotic minimax error.

2) *Minimizing error of current agent:* We now assume that every agent is acting to minimize its local minimax error probability, and that each agent past the first does not know which position it is in. This is true in general for most social networks, where it is difficult to find the root of any information spread. Hence, users typically would not know how many hops information has been propagated from its source.

We find the decision rule to minimize the maximum error probability by allowing each agent to consider the maximum error probability for every possible position it might be in, then finding the decision rule that minimizes this value. Like in the previous subsection, we also assume that Assumptions 1 and 3 are in effect. More specifically, for each agent  $k \geq 2$ , we wish to find  $\phi$  that minimizes

$$P_{\max}^{SL}(\phi) = \sup_{k \geq 2} P_k^{SL}(\phi | (\phi_1, \phi)), \quad (18)$$

where  $\phi_1$  is the optimal decision rule that minimizes  $P_1^{SL}(\cdot)$ . We will show that for the optimal decision rule  $\phi$ , the maximum error probability occurs either in the second position or at the asymptotic limit, as defined in (12).

**Theorem 8.** *Suppose that Assumptions 1 and 3 hold. Then,  $\phi_* = \arg \min_{\phi} P_{\max}^{SL}(\phi)$  is a randomized likelihood ratio test between the LFDs  $Q_0$  and  $Q_1$ , and*

$$P_{\max}^{SL}(\phi_*) = \max \{P_2^{SL}(\phi_* | \phi_1), P_{\infty}^{DD}(\phi_*)\},$$

where  $P_{\infty}^{DD}(\phi_*)$  is as defined in (12).

*Proof:* See Appendix A-D.  $\blacksquare$

Finding an analytical form for  $\phi_*$  in Theorem 8 is difficult. However, as  $\phi_*$  is known to be a randomized likelihood ratio test of  $Q_0$  and  $Q_1$ , this can be done numerically by minimizing  $\max \{P_2^{SL}(\phi_* | \phi_1), P_{\infty}^{DD}(\phi_*)\}$  with respect to the thresholds for  $\phi_*$ .

## V. NUMERICAL RESULTS

In this section, we provide numerical results to illustrate part of our theoretical contributions in Section IV-C for the case where the uncertainty classes of each agent are  $\epsilon$ -contamination uncertainty classes in (5). Our numerical results show that even if an agent  $k$ 's position is unknown, where  $k \geq 2$ , its optimal decision rule  $\phi_*$  can be chosen to be a randomized likelihood ratio test between the LFDs  $Q_0$  and  $Q_1$ . This is true whether the agent is collaborating with others

to minimize the asymptotic error (decentralized detection) or is trying to minimize its own error probability (social learning). We have shown this in Theorems 7 and 8 respectively. The form of this randomized likelihood ratio test  $\phi_*$  is given in (13). Now, we can rewrite this in terms of an optimization problem.

First, given the nominal distributions  $P_0^*$  and  $P_1^*$ , as well as the contamination values  $\epsilon_{0,k} = \epsilon_0$  and  $\epsilon_{1,k} = \epsilon_1$  for all  $k \geq 1$ , we can use a binary search to compute  $c'$  and  $c''$  in (6). From the definition of the LFDs, we know that the range of thresholds we have to optimize over is bounded between  $bc'$  and  $bc''$ , where  $b = \frac{1 - \epsilon_1}{1 - \epsilon_0}$ . Then, we can derive the LFDs  $Q_0$  and  $Q_1$  and obtain

$$Q_0(\phi_*(Y, 1) = 1) = Q_0(l^*(Y) > t_1) + (1 - p)Q_0(l^*(Y) = t_1)$$

and

$$Q_0(\phi_*(Y, 0) = 1) = Q_0(l^*(Y) > t_0) + (1 - q)Q_0(l^*(Y) = t_0).$$

Similarly, we have

$$Q_1(\phi_*(Y, 1) = 0) = Q_1(l^*(Y) < t_1) + pQ_1(l^*(Y) = t_1)$$

and

$$Q_1(\phi_*(Y, 0) = 0) = Q_1(l^*(Y) < t_0) + qQ_1(l^*(Y) = t_0).$$

In the case where agents collaborate to minimize the asymptotic maximum error  $P_{\max}^{\text{DD}}$ , we minimize the expression in Lemma 3 to obtain the optimal decision rule  $\phi_*$ . From Theorem 7, since  $\phi_*$  is a randomized likelihood ratio test, we can perform the optimization over  $t_1$ ,  $t_0$ ,  $p$  and  $q$  in (13).

Similarly, in the case where agents minimize their local maximum error probability  $P_{\max}^{\text{SL}}$ , we minimize the expression in Theorem 8 over  $t_1$ ,  $t_0$ ,  $p$  and  $q$ .

We now present a numerical example using the exponential distribution. First, we fix  $P_0^*$  as an exponential distribution with mean 1 and let  $P_1^*$  be an exponential distribution with mean 2. We set  $\epsilon_0 = \epsilon_1 = 0.01$ , and  $\pi_0 = \pi_1 = 0.5$ . Then, the optimal decision rule  $\phi_1$  for agent 1 has the form:

$$u_1 = \begin{cases} 0 & \text{if } l^*(Y_1) < 1 \\ 1 & \text{if } l^*(Y_1) \geq 1. \end{cases}$$

Note that for this case, if  $t_1 \neq c'$  and  $t_0 \neq c''$ , then it does not matter what  $p$  and  $q$  are. This is because when the nominal distributions are both continuous,  $Q_0(l^*(Y_k) = x)$  and  $Q_1(l^*(Y_k) = x)$  are both zero unless  $x = c'$  or  $x = c''$ . We plot  $P_k^{\text{SL}}(\phi | (\phi_1, \phi))$  against  $k$  when  $\phi = \phi_A$  and  $\phi_B$ , which are randomized likelihood ratio tests of the form (13). Both rules have  $t_1 = c'$  and  $p = 1$ , but  $\phi_A$  has  $t_0 = 5$  and  $\phi_B$  has  $t_0 = 1.1$ .

Figure 2 shows that the total error probability is decreasing over  $k$  for  $\phi_A$ . Hence, the maximum error probability occurs when  $k = 2$ . For  $\phi_B$ , the maximum occurs at the asymptotic limit  $k \rightarrow \infty$ . This is in line with our conclusion in Theorem 8.

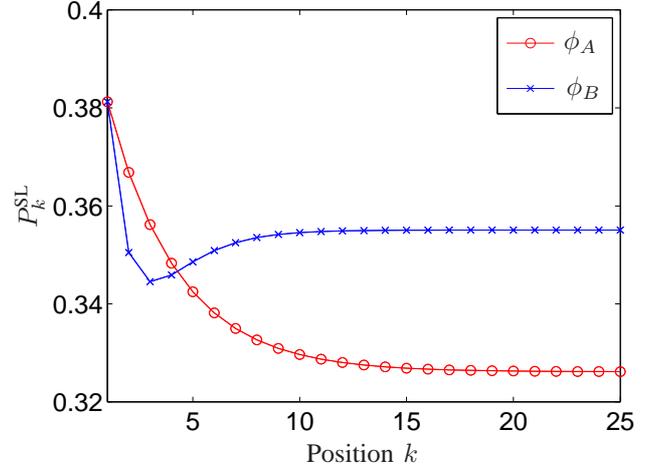


Fig. 2. Comparison of decision rules.

Next, we let  $P_0^*$  be an exponential distribution with mean 1 and  $P_1^*$  be an exponential distribution with variable mean. We let  $\epsilon_0 = \epsilon_1 = 0.01$ . We denote

$$\phi_*^{\text{SL}} = \arg \min_{\phi} P_{\max}^{\text{SL}}(\phi)$$

and

$$\phi_*^{\text{DD}} = \arg \min_{\phi} P_{\infty}^{\text{DD}}(\phi),$$

where  $P_{\max}^{\text{SL}}(\phi)$  and  $P_{\infty}^{\text{DD}}(\phi)$  are defined as in (12) and (18), respectively.

Figure 3 shows that as the mean of  $P_1^*$  increases, both  $P_{\max}^{\text{SL}}(\phi_*^{\text{SL}})$  and  $P_{\infty}^{\text{DD}}(\phi_*^{\text{DD}})$  decrease. This is intuitive as the Kullback-Leibler divergence of  $P_1^*$  from  $P_0^*$  increases as the mean of  $P_1^*$  increases. As the nominal distributions become easier to differentiate, the asymptotic error probability decreases.

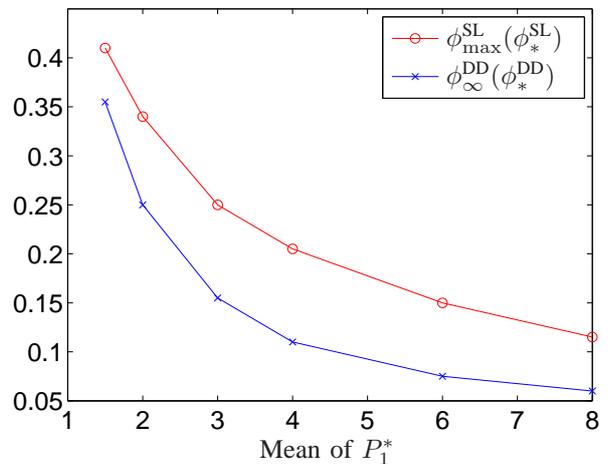


Fig. 3. Error probability versus mean of  $P_1^*$ .

Figure 4 shows that both  $P_{\max}^{\text{SL}}(\phi_*^{\text{SL}})$  and  $P_{\infty}^{\text{DD}}(\phi_*^{\text{DD}})$  increases as  $\epsilon_0 = \epsilon_1$  increases. For this graph,  $P_0^*$  and  $P_1^*$  are kept constant as exponential distributions with means 1 and 2

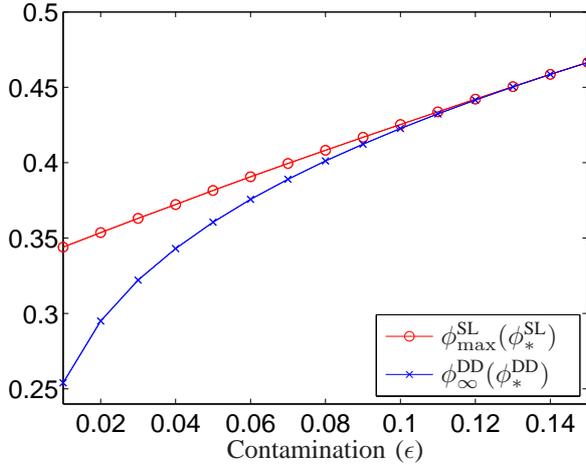


Fig. 4. Error probability versus contamination value.

respectively. As expected, as the contamination increases, so does the asymptotic error. When the contamination is large enough, both  $P_{\max}^{\text{SL}}(\phi_*^{\text{SL}})$  and  $P_{\infty}^{\text{DD}}(\phi_*^{\text{DD}})$  converge towards 0.5. Furthermore, as contamination increases, the gap between  $P_{\max}^{\text{SL}}(\phi_*^{\text{SL}})$  and  $P_{\infty}^{\text{DD}}(\phi_*^{\text{DD}})$  decreases. In a network with a lot of uncertainty in the observation distributions, there is not much incentive in trying to get agents to collaborate, as agents selfishly trying to minimize their own error probability leads to similar error performance.

## VI. CONCLUSION

We have shown that in a tandem network where agents' observation distributions are not known exactly, and belong to uncertainty classes defined by 2-alternating capacities, the minimax error probability is obtained by assuming that each observation is distributed according to the LFDs of the uncertainty classes. In the case where agents know their positions in the tandem network, asymptotically learning the true hypothesis is possible only when the log-likelihood ratio between the LFDs of the uncertainty classes is almost surely unbounded from either above or below. To achieve asymptotic learning of the true hypothesis in social learning, we require that the log-likelihood ratio is almost surely unbounded from both above and below. When uncertainty classes are allowed to vary between agents, learning occurs only if subsequences with log-likelihood ratios that are almost surely unbounded from above and below, respectively, can be found. If, in addition, the uncertainty classes converge to singletons, then asymptotic learning occurs. In the case where agents do not know their positions in the tandem, asymptotic learning of the true hypothesis becomes impossible even if the true distributions of the agents' observations are known. We characterized the minimax error performance in this case, which provided a way to determine the optimal decision rules for the agents.

In this work, we have restricted our attention to the tandem network. It would be of interest to extend some of our results to tree networks, and even to loopy general graphs. Another future research direction would be to consider the robust

detection problem with more than two hypotheses. A possible approach to this problem could be to focus on the LFDs on each possible pair of uncertainty class. Lastly, for the problem where each agent does not know its position in the network, we could instead consider each agent having partial knowledge of its position in the network, and find conditions under which learning the true hypothesis asymptotically is possible.

## APPENDIX A PROOFS OF MAIN RESULTS

### A. Proof of Lemma 1

We will prove this lemma using mathematical induction on  $N$ . The likelihood ratio test for agent  $k \geq 1$  is of the form

$$U_i = \begin{cases} 1 & \text{if } l^*(Y_k, U_{k-1}) > t_k, \\ 0 & \text{otherwise,} \end{cases}$$

where  $t_k$  is some chosen threshold. From the definition of the LFDs, we have

$$\begin{aligned} Q_1^{(1)}(U_1 = 1) &= Q_1^{(1)}(l^*(Y_1) \geq t_1) \\ &\geq Q_0^{(1)}(l^*(Y_1) \geq t_1) \\ &= Q_0^{(1)}(U_1 = 1), \end{aligned}$$

so that the lemma holds for  $N = 1$ .

Assume that the lemma is true for  $N = k - 1$ . Each agent's observation is independent of the previous agent's decision. Hence, we have

$$l^*(Y_k, U_{k-1}) = l^*(Y_k)l^*(U_{k-1}),$$

and the likelihood ratio test can be rewritten in the form

$$U_k = \begin{cases} 1 & \text{if } l^*(Y_k) \geq t_k^0, \\ U_{k-1} & \text{if } t_k^1 \leq l^*(Y_k) < t_k^0, \\ 0 & \text{if } t_k^1 > l^*(Y_k), \end{cases}$$

where  $t_k^i = t_k/l^*(U_{k-1} = i)$ . We obtain

$$\begin{aligned} Q_1^{(k)}(U_k = 1) &= Q_{1,k}(U_k = 1 | U_{k-1} = 1)Q_1^{(k-1)}(U_{k-1} = 1) \\ &\quad + Q_{1,k}(U_k = 1 | U_{k-1} = 0)Q_1^{(k-1)}(U_{k-1} = 0) \\ &= Q_{1,k}(l^*(Y_k) \geq t_k^1)Q_1^{(k-1)}(U_{k-1} = 1) \\ &\quad + Q_{1,k}(l^*(Y_k) \geq t_k^0)(1 - Q_1^{(k-1)}(U_{k-1} = 1)). \end{aligned}$$

From the induction hypothesis, we then have

$$\begin{aligned} &Q_{1,k}(l^*(Y_k) \geq t_k^1)Q_1^{(k-1)}(U_{k-1} = 1) \\ &\quad + Q_{1,k}(l^*(Y_k) \geq t_k^0)(1 - Q_1^{(k-1)}(U_{k-1} = 1)) \\ &\geq Q_{1,k}(l^*(Y_k) \geq t_k^1)Q_0^{(k-1)}(U_{k-1} = 1) \\ &\quad + Q_{1,k}(l^*(Y_k) \geq t_k^0)(1 - Q_0^{(k-1)}(U_{k-1} = 1)) \\ &\geq Q_{0,k}(l^*(Y_k) \geq t_k^1)Q_0^{(k-1)}(U_{k-1} = 1) \\ &\quad + Q_{0,k}(l^*(Y_k) \geq t_k^0)(1 - Q_0^{(k-1)}(U_{k-1} = 1)) \\ &= Q_0^{(k)}(U_k = 1), \end{aligned}$$

where the second inequality follows from the definition of the LFDs. The proof of the lemma is now complete.

### B. Proof of Theorem 4

Without loss of generality, we assume that the log-likelihood ratio of  $Q_0$  and  $Q_1$  is almost surely unbounded from above.

Using the scheme proposed in [6], we can show that we can make the maximum error arbitrarily small as the number of agents tends to infinity. We denote this scheme as  $\Phi_\delta$ . The decision rules of  $\Phi_\delta$  are as follows:

For a given  $N^*$  and a threshold  $t$ ,

$$U_1 = \begin{cases} 0 & \text{for } l^*(Y_1) < t \\ 1 & \text{for } l^*(Y_1) \geq t. \end{cases}$$

For  $1 < k < N^*$ ,

$$U_k = \begin{cases} 0 & \text{for } l^*(Y_k) < t \text{ and } U_{k-1} = 0 \\ 1 & \text{otherwise.} \end{cases}$$

For  $k \geq N^*$ ,

$$U_k = U_{k-1}.$$

Note that  $\Phi_\delta$  is a sequence of likelihood ratio tests between  $Q_1$  and  $Q_0$ . It was shown in [6] that by choosing a suitable  $t$  and  $N^*$ , we can get an arbitrarily small error rate if  $P_1^{(N)} = Q_1^N$ . To see this, consider a point on the ROC curve of an agent using the likelihood ratio test between  $Q_0$  and  $Q_1$  with its tangent to the ROC curve having slope  $t$ . This point is  $(Q_0(l^*(Y) \geq t), Q_1(l^*(Y) \geq t))$ . From our theorem assumption, the initial slope of the ROC curve is  $\infty$ , and so such a point always exists for any  $t$ . From the concavity of the ROC curve, we have

$$Q_0(l^*(Y) \geq t) < \frac{Q_1(l^*(Y) \geq t)}{t}.$$

The asymptotic miss detection probability using the decision rules outlined above is thus

$$\begin{aligned} \lim_{N \rightarrow \infty} P_{M,N}(\Phi_\delta, Q_1^N) &= Q_1(U_{N^*} = 0) \\ &= (1 - Q_1(l^*(Y) \geq t))^{N^*}. \end{aligned}$$

Similarly, the asymptotic false alarm probability is

$$\begin{aligned} \lim_{N \rightarrow \infty} P_{F,N}(\Phi_\delta, Q_0^N) &= 1 - (1 - Q_0(l^*(Y) \geq t))^{N^*} \\ &< 1 - (1 - \frac{Q_1(l^*(Y) \geq t)}{t})^{N^*}. \end{aligned}$$

Choose an arbitrary  $\delta > 0$ . To get the asymptotic miss detection probability smaller than  $\delta$ , we have

$$(1 - Q_1(l^*(Y) \geq t))^{N^*} < \delta$$

and so

$$N^* > \frac{\log(\delta)}{\log(1 - Q_1(l^*(Y) \geq t))}. \quad (19)$$

Similarly, to get the asymptotic false alarm probability smaller than  $\delta$ , we have

$$\frac{\log(1 - \delta)}{\log(1 - \frac{Q_1(l^*(Y) \geq t)}{t})} > N^*. \quad (20)$$

We now make use of the following lemma.

**Lemma A.1.** For any probability distribution  $Q_1 \in \mathcal{R}$ ,

$$\lim_{t \rightarrow \infty} \frac{\log(1 - Q_1(l^*(Y) \geq t))}{\log(1 - \frac{Q_1(l^*(Y) \geq t)}{t})} = \infty.$$

*Proof:* Let  $g(x) = \log(1 - x)$ , a concave function. For  $0 < x < 1$ , we have  $g(x) < 0$ .

For any fixed  $t > 0$ , by Jensen's Inequality,

$$\frac{1}{t}g(x) + \frac{t-1}{t}g(0) \leq g(\frac{x}{t}).$$

As  $g(0) = 0$ , we have

$$\frac{g(x)}{g(\frac{x}{t})} \geq t.$$

Letting  $x = Q_1(l^*(Y) \geq t)$ ,

$$\frac{\log(1 - Q_1(l^*(Y) \geq t))}{\log(1 - \frac{Q_1(l^*(Y) \geq t)}{t})} \geq t.$$

Hence,

$$\lim_{t \rightarrow \infty} \frac{\log(1 - Q_1(l^*(Y) \geq t))}{\log(1 - \frac{Q_1(l^*(Y) \geq t)}{t})} = \infty. \quad \blacksquare$$

From Lemma A.1, by choosing a large enough  $t$ , we have

$$\frac{\log(1 - Q_1(l^*(Y) \geq t))}{\log(1 - \frac{Q_1(l^*(Y) \geq t)}{t})} > 2 \frac{\log(\delta)}{\log(1 - \delta)}$$

and so

$$\frac{\log(1 - \delta)}{\log(1 - \frac{Q_1(l^*(Y) \geq t)}{t})} > 2 \frac{\log(\delta)}{1 - \log(Q_1(l^*(Y) \geq t))}.$$

For a sufficiently large  $t$ ,

$$\frac{\log(\delta)}{1 - \log(Q_1(l^*(Y) \geq t))} > 1.$$

Hence we can find an integer  $N^*$  that lies between the bounds in (19) and (20). Then,

$$\begin{aligned} \lim_{N \rightarrow \infty} \inf_{\phi^{(N)}} P_N^{\text{DD}}(\phi^{(N)}) &\leq \lim_{N \rightarrow \infty} P_N^{\text{DD}}(\Phi_\delta) \\ &= \lim_{N \rightarrow \infty} \sup_{P_0^{(N)} \in \mathcal{P}_0^{(N)}} \pi_0 P_{F,N}(\Phi_\delta, P_0^{(N)}) \\ &\quad + \pi_1 \sup_{P_1^{(N)} \in \mathcal{P}_1^{(N)}} P_{M,N}(\Phi_\delta, P_1^{(N)}) \\ &= \lim_{N \rightarrow \infty} \pi_0 P_{F,N}(\Phi_\delta, Q_0^{(N)}) + \pi_1 P_{M,N}(\Phi_\delta, Q_1^{(N)}) \quad (21) \\ &< \pi_0 \delta + \pi_1 \delta \\ &= \delta, \end{aligned}$$

where (21) is due to Theorem 2. Since  $\delta$  was arbitrarily chosen, we can make the maximum error arbitrarily small.

To show necessity, assume that the log-likelihood of  $Q_1$  and  $Q_0$  is bounded. Then

$$\begin{aligned} \lim_{N \rightarrow \infty} \inf_{\phi^{(N)}} \sup_{(P_0^{(N)}, P_1^{(N)})} P_{e,N}(\phi^{(N)}, P_0^{(N)}, P_1^{(N)}) \\ \geq \lim_{N \rightarrow \infty} \inf_{\phi^{(N)}} P_{e,N}(\phi^{(N)}, Q_0^N, Q_1^N), \end{aligned}$$

which is known to be bounded above zero as  $N \rightarrow \infty$  as the log-likelihood of  $Q_1$  and  $Q_0$  is bounded from both above and below.

### C. Proof of Theorem 6

We will begin by proving that the necessary condition is required for asymptotic learning to occur. Assume that such subsequences do not exist. Without loss of generality, assume that  $\log l^*(Y_k)$  is bounded above zero. Let  $\inf_{k,y} l^*(Y_k = y) = \delta > 0$ .

Assume now that  $\pi_0 Q_{F,k} + \pi_1 Q_{M,k} \rightarrow 0$  as  $k \rightarrow 0$ . Then both  $Q_{F,k} \rightarrow 0$  and  $Q_{M,k} \rightarrow 0$ . Choose  $j$  such that for all  $k \geq j$ ,  $Q_{F,k} + Q_{M,k} \leq 1$  and  $\pi_0 Q_{F,k} + \delta \pi_1 Q_{M,k} \leq \delta \pi_1$ . We have from (11),

$$\begin{aligned} & Q_{F,j+1} \\ &= Q_{F,j} \cdot Q_{0,j+1} \left( l^*(Y_{j+1}) \geq \frac{\pi_0 Q_{F,j}}{\pi_1 (1 - Q_{M,j})} \right) \\ &\quad + (1 - Q_{F,j}) \cdot Q_{0,j+1} \left( l^*(Y_{j+1}) \geq \frac{\pi_0 (1 - Q_{F,j})}{\pi_1 Q_{M,j}} \right) \\ &\geq Q_{F,j} \cdot Q_{0,j+1} (l^*(Y_{j+1}) \geq \delta) \\ &= Q_{F,j}. \end{aligned}$$

By induction,  $Q_{F,k} \geq Q_{F,j}$  for all  $k \geq j$ , a contradiction. Hence, the necessary condition is required for asymptotic learning to occur.

To show the sufficient condition, we first show a series of lemmas.

**Lemma A.2.** *Suppose that the LFDs for a pair of uncertainty classes  $(\mathcal{P}_0, \mathcal{P}_1)$  are  $(Q_0, Q_1)$ . Then, for any random variable  $Y$  with distributions belonging to these uncertainty classes, we have for all  $t > 0$ ,*

$$Q_1(l^*(Y) \leq t) \leq t Q_0(l^*(Y) \leq t) - \frac{t}{2} Q_0(l^*(Y) \leq t/2),$$

and

$$Q_0(l^*(Y) \geq t) \leq \frac{1}{t} Q_1(l^*(Y) \geq t) - \frac{1}{2t} Q_1(l^*(Y) \geq 2t).$$

Furthermore, the above two inequalities hold when  $l^*(Y) \leq t$  and  $l^*(Y) \geq t$  are replaced by  $l^*(Y) < t$  and  $l^*(Y) > t$ , respectively throughout.

*Proof:* To show the first inequality, we observe that

$$\begin{aligned} Q_1(l^*(Y) \leq t) &= t Q_0(l^*(Y) \leq t) \\ &\quad - \int_{x=0}^t (t-x) dQ_0(l^*(Y) = x) \\ &\leq t Q_0(l^*(Y) \leq t) \\ &\quad - \int_{x=0}^{\frac{t}{2}} (t-x) dQ_0(l^*(Y) = x) \\ &\leq t Q_0(l^*(Y) \leq t) \\ &\quad - \int_{x=0}^{\frac{t}{2}} (t - \frac{t}{2}) dQ_0(l^*(Y) = x) \\ &= t Q_0(l^*(Y) \leq t) - \frac{t}{2} Q_0(l^*(Y) \leq t/2). \end{aligned}$$

The second inequality follows by interchanging  $Q_0$  and  $Q_1$ , and replacing  $t$  by  $1/t$ . The proof is now complete.  $\blacksquare$

**Lemma A.3.** *Suppose that  $Q_{F,k} + Q_{M,k}$  is bounded away from zero, and that  $\log l^*(Y_k)$  is almost surely unbounded from*

*either above or below. Then for any integer  $n$ , there exists some  $k' \geq n$  such that*

$$\begin{aligned} & \pi_0 Q_{F,k'} + \pi_1 Q_{M,k'} \\ & \leq \pi_0 Q_{F,k'-1} + \pi_1 Q_{M,k'-1} \\ & \quad - \frac{\pi_1 Q_{M,k'-1}}{2} \cdot Q_{1,k'} \left( l^*(Y_{k'}) \geq \frac{2\pi_0(1 - Q_{F,k'-1})}{\pi_1 Q_{M,k'-1}} \right) \\ & \quad - \frac{\pi_0 Q_{F,k'-1}}{2} \cdot Q_{0,k'} \left( l^*(Y_{k'}) < \frac{\pi_0 Q_{F,k'-1}}{2\pi_1(1 - Q_{M,k'-1})} \right). \end{aligned}$$

*Proof:* For any  $n$ , choose  $k' \geq n$  such that  $Q_{1,k'} \left( l^*(Y_{k'}) \geq \frac{2\pi_0(1 - Q_{F,k'-1})}{\pi_1 Q_{M,k'-1}} \right) > 0$  or  $Q_{0,k'} \left( l^*(Y_{k'}) < \frac{\pi_0 Q_{F,k'-1}}{2\pi_1(1 - Q_{M,k'-1})} \right) > 0$ . This is possible as we assume either  $Q_{F,k}$  or  $Q_{M,k}$  is bounded away from 0, and from our lemma assumption, the lower or upper bound on  $l^*(Y_k)$  converges to 0 or  $\infty$ , respectively (cf. (6)). Then, we have

$$\begin{aligned} & \pi_0 Q_{F,k'} + \pi_1 Q_{M,k'} \\ &= \pi_0 Q_{F,k'-1} \cdot Q_{0,k'} \left( l^*(Y_{k'}) \geq \frac{\pi_0 Q_{F,k'-1}}{\pi_1 (1 - Q_{M,k'-1})} \right) \\ &\quad + \pi_0 (1 - Q_{F,k'-1}) \cdot Q_{0,k'} \left( l^*(Y_{k'}) \geq \frac{\pi_0 (1 - Q_{F,k'-1})}{\pi_1 Q_{M,k'-1}} \right) \\ &\quad + \pi_1 Q_{M,k'-1} \cdot Q_{1,k'} \left( l^*(Y_{k'}) < \frac{\pi_0 (1 - Q_{F,k'-1})}{\pi_1 Q_{M,k'-1}} \right) \\ &\quad + \pi_1 (1 - Q_{M,k'-1}) \\ &\quad \cdot Q_{1,k'} \left( l^*(Y_{k'}) < \frac{\pi_0 Q_{F,k'-1}}{\pi_1 (1 - Q_{M,k'-1})} \right) \\ &\leq \pi_0 Q_{F,k'-1} \cdot Q_{0,k'} \left( l^*(Y_{k'}) \geq \frac{\pi_0 Q_{F,k'-1}}{\pi_1 (1 - Q_{M,k'-1})} \right) \\ &\quad + \pi_1 Q_{M,k'-1} \cdot Q_{1,k'} \left( l^*(Y_{k'}) \geq \frac{\pi_0 (1 - Q_{F,k'-1})}{\pi_1 Q_{M,k'-1}} \right) \\ &\quad - \frac{\pi_1 Q_{M,k'-1}}{2} \cdot Q_{1,k'} \left( l^*(Y_{k'}) \geq \frac{2\pi_0(1 - Q_{F,k'-1})}{\pi_1 Q_{M,k'-1}} \right) \\ &\quad + \pi_1 Q_{M,k'-1} \cdot Q_{1,k'} \left( l^*(Y_{k'}) < \frac{\pi_0 (1 - Q_{F,k'-1})}{\pi_1 Q_{M,k'-1}} \right) \\ &\quad + \pi_0 Q_{F,k'-1} \cdot Q_{0,k'} \left( l^*(Y_{k'}) < \frac{\pi_0 Q_{F,k'-1}}{\pi_1 (1 - Q_{M,k'-1})} \right) \\ &\quad - \frac{\pi_0 Q_{F,k'-1}}{2} \cdot Q_{0,k'} \left( l^*(Y_{k'}) < \frac{\pi_0 Q_{F,k'-1}}{2\pi_1 (1 - Q_{M,k'-1})} \right) \end{aligned}$$

where the inequality follows from Lemma A.2. The lemma is now proved.  $\blacksquare$

We now return to the proof of Theorem 6. If  $Q_{F,k} + Q_{M,k} \rightarrow 0$ , the theorem holds trivially. Therefore we assume otherwise. Since  $P_k^{\text{SL}}(\phi_k | \phi^{(k-1)})$  is bounded and non-increasing, it converges. Suppose that  $\pi_0 Q_{F,k} + \pi_1 Q_{M,k} \rightarrow C$ , for some  $C > 0$ . Either  $\limsup_{k_n} (Q_{F,k_n}) > 0$  or  $\limsup_{k_n} (Q_{M,k_n}) > 0$ . Without loss of generality, let  $\limsup_{k_n} (Q_{F,k_n}) = C' > 0$ . Then there exists a subsequence of agents  $(k_{n_\alpha})_{\alpha \geq 1}$  such that  $Q_{F,k_{n_\alpha}} \rightarrow C'$ . Choose  $N$  such that  $Q_{F,k_{n_\alpha}} > \frac{C'}{2}$  for all

$\alpha' \geq \alpha$ . Then, from Lemma A.2 and Lemma A.3, we have

$$\begin{aligned} & P_{k_{n_\alpha+1}}^{\text{SL}}(\phi_{k_{n_\alpha+1}} | \phi^{(k_{n_\alpha+1}-1)}) \\ & \leq P_{k_{n_\alpha+1}}^{\text{SL}}(\phi_{k_{n_\alpha+1}} | \phi^{(k_{n_\alpha})}) \\ & \leq P_{k_{n_\alpha}}^{\text{SL}}(\phi_{k_{n_\alpha}} | \phi^{(k_{n_\alpha}-1)}) - \frac{\pi_0 Q_{F,k_{n_\alpha}-1}}{2} \\ & \quad \cdot Q_{0,k_{n_\alpha}} \left( l^*(Y_{k_{n_\alpha}}) < \frac{\pi_0 Q_{F,k_{n_\alpha}-1}}{2\pi_1(1-Q_{M,k_{n_\alpha}-1})} \right) \\ & \leq P_{k_{n_\alpha}}^{\text{SL}}(\phi_{k_{n_\alpha}} | \phi^{(k_{n_\alpha}-1)}) - \frac{\pi_0 C'}{4} \\ & \quad \cdot Q_{0,k_{n_\alpha}} \left( l^*(Y_{k_{n_\alpha}}) < \frac{C'}{4\pi_1} \right). \end{aligned}$$

From our theorem assumption that  $\mathcal{P}_{0,k_n}$  converges to a singleton, by letting  $\alpha \rightarrow \infty$  on both sides of the above inequality, we obtain a contradiction. The proof of the theorem is now complete.

#### D. Proof of Theorem 8

The decision rule of agent 1,  $\phi_1$ , is of the form

$$u_1 = \begin{cases} 0 & \text{if } l^*(Y_1) < \frac{\pi_0}{\pi_1} \\ 1 & \text{if } l^*(Y_1) \geq \frac{\pi_0}{\pi_1}. \end{cases}$$

Hence, the minimax error probability for the first agent is  $\pi_0 Q_0(l^*(Y_1) \geq \frac{\pi_0}{\pi_1}) + \pi_1 Q_1(l^*(Y_1) \geq \frac{\pi_0}{\pi_1})$ , which is at least  $\inf_{\phi} P_{\max}^{\text{SL}}(\phi)$ . This is because using the trivial rule of ignoring one's own private observation and simply passing on the decision of the previous agent, we get a constant maximum error probability of  $\pi_0 Q_0(l^*(Y_1) \geq \frac{\pi_0}{\pi_1}) + \pi_1 Q_1(l^*(Y_1) \geq \frac{\pi_0}{\pi_1})$ . Hence, we have

$$\begin{aligned} \inf_{\phi} \sup_{k \geq 2} P_k^{\text{SL}}(\phi | (\phi_1, \phi)) & \leq \pi_0 Q_0(l^*(Y_1) \geq \frac{\pi_0}{\pi_1}) \\ & \quad + \pi_1 Q_1(l^*(Y_1) \geq \frac{\pi_0}{\pi_1}). \end{aligned} \quad (22)$$

We next prove the following lemma.

**Lemma A.4.** *Suppose that Assumptions 1 and 3 hold. There is no loss of optimality if the optimal decision rule  $\phi_*$  for all agents  $k \geq 2$  is restricted to be a randomized likelihood ratio test between  $Q_0$  and  $Q_1$ . Furthermore, we have*

$$P_{\max}^{\text{SL}}(\phi_*) = \sup_{k \geq 2} P_{e,k}((\phi_1, \phi_*), Q_0^k, Q_1^k).$$

*Proof:* The proof is similar to that of Lemma 4. From the proof of Lemma 4 it is shown that if  $\phi_*$  is not a randomized likelihood ratio test between  $Q_0$  and  $Q_1$ , there is some decision rule  $\phi'_*$  which is a randomized likelihood ratio test between  $Q_0$  and  $Q_1$  such that for any  $k \geq 2$ ,  $P_0^{(k)}, P_1^{(k)}$ ,

$$P_{e,k}((\phi_1, \phi_*), P_0^{(k)}, P_1^{(k)}) \geq P_{e,k}((\phi_1, \phi'_*), P_0^{(k)}, P_1^{(k)}),$$

and so there is no loss in optimality in assuming that  $\phi_*$  is a likelihood ratio test between  $Q_0$  and  $Q_1$ .

From Theorem 2, if  $\phi_*$  is a randomized likelihood ratio test between  $Q_0$  and  $Q_1$ , then for any  $k \geq 2$ ,  $P_0^{(k)}, P_1^{(k)}$ , we have

$$P_{e,k}((\phi_1, \phi_*), P_0^{(k)}, P_1^{(k)}) \leq P_{e,k}((\phi_1, \phi_*), Q_0^k, Q_1^k).$$

Hence, we obtain

$$\begin{aligned} P_{\max}^{\text{SL}}(\phi_*) & = \sup_{k \geq 2} \sup_{P_0 \in \mathcal{P}_0, P_1 \in \mathcal{P}_1} P_{e,k}((\phi_1, \phi_*), P_0^{(k)}, P_1^{(k)}) \\ & = \sup_{k \geq 2} P_{e,k}((\phi_1, \phi_*), Q_0^k, Q_1^k), \end{aligned}$$

and the lemma is proved.  $\blacksquare$

We now return to the proof of Theorem 8. From (22), we have

$$\begin{aligned} P_2^{\text{SL}}(\phi_* | \phi_1) & \leq \sup_{k \geq 2} P_k^{\text{SL}}(\phi | (\phi_1, \phi_*)) \\ & \leq \pi_0 Q_0(l^*(Y_1) \geq \frac{\pi_0}{\pi_1}) + \pi_1 Q_1(l^*(Y_1) < \frac{\pi_0}{\pi_1}) \end{aligned} \quad (23)$$

Let

$$\begin{aligned} \alpha & = P_{F,2}((\phi_1, \phi_*), Q_0^2) - \lim_{k \rightarrow \infty} P_{F,k}((\phi_1, \phi_*), Q_0^k), \\ \beta & = P_{M,2}((\phi_1, \phi_*), Q_1^2) - \lim_{k \rightarrow \infty} P_{M,k}((\phi_1, \phi_*), Q_1^k). \end{aligned}$$

and define the function

$$\begin{aligned} g(x) & = \lim_{k \rightarrow \infty} P_{e,k}((\phi_1, \phi_*), Q_0^k, Q_1^k) \\ & \quad + \pi_0 \alpha (Q_0(\phi_*(Y_1, 1) = 1) - Q_0(\phi_*(Y_1, 0) = 1))^{x-1} \\ & \quad + \pi_1 \beta (Q_1(\phi_*(Y_1, 0) = 0) - Q_1(\phi_*(Y_1, 1) = 0))^{x-1}. \end{aligned}$$

From the recurrence relation in Lemma 3, we see that for integer values of  $x$ , this function gives  $P_{e,x}((\phi_1, \phi_*), Q_0^x, Q_1^x)$ . Differentiating with respect to  $x$ , we observe that  $g(x)$  has at most one stationary point for  $x > 0$ . From (23), we have  $g(2) \geq g(1)$ . Hence, the stationary point, if any exists, must be a minimum. This means that for  $x \geq 2$ , the maximum value of  $g(x)$  must be either at  $x = 2$  or at its asymptotic limit. Thus, the theorem is now proven.

#### REFERENCES

- [1] R. Viswanathan, S. C. A. Thomopoulos, and R. Tumuluri, "Optimal serial distributed decision fusion," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 24, no. 4, pp. 366–376, Jul. 1988.
- [2] Z.-B. Tang, K. Pattipati, and D. Kleinman, "Optimization of detection networks. Part I. Tandem structures," *IEEE Trans. Syst., Man, Cybern.*, vol. 21, no. 5, pp. 1044–1059, Sep. 1991.
- [3] T. M. Cover, "Hypothesis testing with finite statistics," *Ann. Math. Statist.*, vol. 40, pp. 825–83, Jun. 1969.
- [4] M. Hellman and T. Cover, "Learning with finite memory," *Ann. Math. Statist.*, vol. 41, pp. 765–782, Jun. 1970.
- [5] J. Kopolowitz, "Necessary and sufficient memory size for m-hypothesis testing," *IEEE Trans. Inf. Theory*, vol. 21, no. 1, pp. 44–46, Jan. 1975.
- [6] D. Papastavrou and M. Athans, "Distributed detection by a large team of sensors in tandem," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 28, pp. 639–653, Jul. 1992.
- [7] P. Swaszek, "On the performance of serial networks in distributed detection," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 29, no. 1, pp. 254–260, Jan. 1993.
- [8] W. P. Tay, J. N. Tsitsiklis, and M. Z. Win, "On the sub-exponential decay of detection error probabilities in long tandems," *IEEE Trans. Inf. Theory*, vol. 54, pp. 4767–4771, Oct. 2008.
- [9] D. Acemoglu, M. A. Dahleh, I. Lobel, and A. Ozdaglar, "Bayesian learning in social networks," *Review of Economic Studies*, vol. 78, no. 4, pp. 1201–1236, Mar. 2011.
- [10] K. Drakopoulos, A. Ozdaglar, and J. Tsitsiklis, "On learning with finite memory," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6859–6872, May 2013.
- [11] W. P. Tay, "Whose opinion to follow in multihypothesis social learning? A large deviations perspective," *IEEE J. Sel. Topics Signal Process.*, vol. 9, no. 2, pp. 344–359, Mar. 2015.

- [12] L. Smith and P. Sorensen, "Pathological outcomes of observational learning," *Econometrica*, vol. 68, no. 2, pp. 371–398, Mar. 2000.
- [13] A. V. Banerjee, "A simple model of herd behavior," *The Quarterly Journal of Economics*, vol. 107, no. 3, pp. 797–817, Aug. 1992.
- [14] A. Banerjee and D. Fudenberg, "Word-of-mouth learning," *Games and Economic Behaviour*, vol. 46, pp. 1–22, Jan. 2004.
- [15] 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.
- [16] Z. Zhang, A. Pezeshki, W. Moran, S. D. Howard, and E. K. Chong, "Error probability bounds for balanced binary relay trees," *IEEE Trans. Inf. Theory*, vol. 58, pp. 3548–3563, Jun. 2012.
- [17] Z. Zhang, E. Chong, A. Pezeshki, W. Moran, and S. Howard, "Learning in hierarchical social networks," *IEEE J. Sel. Topics in Signal Process.*, vol. 7, no. 2, pp. 305–317, Apr. 2013.
- [18] J.-S. Lee and B. Hoh, "Sell your experiences: a market mechanism based incentive for participatory sensing," in *IEEE Int. Conf. Pervasive Computing and Communications*, Mar. 2010, pp. 60–68.
- [19] M. Demirbas, M. Bayir, C. Akcora, Y. Yilmaz, and H. Ferhatosmanoglu, "Crowd-sourced sensing and collaboration using Twitter," in *IEEE Int. Symp. on a World of Wireless Mobile and Multimedia Networks*, Jun. 2010, pp. 1–9.
- [20] D. Wang, L. Kaplan, T. Abdelzaher, and C. Aggarwal, "On credibility estimation tradeoffs in assured social sensing," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 6, pp. 1026–1037, Jun. 2013.
- [21] D. Wang, L. Kaplan, H. Le, and T. Abdelzaher, "On truth discovery in social sensing: a maximum likelihood estimation approach," in *Proc. Int. Conf. on Information Processing in Sensor Networks*, 2012, pp. 233–244.
- [22] INRIX Inc., INRIX Traffic App, [Online]. Available: <http://http://www.inrixtraffic.com/>.
- [23] J. N. Tsitsiklis, "Decentralized detection," pp. 297–344, Nov. 1993.
- [24] P. Varshney, *Distributed Detection and Data Fusion*. New York: Springer-Verlag, 1997.
- [25] Z. Chair and P. Varshney, "Optimal data fusion in multiple sensor detection systems," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-22, no. 1, pp. 98–101, Jan. 1986.
- [26] R. Viswanathan and P. Varshney, "Distributed detection with multiple sensors I. fundamentals," *Proc. of the IEEE*, vol. 85, no. 1, pp. 54–63, Jan. 1997.
- [27] Y. Lin, B. Chen, and P. Varshney, "Decision fusion rules in multi-hop wireless sensor networks," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 41, no. 2, pp. 475–488, Apr. 2005.
- [28] Z. Zhang, E. Chong, A. Pezeshki, and W. Moran, "Hypothesis testing in feedforward networks with broadcast failures," *IEEE J. Sel. Topics in Signal Process.*, vol. 7, no. 5, pp. 797–810, Oct. 2013.
- [29] P. J. Huber, "A robust version of the probability ratio test," vol. 36, no. 6, pp. 1753–1758, Dec. 1965.
- [30] S. Kassam and H. Poor, "Robust techniques for signal processing: A survey," *Proc. of the IEEE*, vol. 73, no. 3, pp. 433–481, Mar. 1985.
- [31] V. V. Veeravalli, T. Basar, and H. V. Poor, "Minimax robust decentralized detection," *IEEE Trans. Inf. Theory*, vol. 40, pp. 35–40, Jan. 1994.
- [32] E. Geraniotis and Y. Chau, "Robust data fusion for multisensor detection systems," *IEEE Trans. Inf. Theory*, vol. 36, no. 6, pp. 1265–1279, Nov. 1990.
- [33] P. J. Huber and V. Strassen, "Minimax tests and the Neyman-Pearson lemma for capacities," *Ann. Statist.*, vol. 1, no. 2, pp. 251–263, Mar. 1973.



**Jack Ho** (S'13) received the B.S. degree in Mathematics from Imperial College London, London, England, in 2011. He is currently working toward the Ph.D. degree in Electrical and Electronic Engineering at Nanyang Technological University.

His research interests include statistical signal processing, estimation theory and detection theory.



**Wee Peng Tay** (S'06 M'08 SM'14) received the B.S. degree in Electrical Engineering and Mathematics, and the M.S. degree in Electrical Engineering from Stanford University, Stanford, CA, USA, in 2002. He received the Ph.D. degree in Electrical Engineering and Computer Science from the Massachusetts Institute of Technology, Cambridge, MA, USA, in 2008. He is currently an Assistant Professor in the School of Electrical and Electronic Engineering at Nanyang Technological University, Singapore. His research interests include distributed detection and

estimation, distributed signal processing, sensor networks, social networks, information theory, and applied probability.

Dr. Tay received the Singapore Technologies Scholarship in 1998, the Stanford University President's Award in 1999, the Frederick Emmons Terman Engineering Scholastic Award in 2002, and the Tan Chin Tuan Exchange Fellowship in 2015. He is the coauthor of the best student paper award at the Asilomar conference on Signals, Systems, and Computers in 2012. He currently serves on the MLSP TC of the IEEE Signal Processing Society, and as the chair of DSNIG in IEEE MMTC. He has also served as a technical program committee member for various international conferences.



**Tony Q.S. Quek** (S'98-M'08-SM'12) received the B.E. and M.E. degrees in Electrical and Electronics Engineering from Tokyo Institute of Technology, Tokyo, Japan, respectively. At Massachusetts Institute of Technology, he earned the Ph.D. in Electrical Engineering and Computer Science. Currently, he is an Assistant Professor with the Information Systems Technology and Design Pillar at Singapore University of Technology and Design (SUTD). He is also a Scientist with the Institute for Infocomm Research.

His main research interests are the application of mathematical, optimization, and statistical theories to communication, networking, signal processing, and resource allocation problems. Specific current research topics include heterogeneous networks, green communications, smart grid, wireless security, internet-of-things, big data processing, and cognitive radio.

Dr. Quek has been actively involved in organizing and chairing sessions, and has served as a member of the Technical Program Committee as well as symposium chairs in a number of international conferences. He is serving as the technical chair for the PHY & Fundamentals Track for IEEE WCNC in 2015, the Communication Theory Symposium for IEEE ICC in 2015, the PHY & Fundamentals Track for IEEE EuCNC in 2015, and the Communication and Control Theory Symposium for IEEE ICC in 2015. He is currently an Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS, the IEEE WIRELESS COMMUNICATIONS LETTERS, and an Executive Editorial Committee Member for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS. He was Guest Editor for the IEEE SIGNAL PROCESSING MAGAZINE (Special Issue on Signal Processing for the 5G Revolution) in 2014, and the IEEE WIRELESS COMMUNICATIONS MAGAZINE (Special Issue on Heterogeneous Cloud Radio Access Networks) in 2015.

Dr. Quek was honored with the 2008 Philip Yeo Prize for Outstanding Achievement in Research, the IEEE Globecom 2010 Best Paper Award, the CAS Fellowship for Young International Scientists in 2011, the 2012 IEEE William R. Bennett Prize, the IEEE SPAWC 2013 Best Student Paper Award, and the IEEE WCSP 2014 Best Paper Award.



**Edwin K. P. Chong** (F'04) received the B.E. degree with First Class Honors from the University of Adelaide, South Australia, in 1987; and the M.A. and Ph.D. degrees in 1989 and 1991, respectively, both from Princeton University, where he held an IBM Fellowship. He joined the School of Electrical and Computer Engineering at Purdue University in 1991, where he was named a University Faculty Scholar in 1999. Since August 2001, he has been a Professor of Electrical and Computer Engineering and Professor of Mathematics at Colorado State

University. His current research interests span the areas of stochastic modeling and control, optimization methods, and communication and sensor networks. He coauthored the best-selling book, *An Introduction to Optimization* (4th Edition, Wiley-Interscience, 2013). He received the NSF CAREER Award in 1995 and the ASEE Frederick Emmons Terman Award in 1998. He was a co-recipient of the 2004 Best Paper Award for a paper in the journal *Computer Networks*. In 2010, he received the IEEE Control Systems Society Distinguished Member Award.

Prof. Chong was the founding chairman of the IEEE Control Systems Society Technical Committee on Discrete Event Systems, and served as an IEEE Control Systems Society Distinguished Lecturer. He is currently a Senior Editor of the IEEE TRANSACTIONS ON AUTOMATIC CONTROL, and has also served on the editorial boards of *Computer Networks* and the *Journal of Control Science and Engineering*. He was a member of the IEEE Control Systems Society Board of Governors and serve as Vice President for Financial Activities until 2014. He has also served on the organizing committees of several international conferences. He has been on the program committees for the IEEE Conference on Decision and Control, the American Control Conference, the IEEE International Symposium on Intelligent Control, IEEE Symposium on Computers and Communications, and the IEEE Global Telecommunications Conference. He has also served in the executive committees for the IEEE Conference on Decision and Control, the American Control Conference, the IEEE Annual Computer Communications Workshop, the International Conference on Industrial Electronics, Technology & Automation, and the IEEE International Conference on Communications. He was the Conference (General) Chair for the Conference on Modeling and Design of Wireless Networks, part of SPIE ITCOM 2001. He was the General Chair for the 2011 Joint 50th IEEE Conference on Decision and Control and European Control Conference.