Whose Opinion to follow in Multihypothesis Social Learning? A Large Deviation Perspective

Wee Peng Tay, Member, IEEE

Abstract

We consider a multihypothesis social learning problem in which an agent has access to a set of private observations and chooses an opinion from a set of experts to incorporate into its final decision. To model individual biases, we allow the agent and experts to have general loss functions and possibly different decision spaces. We characterize the loss exponents of both the agent and experts, and provide an asymptotically optimal method for the agent to choose the best expert to follow. We show that up to asymptotic equivalence, the worst loss exponent for the agent is achieved when it adopts the 0-1 loss function, which assigns a loss of 0 if the true hypothesis is declared and a loss of 1 otherwise. On the other hand, if experts have the same decision space as the agent, then the best loss exponent for the agent is achieved when there is an expert who utilizes the 0-1 loss function.

Index Terms

Social learning, decentralized detection, error exponent, social network, Internet of Things.

I. INTRODUCTION

In an increasingly connected world, our opinions on a phenomenon of interest or event are often not only influenced by our direct independent observations, but also by other people’s public opinions on related events. In an online social network like Twitter or Facebook, users’ opinions and postings are often influenced by the opinions of those they are connected to or are “following” in the social network [1]–[4]. For example, in viral marketing using social networks, marketing companies often target a few

The author is with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore. E-mail: wptay@ntu.edu.sg
influential nodes in the network to help them push a product [5], [6]. Similarly, widespread online access has made it easier for us to follow the opinions of experts like celebrities and industry insiders, who may have access to private information that we are unaware of. For example, we may be interested to determine the financial health of a publicly listed company. In addition to our own observations about the company through its annual financial reports and stock prices, we may also choose to incorporate the “expert” opinion provided by financial blogs like [7] and [8]. In all these examples, inference about a phenomenon of interest is not only based on direct observations but also the opinions of other entities. This is known as social learning [9], [10].

A further example is the Internet of Things (IoT) framework [11], [12]. Sensors each make their own private observations but collaborate by exchanging public information. This can be viewed as a “physical social network” of devices, cooperating to learn about a phenomenon of interest. Sensors originally designed for a specific purpose may collaborate with other sensors to perform inference on a phenomenon they were not specifically designed for. However, in order to ensure energy efficiency, each sensor needs to intelligently choose which other sensors to collaborate with since not all sensors may provide information relevant to it. For example, a sensor trying to estimate the temperature in a particular room of a building may choose to incorporate information from other temperature sensors or sensors tracking the number of occupants in the building, instead of information from a vibration monitoring sensor.

In this paper, we investigate the problem of multihypothesis social learning in which an agent can select an expert opinion from a group of experts, to incorporate into its final inference. In the social network example, an expert is an influential node that may have access to privileged or insider information. In the publicly listed company example, experts include financial commentators and bloggers, and professional investment analysts. In the IoT example, every sensor is an expert in making inferences about the phenomenon of interest they are designed for. Specifically, we consider an agent 0 who wishes to choose from a set of $M$ hypotheses based on its own observations as well as the opinion of an expert chosen from a set of $K$ possible experts. Each expert has access to a set of private observations, which they use to form their own opinions about the true hypothesis, subject to their own local loss functions or biases, which may differ from those of agent 0. By taking into account the experts’ individual biases, our goal is to characterize the rate of decay in the expected loss or the loss exponent incurred by agent 0, and to find an asymptotically optimal policy for that agent.
A. Related Works

The problem of selecting the best expert opinion to follow is related to the problem of decentralized detection or decentralized hypothesis testing, which has been extensively studied in [13]–[23] and the references therein. In the decentralized detection problem, each agent has its own private observations, but cooperate with each other so that the whole network of agents reaches a decision regarding a common underlying phenomenon of interest. Here, which agent passes information to which other agent is determined by a known network topology like the parallel configuration [24], tandem network [25]–[27], and tree architectures [19]–[22]. Therefore, all these works do not consider the problem of which agent’s opinion to follow. Furthermore, in a decentralized detection problem, all agents are assumed to have the same hypotheses and loss functions for declaring the wrong hypothesis, or have the common goal of minimizing the loss at a last agent known as the fusion center.

We study a related but somewhat different problem from the decentralized detection problem. We consider the scenario where experts may be concerned about different phenomena of interest, and made decisions according to their own biases, instead of minimizing the loss of a particular agent or fusion center. Once an expert has been fixed, our problem is similar to decentralized detection in the tandem or tree network considered in [20], [26], which however do not allow the experts to have individual biases in their decision making. In [9], the authors consider binary hypothesis testing in a social network, in which agents sequentially observe the opinions of a stochastically generated neighborhood of agents in the network. Each agent has the same 0-1 loss function,\(^1\) and conditions are derived for the Bayesian error probability of the \(n\)th agent to approach to zero as \(n\) becomes large. The network model we consider in this paper is a two-layer hierarchical tree network, which is much simpler compared to that studied in [9]. This is because our goal is to analyze how an expert’s opinion impacts that of a particular agent. In addition, we consider a multihypothesis testing problem in which each agent has different loss functions or even different number of hypotheses.

The reference [28] considers a multihypothesis testing problem in a parallel configuration with a single agent, which is allowed to take actions that affect the observations it makes. The distribution of the observation at each time step depends on the action of the agent in the previous time step, and has bounded Kullback-Liebler divergences under any pair of hypotheses. The agent’s aim is to minimize the Bayesian error exponent. Our work can be viewed as a generalization of a result in [28] (which

\(^1\)The 0-1 loss function assigns a loss of 0 if the agent declares the same hypothesis as the true underlying hypothesis, and a loss of 1 otherwise.
also considers the sequential detection problem that we do not study here) to the case where one of the agent observations is the opinion of an expert, which itself has asymptotically many private observations leading to unbounded Kullback-Liebler divergences for its opinion. Furthermore, we consider general loss functions and the corresponding loss exponent, which is a generalization of the Bayesian error exponent. The characterization of the Bayesian error exponents in multihypothesis testing has also been studied in [29], [30], which derive the achievable error exponents region.

B. Our Contributions

In the following, we summarize our main contributions. Since the decentralized decision making problem is NP-complete [31], our goal is to optimize the loss exponent of agent 0, the absolute value of which is the rate of decay of the expected loss incurred by agent 0 as the number of private observations grows large. We suppose that agent 0 has a decision space consisting of \( M \) hypotheses, and has a choice of \( K \) expert opinions to follow.

(i) We consider a general \( M \)-ary multihypothesis social learning framework in which an agent 0 and every expert have loss functions that depend on the number of private observations the agent or expert has access to. Each expert makes a decision that takes values from a decision space, the size of which may not be the same as \( M \). We characterize the loss exponent of each expert, and provide an asymptotically optimal policy for an expert to achieve its optimal loss exponent (Theorem 4).

(ii) We derive the optimal loss exponent for agent 0 after it has incorporated the opinion of a particular expert, and provide an asymptotically optimal method for agent 0 to choose the best expert to follow (Theorem 8). We also show that the worst loss exponent, up to asymptotic equivalence,\(^2\) for agent 0 is achieved when agent 0 adopts the 0-1 loss function (cf. Remark 2 of Section IV-B).

(iii) We show that if all experts have the same decision space as agent 0, then the best loss exponent, up to asymptotic equivalence, for agent 0 is achieved if it chooses an expert who uses the 0-1 loss function, regardless of the loss function adopted by agent 0 (Proposition 9). This is somewhat surprising as conventional wisdom seems to indicate that agent 0 should choose an expert with similar loss function as itself. On the other hand, under the simplifying assumption that based only on its own private observations, agent 0 discriminates equally well between any pair of hypotheses and has 0-1 loss, we show that if an expert has a decision space with number of states strictly less

\(^2\)See Section II-A for the definition of asymptotic equivalence.
than $M$, then its opinion is ignored by agent 0 (Proposition 10). In this case, additional information is useless, which is again somewhat unexpected.

Our conclusions show that under mild assumptions and up to asymptotic equivalence, the worst loss function for agent 0 is the 0-1 loss, while the best expert to be chosen has 0-1 loss function. This leads to the interesting question of whether there exists an equilibrium in the choice of loss functions in a network of agents who can incorporate opinions from each other. This is however out of the scope of the current work, and will be addressed in our future research.

The rest of this paper is organized as follows. In Section II, we describe our system model, problem formulation and assumptions. We characterize the agents’ loss exponents in Section III, and provide asymptotically optimal policies to achieve the optimal loss exponents. In Section IV, we discuss the optimal expert choice for agent 0, and derive insights into this choice by making simplifying assumptions. We conclude in Section V. Appendix A contains a brief review of the main definitions of large deviation theory, and we defer all proofs to Appendix B. Appendix C contains a characterization of the asymptotic decision regions, which are introduced in Section III.

II. Problem Formulation

In this section, we define our system model, assumptions and some notations. We consider an underlying measurable space $(\Omega, \mathcal{F})$ on which all random variables in this paper are defined. We adopt the following notations throughout this paper. Let $\mathbb{R}$ be the space of real numbers, and let $\langle t, z \rangle$ be the inner product of the vectors $t$ and $z$ in $\mathbb{R}^{M-1}$. For $z \in \mathbb{R}^{M-1}$, let $z^0 = (0, z) \in \mathbb{R}^M$ be the vector augmented with a zero as the first element. The range of integers $a, a+1, \ldots, b$ is denoted as $[a, b]$. The notations $x[1:n]$ and $(x_i)_{i=1}^n$ are used to represent the vectors or sequences $x = (x[1], x[2], \ldots, x[n])$ and $(x_1, x_2, \ldots, x_n)$, respectively. We also make use of $x[i]$ to denote the $i$th element in the vector $x$. We use $B_\epsilon(u)$ to denote an open sphere of radius $\epsilon$ around $u$, in a metric space that will be clear from the context. The notation $\text{Tr}(A)$ denotes the trace of the matrix $A$, $\nabla_t f(t)$ is the gradient of $f(t)$ with respect to (w.r.t.) the vector $t$, and $\nabla^2_t f(t)$ is the Hessian w.r.t. $t$ of the function $f(t)$.

A. Learning from an Expert

Suppose that an agent 0 wishes to determine the underlying hypothesis $H$ associated with a phenomenon of interest, which from agent 0’s frame of reference can be modeled by a set of probability measures $\{\mathbb{P}_m : m = 0, 1, \ldots, M-1\}$ on the space $(\Omega, \mathcal{F})$. Let $\mathbb{E}_m$ be the mathematical expectation under $\mathbb{P}_m$, and let $H = m \in [0, M-1]$ if all agents’ private observations have distributions derived from the probability
measure $P_m$. We suppose that $H = m$ has prior probability $\pi_m \in (0, 1)$. The agent 0 can choose to incorporate the opinion of an expert, chosen from a set of expert agents $\{1, 2, \ldots, K\}$ (see Figure 1). If there is no need to distinguish between agent 0 and the experts, we use the generic term “agent” to refer to either of them.

Each agent $k \in \{0, \ldots, K\}$ has access to a set of private observations $Y_k[1 : n_k] = (Y_k[1], \ldots, Y_k[n_k])$. Conditioned on $H = m$, the conditional distribution of $Y_k[l]$, for $l = 1, \ldots, n_k$, belongs to a set of distributions $\{P_{\gamma m} : \gamma \in \Gamma\}$, where $\Gamma$ is a finite index set. For simplicity, we assume that the set of conditional distributions $\{P_{\gamma m} : \gamma \in \Gamma\}$ is the same for all agents, but our results can be easily generalized to the case where different agents have different sets of conditional distributions in order to model the difference in quality of information that each agent may have access to. We assume that for each $\gamma \in \Gamma$, the probability measures $P_{\gamma m}$ for $m = 0, \ldots, M - 1$, are absolutely continuous w.r.t. each other. We further assume that conditioned on $H$, all the random variables $(Y_k[1 : n_k])_{k=0}^K$ are independent.

To motivate our setup, consider the example alluded to in Section I, where an agent 0 is interested to determine the financial health of a large publicly listed conglomerate. In addition to its own observations $Y_0[1 : n_0]$ about the company through annual financial reports and other publicly available indicators like stock prices, the agent 0 may also choose to incorporate the expert opinion of an investment analyst. Because of limited financial resources (since most analyst reports are not free), agent 0 can only choose to subscribe to one analyst, and has to make an optimal choice of which analyst to use. In most of this paper, for simplicity, we restrict ourselves to the case where the opinion of a single expert is considered in the decision making of agent 0, and show how to generalize our results to a finite set of experts later in Remark 4 of Section IV.

Let $x_k = (x_k[\gamma])_{\gamma \in \Gamma}$ be a vector of non-negative weights summing to one, where $x_k[\gamma]$ is the fraction of private observations of agent $k$ that have conditional distribution $P_{\gamma m}$ when the hypothesis $H = m$. 
We say that \( x_k \) is a policy for agent \( k \). Let \( S(\Gamma) \) be the simplex consisting of all agent policies. Based on its observations, each expert \( k \geq 1 \) makes a decision \( D_k = \gamma_k(Y_k[1 : n_k]) \in [0, d_k - 1] \) by minimizing a local loss criterion. We assume that every expert \( k \geq 1 \) chooses its policy \( x_k \) and decision rule \( \gamma_k \) so that

\[
\mathbb{E}[C_k(H, D_k, n_k)]
\]

is minimized, where \( C_k(m, d, n_k) \) is a non-negative loss incurred if the decision of agent \( k \) is \( d \) when the true hypothesis is \( H = m \). We can think of \( C_k(\cdot, \cdot, \cdot) \) as encoding the “bias” of expert \( k \). We call this the \textit{loss function} of agent \( k \),\(^3\) and

\[
\limsup_{n_k \to \infty} \frac{1}{n_k} \log \mathbb{E}[C_k(H, D_k, n_k)]
\]

the \textit{loss exponent} of agent \( k \). An example of a loss function is the 0-1 loss function: \( C_k(m, m, n_k) = 0 \), and \( C_k(m, d, n_k) = 1 \) for all \( n_k \geq 1 \) if \( d \neq m \), which results in (1) being the Bayesian error probability considered in [32]. In this case, the loss exponent (2) is also known as the Bayesian error exponent.

By allowing the loss function to depend on the number of private observations, we can model various practical applications. For example, the expert \( k \) may itself be part of a decentralized detection network like a tree configuration [21], in which case its goal is to minimize the Bayesian error probability at a fusion center. Then, its loss function decays exponentially fast in the number of private observations \( n_k \). In the publicly listed conglomerate example, suppose that one of the hypotheses corresponds to the conglomerate being financially bankrupt. The loss associated with a missed detection of this hypothesis can be modeled to be exponentially larger than the loss associated with a missed detection of another more benign hypothesis.

In our model, agent \( k \)’s decision space consists of \( d_k \) states, where \( d_k \) may not equal to \( M \), the number of hypotheses that agent 0 is interested in. This allows us to model scenarios where different agents may have different models for the underlying state of the world. Using the conglomerate example described above, agent 0 may be interested in the financial health of the whole conglomerate, while a particular analyst may only be interested in, say the real estate arm, of the conglomerate. Nevertheless, the performance of the real estate arm has a bearing on the overall health of the conglomerate. In particular, expert \( k \) may not distinguish between two hypotheses, say \( M - 2 \) and \( M - 1 \). This can be modeled by assuming that the expert declares the decision \( d = m \) when \( H = m < M - 2 \) and \( d = M - 2 \) if

\(^3\)To be technically correct, \((C_k(\cdot, \cdot, n_k))_{n_k \geq 1}\) should be called a sequence of loss functions, but for simplicity, we will not adopt this terminology.
\( H = M - 2 \) or \( M - 1 \). In this case, we let \( C_k(M - 2, d, n_k) = C_k(M - 1, d, n_k) = 1 \) for all \( d < M - 2 \) and \( C_k(M - 2, M - 2, n_k) = C_k(M - 1, M - 2, n_k) = 0 \). This model is also general enough to model different decision spaces that may not map directly to any of the hypotheses of agent 0.

On the other hand, we note that without loss of generality, there is no need to consider the case where an expert uses a model in which there are more than \( M \) hypotheses, since agent 0 can always choose a \( M \) that is sufficiently large in its model. In addition, since our analysis is based on the frame of reference of agent 0, it has no knowledge of any additional hypotheses, and in practical applications, it simply assumes that expert \( k \) minimizes the expected loss given in (1). We assume that based on the publicized expertise of each expert, agent 0 knows its loss function decay rates, defined in Assumption 1 below.

Let \( D_0(k) = \gamma_0(Y_0[1 : n_0], D_k) \) be the decision made by agent 0 after incorporating the opinion \( D_k \) of agent \( k \). The policy \( x_0 \) and decision rule \( \gamma_0 \) are chosen to minimize

\[
\mathbb{E}[C_0(H, D_0(k), n_0)],
\]

where \( C_0(m, d, n_0) \) is the non-negative loss incurred by agent 0 if it decides in favor of hypothesis \( d \) when the true hypothesis is \( H = m \), and the number of private observations is \( n_0 \). We make the following assumptions regarding the loss function of each agent.

**Assumption 1.**

(i) For all \( n_0 \geq 1 \) and \( m \in [0, M - 1] \), we have \( C_0(m, m, n_0) = 0 \), and for every \( d \in [0, M - 1] \) such that \( d \neq m \), we have \( 0 < C_0(m, d, n_0) \leq 1 \) with

\[
- \lim_{n_0 \to \infty} \frac{1}{n_0} \log C_0(m, d, n_0) = c_0(m, d) \equiv c_0(m) \in [0, \infty).
\]

(ii) For each agent \( k \geq 1 \), we have for every \( m \in [0, M - 1] \) and \( d \in [0, d_k - 1] \),

\[
- \lim_{n_k \to \infty} \frac{1}{n_k} \log C_k(m, d, n_k) = c_k(m, d) \geq 0.
\]

We assume that agent 0 knows the loss decay rates \( c_k(\cdot, \cdot) \), for all \( k = 1, \ldots, K \), based on the publicized expertise of each expert.

We note that Assumption 1 does not impose significant loss in generality. Since the number of agents and decision states are finite, we can normalize \( C_k(m, d, n_k) \) by \( \max_{m,d} C_k(m, d, n_k) \) or \( \sum_d C_k(m, d, n_k) \) so that for each agent \( k \), \( C_k(m, d, n_k) \leq 1 \) for all \( m \) and \( d \). In Assumption 1(i), we assume that \( -(1/n_0) \log C_0(m, d, n_0) \) for all \( d \neq m \) converge to the same \( c_0(m) \). This is because otherwise, easy examples can be constructed in which agent 0 declares \( H = m \) for \( n_0 \) sufficiently large even though the true underlying hypothesis is \( H = m' \neq m \).
Consider two loss functions with loss decay rates $c_k(m,d)$ and $c'_k(m,d)$ respectively. If $c_k(m,d)$ and $c'_k(m,d)$ differ by the same constant for all $m$ and $d$, then their corresponding loss exponents (2) are different but the optimal policies to minimize (2) are the same. We therefore say that two loss functions are asymptotically equivalent if their respective loss decay rates $c_k(m,d)$ and $c'_k(m,d)$ differ by the same constant for all $m$ and $d$. It can be shown that each loss function belongs to an equivalence class, with the equivalence relation being asymptotic equivalence. We call a loss function with $\min_{m,d} c_k(m,d) = 0$ a canonical loss function of its equivalence class. For example, the 0-1 loss function is a canonical loss function. In another example, consider the set of all loss functions $C_k(m,d,n_k)$ with $\sum_d C_k(m,d,n_k) = 1$ for each $m$ and $n_k$, which imposes a total loss of 1 for each hypothesis. Then, it can be shown that each of these loss functions is a canonical loss function.

We say that two loss exponents are asymptotically equivalent if they have the same policy and their loss functions are asymptotically equivalent. For fair comparison of loss exponents, we will always assume canonical loss functions. We are interested to characterize the optimal loss exponent

$$\min_{1 \leq k \leq K} \limsup_{n_0 \to \infty} \frac{1}{n_0} \log \mathbb{E}[C_0(H,D_0(k),n_0)],$$

when the number of private observations of agent 0 becomes large.

We assume that for all experts $k \geq 1$, $q_k = \lim_{n_0 \to \infty} n_k/n_0$ exists as a limit. If the number $n_k$ of private observations of expert $k$ is such that $q_k = 0$, agent 0 will ignore the opinion of expert $k$ as its opinion becomes asymptotically negligible compared to agent 0’s private observations. Therefore, without loss of generality, we assume that $q_k > 0$ for all $k \geq 1$.

**B. Technical Preliminaries**

For any given random variable $X$ with marginal distribution $\mathbb{P}_i^X$ under hypothesis $H = i$ for all $i \geq 0$, we abuse notation by letting $\ell_{ij}(X)$ be the Radon-Nikodym derivative (or likelihood ratio) of $\mathbb{P}_i^X$ w.r.t. $\mathbb{P}_j^X$. Note that $\ell_{ij}(X)$ is a random variable that depends on the distributions and realization of $X$. By convention, we let $\ell_{ii}(X) = 1$ for all values of $i$ and all realizations of $X$. In addition, for simplicity, we let $\ell_{ij}^\gamma = d\mathbb{P}_i^\gamma / d\mathbb{P}_j^\gamma$ be the Radon-Nikodym derivative of $\mathbb{P}_i^\gamma$ w.r.t. $\mathbb{P}_j^\gamma$, for each $\gamma \in \Gamma$.

Let $Z^\gamma = (\log \ell_{m0}^\gamma)_{m=1}^{M-1}$ be a vector of log likelihood ratios, and for a policy $x = (x[\gamma])_{\gamma \in \Gamma}$, let the log moment generating function of $Z^\gamma$ under $H = m$ be

$$\xi_m(\gamma,t) = \log \mathbb{E}_m[\exp(\langle t, Z^\gamma \rangle)],$$

Note that because of Assumption 1, the loss exponent is negative, with a more negative loss exponent corresponding to a faster loss decay rate.
for all $t \in \mathbb{R}^{M-1}$. The weighted log moment generating function is then given by

$$\varphi_m(t, x) = \sum_{\gamma \in \Gamma} x[\gamma] \xi_m(\gamma, t),$$

and its Fenchel-Legendre transform [33] is

$$\Phi^*_m(z, x) = \sup_{t \in \mathbb{R}^{M-1}} \{\langle t, z \rangle - \varphi_m(t, x)\},$$

where $z \in \mathbb{R}^{M-1}$.

In the special case where the vector $t = (t[1], \ldots, t[M - 1])$ has $t[j] = -t[i] = s$ for some $s \in \mathbb{R}$, $i, j \in [1, M - 1]$, and $t[m] = 0$ for all other $m \neq i, j$, we write

$$\Lambda_{ij}(s, x) = \varphi_i(t, x) = \sum_{\gamma \in \Gamma} x[\gamma] \log \mathbb{E}_i \left[ (\ell_{ji}^\gamma)_s \right].$$

A similar definition for $\Lambda_{0j}(s, x)$ applies when $t[j] = s$ and $t[m] = 0$ for all $m \neq j$.

We assume that the distributions $\{\mathbb{P}_m^\gamma : \gamma \in \Gamma\}$ are well-behaved for all $m \in [0, M - 1]$ in the following assumption. This assumption holds, for example in the case where all private observations of the agents take values from a finite set.

**Assumption 2.** For all $m \in [0, M - 1]$ and all $\gamma \in \Gamma$, $\xi_m(\gamma, t) < \infty$ for all $t \in \mathbb{R}^{M-1}$.

The above assumption implies that the second moments of the log moment generating functions of $Z^\gamma$ are bounded, as shown in the following elementary lemma, whose proof is provided in Appendix B.

**Lemma 1.** Suppose that Assumption 2 holds. Then there exists a non-decreasing function $G(r)$, finite for each $r \geq 0$, such that for all $t \in B_r(0)$, and all $\gamma \in \Gamma$, we have

$$0 \leq \text{Tr} (\nabla^2 \xi_m(\gamma, t)) \leq G(r).$$

**III. Optimal Loss Exponents**

In this section, we first characterize the loss exponents of the experts $k = 1, 2, \ldots, K$, which leads to an asymptotically optimal policy for each expert. We then characterize the loss exponent of agent 0, assuming that it is following the opinion of some expert $k$. In the following proofs, the loss exponents are typically characterized by first finding a lower bound for the exponent, and then constructing a policy that achieves the exponent.
A. Loss Exponents of Experts

Consider an expert $k$, where $k = 1, \ldots, K$. By conditioning on the observations $Y_k[1 : n_k]$, it can be shown (Proposition 2.3 of [24]) that the optimal decision rule for expert $k$ is given by

$$D_k = \arg \min_{0 \leq d \leq d_k - 1} \sum_{m=0}^{M-1} \pi_m C_k(m, d, n_k) \ell_{m0}(Y_k[1 : n_k]),$$

where $\ell_{m0}(Y_k[1 : n_k]) = \prod_{l=1}^{n_k} \ell_{m0}(Y_k[l])$. In general, the right hand side of (8) may have multiple minimizers, which complicate our analysis. We therefore make the following simplifying technical assumption.

**Assumption 3.** For every agent $k = 1, \ldots, K$, and for every $n_k$, there exists a unique $d \in [0, d_k - 1]$ such that $D_k$ in (8) equals to $d$ with probability one.

The above assumption is satisfied if the joint probability distribution of $(\ell_{m0})_{m \geq 1}$ is absolutely continuous with respect to Lebesgue measure under any hypothesis $H = 0, \ldots, M - 1$. This can be shown via an easy inductive argument, which shows that $(\ell_{m0}(Y_k[1 : n_k]))_{m \geq 1}$ is in turn absolutely continuous with respect to Lebesgue measure under any hypothesis $H$. The assumption then follows by the same argument in Lemma 5.1 of [13], which we refer the reader to.

If agent 0 follows expert $k$, its loss exponent depends on the loss exponent (2) of the expert $k$, which is in turn related to the probability exponents

$$\limsup_{n_k \to \infty} \frac{1}{n_k} \log \mathbb{P}_m(D_k = d),$$

where $m \in [0, M - 1]$ and $d \in [0, d_k - 1]$. In the following, our aim is to characterize the probability exponents (9). To do that, we let

$$\tilde{f}_k(z, d, n_k) = \frac{1}{n_k} \log \sum_{m=0}^{M-1} \pi_m e^{\log C_k(m, d, n_k) + n_k z[m]},$$

and

$$f_k(z, d, n_k) = \tilde{f}_k(z, d, n_k) - \min_{d' \neq d} \tilde{f}_k(z, d', n_k),$$

where $z = (z[0], z[1], \ldots, z[M - 1]) \in \mathbb{R}^M$. Suppose that agent $k$ adopts the policy $x_{k, n_k}$ when it has access to $n_k$ private observations. Then from (8), we obtain

$$\mathbb{P}_m(D_k = d) = \mathbb{P}_m(f_k(\tilde{Z}_{n_k}^0(x_{k, n_k}), d, n_k) < 0),$$
where
\[
\bar{Z}_{n_k}(x_{k,n_k}) = \left( \frac{1}{n_k} \log \ell_{m0}(Y_k[1:n_k]) \right)_{m=1}^{M-1}
\] (11)
is a vector in \( \mathbb{R}^M \) of log likelihood ratios, and \( \bar{Z}_{n_k}^0(x_{k,n_k}) = (0, \bar{Z}_{n_k}(x_{k,n_k})) \). In the following, we show that \( f_k(z, d, n_k) \) converges uniformly in \( z \) to
\[
f_k(z, d) = \tilde{f}_k(z, d) - \min_{d' \neq d} \tilde{f}_k(z, d'),
\] (12)
where
\[
\tilde{f}_k(z, d) = \max_{0 \leq m \leq M-1} \{ z[m] - c_k(m, d) \}.\] (13)

The following two lemmas’ proofs are given in Appendix B. The reader is also referred to Appendix A for a review of the definition of a large deviation principle (LDP).

**Lemma 2.** Suppose that Assumption 1 holds. For all \( k = 1, \ldots, K \), and all \( d \in [0, d_k - 1] \), \( f_k(z, d, n) \to \tilde{f}_k(z, d) \) and \( f_k(z, d, n) \to f_k(z, d) \) uniformly in \( z \in \mathbb{R}^M \) as \( n \to \infty \).

**Lemma 3.** Suppose that Assumption 2 holds, agent \( k \) adopts the policy \( x_{k,n_k} = (x_{k,n_k}[\gamma])_{\gamma \in \Gamma} \) when it has access to \( n_k \) private observations, and \( x_{k,n_k} \to x_k \) as \( n_k \to \infty \). Then, the sequence of random variables \( (\bar{Z}_{n_k}(x_{k,n_k}))_{n_k \geq 1} \) defined in (11) satisfies a LDP under hypothesis \( H = m \), for every \( m \in [0, M - 1] \), with good rate function \( \Phi_m^*(\cdot, x_k) \).

We are now ready to present our first main result. For \( k \geq 1 \), let
\[
A_k(d) = \left\{ z \in \mathbb{R}^{M-1} : f_k(z^0, d) < 0, \text{ where } z^0 = (0, z) \right\},
\] (14)
and
\[
I_k(x) = \min_{0 \leq m \leq M-1} \left\{ \min_{z \in A_k(d)} \Phi_m^*(z, x) + c_k(m, d) \right\}.
\] (15)
The sets \( A_k(d) \) can be interpreted as the asymptotic decision region for expert \( k \) to declare decision \( d \) based on its sufficient statistics \( \bar{Z}_{n_k}(x_{k,n_k}) \); see Section IV-A for a discussion. We further note that the minimization over \( z \in A_k(d) \) in the definition of \( I_k(x) \) in (15) is not an infimum because \( \Phi_m^*(z, x) \) has compact level sets since it is a good rate function (see Appendix A).

**Theorem 4.** Suppose that Assumptions 1, 2, and 3 hold. Consider an agent \( k \in \{1, \ldots, K\} \) who adopts a sequence of optimal policies \( (x_{k,n_k})_{n_k \geq 1} \) that minimizes (1) for each \( n_k \). Then, the following holds.
(i) The loss exponent of agent $k$ is given by
\[
\lim_{n_k \to \infty} \frac{1}{n_k} \log \mathbb{E} [C_k(H, D_k, n_k)] = - \max_{x \in \mathbb{S}(\Gamma)} I_k(x).
\]

(ii) There is no loss in optimality asymptotically, if we restrict the sequence of policies $(x_{k,n_k})_{n_k \geq 1}$ such that $\lim_{n_k \to \infty} x_{k,n_k} = x_k^*$, for some $x_k^* \in \arg \max_{x \in \mathbb{S}(\Gamma)} I_k(x)$.

(iii) Let $x_k^* \in \arg \max_{x \in \mathbb{S}(\Gamma)} I_k(x)$. For every $m \in [0, M-1]$, and $d \in [0, d_k-1]$, we have
\[
\lim_{n_k \to \infty} \frac{1}{n_k} \log \mathbb{P}_m(D_k = d) = - \min_{z \in \mathcal{A}_k(d)} \Phi_m(z, x_k^*).
\]

Proof: See Appendix B.

Theorem 4 allows each expert to find an asymptotically optimal policy by maximizing (15) (see however Remark 1 in Section IV-B). In addition, we will see later that Theorem 4(iii) allows agent 0 to characterize its own loss exponent and thus optimally choose the expert to follow.

B. Loss Exponent of Agent 0

We first present a generalization of Theorem 5 of [34] (see also [23] for a slightly more updated version). The proof steps are similar to that in [34] and [23], and are provided in Appendix B for completeness.

**Proposition 5.** Suppose that Assumptions 1 and 2 hold, and agent 0 adopts the opinion of agent $k \geq 1$ and policy $x_0$. Let
\[
s_{ij}^* = \arg\max_{s \in [0,1]} \left\{ s \log \frac{C_0(i,j,n_0)\mathbb{P}_i(D_k = d)}{C_0(j,i,n_0)\mathbb{P}_j(D_k = d)} - \Lambda_{ij}(s, x_0) \right\}
\]
where $\Lambda_{ij}(\cdot, \cdot)$ is as defined in (6). Then, for any $\epsilon > 0$, and any $d \in [0, d_k-1]$, there exists $n$ such that for all $n_0 \geq n$, we have for all $i \neq j$,
\[
\frac{1}{n_0} \log \left\{ \min_{j' \neq i} C_0(i,j',n_0)\mathbb{P}_i(D_0(k) \neq i, D_k = d) \right. \\
+ \left. C_0(j,i,n_0)\mathbb{P}_j(D_0(k) = i, D_k = d) \right\} \\
\geq (1 - s_{ij}^*) \left( \frac{1}{n_0} \log \mathbb{P}_i(D_k = d) - c_0(i) \right) \\
+ s_{ij}^* \left( \frac{1}{n_0} \log \mathbb{P}_j(D_k = d) - c_0(j) \right) + \Lambda_{ij}(s_{ij}^*, x_0) - \epsilon. \tag{16}
\]

In the following, we characterize the loss exponent of agent 0 if it chooses expert $k$. Recall that $q_k = \lim_{n_0 \to \infty} n_k/n_0$. 

**Theorem 5.**
Theorem 6. Suppose that Assumptions 1, 2, and 3 hold. Suppose that agent 0 adopts the opinion of expert $k \geq 1$, which has the asymptotically optimal policy $x^*_k$. Then, the loss exponent of agent 0 is
\[
\lim_{n_0 \to \infty} \frac{1}{n_0} \log \mathbb{E}[C_0(H, D_0(k), n_0)] = -\max_{x_0 \in \mathbb{D}(1)} \mathcal{E}_0(k, x_0),
\]
where
\[
\mathcal{E}_0(k, x_0) = \min_{0 \leq d \leq d_k - 1} \max_{s \in [0,1]} \left\{ (1 - s) \left( q_k \min_{z \in A_k(d)} \Phi^*_m(z, x^*_k) + c_0(i) \right) + s \left( q_k \min_{z \in A_k(d)} \Phi^*_m(z, x^*_k) + c_0(j) \right) - \Lambda_{ij}(s, x_0) \right\}.
\]

Proof: See Appendix B.

IV. OPTIMAL CHOICE OF EXPERT

In this section, we address the question of how to choose an optimal expert for agent 0 to follow. In order to interpret our results, we first characterize the regions $A_k(d)$ in (14) as asymptotic decision regions for expert $k$, and then revisit Theorem 6 to derive the optimal loss exponent for agent 0. Finally, we make additional simplifying assumptions in order to provide more insights into our results.

A. Asymptotic Decision Regions

Suppose that agent $k \geq 1$ adopts the sequence of policies $x_{k,n_k} \to x^*_k$ as $n_k \to \infty$. From (10), for large $n_k$, we have for each $m \in [0, M - 1]$, and $d \in [0, d_k - 1]$,
\[
\frac{1}{n_k} \log \mathbb{P}_m(D_k = d) \approx \frac{1}{n_k} \log \mathbb{P}_m(f_k(\tilde{Z}_{n_k}(x_{k,n_k}), d) < 0) \\
\approx - \min_{z \in A_k(d)} \Phi^*_m(z, x^*_k),
\]
where $A_k(d)$ is as defined in (14), and $\Phi^*_m(z, x^*_k)$ is a rate function. In the same spirit as the Gärtner-Ellis Theorem [33], we can interpret the sets $A_k(d)$ as the asymptotic decision region for deciding $D_k = d$ based on $\tilde{Z}_{n_k}(x_{k,n_k})$ as $n_k \to \infty$. Clearly, the regions $A_k(d)$, $d = 0, \ldots, d_k - 1$, are convex, otherwise one can use a randomization argument to produce a decision rule that contradicts Assumption 3. In the following, we list some properties of the rate functions $\Phi^*_m(\cdot, \cdot)$ that will be useful in helping us to

\[5\]To avoid cluttered notations, we let $\min_{i \neq j}$ and $\max_{i \neq j}$ be the minimization or maximization over all unordered pairs $(i, j) \in [0, M - 1]^2$ such that $i \neq j$, respectively.
interpret our results. We also refer the reader to Appendix C for a characterization of \( A_k(d) \) in terms of the union of intersections of multiple half-spaces.

**Lemma 7.** For every \( m \in [0, M - 1] \), \( \Phi^*_m(z, x) \) is non-negative, and is convex in \( z \) and concave in \( x \). Furthermore, for any \( x \), \( \min_z \Phi^*_m(z, x) = 0 \), and the minimum is achieved at \( \bar{z}_m(x) = \sum_{\gamma \in \Gamma} x[\gamma] \mathbb{E}_m[Z^{\gamma}] \), where \( Z^{\gamma} = (\log \ell_{i_0}^{\gamma})_{i=1}^{M-1} \).

**Proof:** See Appendix B.

**B. Choosing an Expert**

If agent 0 does not incorporate the opinion of any expert, we use the notation \( \mathcal{E}_0(0, x_0) \) to denote its absolute loss exponent when using policy \( x_0 \). From Theorem 6, by setting \( q_k = 0 \), we have

\[
\mathcal{E}_0(0, x_0) = \min_{i \neq j} \max_{s \in [0,1]} \{(1 - s)c_0(i) + sc_0(j) - \Lambda_{ij}(s, x_0)\}. \tag{18}
\]

In the case of minimizing the Bayesian error probability at agent 0, without the help of any experts, we can further set \( c_0(i) = c_0(j) = 0 \) to obtain the absolute Bayesian error exponent

\[
\mathcal{E}_{0,B}(0, x_0) = -\max_{i \neq j} \min_{s \in [0,1]} \Lambda_{ij}(s, x_0), \tag{19}
\]

which recovers the result in [32]. Similarly, if agent 0 chooses expert \( k \), and adopts the 0-1 loss function, we let its absolute loss exponent be

\[
\mathcal{E}_{0,B}(k, x_0) = \min_{i \neq j} \max_{0 \leq d \leq d_k - 1} \left\{(1 - s)q_k \min_{z \in A_k(d)} \Phi^*_i(z, x^*_k) + sq_k \min_{z \in A_k(d)} \Phi^*_j(z, x^*_k) - \Lambda_{ij}(s, x_0)\right\}. \tag{20}
\]

The following proposition follows immediately from Theorem 6, and provides a method for agent 0 to optimally choose which expert to follow.

**Theorem 8.** Suppose that Assumptions 1, 2, and 3 hold. The optimal loss exponent of agent 0 is

\[
\min_{1 \leq k \leq K} \lim_{n_0 \to \infty} \frac{1}{n_0} \log \mathbb{E} \left[C_0(H, D_0(k), n_0)\right] = -\max_{1 \leq k \leq K} \max_{x_0 \in \mathbb{S}(\Gamma)} \mathcal{E}_0(k, x_0).
\]

Furthermore, for any \( k \geq 1 \) and policy \( x_0 \), we have

\[
\mathcal{E}_0(k, x_0) \geq \mathcal{E}_{0,B}(k, x_0) \tag{21}
\]

\[
\mathcal{E}_0(k, x_0) \geq \mathcal{E}_0(0, x_0) \geq \mathcal{E}_{0,B}(0, x_0). \tag{22}
\]

**Proof:** The first part of the theorem is a direct consequence of Theorem 6. From Lemma 7, we have

\[
\min_{z \in A_k(d)} \Phi^*_i(z, x^*_k) \geq 0 \text{ for any } i \in [0, M - 1], k \geq 1, d \in [0, d_k - 1], \text{ and policy } x^*_k. \]
Assumption 1 implies that $c_0(i) \geq 0$ for all $i \in [0, M - 1]$. Therefore, the inequalities (21) and (22) follow from (17), (18), (19), and (20), and the proof is complete.

**Remark 1.** Theorem 8 shows that agent 0 should choose an expert $k$ that maximizes $\max_{x_0} E_0(k, x_0)$, which depends on $\min_{z \in A_k(d)} \Phi_m^*(z, x_k^*)$, for $m \in [0, M - 1]$ and $d \in [0, d_k - 1]$. The regions $(A_k(d))_d$ are defined by the loss function decay rates $c_k(m, d)$ for all $m$ and $d$, which can be determined by agent 0 if $c_k(\cdot, \cdot)$ is known publicly (see Appendix C). The rate function $\Phi_m^*(z, x_k^*)$ is determined solely by the statistics of agent $k$’s private observations, and the policy $x_k^*$ it adopts. In some situations, the expert may disclose the composition of its information sources $x_k^*$ directly. If it is unknown, it can be found by maximizing $I_k(x)$ in (15). This however may not be an easy numerical procedure [32] if $M$ and $d_k$ are large. Assuming that all infima and suprema are achievable, a possible method is the following alternating optimization approach:

**Step 1.** Initial guesses $x_k^*(0)$ and $z_{m,d}(0)$ for every $m \in [0, M - 1]$ and $d \in [0, d_k - 1]$ are made. Set $l = 1$.

**Step 2.** For each $m \in [0, M - 1]$ and $d \in [0, d_k - 1]$, find

$$t_{m,d}(l) = \arg \max_{t \in \mathbb{R}^{M-1}} \{ \langle t, z_{m,d}(l-1) \rangle - \varphi_m(t, x_k^*(l-1)) \}.$$ 

**Step 3.** For each $m \in [0, M - 1]$ and $d \in [0, d_k - 1]$, find

$$z_{m,d}(l) = \arg \min_{z \in A_k(d)} \langle t_{m,d}(l), z \rangle.$$ 

From Appendix C, we see that solving the above optimization problem is equivalent to solving a collection of linear programs.

**Step 4.** Find $x_k^*$ by solving the linear program

$$\max_{r, x = (x[\gamma])_{\gamma \in \Gamma}} r$$

subject to

$$r \leq \langle t_{m,d}(l), z_{m,d}(l) \rangle$$

$$- \sum_{\gamma \in \Gamma} x[\gamma] \xi_m(\gamma, t_{m,d}(l)) + c_k(m, d),$$

$$\forall m \in [0, M - 1], \forall d \in [0, d_k - 1],$$

$$x[\gamma] \geq 0, \forall \gamma \in \Gamma,$$
\[
\sum_{\gamma \in \Gamma} x[\gamma] = 1.
\]

Step 5. Set \(l = l + 1\) and repeat Steps 2-4.

Unfortunately, there is no guarantee that the above procedure converges to the correct solution. Moreover, Steps 3 and 4 above become unwieldy when \(M\) or \(d_k\) are large. In the next subsection, we consider special cases that have simplified solutions.

Remark 2. The first inequality in \((22)\) shows that there is no loss in optimality for agent 0 to incorporate the opinion of any expert, verifying the adage that there is no harm in having more information. However, incorporating additional information does not necessarily improve its loss exponent; see the following Remark 3 and Proposition 10.

The inequality in \((21)\) shows that of all the loss functions satisfying Assumption 1(i), the 0-1 loss is the worst canonical loss function for agent 0. In particular, consider the set of loss functions with \(C_0(m, d, n_0) = C_0(m, n_0)\) for all \(m \in [0, M-1]\) and \(d \neq m\), where the limit \(c_0(m) = -\lim_{n_0 \to \infty} (1/n_0) \log C_0(m, n_0)\) exists. Furthermore, each loss function has a total cost constraint

\[
\sum_{m=0}^{M-1} C_0(m, n_0) = M.
\]

It can be shown that all such loss functions satisfy Assumption 1(i), and are canonical loss functions. The 0-1 loss function belongs to this set, and divides the total cost equally among all types of missed detections, which results in the worst loss exponent for agent 0. This can be explained intuitively by observing that if there exists a \(m\) such that \(c_0(m) > 0\), then missed detection of \(H = m\) incurs an exponentially decaying loss, so that the hypothesis \(m\) can effectively be ignored. Agent 0 then effectively has a smaller set of hypotheses, leading to a lower expected loss.

Remark 3. Suppose that agent 0 adopts the policy \(x_0\), and incorporates the opinion of agent \(k \geq 1\) with corresponding asymptotically optimal policy \(x^*_k\). Suppose further that \((i^*, j^*)\) is such that

\[
\max_{s \in [0,1]} \{(1-s)c_0(i^*) + sc_0(j^*) - \Lambda_{ij}(s, x_0)\} = \mathcal{E}_0(0, x_0),
\]

i.e., the loss of agent 0 is dominated by its error in distinguishing between hypotheses \(i^*\) and \(j^*\). Then, equality in the first inequality of \((22)\) holds if there exists a \(d \in [0, d_k-1]\) such that \(\min_{z \in A_k(d)} \Phi_{i^*}^*(z, x^*_k) = \min_{z \in A_k(d)} \Phi_{j^*}^*(z, x^*_k) = 0\). The asymptotic acceptance region corresponding to \(d\) thus includes the minima of the rate functions corresponding to hypotheses \(i^*\) and \(j^*\). This implies that expert \(k\) cannot distinguish between these two hypotheses with probability of error going to zero exponentially fast in the number of private observations, and is thus useless to agent 0, as verified by Theorem 6.
Remark 4. It is easy to generalize Theorem 6 or Theorem 8 to the case where more than one agent can be chosen. In particular, if all $K$ experts’ opinions are adopted by agent 0, Theorem 6 holds with absolute loss exponent

$$\min_{i \neq j} \max_{0 \leq d \leq d_k - 1} \left\{ \left( 1 - s \right) \left( \sum_{k=1}^{K} q_k \min_{z \in A_k(d)} \Phi_i^*(z, x_k^*) + c_0(i) \right) + s \left( \sum_{k=1}^{K} q_k \min_{z \in A_k(d)} \Phi_j^*(z, x_k^*) + c_0(j) \right) - \Lambda_{ij}(s, x_0) \right\}.$$ 

In the following subsection, we consider the special cases where the size of each expert’s decision space is less than or equal to $M$ under additional simplifying assumptions.

C. Special Cases

We first consider the case where experts have the same decision space as agent 0, and has zero loss if they decide on the true underlying hypothesis. Surprisingly, the following result shows that agent 0 should choose an expert who utilizes the 0-1 loss (if any), instead of an expert who adopts the same canonical loss function as itself. This indicates that agent 0 receives more information from an “unbiased” expert than one who has the same “bias” as itself. Assuming that independent news agencies are in general “unbiased,” our result suggests that such news agencies are unlikely to be replaced by social news reporting or social blogs in a rational world. In designing a cooperative IoT network, each sensor should be designed to make two inferences: one based on its local loss functions, and another based on the 0-1 loss function, which is then broadcast to its neighbors.

Proposition 9. Suppose that Assumptions 1, 2, and 3 hold, and every agent $k \geq 1$ has $d_k = M$, $C_k(m, m, n_k) = 0$, and $C_k(m, d, n_k) > 0$ for all $m, d \in [0, M - 1]$ with $d \neq m$. Suppose that agent 0 follows the opinion of expert $k \geq 1$, who adopts the policy $x_k^*$. Then, the loss exponent of agent 0 satisfies

$$\lim_{n_0 \to \infty} \frac{1}{n_0} \log \mathbb{E} \left[ C_0(H, D_0(k), n_0) \right] \geq - \max_{x_0 \in \mathcal{S}(\Gamma)} \hat{\mathcal{E}}_0(k, x_0),$$

where

$$\hat{\mathcal{E}}_0(k, x_0) = \min_{i,j} \max_{t \in [0,1]} \left\{ (1 - s)c_0(i) + sc_0(j) - sq_k \min_{t \in [0,1]} \Lambda_{ij}(t, x_k^*) - \Lambda_{ij}(s, x_0) \right\}.$$ 

Furthermore, equality holds in the inequality (23) if the loss function of expert $k$ is asymptotically equivalent to the 0-1 loss function.

$^6$The notation $\min_{i,j} \min_{x \in [0,1]}$ means minimization over all ordered pairs $(i, j)$ with $i, j \in [0, M - 1]$ and $i \neq j.$
Proof: See Appendix B.

In the following, we make two additional assumptions in order to simplify the loss exponent of agent 0.

**Assumption 4.** Agent 0 adopts the 0-1 loss criterion, i.e., $c_0(m) = 0$ for all $m \in [0, M - 1]$. Furthermore, agent 0 adopts the policy $x_0$ such that for any pair of distinct hypotheses $(i, j)$, $\Lambda_{ij}(\cdot, x_0) = \Lambda(\cdot, x_0)$.

**Assumption 5.** Every agent $k \geq 1$ has $d_k = M$, with loss decay rates

$$c_k(m, d) = \begin{cases} c_k(m) & \text{if } d \neq m, \\ \infty & \text{if } d = m, \end{cases}$$

where $c_k(m) \geq 0$ for all $m \in [0, M - 1]$.

**Proposition 10.** Suppose that Assumptions 1 - 3 hold.

(i) Suppose further that Assumption 4 holds. For any agent $k \geq 1$, if $d_k < M$, we have $\mathcal{E}_0(k, x_0) = \mathcal{E}_{0,B}(0, x_0)$, i.e., agent 0 ignores the opinion of agent $k$.

(ii) Suppose further that Assumption 5 holds. Then, the loss exponent of agent 0 is

$$\tilde{\mathcal{E}}_0(k, x_0) = \min_{1 \leq k \leq K} \max_{i, j: i \neq j, s \in [0,1]} \{(1 - s)c_0(i) + sc_0(j) - sq_k\Lambda_{ij}(s_{ij}, x^*_k) - \Lambda(s, x_0)\},$$

and

$$(x^*_k, s_{ij}) = \arg\max_{2^k \in \mathfrak{F}, s \in [0,1]} \{(1 - s)c_k(i) + sc_k(j) - \Lambda_{ij}(s, x_k^*)\}.$$

Proof: See Appendix B.

Proposition 10(i) can be explained intuitively as follows. If an expert $k$ has $d_k < M$, then there exists two hypotheses $i$ and $j$ of agent 0 that it does not discriminate between. The probability of making an error between these two hypotheses by the expert is therefore bounded away from zero, and the expert’s opinion does not help agent 0 to differentiate between hypotheses $i$ and $j$. The error probability of agent 0 declaring $H = i$ when the true hypothesis is $H = j$ or vice versa, thus dominates and is the same as when agent 0 ignores the expert’s opinion.

**V. Conclusion**

We have studied a multihypothesis social learning problem in which an agent makes a decision with the help of a chosen expert. We have considered a general framework that allows the agent and experts to have different loss functions (biases), and different decision spaces. We have characterized the loss
exponent of the agent in terms of the chosen expert’s probability error exponent and loss exponent, which
allows us to choose the asymptotically optimal expert as well as the agent’s policy. We have shown that if
the experts have the same decision space as the agent, then the agent should choose an expert with a loss
function asymptotically equivalent to the 0-1 loss function, regardless of the agent’s own loss function.
Moreover, the worst canonical loss function for the agent is the 0-1 loss function. In this case, if the
agent’s error exponents for discriminating between any pair of hypotheses are the same, then it ignores
any expert with a decision space smaller than the number of hypotheses.

In this paper, we have assumed that the agent knows the experts’ loss decay rates, which may not
be valid in some practical scenarios. For example, expert opinions may very well depend on the mood
of the expert at the time the opinion is publicized. Therefore, it is of interest to consider minimax loss
exponents in which the expert’s loss function or observation probability distributions are drawn from
uncertainty classes. Minimax decentralized hypothesis testing has been studied in [35], and in our recent
work [36] in which we consider robust social learning in a tandem network. However, additional research
is required to address more complex network architectures and applications like that considered in this
paper.

**APPENDIX A**

**MATHEMATICAL PRELIMINARIES**

In this appendix, we briefly review some basic definitions and a result from large deviation theory.
The reader is referred to [33], [37] for details.

Let $\mathcal{X}$ be a Polish space. The function $\Phi : \mathcal{X} \rightarrow [0, \infty]$ is said to be a good rate function if $\Phi$ is
lower semicontinuous and has compact level sets $\{x \in \mathcal{X} : \Phi(x) \leq c\}$ for all $c \geq 0$. Let $(P_n)_{n \geq 1}$ be a
sequence of probability measures on $\mathcal{X}$. This sequence of probability measures is said to satisfy a
large deviation principle (LDP) if for some good rate function $\Phi$, and for all closed sets $C \subset \mathcal{X}$, we have

$$\limsup_{n \to \infty} \frac{1}{n} \log P_n(C) \leq - \inf_{x \in C} \Phi(x),$$

and for all open sets $O \subset \mathcal{X}$, we have

$$\liminf_{n \to \infty} \frac{1}{n} \log P_n(O) \geq - \inf_{x \in O} \Phi(x).$$

A sequence of random variables $(Z_n)_{n \geq 1}$ is said to satisfy a LDP if the sequence of marginal distributions
$P_n(\cdot) = P(Z_n \in \cdot)$ satisfies a LDP. The celebrated Gärtner-Ellis Theorem [33] provides sufficient
conditions for a sequence of random variables to satisfy a LDP.
When not all the conditions required by the Gärtner-Ellis Theorem hold (as is the case in some of our proofs), we instead make use of a uniform lower bound, which is a generalization of Theorem 1.3.13 of [37]. The proof is omitted here for brevity.

**Lemma A.1.** Let $Z_1, Z_2, \ldots, Z_n$ be independent $\mathbb{R}^M$-valued random variables, with $x[\gamma]$ fraction of them having distribution $\mathbb{P}^\gamma$, for each $\gamma \in \Gamma$. Let $Z^\gamma$ have distribution $\mathbb{P}^\gamma$.

$$
\varphi(t) = \sum_{\gamma \in \Gamma} x[\gamma] \log \mathbb{E} [\exp(\langle t, Z^\gamma \rangle)],
$$

and

$$
\Phi^*(z) = \sup_t \{\langle t, z \rangle - \varphi(t)\}.
$$

Suppose that $\mathbb{E} [\exp(\langle t, Z^\gamma \rangle)] < \infty$ for all $t \in \mathbb{R}^M$, then for $t_z$ such that $\nabla_t \varphi(t_z) = z$, and any $\epsilon > 0$, we have

$$
\frac{1}{n} \log \mathbb{P}\left( \frac{1}{n} \sum_{l=1}^{n} Z_l \in B_\epsilon(z) \right) \geq -\Phi^*(z) - \|t_z\|\epsilon + \frac{1}{n} \log \left( 1 - \frac{1}{ne^2} \text{Tr}\left( \nabla_t^2 \varphi(t_z) \right) \right).
$$

**APPENDIX B**

**Proofs of Main Results**

**A. Proof of Lemma 1**

The lower inequality in (7) holds because $\xi_m(\gamma, t)$ is convex in $t$ for each $\gamma$ (see Lemma 2.2.31 of [33]). Furthermore, we can define

$$
G(r) = \max_{t \in B_r(0)} \text{Tr}\left( \nabla_t^2 \xi_m(\gamma, t) \right),
$$

which is finite because of Assumption 2, and the fact that $\text{Tr}\left( \nabla_t^2 \xi_m(\gamma, t) \right)$ is continuous on the compact set $B_r(0)$. The proof of the lemma is now complete.

**B. Proof of Lemma 2**

Let $\epsilon$ be a positive number. From Assumption 1(ii), for $n$ sufficiently large, we have $(\log M)/n \leq \epsilon/2$, $\min_m (\log \pi_m)/n \geq -\epsilon/2$, and

$$
\left| \frac{1}{n} \log C_k(m, d, n) + c_k(m, d) \right| \leq \frac{\epsilon}{2},
$$
for all \( m \in [0, M - 1] \). This implies that for all \( z \in \mathbb{R}^M \), we have
\[
\tilde{f}_k(z, d, n) \leq \max_m \left\{ \frac{1}{n} \log C_k(m, d, n) + z[m] \right\} + \frac{1}{n} \log M
\]
\[
\leq \max_m \left\{ -c_k(m, d) + z[m] \right\} + \epsilon
\]
\[
= \tilde{f}_k(z, d) + \epsilon,
\]
and
\[
\tilde{f}_k(z, d, n) \geq \max_m \left\{ \frac{1}{n} \log C_k(m, d, n) + z[m] \right\} + \min_m \left\{ \frac{1}{n} \log \pi_m \right\} + \min_m \left\{ \frac{1}{n} \log \pi_m \right\} \log \pi_m
\]
\[
\geq \max_m \left\{ -c_k(m, d) + z[m] \right\} - \epsilon
\]
\[
= \tilde{f}_k(z, d) - \epsilon,
\]
which shows that \( \tilde{f}_k(z, d, n) \to \tilde{f}_k(z, d) \) uniformly in \( z \). Since the minimum of a finite set of uniformly convergent functions is also uniformly convergent, the lemma follows.

C. Proof of Lemma 3

We apply the Gärtner-Ellis Theorem [33] to prove the lemma. Let \( Z_i = (\log \ell_{m0}(Y_k[i]))_{m=1}^{M-1} \). We have
\[
n_k Z_{n_k}(x_k, n_k) = \sum_{i=1}^{n_k} Z_i \quad \text{since} \ Z_i; i = 1 \ldots, n_k \ \text{are independent.}
\]
For every \( t \in \mathbb{R}^{M-1} \), we obtain
\[
\frac{1}{n_k} \log \mathbb{E}_m \left[ \exp \left( \langle n_k t, Z_{n_k}(x_k, n_k) \rangle \right) \right]
\]
\[
= \frac{1}{n_k} \log \mathbb{E}_m \left[ \exp \left( \langle t, \sum_{i=1}^{n_k} Z_i \rangle \right) \right]
\]
\[
= \frac{1}{n_k} \sum_{i=1}^{n_k} \log \mathbb{E}_m \left[ \exp \left( \langle t, Z_i \rangle \right) \right]
\]
\[
= \sum_{\gamma \in \Gamma} x_{k,n_k}[\gamma] \log \mathbb{E}_m \left[ \exp \left( \langle t, (\log \ell_{m0})^{M-1} \rangle \right) \right]
\]
\[
\to \varphi_m(t, x_k),
\]
as \( n_k \to \infty \). From Assumption 2 and Lemma 2.3.9 of [33], we have \( \Phi^*_m(z, x_k) \) is a good rate function, and the lemma follows from the Gärtner-Ellis Theorem.
D. Proof of Theorem 4

We prove claim (i) by first deriving a lower bound for the loss exponent of agent $k$, and showing that this bound is achievable. We have for any $\epsilon > 0$, and $n_k$ sufficiently large,

\[
\frac{1}{n_k} \log \mathbb{E} \left[ C_k(H, D_k, n_k) \right] = \frac{1}{n_k} \log \sum_{m=0}^{M-1} \sum_{d=0}^{d_k-1} \pi_m C_k(m, d, n_k) \mathbb{P}_m(D_k = d) \geq \max_{0 \leq m \leq M-1} \left\{ \frac{1}{n_k} \log C_k(m, d, n_k) + \frac{1}{n_k} \log \mathbb{P}_m(D_k = d) \right\} + \min_{m} \frac{1}{n_k} \log \pi_m \geq \max_{0 \leq m \leq M-1} \left\{ \frac{1}{n_k} \log \mathbb{P}_m(D_k = d) - c_k(m, d) \right\} - \epsilon,
\]

which can be further lower bounded by lower bounds on the probability exponents. For each $z \in \mathbb{R}^{M-1}$, let $t_z$ be the solution to the equation

\[
\nabla t \varphi_i(t, x_k, n_k) = z,
\]

if the solution exists. For each $r > 0$, let $H_r = \{ z \in \mathbb{R}^{M-1} : \| t_z \| \leq r \}$, and $A_k(d, r, \epsilon) = \{ z \in \mathbb{R}^{M-1} : f_k(z^0, d) < -\epsilon/2, \text{ where } z^0 = (0, z) \} \cap H_r$. From (10) and (11), we then have for any $u < -\epsilon$,

\[
\frac{1}{n_k} \log \mathbb{P}_m(D_k = d) = \frac{1}{n_k} \log \mathbb{P}_m(f_k(\tilde{Z}_{n_k}^0(x_k, n_k), d, n_k) < 0) \geq \frac{1}{n_k} \log \mathbb{P}_m(f_k(\tilde{Z}_{n_k}^0(x_k, n_k), d, n_k) \in B_\epsilon(u)) \geq \frac{1}{n_k} \log \mathbb{P}_m(f_k(\tilde{Z}_{n_k}^0(x_k, n_k), d) \in B_{\epsilon/2}(u))
\]

where the last inequality follows from Lemma 2 for $n_k$ sufficiently large. Since $u < -\epsilon$ is arbitrary, we obtain for $n_k$ sufficiently large,

\[
\frac{1}{n_k} \log \mathbb{P}_m(D_k = d) \geq \sup_{z : B_\epsilon(z) \subseteq A_k(d, r, \epsilon)} \frac{1}{n_k} \log \mathbb{P}_m(\tilde{Z}_{n_k}(x_k, n_k) \in B_\epsilon(z)).
\]
From Lemma A.1, we can further lower bound the right hand side of (26) to obtain

\[
\frac{1}{n_k} \log \mathbb{P}_m(D_k = d) \\
\geq - \inf_{z: B(z) \subset A_k(d, r, \epsilon)} \left\{ \Phi^*_m(z, x_k, n_k) + \|t_z\| \epsilon \\
- \frac{1}{n_k} \log \left( 1 - \frac{1}{n_k \epsilon^2} \text{Tr} \left( \nabla^2 \phi_i(t_z, x_k, n_k) \right) \right) \right\}
\]

\[
\geq - \inf_{z: B(z) \subset A_k(d, r, \epsilon)} \Phi^*_m(z, x_k, n_k) + r \epsilon \\
- \frac{1}{n_k} \log \left( 1 - \frac{1}{n_k \epsilon^2} G(r) \right),
\]

(27)

where the last inequality follows from Lemma 1. Combining (27) with (24), and letting \(n_k \to \infty\), and then taking \(r = 1/\sqrt{\epsilon}\) and \(\epsilon \to 0\), we have

\[
\lim inf_{n_k \to \infty} \frac{1}{n_k} \log \mathbb{E}[C_k(H, D_k, n_k)] \geq - \sup_{x \in \mathbb{S}(\Gamma)} I_k(x).
\]

(28)

Since \(\Phi^*_m(z, x)\) is continuous in \(x\), \(\min_{z \in A_k(d)} \Phi^*_m(z, x)\) is upper semi-continuous in \(x\), and \(I_k(x)\) is upper semi-continuous in \(x\). In addition, \(\mathbb{S}(\Gamma)\) is compact, therefore the supremum over \(x\) on the right hand side of (28) is a maximization. Consider the policy \(x_k^* = \arg \max_{x \in \mathbb{S}(\Gamma)} I_k(x)\). For each \(n_k \geq 1\), let agent \(k\) use the policy \(x_{k, n_k}\) where \(x_{k, n_k}[\gamma] = \lfloor x_k^*[\gamma] n_k \rfloor / n_k\) for all \(\gamma \in \Gamma\), and if \(x_{k, n_k}[\gamma]\) do not sum to 1 over \(\gamma \in \Gamma\), we simply choose the remaining private observations from an arbitrary distribution, and ignore them when making the decision for agent \(k\). We have \(x_{k, n_k} \to x_k^*\) as \(n_k \to \infty\). From Lemma 2, we have for each \(d \in [0, d_k - 1]\),

\[
\lim sup_{n_k \to \infty} \sup_{z} |f_k(z, d, n_k) - f_k(z, d)| = 0,
\]

and applying Theorem 4.2.23 of [33], we obtain from (25) and Lemma 3 that

\[
\lim sup_{n_k \to \infty} \frac{1}{n_k} \log \mathbb{P}_m(D_k = d) \leq - \min_{z \in A_k(d)} \Phi^*_m(z, x_k^*).
\]

(29)

We then have

\[
\lim sup_{n_k \to \infty} \frac{1}{n_k} \log \mathbb{E}[C_k(H, D_k, m_l)] \\
\leq \lim sup_{n_k \to \infty} \max_{0 \leq m \leq M - 1} \left\{ \frac{1}{n_k} \log \mathbb{P}_m(D_k = d) - c_k(m, d) \right\} \\
\leq -I_k(x_k^*),
\]

(30)

where the last inequality follows from (29). Finally, (28) together with (30) gives us claim (i).
To show claim (ii), fix any \( x_k^* \in \arg\max_{x \in \mathcal{S}(T)} I_k(x) \). We note that if there exists a subsequence of policies \( (x_{km})_{l \geq 1} \) with \( \lim_{l \to \infty} x_{km} = x_k \) and \( I_k(x_k) < I_k(x_k^*) \), then using the same arguments that lead to (28), we have

\[
\liminf_{n_k \to \infty} \frac{1}{n_k} \log \mathbb{E} [C_k(H, D_k, n_k)] \geq -I_k(x_k) > -I_k(x_k^*),
\]

a contradiction to (30). Therefore, each policy subsequence converges to some \( x_k \) with \( I_k(x_k) = I_k(x_k^*) \), and there is no loss in optimality if we restrict the sequence of policies to converge to \( x_k^* \).

Finally, to show claim (iii), we have from (27) that

\[
\liminf_{l \to \infty} \frac{1}{n_k} \log \mathbb{P}_m(D_k = d) \geq - \min_{z \in \mathcal{A}_k(d)} \Phi_m^*(z, x_k^*),
\]

since \( \Phi_m^*(z, x) \) is continuous in \( x \). Together with (29), the claim now follows, and the theorem is proved.

E. Proof of Proposition 5

From Theorem 5 of [34] (or Proposition A.2 of [23]), we have for \( i, j \in [0, M - 1] \), with \( j \neq i \), and every \( s \in [0, 1] \), either

\[
\mathbb{P}_i(D_0(k) \neq i \mid D_k = d) \geq \frac{1}{4} \exp \left( n_0 \Lambda_{ij}(s, x_0) - s n_0 \frac{d}{ds} \Lambda_{ij}(s, x_0) - s \sqrt{2 n_0 \frac{d^2}{ds^2} \Lambda_{ij}(s, x_0)} \right),
\]

or

\[
\mathbb{P}_j(D_0(k) = i \mid D_k = d) \geq \frac{1}{4} \exp \left( n_0 \Lambda_{ij}(s, x_0) + (1 - s) n_0 \frac{d}{ds} \Lambda_{ij}(s, x_0) - (1 - s) \sqrt{2 n_0 \frac{d^2}{ds^2} \Lambda_{ij}(s, x_0)} \right).
\]

If \( s^*_{ij} \in (0, 1) \), we have

\[
\frac{d}{ds} \Lambda_{ij}(s^*_{ij}, x_0) = \frac{1}{n_0} \log \frac{C_0(i, j, n_0) \mathbb{P}_i(D_k = d)}{C_0(j, i, n_0) \mathbb{P}_j(D_k = d)},
\]
and using Lemma 1, (32) and (33), we obtain
\[
\frac{1}{n_0} \log \left\{ \min_{j' \neq i} C_0(i, j', n_0) \mathbb{P}_i (D_0(k) \neq i, D_k = d) \right. \\
+ C_0(j, i, n_0) \mathbb{P}_j (D_0(k) = i, D_k = d) \\
+ \frac{1 - s_{ij}^*}{n_0} \log \left( \min_{j' \neq i} C_0(i, j', n_0) \mathbb{P}_i (D_k = d) \right) \\
+ \frac{s_{ij}^*}{n_0} \log (C_0(j, i, n_0) \mathbb{P}_j (D_k = d)) \\
+ \Lambda_{ij}(s_{ij}^*, x_0) - \frac{1}{n_0} \log 4 - \sqrt{\frac{2G(2)}{n_0}} \\
\geq (1 - s_{ij}^*) \left( \frac{1}{n_0} \log \mathbb{P}_i (D_k = d) - c_0(i) \right) \\
+ s_{ij}^* \left( \frac{1}{n_0} \log \mathbb{P}_j (D_k = d) - c_0(j) \right) + \Lambda_{ij}(s_{ij}^*, x_0) - \epsilon,
\]
where the last inequality follows from Assumption 1 for $n_0$ sufficiently large. On the other hand, if $s_{ij}^* = 0$, we have
\[
\frac{d}{ds} \Lambda_{ij}(0, x_0) \geq \frac{1}{n_0} \log \frac{C_0(i, j, n_0) \mathbb{P}_i (D_k = d)}{C_0(j, i, n_0) \mathbb{P}_j (D_k = d)},
\]
since $\Lambda_{ij}(s, x_0)$ is convex $s$. The inequality (16) then holds trivially. A similar argument holds for $s_{ij}^* = 1$, and the proposition is proved.

**F. Proof of Theorem 6**

Suppose that agent 0 adopts the policy $x_0$. For any $\epsilon > 0$, and for $n_0$ sufficiently large, we have
\[
\frac{1}{n_0} \log \mathbb{E} [C_0(H, D_0(k), n_0)] \\
= \frac{1}{n_0} \log \sum_{d=0}^{d_k-1} \sum_{i \neq j} \pi_i C_0(i, j, n_0) \mathbb{P}_i (D_0(k) = j, D_k = d) \\
\geq \max_{0 \leq d \leq d_k-1} \frac{1}{n_0} \log \left\{ \min_{j' \neq i} C_0(i, j', n_0) \right. \\
\left. \mathbb{P}_i (D_0(k) \neq i, D_k = d) \right. \\
+ C_0(j, i, n_0) \mathbb{P}_j (D_0(k) = i, D_k = d) \\
- \frac{1}{n_0} \log 2 + \min_m \frac{1}{n_0} \log \pi_m \\
\geq \max_{0 \leq d \leq d_k-1} \min_{s \in [0,1]} \left\{ (1 - s) \left( \frac{1}{n_0} \log \mathbb{P}_i (D_k = d) - c_0(i) \right) \\
+ s \left( \frac{1}{n_0} \log \mathbb{P}_j (D_k = d) - c_0(j) \right) + \Lambda_{ij}(s, x_0) \right\} - \epsilon, \quad (34)
\]
where the last inequality follows from Proposition 5. By letting \( n_0 \to \infty \) and \( \epsilon \to 0 \) in (34), we obtain

\[
\limsup_{n_0 \to \infty} \frac{1}{n_0} \log \mathbb{E} [C_0(H, D_0(k), n_0)] \\
\geq \max_{i \neq j} \min_{0 \leq d \leq d_k - 1} \left\{ (1 - s) \left( \lim_{n_0 \to \infty} \frac{1}{n_0} \log \mathbb{P}_i(D_k = d) - c_0(t) \right) \right. \\
+ \left. s \left( \lim_{n_0 \to \infty} \frac{1}{n_0} \log \mathbb{P}_j(D_k = d) - c_0(j) \right) + \Lambda_{ij}(s, x_0) \right\},
\]

(35)

and the lower bound (17) follows from Theorem 4(iii).

We next show that there exists a decision rule for agent 0 that achieves \( \mathcal{E}_0(x_0) \) in (17). Given \( D_k = d \), consider the following rule to differentiate between hypotheses \( H = i \) and \( H = j \) for \( i \neq j \), which declares \( H = i \) iff

\[
\frac{1}{n_0} \log \ell_{ji}(Y_0[1 : n_0]) \leq h_{ji},
\]

where

\[
h_{ji} \triangleq -q_k \min_{z \in A_k(d)} \Phi_i^*(z, x_k^*) + q_k \min_{z \in A_k(d)} \Phi_j^*(z, x_k^*) - c_0(i) + c_0(j).\]

By a simple generalization of Cramér’s Theorem\(^7\) [33], and Theorem 4(iii), we have for every \( \epsilon > 0 \) and all \( n_0 \) sufficiently large,

\[
\frac{1}{n_0} \log \left( C_0(i, j, n_0) \mathbb{P}_i(D_0(k) = j, D_k = d) \right) = \frac{1}{n_0} \log \mathbb{P}_0(i, j, n_0) + \frac{1}{n_0} \log \mathbb{P}_i(D_k = d) \\
+ \frac{1}{n_0} \log \mathbb{P}_j(D_0(k) = j \mid D_k = d) \\
\leq -c_0(i) - q_k \min_{z \in A_k(d)} \Phi_i^*(z, x_k^*) \\
- \max_{s \in [0, 1]} \left\{ s h_{ji} - \Lambda_{ij}(s, x_0) \right\} + \epsilon \\
\leq - \max_{s \in [0, 1]} \left\{ (1 - s) \left( q_k \min_{z \in A_k(d)} \Phi_i^*(z, x_k^*) + c_0(i) \right) \\
+ s \left( q_k \min_{z \in A_k(d)} \Phi_j^*(z, x_k^*) + c_0(j) \right) - \Lambda_{ij}(s, x_0) \right\} + \epsilon,
\]

from which we obtain

\[
\frac{1}{n_0} \log \mathbb{E} [C_0(H, D_0(k), n_0)] \leq \max_{i \neq j} \min_{0 \leq d \leq d_k - 1} \left\{ \frac{1}{n_0} \log \left( C_0(i, j, n_0) \mathbb{P}_i(D_0(k) = j, D_k = d) \right) + \frac{1}{n_0} \log M \right\} \\
\leq -\mathcal{E}_0(k, x_0) + \epsilon.
\]

By taking \( n_0 \to \infty \) and \( \epsilon \to 0 \), we obtain the theorem by maximizing \( \mathcal{E}_0(x_0) \) over all policies \( x_0 \). The proof is now complete.

\(^7\)Cramér’s Theorem applies to independent and identically distributed (i.i.d.) observations. The private observations \( Y_0[1 : n_0] \) are not i.i.d., but are independent and can be divided into groups of i.i.d. observations.
G. Proof of Lemma 7

The non-negativity and convexity of $\Phi_m(z, x)$ follows from Lemma 2.2.31 of [33]. From Jensen’s inequality, for any $t \in \mathbb{R}^{M-1}$, we have

$$ \varphi_m(t, x) \geq \sum_{\gamma \in \Gamma} x[\gamma] E_m[(t, Z^\gamma)] = \langle t, \tilde{z}_m(x) \rangle, $$

which implies that $\Phi_m(\tilde{z}, x) = 0$, and the lemma is proved.

H. Proof of Proposition 9

Since the proof is similar to that of Theorem 6, we provide only an outline here. From the proposition assumptions, we have for every $i \in [0, M - 1]$, $P_i(D_k = i)$ is bounded away from zero, i.e.,

$$ \lim_{n_k \to \infty} (1/n_k) \log P_i(D_k = i) = 0 $$

because otherwise the expected loss of agent $k$ can be decreased. A similar argument that yields (35) now shows that

$$ \lim_{n_0 \to \infty} \frac{1}{n_0} \log E[C_0(H, D_0(k), n_0)] $$

is lower bounded by the maximum of

$$ \min_{s \in [0, 1]} \left\{ -(1 - s)c_0(i) - sc_0(j) + s \lim_{n_0 \to \infty} \frac{1}{n_0} \log P_j(D_k = i) + \Lambda_{ij}(s, x_0) \right\}, \quad (36) $$

and

$$ \min_{s \in [0, 1]} \left\{ -(1 - s)c_0(i) - sc_0(j) + (1 - s) \lim_{n_0 \to \infty} \frac{1}{n_0} \log P_i(D_k \neq i) + \Lambda_{ij}(s, x_0) \right\}, \quad (37) $$

over all pairs $(i, j)$ with $i, j \in [0, M - 1]$ and $i \neq j$. We also have from Theorem 5 of [34] that for $t_{ij} = \arg \min_{t \in [0, 1]} \Lambda_{ij}(t, x_k)$, either

$$ \lim_{n_k \to \infty} \frac{1}{n_k} \log P_i(D_k \neq i) \geq \Lambda_{ij}(t_{ij}, x_k), $$

or

$$ \lim_{n_k \to \infty} \frac{1}{n_k} \log P_j(D_k = i) \geq \Lambda_{ij}(t_{ij}, x_k). $$

Together with (36) and (37), and the fact that $\Lambda_{ji}(s, x_0) = \Lambda_{ij}(1 - s, x_0)$, we obtain the lower bound (23). Finally, the same achievability proof in Theorem 6 for the case where expert $k$ uses the 0-1 loss function shows that this lower bound is achievable, and the proposition is proved.

I. Proof of Proposition 10

We first show claim (i). Suppose that agent $k$ adopts the policy $x_k$. From the pigeonhole principle, if $d_k < M$, there exists a region $A_k(d)$ in which both $\Phi_i(\cdot, x_k)$ and $\Phi_j(\cdot, x_k)$ achieve their minimum value of 0, for some $i \neq j$. From Assumption 4, (17), and (19), we have

$$ E_0(k, x_0) = -\min_{s \in [0, 1]} \Lambda(s, x_0) = E_{0,B}(k, x_0), $$

where

$$ E_{0,B}(k, x_0) = -\min_{s \in [0, 1]} \Lambda_{ij}(s, x_0), $$

for some $i \neq j$. The same argument as in Proposition 9 yields

$$ \lim_{n_k \to \infty} \frac{1}{n_k} \log P_i(D_k \neq i) \geq \Lambda_{ij}(t_{ij}, x_k), $$

or

$$ \lim_{n_k \to \infty} \frac{1}{n_k} \log P_j(D_k = i) \geq \Lambda_{ij}(t_{ij}, x_k). $$

Together with (36) and (37), and the fact that $\Lambda_{ji}(s, x_0) = \Lambda_{ij}(1 - s, x_0)$, we obtain the lower bound (23). Finally, the same achievability proof in Theorem 6 for the case where expert $k$ uses the 0-1 loss function shows that this lower bound is achievable, and the proposition is proved.
and the claim is proved.

We next show claim (ii). Suppose that agent 0 adopts the opinion of agent $k$. The same argument we used to derive (18) shows that the loss exponent of agent $k$ under hypothesis $H = i$ when it declares hypothesis $j$ is

$$
\max_{s \in \mathbb{S}(1)} \left\{ (1 - s) c_k(i) + sc_k(j) - \Lambda_{ij}(s, x_k^*) \right\},
$$

and using the same argument as in Proposition 9, we obtain the proposition.

**APPENDIX C**

**CHARACTERIZATION OF ASYMPTOTIC DECISION REGIONS**

In this appendix, we give a characterization for the asymptotic decision region $A_k(d)$ for an agent $k$, and $d \in [0, d_k - 1]$. For $i, j \in [0, M - 1]$ and $p, q \in [0, d_k]$, define the half-space

$$
B_k(i, p, j, q) = \{ z = (z[m])_{1 \leq m \leq M - 1} \in \mathbb{R}^{M-1} : z[i] - z[j] \geq c_k(i, p) - c_k(j, q) \},
$$

(38)

where $z[0] = 0$. For each $p \in [0, d_k - 1]$, let $m_p \in [0, M - 1]$ be a chosen corresponding index. From (13), we have $z \in \cap_{i \neq m_p} B_k(m_p, p, i, p)$ if $f_k(z, p) = z[m_p] - c_k(m_p, p)$.

Consider a $z \in \mathbb{R}^{M-1}$ such that $f(z, d) < 0$. Then, there exists a sequence $(m_p)_{0 \leq p \leq d_k - 1} \in [0, M - 1]^{d_k}$ such that $f_k(z, p) = z[m_p] - c_k(m_p, p)$ for all $p \in [0, d_k - 1]$ and

$$
\tilde{f}_k(z, p) = z[m_p] - z[m_d] - c_k(m_p, p) + c_k(m_d, d) > 0
$$

for all $p \neq d$, i.e.,

$$
z \in H_d((m_p)_{p=0}^{d_k-1}) = \bigcap_{p=0}^{d_k-1} \bigcap_{i \neq m_p} B_k(m_p, p, i, p) \bigcap_{p \neq d} B_k(m_p, p, m_d, d).
$$

On the other hand, if such a sequence $(m_p)_{p=0}^{d_k-1}$ exists, then $z \in A_k(d)$. Therefore, the set $A_k(d)$ is the union over all sequences $(m_p)_{p=0}^{d_k-1} \in [0, M - 1]^{d_k}$ of $H_d((m_p)_{p=0}^{d_k-1})$.

**REFERENCES**

[1] S. Aral and D. Walker, “Identifying influential and susceptible members of social networks,” *Science*, vol. 337, no. 6092, pp. 337–341, 2012.

[2] A. Mitchell, J. Kiley, J. Gottfried, and E. Guskin. (2013, Oct.) The role of news on Facebook. Pew Research Center. [Online]. Available: http://www.journalism.org/files/2013/10/facebook_news_10-24-2013.pdf

[3] E. Bakshy, J. M. Hofman, W. A. Mason, and D. J. Watts, “Everyone’s an influencer: quantifying influence on Twitter,” in *Proc. 4th ACM International Conference on Web Search and Data Mining*, 2011.
[4] W. Luo, W. P. Tay, and M. Leng, “Identifying infection sources and regions in large networks,” IEEE Trans. Signal Process., vol. 61, no. 11, pp. 2850–2865, Jun. 2013.

[5] J. Leskovec, L. A. Adamic, and B. A. Huberman, “The dynamics of viral marketing,” ACM Trans. Web, vol. 1, May 2007.

[6] Y. Zhang, Z. Wang, and C. Xia, “Identifying key users for targeted marketing by mining online social network,” in Proc. IEEE International Conference on Advanced Information Networking and Applications Workshops, 2010.

[7] Zero Hedge. [Online]. Available: http://www.zerohedge.com/

[8] DealBook. [Online]. Available: http://dealbook.nytimes.com/

[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] Y. Kanoria and O. Tamuz, “Tractable Bayesian social learning on trees,” IEEE J. Sel. Areas Commun., vol. 31, no. 4, pp. 756–765, April 2013.

[11] G. Kortuem, F. Kawsar, D. Fitton, and V. Sundramoorthy, “Smart objects as building blocks for the Internet of Things,” IEEE Internet Computing, vol. 14, no. 1, pp. 44–51, 2010.

[12] Y. Ding, Y. Jin, L. Ren, and K. Hao, “An intelligent self-organization scheme for the Internet of Things,” IEEE Computational Intelligence Magazine, vol. 8, no. 3, pp. 41–53, 2013.

[13] J. N. Tsitsiklis, “Extremal properties of likelihood-ratio quantizers,” IEEE Trans. Commun., vol. 41, no. 4, pp. 550–558, 1993.

[14] B. Chen and P. K. Varshney, “A Bayesian sampling approach to decision fusion using hierarchical models,” IEEE Trans. Signal Process., vol. 50, no. 8, pp. 1809–1818, Aug. 2002.

[15] J.-F. Chamberland and V. V. Veeravalli, “Decentralized detection in sensor networks,” IEEE Trans. Signal Process., vol. 51, no. 2, pp. 407–416, 2003.

[16] Y. Lin, B. Chen, and P. K. Varshney, “Decision fusion rules in multi-hop wireless sensor networks,” IEEE Trans. Aerosp. Electron. Syst., vol. 41, no. 2, pp. 475–488, Apr. 2005.

[17] P. Willett, P. Swaszek, and R. Blum, “The good, bad and ugly: distributed detection of a known signal in dependent Gaussian noise,” IEEE Trans. Signal Process., vol. 48, no. 12, pp. 3266 – 3279, Dec. 2000.

[18] H. Chen, B. Chen, and P. Varshney, “A new framework for distributed detection with conditionally dependent observations,” IEEE Trans. Signal Process., vol. 60, no. 3, pp. 1409 –1419, Mar. 2012.

[19] W. P. Tay, J. N. Tsitsiklis, and M. Z. Win, “Data fusion trees for detection: Does architecture matter?” IEEE Trans. Inf. Theory, vol. 54, no. 9, pp. 4155–4168, Sep. 2008.

[20] ———, “On the impact of node failures and unreliable communications in dense sensor networks,” IEEE Trans. Signal Process., vol. 56, no. 6, pp. 2535–2546, Jun. 2008.

[21] ———, “Bayesian detection in bounded height tree networks,” IEEE Trans. Signal Process., vol. 57, no. 10, pp. 4042–4051, Oct. 2009.

[22] Z. Zhang, A. Pezeshki, W. Moran, S. Howard, and E. K. P. Chong, “Error probability bounds for balanced binary relay trees,” IEEE Trans. Inf. Theory, vol. 58, no. 6, pp. 3548–3563, 2012.

[23] W. P. Tay, “The value of feedback in decentralized detection,” IEEE Trans. Inf. Theory, vol. 58, no. 12, pp. 7226–7239, Dec. 2012.

[24] J. N. Tsitsiklis, “Decentralized detection,” Advances in Statistical Signal Processing, vol. 2, pp. 297–344, 1993.

[25] J. D. Papastavrou and M. Athans, “Distributed detection by a large team of sensors in tandem,” IEEE Trans. Aerosp. Electron. Syst., vol. 28, no. 3, pp. 639–653, 1992.
[26] 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, no. 10, pp. 4767–4771, Oct. 2008.

[27] K. Drakopoulos, A. Ozdaglar, and J. Tsitsiklis, “On learning with finite memory,” IEEE Trans. Inf. Theory, vol. 59, no. 10, pp. 6859–6872, Oct 2013.

[28] S. Nitinawarat, G. Atia, and V. Veeravalli, “Controlled sensing for multihypothesis testing,” IEEE Trans. Autom. Control, vol. 58, no. 10, pp. 2451–2464, Oct 2013.

[29] E. Tuncel, “On error exponents in hypothesis testing,” IEEE Trans. Inf. Theory, vol. 51, no. 8, pp. 2945–2950, Aug 2005.

[30] N. Grigoryan and A. Harutyunyan, “Error exponents in multiple hypothesis testing for arbitrarily varying sources,” in IEEE Information Theory Workshop, Aug 2010, pp. 1–5.

[31] J. N. Tsitsiklis and M. Athans, “On the complexity of decentralized decision making and detection problems,” IEEE Trans. Autom. Control, vol. 30, pp. 440–446, 1985.

[32] J. N. Tsitsiklis, “Decentralized detection by a large number of sensors,” Math. Control, Signals, Syst., vol. 1, pp. 167–182, 1988.

[33] A. Dembo and O. Zeitouni, Large Deviations Techniques and Applications. New York, NY: Springer-Verlag, 1998.

[34] C. E. Shannon, R. G. Gallager, and E. R. Berlekamp, “Lower bounds to error probability for coding on discrete memoryless channels, I,” Information and Control, vol. 10, pp. 65–103, 1967.

[35] V. Veeravalli, T. Basar, and H. Poor, “Minimax robust decentralized detection,” in IEEE Trans. on Inf. Theory, vol. 40, 1994, pp. 35–40.

[36] J. Ho, W. P. Tay, and T. Q. S. Quek, “Robust detection and social learning in tandem networks,” in Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing, Florence, Italy, May 2014.

[37] D. Stroock, Probability Theory: An Analytic View. Cambridge, UK: Cambridge University Press, 1993.