Complexity of Bayesian Belief Exchange over a Network

M. Amin Rahimian, Ali Jadbabaie, Elchanan Mossel *

***Abstract*—Many important real-world decision making problems involve group interactions among individuals with purely informational externalities, such situations arise for example in jury deliberations, expert committees, medical diagnosis, etc. In this paper, we will use the framework of iterated eliminations to model the decision problem as well as the thinking process of a Bayesian agent in a group decision/discussion scenario. We model the purely informational interactions of rational agents in a group, where they receive private information and act based upon that information while also observing other people's beliefs. As the Bayesian agent attempts to infer the true state of the world from her sequence of observations which include her neighbors' beliefs as well as her own private signal, she recursively refines her belief about the signals that other players could have observed and beliefs that they would have hold given the assumption that other players are also rational. We further analyze the computational complexity of the Bayesian belief formation in groups and show that it is $\mathcal{NP}$-hard. We also investigate the factors underlying this computational complexity and show how belief calculations simplify in special network structures or cases with strong inherent symmetries. We finally give insights about the statistical efficiency (optimality) of the beliefs and its relations to computational efficiency.**

***Index Terms*—Rational Choice theory, Computational social choice, Distributed hypothesis testing, Bayesian inference, Computational complexity, Bayesian computations**

I. Introduction

In decision theory, the seminal work of Aumann [1] studies the interactions of two rational agents with common prior beliefs and concludes that if the values of their posterior beliefs are common knowledge between the two agents, then the two values should be the same: rational agents cannot agree to disagree. The later work of Geanakoplos and Polemarchakis [2] investigates how rational agents reach an agreement by communicating back and forth and refining their information partitions. Following [1] and [2], a large body of literature studies strategic interaction of agents in a social network, where they receive private information and act based upon that information while also observing each other's actions [3], [4]. Although this model is motivated by the interactions of people in a group, the so-called group decision process (GDP), the same framework can be applied in the analysis of distributed estimation and consensus problems [5], [6], [7].

When making decisions under uncertainty, the agents form beliefs about unknown parameters of interest and their actions and decisions reflect their beliefs. In this paper, we focus on a case where agents announce their beliefs truthfully at every decision epoch; in practice, jurors may express their belief about the probability of guilt in a criminal case or more generally people may make statements that are expressive of their beliefs. In [8] Eyster and Rabin explain that rich-enough action spaces can reveal the underlying beliefs that lead to actions; subsequently, an individual's action is a fine reflection of her beliefs.

The rational approach advocates formation of Bayesian posterior beliefs by application of Bayes rule to the entire sequence of observations successively at every step. Calculating the Bayesian posterior belief in partially observed settings such as GDP is notoriously difficult as the Bayesian agent should reason about the third party interactions that precede (and are the causes of) her observations and are not observable to her. Moreover, when a rational agent observes her neighbors in a network, she should compensate for the repetitions in the sources of her information: several of the neighboring agents may have been influenced by the same source of information (type-I repetitions) and neighboring actions may have been affected by the past actions of the agent herself (type-II repetitions); all these lead to inter-dependencies and correlations in the information received from different neighbors. Failure to account for such structural dependencies subjects the agents to mistakes and inefficiencies such as redundancy neglect [8], and data incest [9], caused by type-I and type-II repetitions respectively.

In recent works, recursive techniques have been applied to analyze Bayesian decision problems with partial success [10], [11]. In this work, we are particularly interested in the computations that the Bayesian agent should undertake to achieve her goal of producing best recommendations/beliefs at every decision epoch during a group discussion. In this note, we will follow the framework of iterated eliminations to model the thinking process of a Bayesian agent in a group decision-making scenario as she refines her beliefs with the increasing history of observations. The iterated elimination approach that is describe in Section II curbs some of the complexities of GDP, but only to a limited extent. In a group decision scenario, the initial private signals of the agents constitute a search space that is exponential in the size of the network. The ultimate goal of the agents is to get informed about the private signals of each other, information that is reflected in their beliefs. The Bayesian agent is initially informed of only her own signal; however, as her history of interactions with other group members becomes enriched, her knowledge of the possible private signals that others may have observed also gets refined; thus enabling her to form more refined beliefs.

These calculations of the Bayesian agent can be cast in the framework of a partially observed Markov decision process (POMDP). Accordingly, the private signals of all agents con-

*Institute for Data, Systems, and Society (IDSS), Massachusetts Institute of Technology (MIT), Cambridge, MA 02139, USA. (emails: rahimian, jadbabai, elmos @mit.edu).

Jadbabaie and Rahimian are supported by ARO grant MURI W911NF-12-1-0509. Mossel is partially supported by NSF grant CCF 1665252, DOD ONR grant N00014-17-1-2598, and NSF grant DMS- 1737944.

stitute the state space of the problem and the decision maker only has access to a deterministic function of the state, the so-called partial observations. In GDP the beliefs of the neighbors constitute each agent's partial observations. The partially observed problem and its relations to the decentralized and team decision problems have been the subject of major classical contributions [12], [13]; in particular, the partially observed problem is known to be PSPACE-hard in the worst case [14, Theorem 6]. However, unlike the general POMDP, the state (private signals) in a GDP do not undergo Markovian jumps as they are fixed at the initiation of GDP. Hence, determining the complexity of GDP requires a different analysis. To address this requirement, in Section III we introduce and analyze a structural property of the graph which we refer to as transparency and it plays a critical role in characterizing the hardness and determining the computations of a Bayesian agent when forming her posterior beliefs. In the extended online version of our work [15], we provide our $\mathcal{NP}$-hardness proof, showing that well-known $\mathcal{NP}$-complete problems are special cases of GDP. This result complements and informs the existing literature on Bayesian learning over networks; in particular, those which offer efficient algorithms for special settings such as Gaussian signals and state space [16], or with binary actions in a complete graph [17].

II. Formation of Bayesian Beliefs: The GROUP-DECISION Problem

Consider a society of n agents that are labeled by $[n] = \{1, \dots, n\}$ and interact according to a fixed directed graph $\mathcal{G}$. For each agent $i \in [n]$, $\mathcal{N}_i$ denotes a fixed neighborhood $\mathcal{N}_i \subset [n]$ that is the set of all agents observed by agent i. We use $\delta(j, i)$ to denote the length (number of edges) of the shortest path connecting j to i. The agents engage in discussion by repeatedly exchanging their beliefs about an issue of common interest. We model the topic of the discussion/group-decision process by a state θ belonging to a finite set Θ. For example in the course of a political debate, Θ can be the set of all political parties and it would take a binary value in a bipartisan system. The value/identity of θ is not known to the agents but they each receive a private signal about the unknown θ. Each signal $\mathbf{s}_i$ belongs to a finite set $\mathcal{S}_i$ and its distribution conditioned on θ is given by $\mathbb{P}_{i,\theta}(\cdot)$ which is referred to as the *signal structure* of agent i. We use $\mathbb{P}_\theta(\cdot)$ to denote the joint distribution of the private signals of all agents, signals being independent across the agents. Starting from a full-support prior belief $\nu(\cdot)$, at any time $t \in \{0, 1, 2, \dots\}$ the agent holds a belief $\boldsymbol{\mu}_{i,t}$, which is her Bayesian posterior on Θ given her knowledge of the signal structure and priors as well as her history of observations, which include her initial private signal as well as the beliefs that she has observed in her neighbors throughout past times $\tau < t$. Throughout the paper, any random variable is denoted in boldface letter, vectors are represented in lowercase letters with a bar over them.

Building on the prior works [11], we now describe in detail the Bayesian calculations that take place among the group members. To calculate her Bayesian posterior, each agent keeps track of a list of possible combinations of private signals of all other agents, and at each iteration after simulating the network at every possible combination of the private signals, she crosses out the entries in her list that are inconsistent with the beliefs that she has observed from her neighbors up to that point. In this scheme, the agent not only needs to keep track of the list of private signals that are consistent with her observations, but also to evaluate the consistency of each of these signal profiles she needs to consider what other agents regard as consistent with their own observations under the particular set of initial signals. The latter consideration enables the decision maker to calculate beliefs of other agents under any circumstances that arise at a fixed profile of initial signals, as she tries to evaluate the feasibility of that particular signal profile. Hence, the list of private signals can be regarded as the information set representing the current understanding of the agent about her environment and the way additional observations are informative is by trimming the current information set and reducing the ambiguity in the set of initial signals that have caused the agent's history of past observations.

To proceed, let $\overline{s} = (s_1, \dots, s_n) \in \mathcal{S}_1 \times \dots \times \mathcal{S}_n$ be a typical profile of initial signals observed by each agent across the network, and denote the set of all private signal profiles that agent i regards as feasible at time t, i.e. her information set at time t, by $\boldsymbol{\mathcal{I}}_{i,t} \subset \mathcal{S}_1 \times \dots \times \mathcal{S}_n$; this set is a random set, as it is determined by the random observations of agent i up to time t. Starting form $\boldsymbol{\mathcal{I}}_{i,0} = \{\mathbf{s}_i\} \times \prod_{j \neq i} \mathcal{S}_j$, at every decision epoch t agent i removes those signal profiles in $\boldsymbol{\mathcal{I}}_{i,t-1}$ that are not consistent with her history of observations up to time t and comes up with a trimmed set of signal profiles $\boldsymbol{\mathcal{I}}_{i,t} \subset \boldsymbol{\mathcal{I}}_{i,t-1}$ to form her Bayesian posterior belief and make her decision at time t.

The set of feasible signals $\boldsymbol{\mathcal{I}}_{i,t}$ is mapped to a Bayesian posterior for agent i at time t as follows:

$$\boldsymbol{\mu}_{i,t}(\theta) = \frac{\sum_{\overline{s} \in \boldsymbol{\mathcal{I}}_{i,t}} \mathbb{P}_\theta(\overline{s}) \nu(\theta)}{\sum_{\theta' \in \Theta} \sum_{\overline{s} \in \boldsymbol{\mathcal{I}}_{i,t}} \mathbb{P}_{\theta'}(\overline{s}) \nu(\theta')}. \tag{1}$$

In this sense, the Bayesian posterior is a sufficient statistic for the history of observations and unlike the observation history, it does not grow in dimension. For example at time 0 agent i learns her private signal $\mathbf{s}_i$, this enables her to initialize her list of feasible signals: $\boldsymbol{\mathcal{I}}_{i,0} = \{\mathbf{s}_i\} \times \prod_{k \in [n] \setminus \{i\}} \mathcal{S}_k$. Subsequently, her Bayesian posterior at time 0 is given by:

$$\boldsymbol{\mu}_{i,0}(\theta) = \frac{\sum_{\overline{s} \in \boldsymbol{\mathcal{I}}_{i,0}} \mathbb{P}_\theta(\overline{s}) \nu(\theta)}{\sum_{\theta' \in \Theta} \sum_{\overline{s} \in \boldsymbol{\mathcal{I}}_{i,0}} \mathbb{P}_{\theta'}(\overline{s}) \nu(\theta')} = \frac{\mathbb{P}_{i,\theta}(\mathbf{s}_i) \nu(\theta)}{\sum_{\theta' \in \Theta} \mathbb{P}_{i,\theta'}(\mathbf{s}_i) \nu(\theta')}$$

At time 1 having observed her neighbor's beliefs from time 0 agent i learns the likelihood of their private signals of each of her neighbors $\{\mathbf{s}_j, j \in \mathcal{N}_i\}$. Subsequently, she crosses out any signal profile $\overline{s} \in \mathcal{I}_{i,0}$ for which s_j does not satisfy

$$\boldsymbol{\mu}_{j,0}(\theta) = \frac{\mathbb{P}_{j,\theta}(s_j) \nu(\theta)}{\sum_{\theta' \in \Theta} \mathbb{P}_{j,\theta'}(s_j) \nu(\theta')},$$

thus pruning $\boldsymbol{\mathcal{I}}_{i,0}$ into the smaller set $\boldsymbol{\mathcal{I}}_{i,1}$. She then updates her Bayesian posterior $\boldsymbol{\mu}_{i,1}$ according to (1). Note that the initial belief exchanges between neighbors reveal the likelihoods of the private signals in the neighboring agents. Hence, from her observations of the beliefs of her neighbors at time zero $\{\boldsymbol{\mu}_{j,0}, j \in \mathcal{N}_i\}$, agent i learns all that she ever needs to know regarding the private signals of her neighbors so far as their influence on her beliefs about the unknown state θ is concerned. At time 2, the agent observes her neighbors' beliefs $\{\boldsymbol{\mu}_{j,1}, j \in \mathcal{N}_i\}$ for a second time. The second interaction informs her about what beliefs her neighbor's neighbors may have hold at time 0 and in turn what private signals they have observed at time 0. To proceed, we introduce the notation $\mathcal{N}_i^{\tau}$ as the τ-th order neighborhood of agent i comprising entirely of those agents who are connected to agent i through a walk of length τ: $\mathcal{N}_i^{\tau} = \{j \in [n] : j \in \mathcal{N}_{i_1}, i_1 \in \mathcal{N}_{i_2}, \ldots, i_{\tau-1} \in \mathcal{N}_i, \text{ for some } i_1, \ldots, i_{\tau-1} \in [n]\}$; in particular, $\mathcal{N}_i^1 = \mathcal{N}_i$ and we use the convention $\mathcal{N}_i^0 = \{i\}$. We further denote $\bar{\mathcal{N}}_i^t := \cup_{\tau=0}^{t} \mathcal{N}_i^{\tau}$ as the set of all agents who are within distance t of or closer to agent i, i.e. at most t hops away from i. We sometimes refer to $\bar{\mathcal{N}}_i^t$ as her t-radius ego-net.

Considering her neighbors' neighbors for the first time at $t = 2$, agent i calculates the time one beliefs of all of the agents in $\mathcal{N}_i^2$ for each of the signal profiles belonging to $\boldsymbol{\mathcal{I}}_{i,1}$ and use the result to calculate the time two beliefs of all her neighbors for each $\overline{s} \in \boldsymbol{\mathcal{I}}_{i,1}$. Any $\overline{s}$ for which the calculated time 2 belief of some neighbor $j \in \mathcal{N}_i$ does not agree with the observed belief $\boldsymbol{\mu}_{j,2}$ is subsequently removed from $\boldsymbol{\mathcal{I}}_{i,1}$ and the updated list $\boldsymbol{\mathcal{I}}_{i,2}$ is thus obtained. A similar set of calculations is repeated at time three: for every signal profile in $\boldsymbol{\mathcal{I}}_{i,2}$ that have survived the pruning process up until $t = 3$, the agent starts by calculating the beliefs of agents in $\mathcal{N}_i^3$ at time zero (as determined by their private signals fixed in $\overline{s}$). These beliefs along with the rest of the private signal in turn determine the beliefs of the agents in $\mathcal{N}_i^2$ at time one as well those of $\mathcal{N}_i$ at time two; subsequently, the agent can compare the latter calculated beliefs with her most recent observation of her neighbors' beliefs and eliminate the signal profiles for which there is a mismatch. This process can be formalized as an iterated eliminations algorithm for computing the Bayesian posterior belief in a general network at any time t, [15, Algorithm A4]. However, the resultant computations increase exponentially in n and in fact they can be bounded as $O(n^2 M^{2n-1} m)$, where M is an upper bound on the size of the signal space and m is the cardinality of the state space Θ. We now describe the GROUP-DECISION problem formally.In the extended online version of our work [15], we show that for general network structures (and in the worst case) one cannot hope to do much better than an exponential-time algorithm (if $\mathcal{P} \neq \mathcal{NP}$): "GROUP-DECISION is $\mathcal{NP}$-hard".

Consider the finite state space $\Theta = \{\theta_1, \ldots, \theta_m\}$ and for all $2 \leq k \leq m$, let: $\boldsymbol{\lambda}_i(\theta_k) := \log\left(\mathbb{P}_{i,\theta_k}(\mathbf{s}_i)/\mathbb{P}_{i,\theta_1}(\mathbf{s}_i)\right)$, $\boldsymbol{\phi}_{i,t}(\theta_k) := \log\left(\boldsymbol{\mu}_{i,t}(\theta_k)/\boldsymbol{\mu}_{i,t}(\theta_1)\right)$. Here and throughout the rest of the paper, we assume the agents have started from uniform prior beliefs and the size of the state space is $m = 2$, thence we enjoy a slightly simpler notation; otherwise knowing the (non-uniform) priors of each other the agents can always compensate for the effect of the priors as they observe each others' beliefs. Moreover, with a binary state space $\Theta = \{\theta_1, \theta_2\}$, the agents need to only keep track of one set of belief and likelihood ratios corresponding to the pair (θ_1, θ_2), whereas in general the agents should form and calculate $m - 1$ ratio terms for each of the pairs (θ_1, θ_k), $k = 2, \ldots, m$ to have a fully specified belief. For a binary state space with no danger of confusion we can use the simplified notation $\boldsymbol{\lambda}_i = \boldsymbol{\lambda}_i(\theta_2) := \log\left(\mathbb{P}_{i,\theta_2}(\mathbf{s}_i)/\mathbb{P}_{i,\theta_1}(\mathbf{s}_i)\right)$, and $\boldsymbol{\phi}_{i,t} = \boldsymbol{\phi}_{i,t}(\theta_2) = \log\left(\boldsymbol{\mu}_{i,t}(\theta_2)/\boldsymbol{\mu}_{i,t}(\theta_1)\right)$.

Problem 1 (GROUP-DECISION)**.** *At any time t, given the knowledge of graph and signal structures, as well as the private signal $\mathbf{s}_i$ and the history of observed neighboring beliefs $\boldsymbol{\mu}_{j,\tau}, j \in \mathcal{N}_i, \tau \in [t-1]$ determine the Bayesian posterior belief $\boldsymbol{\mu}_{i,t}$.*

In general GROUP-DECISION is a hard problem and we provide the formal $\mathcal{NP}$-hardness proof details in [15, Section 5]. We study the structural features that lead to this hardness in Section III; in particular, we investigate special cases where the calculations simplify and efficient algorithms are possible.

III. Transparent Structures: Statistical and Computational Efficiency

Here, we introduce a special class of structures which play an important role in determining the type of calculations that agent i should undertake to determine her posterior belief (recall from Section II that the t-radius ego-net of agent i, $\bar{\mathcal{N}}_i^t$, is the set of all agents who are within distance t of or closer to agent i, i.e. at most t hops away from i):

Definition 1 (Transparency)**.** *The graph structure $\mathcal{G}$ is transparent to agent i at time t, if for all $j \in \mathcal{N}_i$ and every $\tau \leq t-1$ we have that: $\boldsymbol{\phi}_{j,\tau} = \sum_{k \in \bar{\mathcal{N}}_j^{\tau}} \boldsymbol{\lambda}_k$, for any choice of signal structures and all possible initial signals.*

Remark 1 (Transparency, statistical efficiency, and impartial inference)**.** Such agents j whose beliefs satisfy the equation in Definition 1 at some time τ are said to hold a *transparent* or *efficient* belief; the latter signifies the fact that the such a belief coincides with the Bayesian posterior if agent j were given a direct access to the private signals of every agent in $\bar{\mathcal{N}}_j^{\tau}$. This is indeed the best possible (or statistically efficient) belief that agent j can hope to form given the information available to her at time τ. The same connection to the statistically efficient beliefs arise in the work of Eyster and Rabin who formulate the closely related concept of "impartial inference" in a model of sequential decisions by different players in successive rounds [18]; accordingly, impartial inference ensures that the full informational content of all signals that influence a player's beliefs can be extracted and players can fully (rather than partially) infer their predecessors' signals. In other words, under impartial inference, players' immediate predecessors provide "sufficient statistics" for earlier movers that are indirectly observed [18, Section 3]. Last but not least, it is worth noting that statistical efficiency or impartial inference are properties of the posterior beliefs, and as such the signal structures may be designed so that statistical efficiency or impartial inference hold true for a particular problem setting; on the other hand, transparency is a structural property of the network and would hold true for any choice of signal structures and all possible initial signals.

The following is a sufficient graphical condition for agent i to hold an efficient (transparent) belief at time t: there are no agents $k \in \bar{\mathcal{N}}_i^t$ that has multiple paths to agent i, unless it is among its neighbors (directly observed by it).[1]

Proposition 1 (Graphical condition for transparency)**.** *Agent i will hold a transparent (efficient) belief at time t if there are no $k \in \bar{\mathcal{N}}_i^t \setminus \mathcal{N}_i$ such that for $j \neq j'$, both j and j' belonging to $\mathcal{N}_i$, we have $k \in \bar{\mathcal{N}}_j^{t-1}$ and $k \in \bar{\mathcal{N}}_{j'}^{t-1}$.*

The proof of this proposition can be found in the extended online version of our manuscript [15, Proposition 4.3]. It is worth noting that in the course of the proof, for the structures that satisfy the sufficient condition for transparency, we obtain a simple algorithm for updating beliefs by setting the total innovation at every step equal to the sum of the most recent innovations observed at each of the neighbors, correcting for those neighbors who are being recounted:

1) Initialize: $\boldsymbol{\phi}_{i,0} = \boldsymbol{\lambda}_i$, $\hat{\boldsymbol{\phi}}_{j,0} = \boldsymbol{\phi}_{j,0} = \boldsymbol{\lambda}_j$, $\boldsymbol{\phi}_{i,1} = \sum_{j \in \bar{\mathcal{N}}_i^1} \boldsymbol{\phi}_{j,0}$.
2) For all $t > 1$, set:
   - $\hat{\boldsymbol{\phi}}_{j,t-1} = \boldsymbol{\phi}_{j,t-1} - \boldsymbol{\phi}_{j,t-2}$,
   - $\hat{\boldsymbol{\phi}}_{i,t} = \sum_{j \in \mathcal{N}_i} \left(\hat{\boldsymbol{\phi}}_{j,t-1} - \sum_{k \in \mathcal{N}_i : \delta(k,j) = t-1} \boldsymbol{\phi}_{k,0} \right)$,
   - $\boldsymbol{\phi}_{i,t} = \boldsymbol{\phi}_{i,t-1} + \hat{\boldsymbol{\phi}}_{i,t}$.

Rooted (directed) trees are a special class of transparent structures, which also satisfy the sufficient structural condition of Proposition 1; indeed, in case of a rooted tree for any agent k that is indirectly observed by agent i, there is a unique path connecting k to i. As such the correction terms for the sum of innovations observed in the neighbors is always zero, and we have $\hat{\boldsymbol{\phi}}_{i,t} = \sum_{j \in \mathcal{N}_i} \hat{\boldsymbol{\phi}}_{j,t-1}$, i.e. the innovation at every time step is equal to the total innovations observed in all the neighbors.

Example 1 (Transparent structures)**.**

Fig. 1 illustrates cases of transparent and nontransparent structures. We refer to them as first, second, third, and forth in their respective order from left to right. All structures except the first one are transparent. To see how the transparency is violated in the first structure, consider the beliefs of agent i: $\boldsymbol{\phi}_{i,0} = \boldsymbol{\lambda}_i$, $\boldsymbol{\phi}_{i,1} = \boldsymbol{\lambda}_i + \boldsymbol{\lambda}_{j_1} + \boldsymbol{\lambda}_{j_2}$; at time two, agent 1 observes $\boldsymbol{\phi}_{j_1,1} = \boldsymbol{\lambda}_{j_1} + \boldsymbol{\lambda}_{k_1} + \boldsymbol{\lambda}_{k_2}$ and $\boldsymbol{\phi}_{j_2,1} = \boldsymbol{\lambda}_{j_2} + \boldsymbol{\lambda}_{k_2} + \boldsymbol{\lambda}_{k_3}$. Knowing $\boldsymbol{\phi}_{j_1,0} = \boldsymbol{\lambda}_{j_1}$ and $\boldsymbol{\phi}_{j_2,0} = \boldsymbol{\lambda}_{j_2}$ she can infer the value of the two sub-sums $\boldsymbol{\lambda}_{k_1} + \boldsymbol{\lambda}_{k_2}$ and $\boldsymbol{\lambda}_{k_1} + \boldsymbol{\lambda}_{k_3}$, but there is no way for her to infer their total sum $\boldsymbol{\lambda}_{j_1} + \boldsymbol{\lambda}_{j_2} + \boldsymbol{\lambda}_{k_1} + \boldsymbol{\lambda}_{k_2} + \boldsymbol{\lambda}_{k_3}$. Agent i cannot hold an efficient or transparent belief at time two. The issue is resolved in the second structure by adding a direct link so that agent k_2 is directly observed by agent i; the sufficient structural condition of Proposition 1 is thus satisfied and we have $\boldsymbol{\phi}_{i,2} = \boldsymbol{\lambda}_i + \boldsymbol{\lambda}_{j_1} + \boldsymbol{\lambda}_{j_2} + \boldsymbol{\lambda}_{k_1} + \boldsymbol{\lambda}_{k_2} + \boldsymbol{\lambda}_{k_3}$. In structure three, we have $\boldsymbol{\phi}_{i,2} = \boldsymbol{\lambda}_i + \boldsymbol{\lambda}_{j_1} + \boldsymbol{\lambda}_{j_2} + \boldsymbol{\lambda}_{k_1} + \boldsymbol{\lambda}_{k_2} = \boldsymbol{\lambda}_i + \boldsymbol{\phi}_{j_1,1} + \boldsymbol{\phi}_{j_2,0}$. Structure four is also transparent and we have $\boldsymbol{\phi}_{i,2} = \boldsymbol{\lambda}_i + \boldsymbol{\lambda}_{j_1} + \boldsymbol{\lambda}_{j_2} + \boldsymbol{\lambda}_{k_1} + \boldsymbol{\lambda}_{k_2} + \boldsymbol{\lambda}_{k_3} + \boldsymbol{\lambda}_{k_4} = \boldsymbol{\lambda}_i + \boldsymbol{\phi}_{j_1,1} + \boldsymbol{\phi}_{j_2,1}$ and $\boldsymbol{\phi}_{i,3} = \boldsymbol{\lambda}_i + \boldsymbol{\lambda}_{j_1} + \boldsymbol{\lambda}_{j_2} + \boldsymbol{\lambda}_{k_1} + \boldsymbol{\lambda}_{k_2} + \boldsymbol{\lambda}_{k_3} + \boldsymbol{\lambda}_{k_4} + \boldsymbol{\lambda}_l = \boldsymbol{\lambda}_i + \boldsymbol{\phi}_{j_1,1} + \boldsymbol{\phi}_{j_2,1} + (\boldsymbol{\phi}_{j_1,2} - \boldsymbol{\phi}_{j_1,1})$, where in the last equality we have used the fact that $\boldsymbol{\lambda}_l = (\boldsymbol{\phi}_{j_1,2} - \boldsymbol{\phi}_{j_1,1})$. In particular, note that structures three and four violate the sufficient structural condition laid out in Proposition 1, despite both being transparent.

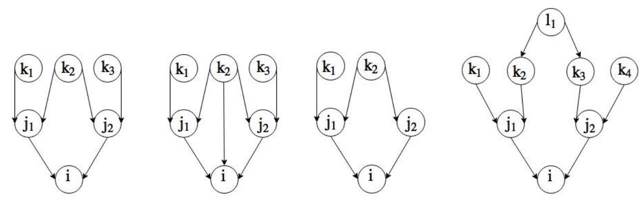

Fig. 1: Only the leftmost structure is not transparent.

When the transparency condition is violated, the neighboring agents' beliefs is a complicated non-linear function of the signal likelihoods of the upstream (indirectly observed) neighbors. Therefore, making inferences about the unobserved private signals from such "*nontransparent*" beliefs is a very complex task: it ultimately leads to agent i reasoning about feasible signal profiles that are consistent with her observations following the iterated elimination procedure described in Section II.

When the transparency conditions are satisfied, the beliefs of the neighboring agents reveal the sum of log-likelihoods for the private signals of other agents within a distance t of agent i. Nevertheless, even when the network is transparent to agent i, cases arise where efficient algorithms for calculating Bayesian posterior beliefs for agent i are unavailable and indeed impossible (if $\mathcal{P} \neq \mathcal{NP}$). In Subsection III-A, we describe the calculations of the Bayesian posterior belief when the transparency condition is satisfied.

A. Belief calculations in transparent structures

Here we describe calculations of a Bayesian agent in a transparent structure. If the network is transparent to agent i, she has access to the following information from the beliefs that she has observed in her neighbors at times $\tau \leq t$, before deciding her belief for time $t+1$:

- Her own signal $\mathbf{s}_i$ and its log-likelihood $\boldsymbol{\lambda}_i$.
- Her observations of the neighboring beliefs: $\{\boldsymbol{\mu}_{j,\tau} : j \in \mathcal{N}_i, \tau \leq t\}$. Due to transparency, these beliefs reveal the following information about sums of log-likelihoods of private signals of subsets of other agents in the network: $\sum_{k \in \bar{\mathcal{N}}_j^{\tau}} \boldsymbol{\lambda}_k = \boldsymbol{\phi}_{i,\tau}$, for all $\tau \leq t$, and any $j \in \mathcal{N}_i$.

From the information available to her, agent i aims to learn as much as possible about the likelihoods of the private signals of others whom she does not observe; indeed, as she has

[1] Jadbabaie would like to acknowledge his private communications about similar conditions that arise in the study of other learning models with P. Molavi.

already learned the likelihoods of the signals that her neighbors have observed from their reported beliefs at time one, at times $t > 1$ she is interested in learning about the agents that are further away from her up to the distance t. Her best hope for time $t+1$ is to learn the sum of $\log$-likelihoods of the signals of all agents that are within distance of at most $t+1$ from her (at most $t+1$-hops away from her in the graph), and to set her posterior belief accordingly; this however is not always possible as demonstrated for agent i in the leftmost graph of Fig. 1. To form her belief, agent i constructs the following system of linear equations in card $(\bar{\mathcal{N}}_{t+1}]) + 1$ unknowns: $\{\boldsymbol{\lambda}_j : j \in \bar{\mathcal{N}}_{t+1}, \text{ and } \bar{\boldsymbol{\lambda}}_{i,t+1}\}$, where $\bar{\boldsymbol{\lambda}}_{i,t+1} = \sum_{j \in \bar{\mathcal{N}}_{t+1}} \boldsymbol{\lambda}_j$ is the best possible (statistically efficient) belief for agent i at time $t+1$, given all the information available to her:

$$\begin{cases} \sum_{k \in \bar{\mathcal{N}}_j^{\tau}} \boldsymbol{\lambda}_k = \boldsymbol{\phi}_{j,\tau}, \text{ for all } \tau \leq t, \text{ and any } j \in \mathcal{N}_i, \\ \sum_{j \in \bar{\mathcal{N}}_i^{t+1}} \boldsymbol{\lambda}_j - \bar{\boldsymbol{\lambda}}_{i,t+1} = 0. \end{cases} \tag{2}$$

She can apply the Gauss-Jordan (factor and solve, Cholesky, QR or any other) method and convert the system of linear equations in card $(\bar{\mathcal{N}}_i^{t+1}) + 1$ variables to its reduced row echelon form. Next if in the reduced row echelon form $\bar{\boldsymbol{\lambda}}_{i,t}$ is a basic variable with fixed value (its corresponding column has a unique non-zero element that is a one, and that one belongs to a row with all zero elements except itself), then she sets her belief optimally such that $\boldsymbol{\phi}_{i,t+1} = \bar{\boldsymbol{\lambda}}_{i,t+1}$; this is the statistically efficient belief at time $t+1$. Recall that in the case of a binary state space, $\log$-belief ratio $\boldsymbol{\phi}_{i,t+1}$ uniquely determines the belief $\boldsymbol{\mu}_{i,t+1}$.

Remark 2 (Statistical and computational efficiency)**.** Having $\boldsymbol{\phi}_{i,t+1} = \bar{\boldsymbol{\lambda}}_{i,t+1}$ signifies the best achievable belief given the observations of the neighboring beliefs as it corresponds to the statistically efficient belief that the agent would have adopted, had she direct access to the private signals of every agent within distance $t+1$ from her (at most $t+1$-hops away from her); notwithstanding the efficient case $\boldsymbol{\phi}_{i,t+1} = \bar{\boldsymbol{\lambda}}_{i,t+1}$ does not necessarily imply that that agent i learns the likelihoods of the signals of other agents in $\bar{\mathcal{N}}_i^{t+1}$; indeed, this was the case for agent i in the forth (transparent) structure of Example 1: agent i learns $\{\boldsymbol{\lambda}_i, \boldsymbol{\lambda}_{j_1}, \boldsymbol{\lambda}_{j_2}, \boldsymbol{\lambda}_{k_1} + \boldsymbol{\lambda}_{k_2}, \boldsymbol{\lambda}_{k_3} + \boldsymbol{\lambda}_{k_4}, \boldsymbol{\lambda}_l\}$ and in particular can determine the efficient beliefs $\bar{\boldsymbol{\lambda}}_{i,2} = \boldsymbol{\lambda}_i + \boldsymbol{\lambda}_{j_1} + \boldsymbol{\lambda}_{j_2} + \boldsymbol{\lambda}_{k_1} + \boldsymbol{\lambda}_{k_2} + \boldsymbol{\lambda}_{k_3} + \boldsymbol{\lambda}_{k_4}$ and $\bar{\boldsymbol{\lambda}}_{i,3} = \boldsymbol{\lambda}_i + \boldsymbol{\lambda}_{j_1} + \boldsymbol{\lambda}_{j_2} + \boldsymbol{\lambda}_{k_1} + \boldsymbol{\lambda}_{k_2} + \boldsymbol{\lambda}_{k_3} + \boldsymbol{\lambda}_{k_4} + \boldsymbol{\lambda}_l$, but she never learns the actual values of the likelihoods $\{\boldsymbol{\lambda}_{k_1}, \boldsymbol{\lambda}_{k_2}, \boldsymbol{\lambda}_{k_3}, \boldsymbol{\lambda}_{k_4}\}$, individually. In other words, it is possible for agent i to determine the sum of $\log$-likelihoods of signals of agents in her higher-order neighborhoods even though she does not learn about each signal likelihood individually. The case where $\bar{\boldsymbol{\lambda}}_{i,t+1}$ can be determined uniquely so that $\boldsymbol{\phi}_{i,t+1} = \bar{\boldsymbol{\lambda}}_{i,t+1}$, is not only statistically efficient but also computationally efficient as complexity of determining the Bayesian posterior belief at time $t+1$ is the same as the complexity of performing Gauss-Jordan steps which is $O(n^3)$ for solving the t.card$(\mathcal{N}_i)$ equations in card$(\bar{\mathcal{N}}_i^{t+1})$ unknowns. Note that here we make no attempt to optimize these computations beyond the fact that their growth is polynomial in n.

Next we consider the case where $\bar{\boldsymbol{\lambda}}_{i,t+1}$ is not a basic variable in the reduced row echelon form of system (2) or it is a basic variable but its value is not fixed by the system and depends on how the free variables are set. In such cases agent i does not have access to the statistically efficient belief $\bar{\boldsymbol{\lambda}}_{i,t+1}$. Instead she has to form her Bayesian posterior belief by inferring the set of all feasible signals for all agents in card$(\bar{\mathcal{N}}_i^{t+1})$ whose likelihoods are consistent with the system (2). To this end, she keeps track of the set of all signal profiles at any time t that are consistent with her information, system (2), at that time. Following the iterated elimination procedure of Section II, let us denote the set of feasible signal profiles for agent i at time t by $\boldsymbol{\mathcal{I}}_{i,t}$. The general strategy of agent i, would be to search over all elements of $\boldsymbol{\mathcal{I}}_{i,t-1}$ and to eliminate (refute) any signal profiles that is inconsistent with (i.e. does not satisfy) the $\mathcal{N}_i$ new equations revealed to her at time t from the transparent beliefs of her neighbors.

Despite the relative simplification that is brought about by transparency, in general there is an exponential number of feasible signal profiles and verifying them for the new $\mathcal{N}_i$ equations would take exponential time. The belief calculations may be optimized by inferring the largest subset of individual likelihood ratios whose summation is fixed by system (2). The verification and refutation process can then be restricted to the remaining signals whose sum of $\log$-likelihoods is not fixed by system (2). For example in leftmost structure of Fig. 1, agent i will not hold a transparent belief at time 2 but she can determine the sub-sum $\boldsymbol{\lambda}_i + \boldsymbol{\lambda}_{j_1} + \boldsymbol{\lambda}_{j_2}$ and her belief would involve a search only over the profile of the signals of the remaining agents $(s_{k_1}, s_{k_2}, s_{k_3})$. At time two, she finds all $(s_{k_1}, s_{k_2}, s_{k_3})$ that agree with the additionally inferred sub-sums $\boldsymbol{\phi}_{j_1,1} - \boldsymbol{\phi}_{j_1,0} = \boldsymbol{\lambda}_{k_1} + \boldsymbol{\lambda}_{k_2}$ and $\boldsymbol{\phi}_{j_2,1} - \boldsymbol{\phi}_{j_2,0} = \boldsymbol{\lambda}_{k_2} + \boldsymbol{\lambda}_{k_3}$; indeed we can express $\boldsymbol{\phi}_{i,2}$ as follows:

$$\boldsymbol{\phi}_{i,2} = \boldsymbol{\lambda}_i + \boldsymbol{\lambda}_{j_1} + \boldsymbol{\lambda}_{j_2} + \\ \log \frac{\sum_{(s_{k_1}, s_{k_2}, s_{k_3}) \in \boldsymbol{\mathcal{I}}_{i,2}} \mathbb{P}_{k_1,\theta_2}(s_{k_1}) \mathbb{P}_{k_2,\theta_2}(s_{k_2}) \mathbb{P}_{k_3,\theta_2}(s_{k_3})}{\sum_{(s_{k_1}, s_{k_2}, s_{k_3}) \in \boldsymbol{\mathcal{I}}_{i,2}} \mathbb{P}_{k_1,\theta_1}(s_{k_1}) \mathbb{P}_{k_2,\theta_1}(s_{k_2}) \mathbb{P}_{k_3,\theta_1}(s_{k_3})},$$

where

$$\boldsymbol{\mathcal{I}}_{i,2} = \{(s_{k_1}, s_{k_2}, s_{k_3}) : \\ \log \frac{\mathbb{P}_{k_1,\theta_2}(s_{k_1})}{\mathbb{P}_{k_1,\theta_1}(s_{k_1})} + \log \frac{\mathbb{P}_{k_2,\theta_2}(s_{k_2})}{\mathbb{P}_{k_2,\theta_1}(s_{k_2})} = \boldsymbol{\lambda}_{k_1} + \boldsymbol{\lambda}_{k_2}, \text{ and} \\ \log \frac{\mathbb{P}_{k_1,\theta_2}(s_{k_1})}{\mathbb{P}_{k_1,\theta_1}(s_{k_1})} + \log \frac{\mathbb{P}_{k_3,\theta_2}(s_{k_3})}{\mathbb{P}_{k_3,\theta_1}(s_{k_3})} = \boldsymbol{\lambda}_{k_2} + \boldsymbol{\lambda}_{k_3}\}.$$

Here we make no attempt in optimizing the computations for the refutation process in transparent structures beyond pointing out that they can increase exponentially with the network size. In the extended online version of this work, we show that well-known $\mathcal{NP}$-complete problems are special cases of the GROUP-DECISION problem and as such the latter is $\mathcal{NP}$-hard [15, Theorem 5.1]; there we also provide examples of special cases where a more positive answer is available and provide an efficient algorithm accordingly.

Theorem 1 (Hardness of GROUP-DECISION)**.** *The GROUP-DECISION (Problem 1) is $\mathcal{NP}$-hard.*

IV. Discussion and Conclusions

We studied the formation of Bayesian posterior beliefs during a group decision process: starting with a private signal, neighboring agents repeatedly observe each others beliefs and form refined opinions based on their observations. We described a process of iterated elimination which requires the agent to repeatedly search over the exponential space of feasible signal profiles and refine it with her increasing observations. This algorithm runs in exponential time and we showed that the formation of Bayesian posterior beliefs is $\mathcal{NP}$-hard, so that one cannot hope for a significant improvement in the running time for general structures. We further, analyzed a graph structural property called "*transparency*", which implies that neighboring agents' beliefs are statistically efficient; in the sense that they reveal the aggregate information regarding the private signals of all agents who have influenced their decisions thus far (directly or indirectly). There is a key difference between the refutation process when the network is transparent to agent i and the pruning that takes place in the iterated elimination process for general (non-transparent) networks. In the latter case, the agent needs to consider the beliefs of other far way agents at any possible signal profile and to simulate the subsequent beliefs of her neighbors conditioned on the particular signal profile. Each signal profile will be rejected and removed from the feasible set if the simulated belief of a neighbor conditioned on that signal profile does not agree with the actual (observed) beliefs at that time. On the other hand, in a transparent structure, the agent does not need to simulate the beliefs of other agents conditioned on a signal profile to investigate its feasibility. She can directly verify whether the individual signals likelihoods satisfy the most recent set of constraints that are revealed to the agent at time t from the transparent beliefs of her neighbors; and if any one of the new equations is violated then, that signal profile will be rejected and removed from the feasible set. This constitutes an interesting bridge between statistical and computational efficiency in the group decision process.

Although determining the posterior beliefs during a GDP is, in general, $\mathcal{NP}$-hard [15, Theorem 5.1], for transparent structures the posterior belief at each step can be computed efficiently using the reported beliefs of the neighbors. Furthermore, the optimality of belief exchange over transparent structures is a unique structural feature of the inference set up in GDP. It provides an interesting and distinct addition to known optimality conditions for inference problems over graphs. In particular, the transparent structures over which efficient and optimal Bayesian belief exchange is achievable include many loopy structures in addition to trees.[1] It would be particularly interesting if one can provide a tight graphical characterization for transparency or provide other useful sufficient conditions that ensure transparency and complement our Proposition 1. More importantly, one would look for other characterizations of the complexity landscape and find other notions of simplicity that are different from transparency.

[1] It is well known that if a Bayesian network has a tree (singly connected or polytree) structure, then efficient inference can be achieved using belief propagation (message passing or sum-product algorithms), cf. [19]. However, in general loopy structures, belief propagation only gives a (potentially useful) approximation of the desired posteriors [20]. Notwithstanding, our Bayesian belief exchange set up also greatly simplifies in the case of tree structures, admitting a trivial sum of innovations algorithm. The authors in [21] study the complexity landscape of inference problems over graphical models in terms of their treewidth. For bounded treewidth structures the junction-tree method (performing belief propagation on the tree decomposition of the graph) works efficiently [22] but there is no class of graphical models with unbounded treewidth in which inference can be performed in time polynomial in treewidth.

References

[1] R. J. Aumann, "Agreeing to disagree," *The annals of statistics*, pp. 1236–1239, 1976.

[2] J. D. Geanakoplos and H. M. Polemarchakis, "We can't disagree forever," *Journal of Economic Theory*, vol. 28, no. 1, pp. 192–200, 1982.

[3] D. Gale and S. Kariv, "Bayesian learning in social networks," *Games and Economic Behavior*, vol. 45, pp. 329–346, 2003.

[4] D. Rosenberg, E. Solan, and N. Vieille, "Informational externalities and emergence of consensus," *Games and Economic Behavior*, vol. 66, no. 2, pp. 979–994, 2009.

[5] V. Borkar and P. Varaiya, "Asymptotic agreement in distributed estimation," *IEEE Transactions on Automatic Control*, vol. 27, no. 3, pp. 650–655, Jun 1982.

[6] J. N. Tsitsiklis and M. Athans, "Convergence and asymptotic agreement in distributed decision problems," *Automatic Control, IEEE Transactions on*, vol. 29, no. 1, pp. 42–50, 1984.

[7] K. Rahnama Rad and A. Tahbaz-Salehi, "Distributed parameter estimation in networks," *49th IEEE Conference on Decision and Control (CDC)*, pp. 5050–5055, 2010.

[8] E. Eyster and M. Rabin, "Naive herding in rich-information settings," *American economic journal: microeconomics*, vol. 2, no. 4, pp. 221–243, 2010.

[9] V. Krishnamurthy and W. Hoiles, "Online reputation and polling systems: Data incest, social learning, and revealed preferences," *IEEE Transactions on Computational Social Systems*, vol. 1, no. 3, pp. 164–179, 2014.

[10] M. Harel, E. Mossel, P. Strack, and O. Tamuz, "The speed of social learning," *arXiv preprint arXiv:1412.7172*, 2014.

[11] Y. Kanoria and O. Tamuz, "Tractable bayesian social learning on trees," *Selected Areas in Communications, IEEE Journal on*, vol. 31, no. 4, pp. 756–765, April 2013.

[12] J. N. Tsitsiklis and M. Athans, "On the complexity of decentralized decision making and detection problems," *IEEE Transactions on Automatic Control*, vol. 30, no. 5, pp. 440–446, 1985.

[13] R. Radner, "Team decision problems," *The Annals of Mathematical Statistics*, vol. 33, no. 3, pp. 857–881, 1962.

[14] C. H. Papadimitriou and J. N. Tsitsiklis, "The complexity of markov decision processes," *Mathematics of operations research*, vol. 12, no. 3, pp. 441–450, 1987.

[15] A. Jadbabaie, E. Mossel, and M. A. Rahimian, "Bayesian group decisions: algorithms and complexity," *arXiv preprint*, 2017.

[16] E. Mossel and O. Tamuz, "Efficient bayesian learning in social networks with gaussian estimators," *arXiv preprint arXiv:1002.0747*, 2010.

[17] ——, "Making consensus tractable," *ACM Transactions on Economics and Computation*, vol. 1, no. 4, p. 20, 2013.

[18] E. Eyster and M. Rabin, "Extensive imitation is irrational and harmful," *The Quarterly Journal of Economics*, p. qju021, 2014.

[19] J. Pearl, *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann, 2014.

[20] K. P. Murphy, Y. Weiss, and M. I. Jordan, "Loopy belief propagation for approximate inference: An empirical study," in *Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence*. Morgan Kaufmann Publishers Inc., 1999, pp. 467–475.

[21] V. Chandrasekaran, N. Srebro, and P. Harsha, "Complexity of inference in graphical models," *arXiv preprint arXiv:1206.3240*, 2012.

[22] R. G. Cowell, P. Dawid, S. L. Lauritzen, and D. J. Spiegelhalter, *Probabilistic networks and expert systems: Exact computational methods for Bayesian networks*. Springer, 2006.