SOCIAL LEARNING WITH PARTIAL INFORMATION SHARING

Virginia Bordignon[★]   *Vincenzo Matta*[†]   *Ali H. Sayed*[★]

[★] School of Engineering, Ecole Polytechnique Fédérale de Lausanne (EPFL)
[†]DIEM, University of Salerno

ABSTRACT

This work studies the learning abilities of agents sharing partial beliefs over social networks. The agents observe data that could have risen from one of several hypotheses and interact locally to decide whether the observations they are receiving have risen from a particular hypothesis of interest. To do so, we establish the conditions under which it is sufficient to share *partial* information about the agents' belief in relation to the hypothesis of interest. Some interesting convergence regimes arise.

Index Terms— Social learning, partial information, Bayesian update, diffusion strategy.

1. INTRODUCTION AND RELATED WORK

Modeling opinion formation over social networks is a subject of great interest, including in modern times with the proliferation of social platforms. Many algorithmic approaches have been conceived for this purpose [1–4], including the non-Bayesian approach, in which agents update their beliefs or opinions by using local streaming observations and by combining information shared by their neighbors. Some of the main studies along these lines rely on consensus and diffusion strategies [5, 6], both with linear and log-exponential belief combinations (see, e.g., [7–12]). In all of these works, it is assumed that agents share with their neighbors their *entire* belief vectors, which can be an unrealistic assumption since in many cases the agents are more likely to exchange their opinions on specific questions/hypotheses.

In this work, we consider that agents share only *partial* beliefs. We motivate the model by means of an example. Consider a collection of agents connected by a topology and monitoring the weather conditions in a certain region of space. Each agent is measuring some local data that is dependent on the weather conditions, say, temperature, pressure, wind speed and direction. The agents wish to interact to decide whether the state of nature is one of three possibilities: sunny, rainy or snowing. At every time instant i, each agent k will have a belief vector, say, $\mu_{k,i}(\theta) = [0.6, 0.2, 0.2]$. These values mean that, at time i, agent k believes it is sunny with probability 0.6. Similarly, all other agents in the network will have their own beliefs. If the weather conditions are sunny, we would like the agents to converge after sufficient iterations to the belief vector $[1, 0, 0]$. This type of convergence behavior has already been established in prior studies [7, 12] for strongly-connected networks (i.e., over networks where there is always a path and a reverse path between any two agents, in addition to at least one agent having a self-loop to reflect confidence in their local data).

But what if we now formulate a different question for the network to answer? Assume the objective is for the network to decide "whether it is sunny," rather than try to discover the true state of nature (which could be sunny, rainy, or snowing. Will the agents still need to share their *entire* belief vectors repeatedly to find out

whether it is sunny or not? What if we devise a cooperation scheme where agents share only, at every iteration, the likelihood they have for the "sunny" condition and ignore the other entries in their belief vector. Will they still be able to learn if it is sunny? That is, can agents still learn if they only share partial information without needing to share simultaneously information about *all* possible hypotheses?

This work answers these questions and provides conditions under which different interesting forms of convergence regimes can occur. We will consider that agents are not willing to share information on all hypotheses, while aiming to answer a binary question regarding one specific hypothesis of interest. To model this choice, we present in Eqs. (2)–(5) below the learning algorithm by which agents share partial information in relation to the particular question that the network is trying to answer.

Theoretical results are provided for two alternatives: i) the *partial* approach, where agents are so focused on the partially shared opinion that they segment also their own belief; ii) and the *self-aware partial* approach where each agent takes into account its own true belief vector. Due to space limitations some proofs are omitted. The analysis of both algorithms points toward interesting behavior, which complements our understanding of social learning systems. It also motivates the enunciation of a third algorithm, which will be merely introduced in this work: it consists of agents sharing their most relevant belief component at each time.

2. PARTIAL INFORMATION

In this work, we consider a strongly connected network of N agents interacting to learn an unknown state of nature $\theta \in \Theta$. Each agent k observes, at instant i, streaming data $\boldsymbol{\xi}_{k,i} \in \mathcal{X}$ (we use bold notation for random variables), which can be either continuous or discrete random variables assumed independent across time. The likelihood of the data corresponding to a given hypothesis θ will be denoted by $L(\xi|\theta)$, for $\xi \in \mathcal{X}$. The data $\boldsymbol{\xi}_{k,i}$ are distributed according to $L(\xi|\theta_0)$, where θ_0 is the true hypothesis.

We model the topology by means of a strongly-connected graph with a left-stochastic combination matrix A whose nonnegative entries $a_{\ell k}$ are equal to zero if $\ell \notin \mathcal{N}_k$, with $\mathcal{N}_k$ being the set of neighbors of agent k. We denote by v the Perron eigenvector of A, which is defined as

$$Av = v, \qquad \mathbb{1}^\top v = 1, \qquad v \succ 0. \tag{1}$$

Assumption 1 (Positive initial beliefs)**.** *In the absence of any prior information, all agents assign positive initial beliefs to all hypotheses, i.e.,* $\boldsymbol{\mu}_{k,0}(\theta) > 0$ *for each agent k and all $\theta \in \Theta$.* □

Assumption 2 (Likelihood ratios)**.** *For each agent k and $i = 1, 2, \ldots,$ the random variables* $\log\left(L(\boldsymbol{\xi}_{k,i}|\theta)/L(\boldsymbol{\xi}_{k,i}|\theta')\right)$ *are integrable for every pair* $\{\theta, \theta'\} \in \Theta.$ □

2.1. Partial approach

Let θ_{TX} denote a hypothesis of interest[1]. The objective for the agents is to verify whether the state of nature agrees with θ_{TX} or not. Thus, we propose the following modified version of the log-linear social learning diffusion construction used in [12, 13]:

$$\boldsymbol{\psi}_{k,i}(\theta) = \frac{\boldsymbol{\mu}_{k,i-1}(\theta)L(\boldsymbol{\xi}_{k,i}|\theta)}{\sum_{\theta'\in\Theta}\boldsymbol{\mu}_{k,i-1}(\theta')L(\boldsymbol{\xi}_{k,i}|\theta')}, \tag{2}$$

$$\widehat{\boldsymbol{\psi}}_{k,i}(\theta) = \begin{cases} \boldsymbol{\psi}_{k,i}(\theta), & \text{for } \theta = \theta_{\mathsf{TX}}, \\ \dfrac{1-\boldsymbol{\psi}_{k,i}(\theta_{\mathsf{TX}})}{H-1}, & \text{for } \theta \neq \theta_{\mathsf{TX}}, \end{cases} \tag{3}$$

$$\boldsymbol{\mu}_{k,i}(\theta) = \frac{\exp\left\{\sum_{\ell=1}^{N} a_{\ell k}\log\widehat{\boldsymbol{\psi}}_{\ell,i}(\theta)\right\}}{\sum_{\theta'\in\Theta}\exp\left\{\sum_{\ell=1}^{N} a_{\ell k}\log\widehat{\boldsymbol{\psi}}_{\ell,i}(\theta')\right\}}. \tag{4}$$

In (2), the agents perform a Bayesian update to construct an intermediate belief vector $\boldsymbol{\psi}_{k,i}$. In order to obtain the final belief vector $\boldsymbol{\mu}_{k,i}$, we consider the usual log-linear combination rule seen in (4). However, each agent k shares only the component $\boldsymbol{\psi}_{k,i}(\theta_{\mathsf{TX}})$. The receiving agents will then split the remaining mass $1-\boldsymbol{\psi}_{k,i}(\theta_{\mathsf{TX}})$ uniformly across the values $\theta \neq \theta_{\mathsf{TX}}$. For this reason, the combination rule is applied to the modified belief vectors $\{\widehat{\boldsymbol{\psi}}_{\ell,i}\}_{\ell=1}^{N}$ in (3).

2.2. Self-aware partial approach

The second approach consists in rewriting the combination step of the algorithm in such a way that agent k combines its neighbors modified beliefs $\{\widehat{\boldsymbol{\psi}}_{\ell,i}\}_{\ell=1,\ell\neq k}^{N}$ with its own true belief $\boldsymbol{\psi}_{k,i}$:

$$\boldsymbol{\mu}_{k,i}(\theta) = \frac{\exp\left\{a_{kk}\log\boldsymbol{\psi}_{k,i}(\theta) + \sum_{\substack{\ell=1\\\ell\neq k}}^{N} a_{\ell k}\log\widehat{\boldsymbol{\psi}}_{\ell,i}(\theta)\right\}}{\sum_{\theta'\in\Theta}\exp\left\{a_{kk}\log\boldsymbol{\psi}_{k,i}(\theta') + \sum_{\substack{\ell=1\\\ell\neq k}}^{N} a_{\ell k}\log\widehat{\boldsymbol{\psi}}_{\ell,i}(\theta')\right\}}. \tag{5}$$

We can distinguish in (5) two terms in its numerator: a first *self-awareness* term and a second term combining the partially true beliefs from neighbors.

3. CONVERGENCE ANALYSIS

3.1. Partial approach

In this approach, agent k chooses to combine its own modified belief vector $\widehat{\boldsymbol{\psi}}_{k,i}$ with its neighbors' equally modified beliefs. Lemma 1 states a result regarding the asymptotic rate of convergence while Theorem 1 establishes the convergence behavior of belief vectors, under different regimes. Note that $D_{\mathsf{KL}}[\cdot]$ refers to the Kullback-Leibler (KL) divergence. In the following, when computing KL divergences the argument ξ of the pertinent distributions is omitted for ease of notation, and a.s. refers to almost sure convergence. Due to space constraints, proofs will be partially introduced.

Lemma 1 (Asymptotic rate of convergence). *Under Assumptions 1 and 2, we have that, for any k and for all $\theta \neq \theta_{\mathsf{TX}}$:*

$$\lim_{i\to\infty}\frac{1}{i}\log\frac{\boldsymbol{\mu}_{k,i}(\theta)}{\boldsymbol{\mu}_{k,i}(\theta_{\mathsf{TX}})} \stackrel{\text{a.s.}}{=} D_{\mathsf{KL}}[L(\theta_0)||L(\theta_{\mathsf{TX}})]$$

$$- D_{\mathsf{KL}}[L(\theta_0)||P(\theta_{\mathsf{TX}}^c)], \tag{6}$$

where we define the average likelihood function distinct from θ as

$$P(\xi|\theta^c) \triangleq \frac{1}{H-1}\sum_{\tau\neq\theta} L(\xi|\tau). \tag{7}$$

Proof of Lemma 1. Note that (4) is equivalent to writing, for every pair $\theta, \theta' \in \Theta$:

$$\log\frac{\boldsymbol{\mu}_{k,i}(\theta)}{\boldsymbol{\mu}_{k,i}(\theta')} = \sum_{\ell=1}^{N} a_{\ell k}\log\frac{\widehat{\boldsymbol{\psi}}_{k,i}(\theta)}{\widehat{\boldsymbol{\psi}}_{k,i}(\theta')}. \tag{8}$$

We see from (3) that, for all $\theta, \theta' \neq \theta_{\mathsf{TX}}$,

$$\log\frac{\boldsymbol{\mu}_{k,i}(\theta)}{\boldsymbol{\mu}_{k,i}(\theta')} = 0 \Rightarrow \boldsymbol{\mu}_{k,i}(\theta) = \boldsymbol{\mu}_{k,i}(\theta'). \tag{9}$$

Considering (9) and the recursion in (8) with $\theta' = \theta_{\mathsf{TX}}$ results in

$$\log\frac{\boldsymbol{\mu}_{k,i}(\theta)}{\boldsymbol{\mu}_{k,i}(\theta_{\mathsf{TX}})} = \sum_{\ell=1}^{N} a_{\ell k}\left[\log\frac{\boldsymbol{\mu}_{\ell,i-1}(\theta)}{\boldsymbol{\mu}_{\ell,i-1}(\theta_{\mathsf{TX}})} + \log\frac{P(\boldsymbol{\xi}_{\ell,i}|\theta_{\mathsf{TX}}^c)}{L(\boldsymbol{\xi}_{\ell,i}|\theta_{\mathsf{TX}})}\right]. \tag{10}$$

Multiplying by $1/i$, taking the limit over i as it goes to infinity, and using limiting arguments as in [11], we get

$$\lim_{i\to\infty}\frac{1}{i}\log\frac{\boldsymbol{\mu}_{k,i}(\theta)}{\boldsymbol{\mu}_{k,i}(\theta_{\mathsf{TX}})} \stackrel{\text{a.s.}}{=} \sum_{\ell=1}^{N}\underbrace{v_\ell}_{=1}\mathbb{E}\left[\log\frac{P(\boldsymbol{\xi}_{\ell,i}|\theta_{\mathsf{TX}}^c)}{L(\boldsymbol{\xi}_{\ell,i}|\theta_{\mathsf{TX}})}\right], \tag{11}$$

which establishes the desired conclusion. ∎

Theorem 1 (Belief collapse). *Under Assumptions 1 and 2, for any k, we have that:*
If $\theta_{\mathsf{TX}} = \theta_0$,

$$D_{\mathsf{KL}}[L(\theta_0)||P(\theta_0^c)] > 0 \Rightarrow \lim_{i\to\infty}\boldsymbol{\mu}_{k,i}(\theta_0) \stackrel{\text{a.s.}}{=} 1. \tag{12}$$

If conversely $\theta_{\mathsf{TX}} \neq \theta_0$,

$$\frac{D_{\mathsf{KL}}[L(\theta_0)||P(\theta_{\mathsf{TX}}^c)]}{D_{\mathsf{KL}}[L(\theta_0)||L(\theta_{\mathsf{TX}})]} > 1 \Rightarrow \lim_{i\to\infty}\boldsymbol{\mu}_{k,i}(\theta_{\mathsf{TX}}) \stackrel{\text{a.s.}}{=} 1, \tag{13}$$

and

$$\frac{D_{\mathsf{KL}}[L(\theta_0)||P(\theta_{\mathsf{TX}}^c)]}{D_{\mathsf{KL}}[L(\theta_0)||L(\theta_{\mathsf{TX}})]} < 1 \Rightarrow \lim_{i\to\infty}\boldsymbol{\mu}_{k,i}(\theta) \stackrel{\text{a.s.}}{=} \frac{1}{H-1}, \tag{14}$$

for all $\theta \neq \theta_{\mathsf{TX}}$.

Proof of Theorem 1. When $\theta_{\mathsf{TX}} = \theta_0$, if the inequality in (12) holds, the convergence rate from Lemma 1 will be strictly negative, which implies that, since $\boldsymbol{\mu}_{k,i}(\theta_0)$ is bounded by 1, for any $\theta \neq \theta_0$, $\lim_{i\to\infty}\boldsymbol{\mu}_{k,i}(\theta) \stackrel{\text{a.s.}}{=} 0$, resulting in (12). When $\theta_{\mathsf{TX}} \neq \theta_0$, the convergence behavior will depend on the sign of the RHS of (6). If the inequality in (13) holds, the sign is negative and we get the convergence in (13). If the inequality in (14) holds, the sign is positive, which implies that $\boldsymbol{\mu}_{k,i}(\theta_{\mathsf{TX}})$ goes to zero, and all $\boldsymbol{\mu}_{k,i}(\theta)$ are equal in view of (9). ∎

Theorem 1 reveals that when the transmitted hypothesis θ_{TX} is equal to the true one, all agents are able to recover the truth. Moreover, in the alternative scenario, two situations can be produced: either all agents will converge to a belief concentrated at the hypothesis θ_{TX}, or they will converge to a uniform splitting across the non-transmitted hypotheses. This behavior admits a very useful interpretation. When (13) is verified, the transmitted hypothesis is more easily confounded with the true one (since their KL divergence is small). In comparison, when (14) is verified, there is no sufficient evidence for choosing θ_{TX}, but since in view of (9) all the unshared beliefs remain equal, the only meaningful choice of the agents is to split equally their beliefs across the remaining $H-1$ hypotheses.
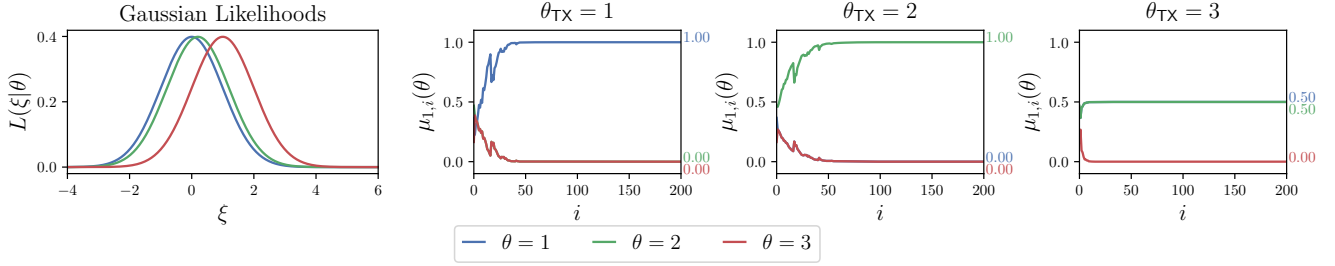
Fig. 2. *Partial approach*: Belief evolution for agent 1 for different choices of θ_{TX}.

3.2. Self-aware partial approach

Now, we consider that agent k will combine its true belief vector $\boldsymbol{\psi}_{k,i}$ with its neighbors' modified belief vectors $\{\widehat{\boldsymbol{\psi}}_{\ell,i}\}_{\ell=1,\ell\neq k}^N$. When $\theta_{\mathsf{TX}} = \theta_0$, we have the result stated in Theorem 2. For this approach, proofs are omitted due to space constraints.

Theorem 2 (Truth learning with self-awareness). *Under Assumptions 1 and 2, when $\theta_{\mathsf{TX}} = \theta_0$, if $D_{\mathsf{KL}}\left[L(\theta_0)||P_q(\theta_0^c)\right] \neq 0$ for any pmf $q(\cdot)$ over $\Theta \backslash \theta_0$ such that we define*

$$P_q(\xi|\theta_0^c) \triangleq \sum_{\tau \neq \theta_0} q(\tau)L(\xi|\tau), \quad (15)$$

then for any k, we have

$$\lim_{i \to \infty} \boldsymbol{\mu}_{k,i}(\theta_0) \stackrel{\text{a.s.}}{=} 1. \quad (16)$$

∎

On the other hand, when $\theta_{\mathsf{TX}} \neq \theta_0$, two situations may arise: the belief concerning the transmitted hypothesis will collapse either to zero or to one, under the sufficient conditions given in Lemmas 2 and 3, respectively.

Lemma 2 (Alternative behavior with self-awareness). *Under Assumption 1 and 2, when $\theta_{\mathsf{TX}} \neq \theta_0$, for any k, if*

$$D_{\mathsf{KL}}\left[L(\theta_0)||L(\theta_{\mathsf{TX}})\right] > \frac{1}{H-1} \sum_{\tau \neq \theta_{\mathsf{TX}}} D_{\mathsf{KL}}\left[L(\theta_0)||L(\tau)\right], \quad (17)$$

then

$$\lim_{i \to \infty} \boldsymbol{\mu}_{k,i}(\theta_{\mathsf{TX}}) \stackrel{\text{a.s.}}{=} 0. \quad (18)$$

∎

Under the sufficient condition in Lemma 2, we discover that the belief concerning θ_{TX} collapses to zero with probability one. As for the remaining elements, it can be shown that they will exhibit an oscillatory behavior according to the recursion

$$\log \frac{\boldsymbol{\mu}_{k,i}(\theta)}{\boldsymbol{\mu}_{k,i}(\theta')} = a_{kk} \log \frac{\boldsymbol{\mu}_{k,i-1}(\theta)}{\boldsymbol{\mu}_{k,i-1}(\theta')} + a_{kk} \log \frac{L(\boldsymbol{\xi}_{k,i}|\theta)}{L(\boldsymbol{\xi}_{k,i}|\theta')} \quad (19)$$

for $\theta, \theta' \neq \theta_{\mathsf{TX}}$. This scenario differs from the partial approach, where the above recursion is given instead by (9), where the beliefs for non-transmitted hypotheses evolve equally.

Before presenting Lemma 3, we introduce an extra assumption on the boundedness of the likelihood functions.

Assumption 3 (Bounded likelihoods). *Let there be $M > 0$, such that, for all $\xi \in \mathcal{X}$ and for any pair $\theta, \theta' \neq \theta_{\mathsf{TX}}$, we have:*

$$-M \leq \log \frac{L(\xi|\theta)}{L(\xi|\theta')} \leq M. \quad (20)$$

□

Lemma 3 (Mislearning with self-awareness). *Under Assumptions 1, 2 and 3, for any k, when $\theta_{\mathsf{TX}} \neq \theta_0$, if*

$$D_{\mathsf{KL}}\left[L(\theta_0)||P(\theta_{\mathsf{TX}}^c)\right] > D_{\mathsf{KL}}\left[L(\theta_0)||L(\theta_{\mathsf{TX}})\right] + M \sum_{\ell=1}^N v_\ell a_{\ell\ell}, \quad (21)$$

then

$$\lim_{i \to \infty} \boldsymbol{\mu}_{k,i}(\theta_{\mathsf{TX}}) \stackrel{\text{a.s.}}{=} 1. \quad (22)$$

∎

Lemma 3 reveals that, for some sufficiently small choice of self-awareness weights $a_{\ell\ell}$ for every $\ell = 1, 2, \ldots, N$, as long as $D_{\mathsf{KL}}\left[L(\theta_0)||P(\theta_{\mathsf{TX}}^c)\right] > D_{\mathsf{KL}}\left[L(\theta_0)||L(\theta_{\mathsf{TX}})\right]$ holds, the belief concerning the transmitted hypothesis θ_{TX} will converge to one, while all others will go to zero. The sufficient condition in (21) is expected to be not tight because Lemma 3 relies on the bounds in (20), which are usually not tight. This condition is nonetheless useful since it highlights the possibility that mislearning occurs.

4. SIMULATION RESULTS

Consider a network, whose topology can be seen in Fig. 1 (self-loops are present at every node and are not displayed) and whose combination matrix A is determined using a parametrized averaging rule [14]:

$$a_{\ell k} = \begin{cases} \lambda, & \text{if } \ell = k \\ \dfrac{1-\lambda}{n_k - 1}, & \text{if } \ell \neq k \text{ and } \ell \in \mathcal{N}_k \\ 0, & \text{otherwise,} \end{cases} \quad (23)$$

where λ is a hyperparameter used to tune self-awareness and n_k is the degree of agent k (including k itself). The set of hypotheses is given by $\Theta = \{1, 2, 3\}$, and the true state of nature corresponds to the first hypothesis, i.e., $\theta_0 = 1$.
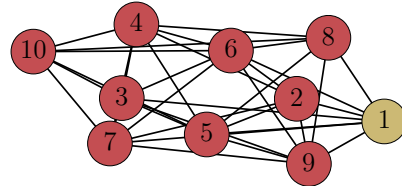

Fig. 1. Network topology, where agent 1 is highlighted.

4.1. Partial approach

We first consider the partial approach, with hyperparameter $\lambda = 0.5$. Assume the same family of unit-variance Gaussian likelihood functions across all agents, whose means are chosen as 0, 0.2 and 1 for $\theta = 1$, $\theta = 2$ and $\theta = 3$ respectively, as seen in the left part of Fig. 2. We will focus on the behavior of agent 1 for illustration purposes. As predicted by Theorem 1, when $\theta_{\mathsf{TX}} = \theta_0 = 1$, we observe that the agent's belief collapses with full confidence
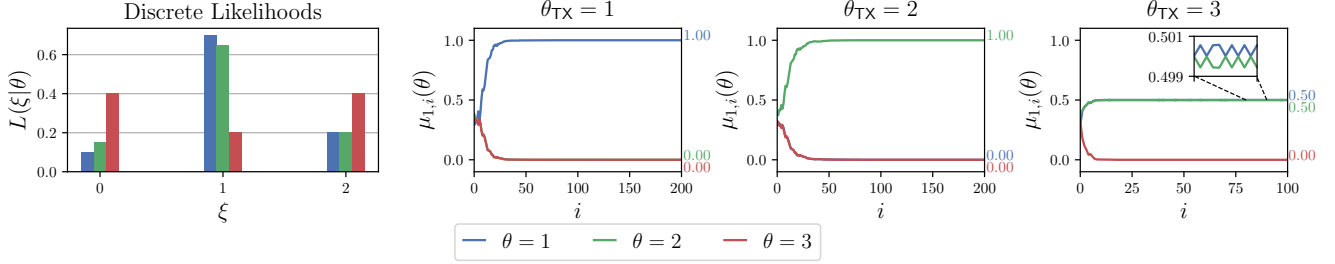
Fig. 3. *Self-aware partial approach:* Belief evolution for agent 1 for different choices of θ_{TX}.

to the true hypothesis as seen in Fig. 2. When $\theta_{\mathsf{TX}} = 2$, since $D_{\mathsf{KL}}[L(\theta_0)||L(\theta_{\mathsf{TX}})] - D_{\mathsf{KL}}[L(\theta_0)||P(\theta_{\mathsf{TX}}^c)] = -0.091 < 0$, the agent's belief vector converges with full confidence to the shared hypothesis. Alternatively, when $\theta_{\mathsf{TX}} = 3$, $D_{\mathsf{KL}}[L(\theta_0)||L(\theta_{\mathsf{TX}})] - D_{\mathsf{KL}}[L(\theta_0)||P(\theta_{\mathsf{TX}}^c)] = 0.494 > 0$, so the agent's belief is equally split among all hypotheses different than the shared one.

4.2. Self-aware partial approach

For this approach, we consider $\lambda = 0.03$ and a family of likelihood functions with discrete support over ξ, i.e. $\xi \in \{0, 1, 2\}$, as seen in the left panel of Fig. 3. In this scenario, Assumption 3 is satisfied, and, more particularly, the constant M from (20) is equal to 3.5.

When $\theta_{\mathsf{TX}} = \theta_0 = 1$, illustrating Theorem 2, we observe almost sure convergence to the true hypothesis, as seen in Fig. 3. Some interesting phenomena arise from observing the evolution of beliefs when $\theta_{\mathsf{TX}} \neq \theta_0$.

When $\theta_{\mathsf{TX}} = 2$, $D_{\mathsf{KL}}[L(\theta_0)||P(\theta_{\mathsf{TX}}^c)] - D_{\mathsf{KL}}[L(\theta_0)||L(\theta_{\mathsf{TX}})] + M \sum_{\ell=1}^{N} v_\ell a_{\ell\ell} = 0.02 > 0$ and therefore Lemma 3 is satisfied. In simulation, for larger values of λ, even though the previous condition no longer holds, we still observe that $\lim_{i\to\infty} \boldsymbol{\mu}_{k,i}(\theta_{\mathsf{TX}}) = 1$, which is due to the fact that Lemma 3 only establishes a sufficient condition. Another interesting observation from the simulation is that when λ is sufficiently large, i.e., the self-awareness term is dominant, we are able to recover the alternative behavior for which $\lim_{i\to\infty} \boldsymbol{\mu}_{k,i}(\theta_{\mathsf{TX}}) = 0$.

When $\theta_{\mathsf{TX}} = 3$, we observe that $D_{\mathsf{KL}}[L(\theta_0)||L(\theta_{\mathsf{TX}})] - \frac{1}{H-1}\sum_{\tau\neq\theta_{\mathsf{TX}}} D_{\mathsf{KL}}[L(\theta_0)||L(\tau)] = 0.094 > 0$ and therefore Lemma 2 holds and no choice of λ results in a change of regime, i.e., $\boldsymbol{\mu}_{k,i}(\theta_{\mathsf{TX}})$ will always converge to zero almost surely. This behavior is seen in Fig. 2, where the beliefs concerning hypotheses $\theta = 1$ and $\theta = 2$ oscillate indefinitely, according to recursion (19).

In this case, the self-awareness parameter λ will influence the behavior of the beliefs concerning the non-transmitted hypotheses, more specifically, it will influence the amplitude of their oscillatory behavior. In Fig. 4, we see how increasing λ makes the oscillatory behavior of (19) increase.
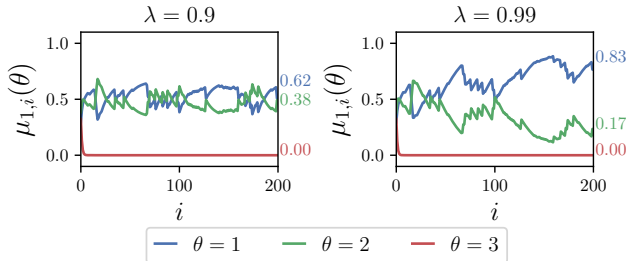

Fig. 4. Belief evolution for agent 1 with $\theta_{\mathsf{TX}} = 2$. *Left:* For $\lambda = 0.9$. *Right:* For $\lambda = 0.99$.

Besides, since the problem is identifiable at every agent, a growing self-awareness parameter brings the agents closer to the truth. Note that in the limit, when $\lambda \to 1$, we obtain a non-cooperative learning framework.

4.3. A third approach

So far we have considered two algorithms where the agents choose *deterministically* to share one hypothesis from among the available ones. Another interesting approach would be to consider instead the following intermediate step:

$$\widehat{\boldsymbol{\psi}}_{k,i}(\theta) = \begin{cases} \boldsymbol{\psi}_{k,i}(\theta), & \text{for } \theta = \boldsymbol{\theta}_{\mathsf{MAX}}^{k,i}, \\ \dfrac{1 - \boldsymbol{\psi}_{k,i}(\boldsymbol{\theta}_{\mathsf{MAX}}^{k,i})}{H - 1}, & \text{for } \theta \neq \boldsymbol{\theta}_{\mathsf{MAX}}^{k,i}, \end{cases} \quad (24)$$

where we defined the random variable $\boldsymbol{\theta}_{\mathsf{MAX}}^{k,i} \triangleq \arg\max_\theta (\boldsymbol{\psi}_{k,i}(\theta))$ as the most relevant hypothesis for agent k at instant i. The behavior of the algorithm without self-awareness, considering (24) as intermediate step can be seen in Fig. 5 for uniform initial beliefs and for random initializations over 100 Monte Carlo runs.
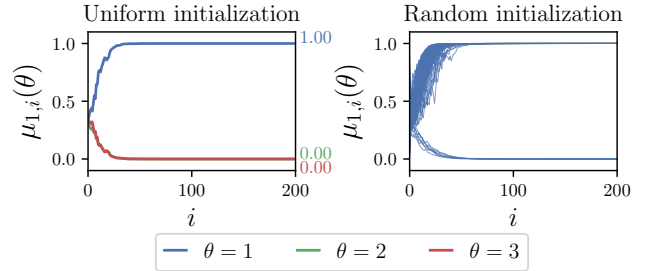

Fig. 5. Social learning with most relevant hypothesis sharing. *Left:* Uniform initialization. *Right:* Random initialization with 100 Monte Carlo experiments.

Observing the convergence results in Fig. 5, the following remarkable behavior emerges. First (left of Fig. 5), under uniform initialization, partial information with max-belief sharing can be enough to learn properly the true hypothesis. Moreover, under random initialization (right of Fig. 5), sometimes mislearning arises. We have verified on the data that, notably, mislearning is associated to realizations with poor initialization. In these realizations, $\boldsymbol{\theta}_{\mathsf{MAX}}^{k,i}$ tends to be equal to $\theta = 2$ in the first iterations, whose behavior matches the case seen in (13) of Theorem 1.

5. CONCLUSION

In this work we discussed social learning algorithms that incorporate the notion of partial information sharing. We have shown that if the presumed hypothesis is the true one, then the network is able to learn correctly even under the regime of partial information sharing. Otherwise, if there is sufficient KL divergence between the true likelihood and the likelihood under investigation, then the network is able to correctly discount the latter. However, and as would be expected, a mislearning scenario is possible if the KL divergence between true and presumed hypotheses is small enough, in which case the presumed hypothesis is confounded with the true one.

6. REFERENCES

[1] C. Chamley, A. Scaglione, and L. Li, "Models for the diffusion of beliefs in social networks: An overview," *IEEE Signal Processing Magazine*, vol. 30, no. 3, pp. 16–29, May 2013.

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

[3] P. Molavi, A. Tahbaz-Salehi, and A. Jadbabaie, "A theory of non-bayesian social learning," *Econometrica*, vol. 86, no. 2, pp. 445–490, 2018.

[4] N. E. Friedkin, A. V. Proskurnikov, R. Tempo, and S. E. Parsegov, "Network science on belief system dynamics under logic constraints," *Science*, vol. 354, no. 6310, pp. 321–326, 2016.

[5] M. H. DeGroot, "Reaching a consensus," *Journal of the American Statistical Association*, vol. 69, no. 345, pp. 118–121, 1974.

[6] A. H. Sayed, S. Tu, J. Chen, X. Zhao, and Z. J. Towfic, "Diffusion strategies for adaptation and learning over networks," *IEEE Signal Processing Magazine*, vol. 30, no. 3, pp. 155–171, May 2013.

[7] 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.

[8] X. Zhao and A. H. Sayed, "Learning over social networks via diffusion adaptation," in *Proc. Asilomar Conference on Signals, Systems and Computers*, 2012, pp. 709–713.

[9] S. Shahrampour, A. Rakhlin, and A. Jadbabaie, "Distributed detection: Finite-time analysis and impact of network topology," *IEEE Transactions on Automatic Control*, vol. 61, no. 11, pp. 3256–3268, 2015.

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

[11] A. Nedić, A. Olshevsky, and C. Uribe, "Fast convergence rates for distributed non-bayesian learning," *IEEE Transactions on Automatic Control*, vol. 62, no. 11, pp. 5538–5553, 2017.

[12] A. Lalitha, T. Javidi, and A. D. Sarwate, "Social learning and distributed hypothesis testing," *IEEE Transactions on Information Theory*, vol. 64, no. 9, pp. 6161–6179, 2018.

[13] V. Matta, A. Santos, and A. H. Sayed, "Exponential collapse of social beliefs over weakly-connected heterogeneous networks," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brighton, UK, 2019, pp. 5267–5271.

[14] A. H. Sayed, "Diffusion adaptation over networks," in *Academic Press Library in Signal Processing*, vol. 3, pp. 323–453. Elsevier, 2014.