SAND2020-3796C

Event-Triggered Distributed Inference

Aritra Mitra, John A. Richards, Saurabh Bagchi and Shreyas Sundaram

I. Introduction

We consider a scenario involving a network of agents, where each agent receives a stream of private signals sequentially over time. The observations of every agent are generated by a common underlying distribution, parameterized by an unknown static quantity which we call the *true state of the world*. The task of the agents is to collectively identify this unknown quantity from a finite family of hypotheses, while relying solely on local interactions. The problem described above arises in a variety of contexts ranging from detection and object recognition using autonomous robots, to statistical inference and learning over multiple processors, to sequential decision-making in social networks. As such, the distributed inference/hypothesis testing problem enjoys a rich history [1]–[7], where a variety of techniques have been proposed primarily with the aim of improving the convergence rate. These include data-aggregation using consensus-based linear [1], [2] and log-linear rules [3]–[5], and the more recent min-protocol [6], [7] - the latter leading to the best known learning rate for this problem. Much less explored, however, is the aspect of communication-efficiency - a theme which is becoming increasingly important as we deal with low-power sensor devices and limited-bandwidth communication channels. Motivated by this gap in the literature, we seek to answer the following questions in this paper. (i) When should an agent exchange information with a neighbor? (ii) What piece of information should the agent exchange?

To address the questions posed above, we will draw on ideas from the theory of event-triggered control. The initial results [8], [9] on this topic centred around stabilizing dynamical systems by injecting control inputs only when needed, as opposed to the traditional approach of periodic control inputs. Since then, the ideas emanating from this line of work have found there way into the design of event-driven control and communication techniques for multi-agent systems; the recent survey [10] provides an excellent overview of such techniques. Notably, the applications of event-triggered ideas to multi-agent settings have primarily focused on either consensus or distributed optimization. To the best of our knowledge, this is the first work which explores event-triggered communication in the context of distributed inference. We summarize our contributions below.

A. Mitra, S. Bagchi and S. Sundaram are with the School of Electrical and Computer Engineering at Purdue University. J. A. Richards is with Sandia National Laboratories. Email: {mitra14, sbagchi@purdue.edu, sundara2}@purdue.edu, jaricha@sandia.gov. This work was supported in part by NSF CAREER award 1653648, and by a grant from Sandia National Laboratories. Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

Contributions: The primary contribution of this paper is the development of a novel event-triggered distributed learning rule along with a detailed theoretical characterization of its performance. Our approach to learning is based on the principle of diffusing low beliefs on each false hypothesis across the network. Building on this principle, we design a trigger condition that carefully takes into account the specific structure of the problem, and enables an agent to decide, using purely local information, whether or not to broadcast its belief on a given hypothesis to a specific neighbor. Specifically, based on our event-triggered strategy, an agent broadcasts only those components of its belief vector that have adequate "innovation", to only those neighbors that are in need of the corresponding pieces of information. In this way, our approach not only reduces the number of communication rounds, but also the amount of information transmitted in each round.

We establish that our proposed event-triggered learning rule enables each agent to learn the true state exponentially fast under standard assumptions on the observation model and the network structure. We characterize the learning rate of our algorithm and identify conditions under which one can achieve the best known learning rate in [7], even when the inter-communication intervals between the agents grow potentially unbounded. In other words, we identify sparse communication regimes where communication-efficiency comes essentially for "free". We further demonstrate, both in theory and in simulations, that our event-triggered scheme has the potential of reducing information flow from uninformative agents to informative agents. Finally, we argue that if asymptotic learning of the true state is the only consideration, then one can allow for arbitrarily sparse communication schemes.

II. Model and Problem Formulation

Network Model: We consider a group of agents $\mathcal{V} = \{1, \ldots, n\}$, and model interactions among them via an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$.[1] An edge $(i, j) \in \mathcal{E}$ indicates that agent i can directly transmit information to agent j, and vice versa. The set of all neighbors of agent i is defined as $\mathcal{N}_i = \{j \in \mathcal{V} : (j, i) \in \mathcal{E}\}$. We say that $\mathcal{G}$ is rooted at $\mathcal{C} \subseteq \mathcal{V}$, if for each agent $i \in \mathcal{V} \setminus \mathcal{C}$, there exists a path to it from some agent $j \in \mathcal{C}$. For a connected graph $\mathcal{G}$, we will use $d(i, j)$ to denote the length of the shortest path between i and j.

[1] The results in this paper can be easily extended to directed graphs.

Observation Model: Let $\Theta = \{\theta_1, \theta_2, \ldots, \theta_m\}$ denote m possible states of the world, with each state representing a hypothesis. A specific state $\theta^\star \in \Theta$, referred to as the true state of the world, gets realized. Conditional on its realization, at each time-step $t \in \mathbb{N}_+$, every agent $i \in \mathcal{V}$ privately observes a signal $s_{i,t} \in \mathcal{S}_i$, where $\mathcal{S}_i$ denotes the signal space of agent i.[2] The joint observation profile so generated across the network is denoted $s_t = (s_{1,t}, s_{2,t}, \ldots, s_{n,t})$, where $s_t \in \mathcal{S}$, and $\mathcal{S} = \mathcal{S}_1 \times \mathcal{S}_2 \times \ldots \mathcal{S}_n$. Specifically, the signal s_t is generated based on a conditional likelihood function $l(\cdot|\theta^\star)$, the i-th marginal of which is denoted $l_i(\cdot|\theta^\star)$, and is available to agent i. The signal structure of each agent $i \in \mathcal{V}$ is thus characterized by a family of parameterized marginals $l_i = \{l_i(w_i|\theta) : \theta \in \Theta, w_i \in \mathcal{S}_i\}$. We make certain standard assumptions [1]–[5]: (i) The signal space of each agent i, namely $\mathcal{S}_i$, is finite. (ii) Each agent i has knowledge of its local likelihood functions $\{l_i(\cdot|\theta_p)\}_{p=1}^m$, and it holds that $l_i(w_i|\theta) > 0, \forall w_i \in \mathcal{S}_i$, and $\forall \theta \in \Theta$. (iii) The observation sequence of each agent is described by an i.i.d. random process over time; however, at any given time-step, the observations of different agents may potentially be correlated. (iv) There exists a fixed true state of the world $\theta^\star \in \Theta$ (unknown to the agents) that generates the observations of all the agents. The probability space for our model is denoted $(\Omega, \mathcal{F}, \mathbb{P}^{\theta^\star})$, where $\Omega \triangleq \{\omega : \omega = (s_1, s_2, \ldots), \forall s_t \in \mathcal{S}, \forall t \in \mathbb{N}_+\}$, $\mathcal{F}$ is the σ-algebra generated by the observation profiles, and $\mathbb{P}^{\theta^\star}$ is the probability measure induced by sample paths in Ω. Specifically, $\mathbb{P}^{\theta^\star} = \prod_{t=1}^{\infty} l(\cdot|\theta^\star)$. We will use the abbreviation a.s. to indicate almost sure occurrence of an event w.r.t. $\mathbb{P}^{\theta^\star}$.

The goal of each agent in the network is to eventually learn the true state $\theta^\star$. However, the key challenge in achieving this objective arises from an *identifiability problem* that each agent might potentially face. To make this precise, define $\Theta_i^{\theta^\star} \triangleq \{\theta \in \Theta : l_i(w_i|\theta) = l_i(w_i|\theta^\star), \forall w_i \in \mathcal{S}_i\}$. In words, $\Theta_i^{\theta^\star}$ represents the set of hypotheses that are *observationally equivalent* to $\theta^\star$ from the perspective of agent i. Thus, if $|\Theta_i^{\theta^\star}| > 1$, it will be impossible for agent i to uniquely learn the true state θ^* without interacting with its neighbors.

In the next section, we will develop a distributed learning algorithm that not only resolves the identifiability problem described above, but does so in a communication-efficient manner. Before describing this algorithm, we first recall the following definition from [6] that will show up in our subsequent developments.

Definition 1. *(**Source agents**) An agent i is said to be a source agent for a pair of distinct hypotheses $\theta_p, \theta_q \in \Theta$ if it can distinguish between them, i.e., if $D(l_i(\cdot|\theta_p)||l_i(\cdot|\theta_q)) > 0$, where $D(l_i(\cdot|\theta_p)||l_i(\cdot|\theta_q))$ represents the KL-divergence [11] between the distributions $l_i(\cdot|\theta_p)$ and $l_i(\cdot|\theta_q)$. The set of source agents for pair (θ_p, θ_q) is denoted $\mathcal{S}(\theta_p, \theta_q)$.* □

Throughout the rest of the paper, we will use $K_i(\theta_p, \theta_q)$ as a shorthand for $D(l_i(\cdot|\theta_p)||l_i(\cdot|\theta_q))$.

[2]We use $\mathbb{N}$ and $\mathbb{N}_+$ to represent the set of non-negative integers and positive integers, respectively.

III. An Event-Triggered Distributed Learning Rule

• **Belief-Update Strategy**: In this section, we develop an event-triggered distributed learning rule that enables each agent to eventually learn the truth, despite infrequent information exchanges with its neighbors. Our approach requires each agent i to maintain a local belief vector $\boldsymbol{\pi}_{i,t}$, and an actual belief vector $\boldsymbol{\mu}_{i,t}$, each of which are probability distributions over the hypothesis set Θ. While agent i updates $\boldsymbol{\pi}_{i,t}$ in a Bayesian manner using only its private signals (see eq. (2)), to formally describe how it updates $\boldsymbol{\mu}_{i,t}$, we need to first introduce some notation. Accordingly, let $\mathbb{1}_{ji,t}(\theta) \in \{0,1\}$ be an indicator variable which takes on a value of 1 if and only if agent j broadcasts $\mu_{j,t}(\theta)$ to agent i at time t. Next, we define $\mathcal{N}_{i,t}(\theta) \triangleq \{j \in \mathcal{N}_i | \mathbb{1}_{ji,t}(\theta) = 1\}$ as the subset of agent i's neighbors who broadcast their belief on θ to i at time t. As part of our learning algorithm, each agent i keeps track of the lowest belief on each hypothesis $\theta \in \Theta$ that it has heard up to any given instant t, denoted by $\bar{\mu}_{i,t}(\theta)$. More precisely, $\bar{\mu}_{i,0}(\theta) = \mu_{i,0}(\theta)$, and $\forall t+1 \in \mathbb{N}_+$,

$$\bar{\mu}_{i,t+1}(\theta) = \min\{\bar{\mu}_{i,t}(\theta), \{\mu_{j,t+1}(\theta)\}_{j \in \{i\} \cup \mathcal{N}_{i,t+1}(\theta)}\}. \quad (1)$$

We are now in position to describe the belief-update rule at each agent: $\boldsymbol{\pi}_{i,t}$ and $\boldsymbol{\mu}_{i,t}$ are initialized with $\pi_{i,0}(\theta) > 0, \mu_{i,0}(\theta) > 0, \forall \theta \in \Theta, \forall i \in \mathcal{V}$ (but otherwise arbitrarily), and subsequently updated as follows $\forall t+1 \in \mathbb{N}_+$.

$$\pi_{i,t+1}(\theta) = \frac{l_i(s_{i,t+1}|\theta)\pi_{i,t}(\theta)}{\sum_{p=1}^{m} l_i(s_{i,t+1}|\theta_p)\pi_{i,t}(\theta_p)}. \quad (2)$$

$$\mu_{i,t+1}(\theta) = \frac{\min\{\bar{\mu}_{i,t}(\theta), \pi_{i,t+1}(\theta)\}}{\sum_{p=1}^{m} \min\{\bar{\mu}_{i,t}(\theta_p), \pi_{i,t+1}(\theta_p)\}}. \quad (3)$$

• **Communication Strategy**: We now focus on specifying *when* an agent broadcasts its belief on a given hypothesis to a neighbor. To this end, we first define a sequence $\mathbb{I} = \{t_k\}$ of *event-monitoring* time-steps, where $t_1 = 1$, and $t_{k+1} - t_k = g(k), \forall k \in \mathbb{N}_+$. Here, $g : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ is a continuous, non-decreasing function that takes on integer values at integers. We will henceforth refer to $g(k)$ as the *event-interval* function. At any given time $t \in \mathbb{N}_+$, let $\hat{\mu}_{ij,t}(\theta)$ represent agent i's belief on θ the last time (excluding time t) it transmitted its belief on θ to agent j. Our communication strategy can now be described as follows. At t_1, each agent $i \in \mathcal{V}$ broadcasts its entire belief vector $\boldsymbol{\mu}_{i,t}$ to every neighbor. Subsequently, at each $t_k, k \geq 2$, i transmits $\mu_{i,t_k}(\theta)$ to $j \in \mathcal{N}_i$ if and only if the following event occurs:

$$\mu_{i,t_k}(\theta) < \gamma(t_k) \min\{\hat{\mu}_{ij,t_k}(\theta), \hat{\mu}_{ji,t_k}(\theta)\}, \quad (4)$$

where $\gamma : \mathbb{N} \rightarrow (0, 1]$ is a non-increasing function, which we will henceforth call the *threshold* function. If $t \notin \mathbb{I}$, then an agent i does not communicate with its neighbors at time t, i.e., all inter-agent interactions are restricted to time-steps in $\mathbb{I}$, subject to the trigger-condition given by (4). Notice that we have not yet specified the functional forms of $g(\cdot)$ and $\gamma(\cdot)$; we will comment on this topic later in Section IV.

Fig. 1. The figure shows a network where only agent 1 is informative. In Section III, we design an event-triggered algorithm under which all upstream broadcasts along the path $3 \rightarrow 2 \rightarrow 1$ stop eventually almost surely. At the same time, all agents learn the true state. We demonstrate these facts both in theory (see Section IV), and in simulations (see Section VI).

• **Summary**: At each time-step $t+1 \in \mathbb{N}_+$, and for each hypothesis $\theta \in \Theta$, the sequence of operations executed by an agent i is summarized as follows. (i) Agent i updates its local and actual beliefs on θ via (2) and (3), respectively. (ii) For each neighbor $j \in \mathcal{N}_i$, it decides whether or not to transmit $\mu_{i,t+1}(\theta)$ to j, and collects $\{\mu_{j,t+1}(\theta)\}_{j \in \mathcal{N}_{i,t+1}(\theta)}$.[3] (iii) It updates $\bar{\mu}_{i,t+1}(\theta)$ via (1) using the (potentially) new information it acquires from its neighbors at time $t+1$.

• **Intuition:** The premise of our belief-update strategy is based on diffusing low beliefs on each false hypothesis. For a given false hypothesis θ, the local Bayesian update (2) will generate a decaying sequence $\pi_{i,t}(\theta)$ for each $i \in \mathcal{S}(\theta^*, \theta)$. Update rules (1) and (3) then help propagate agent i's low belief on θ to the rest of the network.

To build intuition regarding our communication strategy, let us consider the network in Fig 1. Suppose $\Theta = \{\theta_1, \theta_2\}, \theta^* = \theta_1$, and $\mathcal{S}(\theta_1, \theta_2) = 1$, i.e., agent 1 is the only informative agent. Since our principle of learning is based on eliminating each false hypothesis, it makes sense to broadcast beliefs only if they are low enough. Based on this observation, one naive approach to enforce sparse communication could be to set a fixed low threshold, say β, and wait till beliefs fall below such a threshold to broadcast. While this might lead to sparse communication initially, there will come a time beyond which the beliefs of all agents on the false hypothesis θ_2 will always stay below β, leading to dense communication eventually. The obvious fix is to introduce an event-condition that is *state-dependent*. Consider the following candidate strategy: an agent broadcasts its belief on a state θ only if it is sufficiently lower than what it was when it last broadcasted about θ. While an improvement over the "fixed-threshold" strategy, this new scheme has the following demerit: broadcasts are not *agent-specific*. In other words, going back to our example, agent 2 (resp., agent 3) might transmit unsolicited information to agent 1 (resp., agent 2) - information, that agent 1 (resp., agent 2) can do without. To remedy this, one can consider a request/poll based scheme, where an agent receives information from a neighbor only by polling that neighbor. However, now each time agent 2 needs information from agent 1, it needs to place a request, *the request itself incurring extra communication*.

Given the above issues, we ask: Is it possible to devise an event-triggered scheme that eventually stops unnecessary broadcasts from agents 3 to 2, and 2 to 1, while preserving essential information flow from agents 1 to 2, and 2 to 3? This leads us to the event condition in Eq. 4. For each $\theta \in \Theta$, an agent i broadcasts $\mu_{i,t}(\theta)$ to a neighbor $j \in \mathcal{N}_i$ only if $\mu_{i,t}(\theta)$ has adequate "innovation" w.r.t. i's last broadcast about θ to j, *and* j's last broadcast about θ to i. A decreasing threshold function $\gamma(t)$ makes it progressively harder to satisfy the event condition in Eq. 4, leading to fewer broadcasts as time progresses.[4] The rationale behind checking the event condition only at time-steps in $\mathbb{I}$ is twofold. First, it saves computations since the event condition need not be checked all the time. Second, and more importantly, it provides an additional instrument to control communication-sparsity on top of event-triggering.

We close this section by highlighting that our event condition (i) is *θ-specific* since an agent may not be equally informative about all states; (ii) is *agent-specific* and incorporates feedback from neighbors; (iii) can be checked using local information only; and (iv) leverages the structure of the specific problem under consideration.

IV. Main Results

In this section, we state the main results of this paper, and discuss their implications. Proofs of all results are deferred to Section V. To state the first result concerning the convergence of our learning rule, let $G(\cdot)$ be used to denote the integral of $g(\cdot)$, and $G^{-1}(\cdot)$ represent the inverse of $G(\cdot)$. Since $g(\cdot)$ is strictly positive by definition, $G(\cdot)$ is strictly increasing, and hence, $G^{-1}(\cdot)$ is well-defined.

Theorem 1. *Suppose the functions $g(\cdot)$ and $\gamma(\cdot)$ satisfy:*

$$\lim_{t\to\infty} \frac{G(G^{-1}(t)-2)}{t} = \alpha \in (0,1]; \quad \lim_{t\to\infty} \frac{\log(1/\gamma(t))}{t} = 0. \tag{5}$$

Furthermore, suppose the following conditions hold. (i) For every pair of hypotheses $\theta_p, \theta_q \in \Theta$, the source set $\mathcal{S}(\theta_p, \theta_q)$ is non-empty. (ii) The communication graph $\mathcal{G}$ is connected. Then, the event-triggered distributed learning rule governed by (1), (2), (3), and (4) guarantee the following.

- ***(Consistency)**: For each agent $i \in \mathcal{V}$, $\mu_{i,t}(\theta^\star) \rightarrow 1$ a.s.*
- ***(Exponentially Fast Rejection of False Hypotheses)**: For each agent $i \in \mathcal{V}$, and for each false hypothesis $\theta \in \Theta \setminus \{\theta^\star\}$, the following holds:*

$$\liminf_{t\to\infty} -\frac{\log \mu_{i,t}(\theta)}{t} \geq \max_{v \in \mathcal{S}(\theta^\star,\theta)} \alpha^{d(v,i)} K_v(\theta^\star, \theta) \ a.s. \tag{6}$$

□

At this point, it is natural to ask: For what classes of functions $g(\cdot)$ does the result of Theorem 1 hold? The following result provides an answer.

Corollary 1. *Suppose the conditions in Theorem 1 hold.*

(i) *Additionally, suppose $g(x) = x^p, \forall x \in \mathbb{R}_+$, where p is any positive integer. Then, for each $\theta \neq \theta^*$, and $i \in \mathcal{V}$:*

$$\liminf_{t\to\infty} -\frac{\log \mu_{i,t}(\theta)}{t} \geq \max_{v \in \mathcal{S}(\theta^\star,\theta)} K_v(\theta^\star, \theta) \ a.s. \tag{7}$$

[3] If $t+1 \notin \mathbb{I}$, this step gets bypassed, and $\mathcal{N}_{i,t+1}(\theta) = \emptyset, \forall \theta \in \Theta$.

[4] We will see later on (Prop. 2) that for the network in Fig. 1, this scheme provably stops communications from agents 3 to 2, and 2 to 1, eventually.

(ii) *Additionally, suppose* $g(x) = p^x, \forall x \in \mathbb{R}_+$, *where* p *is any positive integer. Then, for each* $\theta \neq \theta^*$, *and* $i \in \mathcal{V}$:

$$\liminf_{t\to\infty} -\frac{\log \mu_{i,t}(\theta)}{t} \geq \max_{v \in \mathcal{S}(\theta^\star,\theta)} \frac{K_v(\theta^\star,\theta)}{p^{2d(v,i)}} \; a.s. \quad (8)$$

□

Proof. The proof follows by directly computing the limit in Eq. (5). For case (i), $\alpha = 1$, and for case (ii), $\alpha = 1/p^2$. □

Clearly, the communication pattern between the agents is at least as sparse as the sequence $\mathbb{I}$. The event-triggering strategy that we employ introduces further sparsity, as we formally establish in the next result.

Proposition 1. *Suppose the conditions in Theorem 1 are met. Then, there exists* $\bar{\Omega} \in \Omega$ *such that* $\mathbb{P}^{\theta^*}(\bar{\Omega}) = 1$, *and for each* $\omega \in \bar{\Omega}$, $\exists T_1(\omega), T_2(\omega) < \infty$ *such that the following hold.*

(i) *At each* $t_k \in \mathbb{I}$ *such that* $t_k > T_1(\omega)$, $\mathbb{1}_{ij,t_k}(\theta^*) \neq 1, \forall i \in \mathcal{V}$ *and* $\forall j \in \mathcal{N}_i$.

(ii) *Consider any* $\theta \neq \theta^*$, *and* $i \notin \mathcal{S}(\theta^*,\theta)$. *Then, at each* $t_k > T_2(\omega)$, $\exists j \in \mathcal{N}_i$ *such that* $\mathbb{1}_{ij,t_k}(\theta) \neq 1$.[5]

□

The following result is an immediate application of the above proposition.

Proposition 2. *Suppose the conditions in Theorem 1 are met. Additionally, suppose* $\mathcal{G}$ *is a tree graph, and for each pair* $\theta_p, \theta_q \in \Theta$, $|\mathcal{S}(\theta_p,\theta_q)| = 1$. *Consider any* $\theta \neq \theta^*$, *and let* $\mathcal{S}(\theta^*,\theta) = v_\theta$. *Then, each agent* $i \in \mathcal{V} \setminus \{v_\theta\}$ *stops broadcasting its belief on* θ *to its parent in the tree rooted at* v_θ *eventually almost surely.* □

Before we proceed to prove the results stated in this section, a few comments are in order.

• **On the nature of** $g(\cdot)$ **and** $\gamma(\cdot)$**:** Intuitively, if the event-interval function $g(\cdot)$ does not grow too fast, and the threshold function $\gamma(\cdot)$ does not decay too fast, one should expect things to fall in place. Theorem 1 makes this intuition precise by identifying conditions on $g(\cdot)$ and $\gamma(\cdot)$ that lead to exponentially fast learning of the truth. In particular, our framework allows for a considerable degree of freedom in the choice of $\gamma(\cdot)$ and $g(\cdot)$. Indeed, from (5), we note that any $\gamma(\cdot)$ that decays sub-exponentially works for our purpose. Moreover, Corollary 1 reveals that up to integer constraints, $g(\cdot)$ can be any polynomial or exponential function.

• **Design trade-offs:** What is the price paid for sparse communication? To answer the above question, we set as benchmark the scenario studied in our previous work [7], where we did not account for communication efficiency. There, we showed that each false hypothesis θ gets rejected exponentially fast by every agent at the *network-independent* rate: $\max_{v\in\mathcal{V}} K_v(\theta^*,\theta)$ - the *best* known rate in the existing literature on this problem. We note from (6) that it is only the event-interval function $g(\cdot)$ that potentially impacts the learning rate, since $\alpha \leq 1$. However, from claim (i) in Corollary 1, we glean that, polynomially growing inter-communication intervals between the agents, coupled with our proposed event-triggering strategy, lead to *no loss in the long-term learning rate relative to the benchmark case in [7]*, i.e., communication-efficiency comes essentially for "free" under this regime. Under exponentially growing event-interval functions, one still achieves exponentially fast learning, albeit at a reduced learning rate that is network-structure dependent (see Eq. 8). The above discussion highlights the practical utility of our results in understanding the trade-offs between sparse communication and the rate of learning.

• **Sparse communication introduced by event-triggering:** Observe that being able to eliminate each false hypothesis is enough for learning the true state. In other words, agents need not exchange their beliefs on the true state (of course, no agent knows apriori what the true state is). Our event-triggering scheme precisely achieves this, as evidenced by claim (i) of Proposition 1: every agent stops broadcasting its belief on θ^* eventually almost surely. In addition, an important property of our event-triggering strategy is that it reduces information flow from uninformative agents to informative agents. To see this, consider any false hypothesis $\theta \neq \theta^*$, and an agent $i \notin \mathcal{S}(\theta^*,\theta)$. Since $i \notin \mathcal{S}(\theta^*,\theta)$, agent i's local belief $\pi_{i,t}(\theta)$ will stop decaying eventually, making it impossible for agent i to lower its actual belief $\mu_{i,t}(\theta)$ without the influence of its neighbors. Consequently, when left alone between consecutive event-monitoring time-steps, i will not be able to leverage its own private signals to generate enough "innovation" in $\mu_{i,t}(\theta)$ to broadcast to the neighbor who most recently contributed to lowering $\mu_{i,t}(\theta)$. The intuition here is simple: an uninformative agent cannot outdo the source of its information. This idea is made precise in claim (ii) of Proposition 1. To further demonstrate this facet of our rule, Proposition 2 stipulates that when the baseline graph is a tree, then all upstream broadcasts to informative agents stop after a finite period of time.

A. Asymptotic Learning of the Truth

If asymptotic learning of the true state is all one cares about, i.e., if exponential convergence is no longer a consideration, then one can allow for arbitrarily sparse communication patterns, as we shall now demonstrate. Accordingly, we first allow the baseline graph $\mathcal{G}(t) = (\mathcal{V}, \mathcal{E}(t))$ to now change over time. To allow for this generality, we set $\mathbb{I} = \mathbb{N}_+$, i.e., the event condition (4) is now monitored at each time-step. Furthermore, we set $\gamma(t) = \gamma \in (0,1], \forall t \in \mathbb{N}$. At each time-step $t \in \mathbb{N}_+$, and for each $\theta \in \Theta$, an agent $i \in \mathcal{V}$ decides whether or not to broadcast $\mu_{i,t}(\theta)$ to an instantaneous neighbor $j \in \mathcal{N}_i(t)$ by checking the event condition (4). While checking this condition, if agent i has not yet transmitted to (resp., heard from) agent j about θ prior to time t, then it sets $\hat{\mu}_{ij,t}(\theta)$ (resp., $\hat{\mu}_{ji,t}(\theta)$) to 1. Update rules (1), (2), (3) remain the same, with $\mathcal{N}_{i,t}(\theta)$ now interpreted as $\mathcal{N}_{i,t}(\theta) \triangleq \{j \in \mathcal{N}_i(t) | \mathbb{1}_{ji,t}(\theta) = 1\}$. Finally, by an union graph over an interval $[t_1, t_2]$, we will imply the graph with vertex set $\mathcal{V}$, and edge set $\cup_{\tau=t_1}^{t_2} \mathcal{E}(\tau)$. With these modifications in place, we have the following result.

Theorem 2. *Suppose for every pair of hypotheses* $\theta_p, \theta_q \in \Theta$, $\mathcal{S}(\theta_p,\theta_q)$ *is non-empty. Furthermore, suppose for each*

[5]In this claim, j might depend on t_k.

$t \in \mathbb{N}_+$, *the union graph over* $[t, \infty)$ *is rooted at* $\mathcal{S}(\theta_p, \theta_q)$. *Then, the event-triggered distributed learning rule described above guarantees* $\mu_{i,t}(\theta^*) \to 1$ *a.s.* $\forall i \in \mathcal{V}$. □

V. Proofs

In this section, we provide proofs of all our technical results. We begin by compiling various useful properties of our update rule which will come handy later on.

Lemma 1. *Suppose the conditions in Theorem 1 hold. Then, there exists a set* $\bar{\Omega} \subseteq \Omega$ *with the following properties. (i)* $\mathbb{P}^{\theta^\star}(\bar{\Omega}) = 1$. *(ii) For each* $\omega \in \bar{\Omega}$, *there exist constants* $\eta(\omega) \in (0,1)$ *and* $t'(\omega) \in (0, \infty)$ *such that*

$$\pi_{i,t}(\theta^\star) \geq \eta(\omega), \bar{\mu}_{i,t}(\theta^\star) \geq \eta(\omega), \forall t \geq t'(\omega), \forall i \in \mathcal{V}. \quad (9)$$

(iii) Consider a false hypothesis $\theta \neq \theta^*$, *and an agent* $i \in \mathcal{S}(\theta^*, \theta)$. *Then on the sample path* ω, *we have:*

$$\liminf_{t\to\infty} -\frac{\log \mu_{i,t}(\theta)}{t} \geq K_i(\theta^\star, \theta)\ . \quad (10)$$

□

Although we consider a modified update rule as compared to that in [7], the proofs of claims (ii) and (iii) in the above Lemma essentially follow the same arguments as that of [7, Lemma 2] and [7, Lemma 3], respectively; we thus omit them here. The following result will be the key ingredient in proving Theorem 1.

Lemma 2. *Consider a false hypothesis* $\theta \in \Theta \setminus \{\theta^\star\}$ *and an agent* $v \in \mathcal{S}(\theta^\star, \theta)$. *Suppose the conditions stated in Theorem 1 hold. Then, the following is true for each agent* $i \in \mathcal{V}$:

$$\liminf_{t\to\infty} -\frac{\log \mu_{i,t}(\theta)}{t} \geq \alpha^{d(v,i)} K_v(\theta^\star, \theta)\ a.s. \quad (11)$$

□

Proof. Let $\bar{\Omega} \subseteq \Omega$ be the set of sample paths for which assertions (i)-(iii) of Lemma 1 hold. Fix a sample path $\omega \in \bar{\Omega}$, an agent $v \in \mathcal{S}(\theta^\star, \theta)$, and an agent $i \in \mathcal{V}$. When $i = v$, the assertion of Eq. (11) follows directly from Eq. (10) in Lemma 1. In particular, this implies that for a fixed $\epsilon > 0$, $\exists t_v(\omega, \theta, \epsilon)$, such that:

$$\mu_{v,t}(\theta) < e^{-(K_v(\theta^\star,\theta)-\epsilon)t}, \forall t \geq t_v(\omega, \theta, \epsilon). \quad (12)$$

Moreover, since $\omega \in \bar{\Omega}$, Lemma 1 guarantees the existence of a time-step $t'(\omega) < \infty$, and a constant $\eta(\omega) > 0$, such that on ω, $\pi_{i,t}(\theta^\star) \geq \eta(\omega), \bar{\mu}_{i,t}(\theta^\star) \geq \eta(\omega), \forall t \geq t'(\omega), \forall i \in \mathcal{V}$. Let $\bar{t}_v(\omega, \theta, \epsilon) = \max\{t'(\omega), t_v(\omega, \theta, \epsilon)\}$. Let $t_q > \bar{t}_v$ be the first even-monitoring time-step in $\mathbb{I}$ to the right of $\bar{t}_v$.[6] Now consider any $t_k \in \mathbb{I}$ such that $k \geq q$. In what follows, we will analyze the implications of agent v deciding whether or not to broadcast its belief on θ to a one-hop neighbor $j \in \mathcal{N}_v$ at t_k. To this end, we consider the following two cases.

[6]We will henceforth suppress the dependence of various quantities on ω, θ, and ϵ for brevity.

Case 1: $\mathbb{1}_{vj,t_k}(\theta) = 1$, i.e., v broadcasts $\mu_{v,t_k}(\theta)$ to j at t_k. Thus, since $v \in \mathcal{N}_{j,t_k}(\theta)$, we have $\bar{\mu}_{j,t_k}(\theta) \leq \mu_{v,t_k}(\theta)$ from (1). Let us now observe that $\forall t \geq t_k + 1$:

$$\begin{aligned}\mu_{j,t}(\theta) &\overset{(a)}{\leq} \frac{\bar{\mu}_{j,t-1}(\theta)}{\sum_{p=1}^{m} \min\{\bar{\mu}_{j,t-1}(\theta_p), \pi_{j,t}(\theta_p)\}} \\ &\overset{(b)}{\leq} \frac{\mu_{v,t_k}(\theta)}{\sum_{p=1}^{m} \min\{\bar{\mu}_{j,t-1}(\theta_p), \pi_{j,t}(\theta_p)\}} \overset{(c)}{<} \frac{e^{-(K_v(\theta^\star,\theta)-\epsilon)t_k}}{\eta}.\end{aligned} \quad (13)$$

In the above inequalities, (a) follows directly from (3), (b) follows by noting that the sequence $\{\bar{\mu}_{j,t}(\theta)\}$ is non-increasing based on (1), and (c) follows from (12) and the fact that all beliefs on $\theta^\star$ are bounded below by η for $t \geq \bar{t}_v$.

Case 2: $\mathbb{1}_{vj,t_k}(\theta) \neq 1$, i.e., v does not broadcast $\mu_{v,t_k}(\theta)$ to j at t_k. From the event condition in (4), it must then be that at least one of the following is true: (a) $\mu_{v,t_k}(\theta) \geq \gamma(t_k)\hat{\mu}_{vj,t_k}(\theta)$, and (b) $\mu_{v,t_k}(\theta) \geq \gamma(t_k)\hat{\mu}_{jv,t_k}(\theta)$. Suppose $\mu_{v,t_k}(\theta) \geq \gamma(t_k)\hat{\mu}_{vj,t_k}(\theta)$. From (12), we then have:

$$\hat{\mu}_{vj,t_k}(\theta) \leq \frac{\mu_{v,t_k}(\theta)}{\gamma(t_k)} < \frac{e^{-(K_v(\theta^\star,\theta)-\epsilon)t_k}}{\gamma(t_k)}. \quad (14)$$

In words, the above inequality places an upper bound on the belief of agent v on θ when it last transmitted its belief on θ to agent j, *prior* to time-step t_k; at least one such transmission is guaranteed to take place since all agents broadcast their entire belief vectors to their neighbors at t_1. Noting that $\bar{\mu}_{j,t}(\theta) \leq \hat{\mu}_{vj,t_k}(\theta), \forall t \geq t_k$, using (3), (14), and arguments similar to those for arriving at (13), we obtain:

$$\mu_{j,t}(\theta) < \frac{e^{-(K_v(\theta^\star,\theta)-\epsilon)t_k}}{\eta\gamma(t_k)} \leq \frac{e^{-(K_v(\theta^\star,\theta)-\epsilon)t_k}}{\eta\gamma(t)}, \forall t \geq t_k + 1, \quad (15)$$

where the last inequality follows from the fact that $\gamma(\cdot)$ is a non-increasing function of its argument. Now consider the case when $\mu_{v,t_k}(\theta) \geq \gamma(t_k)\hat{\mu}_{jv,t_k}(\theta)$. Following the same reasoning as before, we can arrive at an identical upper-bound on $\hat{\mu}_{jv,t_k}(\theta)$ as in (14). Using the definition of $\hat{\mu}_{jv,t_k}(\theta)$, and the fact that agent j incorporates its own belief on θ in the update rule (1), we have that $\bar{\mu}_{j,t}(\theta) \leq \hat{\mu}_{jv,t_k}(\theta), \forall t \geq t_k$. Using similar arguments as before, observe that the bound in (15) holds for this case too.

Combining the analyses of cases 1 and 2, referring to (13) and (15), and noting that $\gamma(t) \in (0,1], \forall t \in \mathbb{N}$, we conclude that the bound in (15) holds for each $t_k \in \mathbb{I}$ such that $t_k > \bar{t}_v$. Now since $t_{k+1} - t_k = g(k)$, for any $\tau \in \mathbb{N}_+$ we have:

$$t_{q+\tau} = t_q + \sum_{z=q}^{q+\tau-1} g(z). \quad (16)$$

Next, noting that $g(\cdot)$ is non-decreasing, observe that:

$$t_q + \int_q^{q+\tau} g(z-1)dz \leq t_{q+\tau} \leq t_q + \int_q^{q+\tau} g(z)dz. \quad (17)$$

The above yields: $l(q,\tau) \triangleq t_q + G(q+\tau-1) - G(q-1) \leq t_{q+\tau} \leq t_q + G(q+\tau) - G(q) \triangleq u(q,\tau)$. Fix any time-step $t >$

$u(q,1)$, let $\tau(t)$ be the largest index such that $u(q,\tau(t)) < t$, and $\bar{\tau}(t)$ be the largest index such that $t_{q+\bar{\tau}(t)} < t$. Observe:

$$\bar{t}_v < t_q < t_{q+1} \leq t_{q+\tau(t)} \leq t_{q+\bar{\tau}(t)} < t. \tag{18}$$

Using the above inequality, the fact that $l(q,\tau(t)) \leq t_{q+\tau(t)}$, and referring to (15), we obtain:

$$\mu_{j,t}(\theta) < \frac{e^{-(K_v(\theta^\star,\theta)-\epsilon)t_{q+\bar{\tau}(t)}}}{\eta\gamma(t)} \leq \frac{e^{-(K_v(\theta^\star,\theta)-\epsilon)l(q,\tau(t))}}{\eta\gamma(t)} \tag{19}$$

From the definition of $\tau(t)$, we have $q+\tau(t) = \lceil G^{-1}(t-t_q+G(q))\rceil - 1$. This yields:

$$\begin{aligned} l(q,\tau(t)) &= t_q + G(\lceil G^{-1}(t-t_q+G(q))\rceil - 2) - G(q-1) \\ &\geq t_q + G(G^{-1}(t-t_q+G(q)) - 2) - G(q-1). \end{aligned} \tag{20}$$

From (19) and (20), we obtain the following $\forall t > u(q,1)$:

$$-\frac{\log \mu_{j,t}(\theta)}{t} > \frac{\tilde{G}(t)}{t}(K_v(\theta^\star,\theta)-\epsilon) - \frac{\log c}{t} - \frac{\log(1/\gamma(t))}{t}, \tag{21}$$

where $\tilde{G}(t) = G(G^{-1}(t-t_q+G(q))-2)$, and $c = e^{-(K_v(\theta^*,\theta)-\epsilon)(t_q-G(q-1))}/\eta$. Now taking the limit inferior on both sides of (21) and using (5) yields:

$$\liminf_{t\to\infty} -\frac{\log \mu_{j,t}(\theta)}{t} \geq \alpha(K_v(\theta^\star,\theta)-\epsilon). \tag{22}$$

Finally, since the above inequality holds for any sample path $\omega \in \bar{\Omega}$, and an arbitrarily small ϵ, it follows that the assertion in (11) is true for every one-hop neighbor j of agent v.

Now consider any agent i such that $d(v,i) = 2$. Clearly, there must exist some $j \in \mathcal{N}_v$ such that $i \in \mathcal{N}_j$. Following an identical line of reasoning as before, it is easy to see that with $\mathbb{P}^{\theta^*}$-measure 1, $\mu_{i,t}(\theta)$ decays exponentially at a rate that is at least α times the rate at which $\mu_{j,t}(\theta)$ decays to zero. From (22), the latter rate is at least $\alpha K_v(\theta^*,\theta)$, and hence, the former is at least $\alpha^2 K_v(\theta^*,\theta)$. This establishes the claim of the lemma for all agents that are two-hops away from agent v. Since $\mathcal{G}$ is connected, given any $i \in \mathcal{V}$, there exists a path $\mathcal{P}(v,i)$ in $\mathcal{G}$ from v to i. One can keep repeating the above argument along the path $\mathcal{P}(v,i)$ and proceed via induction to complete the proof. $\square$

We are now in position to prove Theorem 1.

Proof. (**Theorem 1**) Fix a $\theta \in \Theta \setminus \{\theta^\star\}$. Based on condition (i) of the Theorem, $\mathcal{S}(\theta^\star,\theta)$ is non-empty, and based on condition (ii), there exists a path from each agent $v \in \mathcal{S}(\theta^\star,\theta)$ to every agent $i \in \mathcal{V}\setminus\{v\}$; Eq. (6) then follows from Lemma 2. By definition of a source set, $K_v(\theta^\star,\theta) > 0, \forall v \in \mathcal{S}(\theta^\star,\theta)$; Eq. (6) then implies $\lim_{t\to\infty}\mu_{i,t}(\theta) = 0$ a.s., $\forall i \in \mathcal{V}$. $\square$

Proof. (**Proposition 1**) Let the set $\bar{\Omega}$ have the same meaning as in Lemma 2. Fix any $\omega \in \bar{\Omega}$, and note that since the conditions of Theorem 1 are met, $\mu_{i,t}(\theta^*) \to 1$ on $\omega, \forall i \in \mathcal{V}$. We prove the first claim of the proposition via contradiction. Accordingly, suppose the claim does not hold. Since there are only finitely many agents, this implies the existence of some $i \in \mathcal{V}$ and some $j \in \mathcal{N}_i$, such that i broadcasts its belief on θ^* to j infinitely often, i.e., there exists a sub-sequence $\{t_{p_k}\}$ of $\{t_k\}$ at which the event-condition (4) gets satisfied for θ^*. From (4), $\mu_{i,t_{p_k}}(\theta^*) < \gamma^k \mu_{i,t_{p_0}}(\theta^*), \forall k \in \mathbb{N}_+$, where $\gamma \triangleq \gamma(t_{p_0})$. This implies $\lim_{k\to\infty}\mu_{i,t_{p_k}}(\theta^*) = 0$, contradicting the fact that on ω, $\lim_{t\to\infty}\mu_{i,t}(\theta^*) = 1$.

For establishing the second claim, fix $\omega \in \bar{\Omega}$, $\theta \neq \theta^*$, and $i \notin \mathcal{S}(\theta^*,\theta)$. Since $i \notin \mathcal{S}(\theta^*,\theta)$, there exists $\tilde{t}_1 < \infty$ and $\bar{\eta} > 0$, such that $\pi_{i,t}(\theta) \geq \bar{\eta}, \forall t \geq \tilde{t}_1$. This follows from the fact that since θ is observationally equivalent to θ^* for agent i, the claim regarding $\pi_{i,t}(\theta^*)$ in Eq. (9) applies identically to $\pi_{i,t}(\theta)$. Note also that since the conditions of Theorem 1 are met, $\mu_{i,t}(\theta) \to 0$ on ω. From (1), $\bar{\mu}_{i,t}(\theta) \to 0$ as well. Thus, there must exist some $\tilde{t}_2 < \infty$ such that $\min\{\bar{\mu}_{i,t}(\theta), \pi_{i,t+1}(\theta)\} = \bar{\mu}_{i,t}(\theta), \forall t \geq \tilde{t}_2$. Let $\tilde{t} = \max\{\tilde{t}_1,\tilde{t}_2\}$. Consider any $t_k \in \mathbb{I}, t_k > \tilde{t}$. We claim:

$$\mu_{i,t}(\theta) \geq \bar{\mu}_{i,t_k}(\theta), \forall t \in [t_k+1, t_{k+1}], \text{ and} \tag{23}$$

$$\bar{\mu}_{i,t}(\theta) \geq \bar{\mu}_{i,t_k}(\theta), \forall t \in [t_k, t_{k+1}). \tag{24}$$

To see why the above inequalities hold, consider the update of $\mu_{i,t_k+1}(\theta)$ based on (3). Since $t_k > \tilde{t}_2$, we have $\min\{\bar{\mu}_{i,t_k}(\theta), \pi_{i,t_k+1}(\theta)\} = \bar{\mu}_{i,t_k}(\theta)$. Noting that the denominator of the fraction on the R.H.S. of (3) is at most 1, we obtain: $\mu_{i,t_k+1}(\theta) \geq \bar{\mu}_{i,t_k}(\theta)$. If $t_k+1 = t_{k+1}$, then the claim follows. Else, if $t_k+1 < t_{k+1}$, then since no communication occurs at t_k+1, we have from (1) that $\bar{\mu}_{i,t_k+1}(\theta) = \min\{\bar{\mu}_{i,t_k}(\theta), \mu_{i,t_k+1}(\theta)\} \geq \bar{\mu}_{i,t_k}(\theta)$. We can keep repeating the above argument for each $t \in [t_k+1, t_{k+1}]$ to establish the claim. In words, inequalities (23) and (24) reveal that agent i cannot lower its belief on the false hypothesis θ between two consecutive event-monitoring time-steps when it does not hear from any neighbor. We will make use of this fact repeatedly during the remainder of the proof. Let $t_{p_0} > \tilde{t}$ be the first time-step in $\mathbb{I}$ to the right of $\tilde{t}$. Now consider the following sequence, where $k \in \mathbb{N}$:

$$t_{p_{k+1}} = \inf\{t \in \mathbb{I} : t > t_{p_k}, \bar{\mu}_{i,t}(\theta) < \bar{\mu}_{i,t-1}(\theta)\}. \tag{25}$$

The above sequence represents those event-monitoring time-steps at which $\bar{\mu}_{i,t}(\theta)$ decreases. We first argue that $\{t_{p_k}\}$ is well-defined, i.e., each term in the sequence is finite. If not, then based on (24), this would mean that $\bar{\mu}_{i,t}(\theta)$ remains bounded away from 0, contradicting the fact that $\bar{\mu}_{i,t}(\theta) \to 0$ on ω. Next, for each $k \in \mathbb{N}_+$, let $j_{p_k} \in \operatorname{argmin}_{j\in\mathcal{N}_{i,t_{p_k}}(\theta)\cup\{i\}} \mu_{j,t_{p_k}}(\theta)$. We claim that $i \neq j_{p_k}$. To see why this is true, suppose, if possible, $i = j_{p_k}$. Then, based on the definition of t_{p_k}, we would have $\bar{\mu}_{i,t_{p_k}}(\theta) = \mu_{i,t_{p_k}}(\theta) < \bar{\mu}_{i,t_{p_k}-1}(\theta)$. However, as $t_{p_k} > \tilde{t}_2$, we have from (3) that $\mu_{i,t_{p_k}}(\theta) \geq \bar{\mu}_{i,t_{p_k}-1}(\theta)$, leading to the desired contradiction. In the final step of the proof, we claim that i does not broadcast its belief on θ to j_{p_k} over $[t_{p_k}+1, t_{p_{k+1}}]$.

To establish this claim, we start by noting that based on the definitions of j_{p_k} and t_{p_k}, $\bar{\mu}_{i,t_{p_k}}(\theta) = \mu_{j_{p_k},t_{p_k}}(\theta)$. Let us first consider the case when there are no intermediate event-monitoring time-steps in $(t_{p_k}, t_{p_{k+1}})$, i.e., t_{p_k} and $t_{p_{k+1}}$ are consecutive terms in $\mathbb{I}$. Then, at $t_{p_{k+1}}$, $\hat{\mu}_{j_{p_k}i,t_{p_{k+1}}}(\theta) = \mu_{j_{p_k},t_{p_k}}(\theta)$, since no communication occurs over $(t_{p_k}, t_{p_{k+1}})$. Moreover, using (23), $\mu_{i,t_{p_{k+1}}}(\theta) \geq \bar{\mu}_{i,t_{p_k}}(\theta) = \mu_{j_{p_k},t_{p_k}}(\theta)$. Thus, the event condition (4) gets violated at $t_{p_{k+1}}$, and i does not broadcast its belief on θ to j_{p_k}. Next, consider the scenario when there is exactly one event-monitoring time-step - say $\bar{t} \in \mathbb{I}$ - in the interval $(t_{p_k}, t_{p_{k+1}})$. Since t_{p_k} and $\bar{t}$ are now consecutive terms in $\mathbb{I}$, the fact that $\mathbb{1}_{ij_{p_k},\bar{t}}(\theta) \neq 1$ follows from exactly the same reasoning as earlier. We argue that $\mathbb{1}_{j_{p_k}i,\bar{t}}(\theta) \neq 1$ as well. To see this, suppose that j_{p_k} does in fact broadcast $\mu_{j_{p_k},\bar{t}}(\theta)$ to i at $\bar{t}$. For this to happen, the event condition (4) entails: $\mu_{j_{p_k},\bar{t}}(\theta) < \gamma(\bar{t})\mu_{j_{p_k},t_{p_k}}(\theta) = \gamma(\bar{t})\bar{\mu}_{i,t_{p_k}}(\theta) \leq \bar{\mu}_{i,t_{p_k}}(\theta)$. Since $\bar{\mu}_{i,\bar{t}-1}(\theta) \geq \bar{\mu}_{i,t_{p_k}}(\theta)$ from (24), $\mathbb{1}_{j_{p_k}i,\bar{t}}(\theta) = 1$ would then imply that $\bar{\mu}_{i,\bar{t}}(\theta) < \bar{\mu}_{i,\bar{t}-1}(\theta)$, violating the fact that $\bar{t} < t_{p_{k+1}}$. The above reasoning suggests that $\hat{\mu}_{j_{p_k}i,t}(\theta) = \mu_{j_{p_k},t_{p_k}}(\theta), \forall t \in (t_{p_k}, t_{p_{k+1}}]$. Moreover, since $\bar{\mu}_{i,t}(\theta)$ does not decrease at $\bar{t}$ (as $\bar{t} < t_{p_{k+1}}$), we have from (23) that $\mu_{i,t}(\theta) \geq \bar{\mu}_{i,t_{p_k}}(\theta) = \mu_{j_{p_k},t_{p_k}}(\theta), \forall t \in (t_{p_k}, t_{p_{k+1}}]$. It follows from the preceding discussion that (4) gets violated at $t_{p_{k+1}}$, and hence $\mathbb{1}_{ij_{p_k},t_{p_{k+1}}}(\theta) \neq 1$. The above arguments readily carry over to the case when there are an arbitrary number of event-monitoring time-steps in the interval $(t_{p_k}, t_{p_{k+1}})$. Thus, we omit such details.

We conclude that over each interval of the form $(t_{p_k}, t_{p_{k+1}}], k \in \mathbb{N}_+$, there exists a neighbor $j_{p_k} \in \mathcal{N}_i$ to which agent i does not broadcast its belief on θ. We can obtain one such t_{p_1} for each $i \notin \mathcal{S}(\theta^*, \theta)$, and take the maximum of such time-steps to obtain $T_2(\omega)$. □

Proof. (**Proposition 2**) Let us fix $\theta \neq \theta^*$, and partition the set of agents $\mathcal{V} \setminus \{v_\theta\}$ based on their distances from v_θ. Accordingly, we use $\mathcal{L}_q(\theta)$ to represent level-q agents that are at distance q from v_θ, where $q \in \mathbb{N}_+$. Let the agent(s) that are farthest from v_θ be at level $\bar{q}$. Now consider any agent $i \in \mathcal{L}_{\bar{q}}(\theta)$. Based on the conditions of the proposition, note that $i \notin \mathcal{S}(\theta^*, \theta)$, and the only neighbor of i is its parent in the tree rooted at v_θ, denoted by $p_i(\theta)$. Thus, claim (ii) of Proposition 1 applies to agent i, implying that agent i stops broadcasting its belief on θ to $p_i(\theta)$ eventually almost surely. Next, consider an agent $j \in \mathcal{L}_{\bar{q}-1}(\theta)$. We have already argued that after a finite number of time-steps, j will stop hearing broadcasts about θ from its children in level $\bar{q}$. Thus, for large enough k, $\mathcal{N}_{j,t_k}(\theta)$ can only comprise of $p_j(\theta)$, namely the parent of agent j in level $\bar{q} - 2$. In particular, given that $j \notin \mathcal{S}(\theta^*, \theta)$, the decrease in $\bar{\mu}_{j,t}(\theta)$ at time-steps defined by (25) can only be caused by $p_j(\theta)$. It then readily follows from the proof of Proposition 1 that j will stop broadcasting $\mu_{j,t}(\theta)$ to $p_j(\theta)$ eventually almost surely. To complete the proof, we can keep repeating the above argument till we reach level 1. □

Proof. (**Theorem 2**) The proof of this result is similar in spirit to that of Theorem 1. Hence, we only sketch the essential details. We begin by noting that the claims in Lemma 1 hold under the conditions of the theorem - this can be easily verified. Let $\bar{\Omega}$ have the same meaning as in Lemma 2. Fix $\omega \in \bar{\Omega}$ and an arbitrarily small $\epsilon > 0$. Since $\mathbb{P}^{\theta^*}(\bar{\Omega}) = 1$, to prove the result, it suffices to argue that for each false hypothesis $\theta \neq \theta^*$, $\exists T(\omega, \theta, \epsilon)$ such that on ω, $\mu_{i,t}(\theta) < \epsilon, \forall t \geq T(\omega, \theta, \epsilon), \forall i \in \mathcal{V}$. Recall that based on Lemma 1, there exists a time-step $t'(\omega) < \infty$, and a constant $\eta(\omega) > 0$, such that on ω, $\pi_{i,t}(\theta^\star) \geq \eta(\omega), \bar{\mu}_{i,t}(\theta^\star) \geq \eta(\omega), \forall t \geq t'(\omega), \forall i \in \mathcal{V}$. Set $\bar{\epsilon}(\omega) = \min\{\epsilon, \gamma\eta(\omega)\}$. Also, from Lemma 1, we know that there exists $\bar{t}$ such that $\mu_{i,t}(\theta) < \bar{\epsilon}^{|\mathcal{V}|}, \forall t \geq \bar{t}, \forall i \in \mathcal{S}(\theta^*, \theta)$.[7] Let $\tilde{t}_0 = \max\{t', \bar{t}\}$. Since the union graph over $[\tilde{t}_0, \infty)$ is rooted at $\mathcal{S}(\theta^*, \theta)$, there exists a set $\mathcal{F}_1(\theta) \in \mathcal{V} \setminus \mathcal{S}(\theta^*, \theta)$ of agents such that each agent in $\mathcal{F}_1(\theta)$ has at least one neighbor in $\mathcal{S}(\theta^*, \theta)$ in the union graph. Accordingly, consider any $j \in \mathcal{F}_1(\theta)$, and suppose $j \in \mathcal{N}_i(\tau)$, for some $i \in \mathcal{S}(\theta^*, \theta)$, and some $\tau \geq \tilde{t}_0$. The cases $\mathbb{1}_{ij,\tau}(\theta) = 1$ and $\mathbb{1}_{ij,\tau}(\theta) \neq 1$ can be analyzed exactly as in the proof of Lemma 2 to yield:

$$\mu_{j,t}(\theta) < \frac{\bar{\epsilon}^{|\mathcal{V}|}}{\eta\gamma} \leq \bar{\epsilon}^{(|\mathcal{V}|-1)}, \forall t > \tau, \tag{26}$$

where the last inequality follows by noting that $\bar{\epsilon} \leq \eta\gamma$. Let $\tilde{t}_1 > \tilde{t}_0$ be the first time-step by which every agent in $\mathcal{F}_1(\theta)$ has had at least one neighbor in $\mathcal{S}(\theta^*, \theta)$. Then, based on the above reasoning, $\mu_{j,t}(\theta) < \bar{\epsilon}^{(|\mathcal{V}|-1)}, \forall t > \tilde{t}_1, \forall j \in \mathcal{F}_1(\theta)$. If $\mathcal{V} \setminus \{\mathcal{S}(\theta^*, \theta) \cup \mathcal{F}_1(\theta)\} = \emptyset$, then we are done. Else, given the fact that the union graph over $[\tilde{t}_1, \infty)$ is rooted at $\mathcal{S}(\theta^*, \theta)$, there must exist a non-empty set $\mathcal{F}_2(\theta)$ such that each agent in $\mathcal{F}_2(\theta)$ has at least one neighbor from the set $\mathcal{S}(\theta^*, \theta) \cup \mathcal{F}_1(\theta)$ in the union graph. Reasoning as before, one can conclude that there exists a time-step $\tilde{t}_2 > \tilde{t}_1$ such that $\mu_{j,t}(\theta) < \bar{\epsilon}^{(|\mathcal{V}|-2)}, \forall t > \tilde{t}_2, \forall j \in \mathcal{F}_2(\theta)$. To complete the proof, we can keep repeating the above construction till we exhaust the entire vertex set $\mathcal{V}$. □

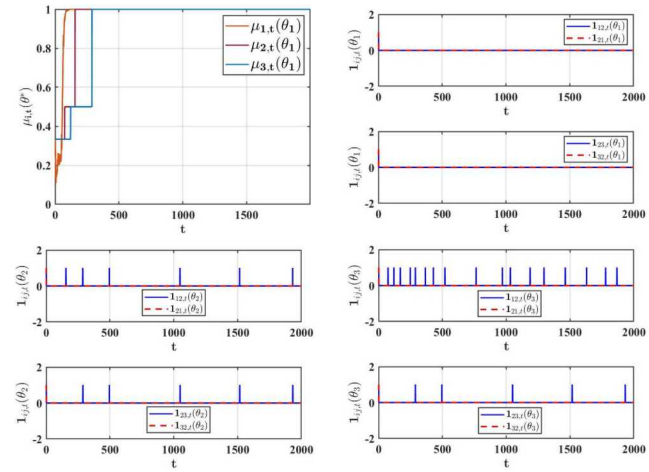

Fig. 2. Plots demonstrating the belief-evolutions and communication patterns for the network in Fig. 1.

VI. SIMULATIONS

In this section, we validate our theoretical findings via a simple simulation example. To do so, we consider the network in Fig. 1. We observe from Fig. 2 that (i) all agent learn the true state; (ii) all agents stop broadcasting about the true state θ_1 after the first time-step; and (iii) all broadcasts along the path $3 \to 2 \to 1$ stop after the first time-step.

[7]As before, we have suppressed dependence of various quantities on ω, θ, and ϵ, since they can be inferred from context.

VII. CONCLUSION

We introduced a new event-triggered distributed learning rule and identified conditions under which it leads to exponentially fast learning of the true state. In particular, we identified sparse communication regimes where one can recover the best known learning rate in the existing literature. We then demonstrated, both in theory and in simulations, that our event-triggered scheme has the ability to reduce information flow from uninformative agents to informative agents in the network.

REFERENCES

[1] A. Jadbabaie, P. Molavi, A. Sandroni, and A. Tahbaz-Salehi, "Non-Bayesian social learning," *Games and Economic Behavior*, vol. 76, no. 1, pp. 210–225, 2012.

[2] A. Jadbabaie, P. Molavi, and A. Tahbaz-Salehi, "Information heterogeneity and the speed of learning in social networks," *Columbia Bus. Sch. Res. Paper*, pp. 13–28, 2013.

[3] S. Shahrampour, A. Rakhlin, and A. Jadbabaie, "Distributed detection: Finite-time analysis and impact of network topology," *IEEE Trans. on Autom. Control*, vol. 61, no. 11, pp. 3256–3268, 2016.

[4] A. Nedić, A. Olshevsky, and C. A. Uribe, "Fast convergence rates for distributed Non-Bayesian learning," *IEEE Trans. on Autom. Control*, vol. 62, no. 11, pp. 5538–5553, 2017.

[5] A. Lalitha, T. Javidi, and A. Sarwate, "Social learning and distributed hypothesis testing," *IEEE Trans. on Info. Theory*, vol. 64, no. 9, 2018.

[6] A. Mitra, J. A. Richards, and S. Sundaram, "A new approach for distributed hypothesis testing with extensions to Byzantine-resilience," in *Proc. of the American Control Conference*, 2019, pp. 261–266.

[7] ——, "A new approach to distributed hypothesis testing and non-bayesian learning: Improved learning rate and byzantine-resilience," *arXiv:1907.03588*, 2019.

[8] P. Tabuada, "Event-triggered real-time scheduling of stabilizing control tasks," *IEEE Transactions on Automatic Control*, vol. 52, no. 9, pp. 1680–1685, 2007.

[9] W. Heemels, K. H. Johansson, and P. Tabuada, "An introduction to event-triggered and self-triggered control," in *2012 IEEE 51st IEEE Conference on Decision and Control (CDC)*. IEEE, 2012, pp. 3270–3285.

[10] C. Nowzari, E. Garcia, and J. Cortés, "Event-triggered communication and control of networked systems for multi-agent consensus," *Automatica*, vol. 105, pp. 1–27, 2019.

[11] T. M. Cover and J. A. Thomas, *Elements of information theory*. John Wiley & Sons, 2012.