Does one Bayesian make a Difference?

Manuel Mueller-Frank†
University of Oxford and Nuffield College

November 8, 2012

Abstract

This paper develops a model of repeated interaction in social networks among agents with differing degrees of sophistication. The focus lies on observational learning, that is each agent receives initial private information and makes inferences regarding the private information of others through the repeated interaction with his neighbors in the network. The main question is how well agents aggregate private information through their local interactions. I show that in finite networks consisting exclusively of non-Bayesian (boundedly rational) agents, who revise their choices by averaging over the previous period’s observed choices, all agents generically fail to aggregate any subset of the privately held information. However, the presence of at least one Bayesian agent in a strongly connected network is shown to be generically sufficient for each agent, that is Bayesian or non-Bayesian, to perfectly aggregate the private information of all agents. The results have implications for the theory of prediction markets and the information value of prices in markets with rational and boundedly rational traders.

†I am grateful to Itai Arieli for valuable comments and discussions. I would also like to thank Eddie Dekel, Shachar Kariv, Margaret Meyer, Stephen Morris, Mallesh Pai, Alessandro Pavan, Toan Phan, Marciano Siniscalchi, Dimitri Vayanos, Peyton Young and seminar audiences at the University of Cambridge, University of Edinburgh, University of Oxford, University of Wisconsin-Madison, the Midwest Economic Theory Conference and the North American Econometric Society meeting. Earlier versions were circulated under "Information Aggregation Properties of Opinion Formation in Social Networks" and "The Efficiency of Social Learning in Networks with Bayesian and Non-Bayesian Agents".

†E-mail: manuel.mueller-frank@economics.ox.ac.uk
1 Introduction

A central question in economic theory is whether and how privately held information of a large number of individuals is aggregated. One particular aggregation device that has been especially prominent is the market price. There is a vast literature on the information value of prices and under which conditions they aggregate the dispersed private information. This discussion started with Keynes [26] and Hayek [23] and continues vividly until the present day. In this paper, I focus on a different mechanism as a means to aggregate private information, which plays an important role in economic reality: social interaction.

The recent internet-based social network revolution that began with the creation of Myspace, and manifests itself in the overwhelming success of Facebook and Twitter, greatly simplified social interaction and communication. As a consequence of these recent developments social networks are becoming an increasingly important factor for the diffusion of information and opinions. The effects of such communication and coordination through social networks can be enormous. For example, anecdotal evidence suggests that Facebook and Twitter played an important role in the recent uprisings in the Middle East which led to the fall of several of the regimes in power. Moreover, the role of networks as an important conduit of information and opinions has long been documented empirically.1

This paper develops a model of repeated interaction in social networks among privately informed agents. The agents are embedded in a social structure and repeatedly interact with their social peers. The observation of behavior facilitates inferences on private information but (in incomplete networks) Bayesian inferences on the private information of others require highly complex considerations (see Gale and Kariv [15]).2 The main goal of this paper is to gain a deeper understanding of how the interaction of agents with differing degrees of sophistication affects the quality of information the respective agents hold in the long run.

Two different types of agents are considered, Bayesian agents and non-Bayesian (boundedly rational) agents who are unable or unwilling to make the complex inferences and instead revise their behavior based on simple learning heuristics. More precisely, the non-Bayesian agents are modeled according to the standard DeGroot [10] model. That is, they revise their choices by taking a weighted average of the previous period’s observed choices. This paper mainly differs from the existing literature in considering network-based repeated interaction among agents of differing degrees of sophistication. The existing literature focuses either on networks consisting only of Bayesian

---

1The importance of inter-personal communication has been (among others) established for the adoption of consumer goods (see Feick and Price [13]), medical innovation (see Coleman, Katz and Menzel [8]), agricultural practices (see Munshi [31], Conley and Udry [9]), and microfinance (see Banerjee, Chandrasekhar, Duflo and Jackson [6]).

2In a complete network each agent is a neighbor of every other agent.
agents, in the following denoted as Bayesian communication structures, and networks consisting only of non-Bayesian agents, henceforth denoted as non-Bayesian communication structures.\footnote{See Gale and Kariv [15], Rosenberg, Solan and Vieille [36], and Mueller-Frank [29] for the analysis of Bayesian communication structures, and DeMarzo, Vayanos and Zwiebel [11], and Golub and Jackson [18] for non-Bayesian communication structures employing the DeGroot model [10].}

I focus on the information aggregation properties of mixed communication structures, networks that consist of Bayesian and non-Bayesian agents, and non-Bayesian communication structures. The main questions of the paper are the following. First, how well, if at all, do non-Bayesian communication structures aggregate information? Second, under which conditions do the choices of all agents in mixed communication structures eventually agree? Third, under which conditions do mixed communication structures perfectly aggregate information? Note that perfect information aggregation is satisfied if the choices of each agent converge to the choice that is optimal conditional on the pooled private information of all agents.\footnote{For any realization of the private information.}

While the focus of the paper lies on information aggregation, the consensus property is of independent interest. In the existing literature on observational learning in social networks, failure of consensus is mainly explained by failure of strong connectedness.\footnote{It has been shown that in Bayesian and non-Bayesian communication structures asymptotic consensus occurs among all agents in strongly connected networks. A network is strongly connected if there exists a directed path connecting any pair of agents.} Therefore, whether or not long-run disagreement in social networks might be due to differing degrees of sophistication, rather than failure of strong connectedness, seems a worthwhile question to address.

The questions are addressed in the following communication model. A finite set of agents are organized in a strongly connected social network. The agents share a common prior over the set of states of the world and each agent has private information represented by a finite partition of the state space.

The timing is as follows. In round $t = 0$ the state of the world is drawn and agents learn the realized cell of their partition. In round $t = 1$ every agent announces her opinion, the posterior probability of some relevant event conditioning on her realized partition cell, to each of her neighbors in the network. The main assumption here is that communication is coarse; agents cannot announce their information but instead communicate a proxy or statistic of their private information, in this case the posterior probability of an uncertain event. The approach of imposing coarseness on communication is standard in the literature on knowledge and consensus. One possible reason for such coarseness might be that communicating private information equates to the agent recounting "every experience and piece of information" he ever had.\footnote{See Geanakoplos and Polemarchakis [17].}

From the second round onward the agents differ conceptually. Non-Bayesian agents use the
standard DeGroot [10] learning heuristic when forming their opinion; the opinion of a non-Bayesian agent in period \( t \) is equal to a weighted average of the opinions announced by his neighbors and himself in period \( t - 1 \). Bayesian agents on the other hand draw fully rational inferences on the private information of the other agents based on the history of opinions of their neighbors. To start, assume that the Bayesian agents commonly know the (i) network, (ii) partitions of all agents, (iii) the type of each agent, Bayesian or non-Bayesian, and (iv) the updating function of each non-Bayesian agent. One example for the model described, would be a situation in which individuals repeatedly communicate their (subjective) probability of the incumbent being reelected president.

1.1 Summary of the Main Results

In order to assess the information aggregation properties of a given communication structure, I introduce the concept of Bayesian equivalence: a communication structure is Bayesian equivalent if, for each agent, there exists a partition of the state space, such that in every state of the world, the limit belief of the agent equals the Bayesian posterior conditioning on the corresponding cell of the respective partition. In a Bayesian equivalent communication structure the limit belief of each agent correctly represents a subset of the private information of all agents.\(^7\)

For a given network consider the updating function of each non-Bayesian agent and denote the tuple of updating functions of all non-Bayesian agents as a weight state. The …rst theorem of the paper establishes that non-Bayesian communication structures fail Bayesian equivalence for almost all weight states. The inability of the non-Bayesian agents to process complex information leads to failure of information aggregation and hence incorrect limit beliefs. Therefore, the presence of at least one Bayesian agent is necessary for social networks to (partially) aggregate private information.

Theorem 1 uncovers a substantial difference in the information aggregation properties of Bayesian and non-Bayesian communication structures. An earlier paper, Mueller-Frank [29], established that the private information of each agent is perfectly incorporated in the eventual consensus beliefs if all agents are Bayesian. Theorem 1 on the other hand establishes that the eventual consensus beliefs fail to correctly represent the private information of any subset of agents, if all agents are non-Bayesian. This stark difference leads to the analysis of opinion formation in networks with Bayesian and non-Bayesian agents.

Theorem 2 states that every mixed communication structure perfectly aggregates the private information of all agents, almost everywhere in the product space of common priors and weight states. Therefore, Theorem 2 answers both of the main questions concerning mixed communi-

\(^7\)Note that Bayesian equivalence is strictly weaker than perfect information aggregation, where the limit belief of each agent incorporates the private information of every agent.
cation structures. First, in the long run the opinions of all agents generically agree in strongly connected networks. Second, independent of the number and location of Bayesian agents in a strongly connected network, the ensuing consensus opinion correctly represents the aggregated private information of all agents.

The results on mixed communication structures are derived under the strong assumption of the Bayesian agents commonly knowing the exact updating function of each non-Bayesian agent. Theorem 4 addresses the robustness of the perfect information aggregation result to relaxing this assumption. Suppose that for each non-Bayesian agent there is a countable set of updating functions that is commonly known to contain his true updating function. This extension is formally captured by an extended state space equaling the product of the fundamental state space and the countable space of possible weight states. Theorem 4 establishes that for almost all pairs of priors and sets of possible weight states any mixed communication structure achieves perfect information aggregation, if the pooled private information of agents never provides certainty regarding the realization of the relevant event. Moreover, this result holds without conditions on the priors the Bayesian agents hold over the set of possible updating functions. Neither common priors nor commonly known heterogeneous priors are necessary.

1.2 Discussion of the Main Result

The results of the paper can be summarized as follows: in arbitrary strongly connected social networks at least one Bayesian agent is necessary and sufficient for repeated interaction to lead to perfect information aggregation. The main result of the paper is somewhat surprising on first glance. In any strongly connected network containing Bayesian and non-Bayesian agents, every agent perfectly learns the private information of all agents, independent of the number and location of the Bayesian agents in the network. Intuitively one might be tempted to believe that in order for each Bayesian agent to infer the private information of all agents and hence drive the non-Bayesian agents to the correct beliefs in the limit, the Bayesians would need to be placed in prominent positions in the network and thereby serve as information hubs, or for the Bayesian agents to satisfy a critical mass condition. Our results show that this is not the case. As long as there exists a single Bayesian agent somewhere in the network, asymptotic consensus and perfect information aggregation occurs.

The reader might wonder about the reliance of the result on the assumption of agents communicating beliefs. While the interaction is described as mere communication of posterior probabilities under absence of utility functions, this setting is generically equivalent to a network interaction

---

8Note that in the following the terms generic and full measure, in the sense of almost everywhere, are used interchangeably.
game in which agents repeatedly select an action under a strict proper scoring utility function and discount the stream of utilities. Perfect information aggregation holds generically in a perfect Bayesian equilibrium of the network interaction game.\(^9\) Therefore, the results carry forward to a fully strategic setting.

The general model has direct implications for specific applications. In particular, the perfect information aggregation result carries forward to market scoring rule models, which were introduced by Hanson \([21, 22]\). Market scoring rule models are used in practice, for example by Microsoft (see Snowberg, Wolfers and Zitzewitz \([38]\)), to implement firm-internal prediction markets for the purpose of aggregating the private information of employees. In the theoretical literature the market scoring rule framework is also used to model financial markets, see for example Ostrovsky \([34]\). The results of this paper imply that in such a market with rational and boundedly rational traders, who employ the DeGroot \([10]\) form of updating, the prices of Arrow-Debreu securities converge to their pooled information value generically.

The rest of the paper is organized as follows. In section 2, I introduce the model and the concept of Bayesian equivalence. In section 3, I consider non-Bayesian communication structures, establish Theorem 1, and briefly discuss an alternative concept to assess information aggregation in non-Bayesian communication structures. Section 4 considers mixed communication structures, presents a simple example of failure of perfect information aggregation and the results on generic perfect information aggregation, Theorem 2, and generic informational equivalence between Bayesian and mixed communication structures, Theorem 3. In section 5, I consider information aggregation under lack of common knowledge of the updating functions of the non-Bayesian agents and present Theorem 4, generic perfect information aggregation in mixed communication structures in a setting of countable sets of commonly considered possible updating functions for each non-Bayesian agent. Section 6 establishes equivalence between the truthful communication model and a network interaction game. In section 7, I consider the implications of the results to the information value of security prices and prediction markets. Section 8 briefly discusses the relation of the paper to the existing literature on learning in networks. Section 9 concludes. The proofs of the theorems are either presented in the main text or in appendix A, in which case the structure of the proof is explained in the main text. Appendix A further provides the proofs of all lemmas. Appendix B provides (i) technical background, (ii) an example of non-Bayesian communication structures to highlight failure of Bayesian equivalence, and (iii) a setting where the asymptotic informational equivalence of Bayesian and Non-Bayesian agents translates into equivalence in expected utility terms within finite time.

\(^9\)See section 6 for details.
2 The Model and Relevant Definitions

A finite set of agents $V = \{1, ..., v\}$ are organized in a strongly connected, directed network $G = (V, E)$. A network is a pair of sets $(V, E)$ such that $E \subset V^2$. The elements of $V$ are nodes of the network, representing the agents, and the elements of $E$ are the edges of the network, representing the direct connections between agents. The neighborhood of agent $i$ is denoted by $N_i$, where $N_i = \{j \in V : ij \in E\}$. A path from $i$ to $j$ in the network equals a sequence of nodes $k_1, ..., k_l$, where $k_1 = i$ and $k_l = j$, such that $k_{f+1} \in N_{k_f}$ for $f = 1, ..., l - 1$. The distance of agent $j$ from agent $i$ in network $G$, denoted by $d_G(i, j)$, equals the length of the shortest path from $i$ to $j$.

The agents face uncertainty, represented by a probability space $(\Omega, \mathcal{F}, p)$, where $\Omega$ is a fundamental state space, $\mathcal{F}$ a $\sigma$-algebra of subsets of $\Omega$ and $p$ a common prior such that each non-empty element of $\mathcal{F}$ has positive probability under $p$. Each agent is endowed with a finite partition $\mathcal{P}_i$ of the fundamental state space $\Omega$.

The state $\omega$ is realized in period 0 and each agent observes the corresponding cell of his partition. In each of countable periods, every agent $i$ communicates with each of his neighbors by announcing her (subjective) posterior probability of a relevant and uncertain event $C \in \mathcal{F}$. This relevant event could for example be the default of a sovereign government, a recession occurring within the next twelve months, or the incumbent being reelection as president.

In the first period all agents $i \in V$ announce their Bayesian posteriors $q_i^1(\omega | p)$ of event $C$ conditioning on their private information,

$$q_i^1(\omega | p) = \frac{p(C \cap \mathcal{P}_i(\omega))}{p(\mathcal{P}_i(\omega))}.$$

There are two types of agents that differ conceptually from the second period onward. The announcement of a non-Bayesian agent, $j \in -B$, in period $t$ equals a weighted average of his own and his neighbors announcements in period $t - 1$. Therefore, non-Bayesian agents behave according to the standard DeGroot [10] model of non-Bayesian learning in networks. For a given network $G$, each non-Bayesian agent $j$ is described by the vector of weights $\theta_j \in \text{ri}\Delta(\lfloor N_j \rfloor + 1)$ he assigns to his neighbors and himself, where $\text{ri}\Delta(\lfloor N_j \rfloor + 1)$ denotes the relative interior of the unit simplex in $\mathbb{R}^{\lfloor N_j \rfloor + 1}$.

Bayesian agents, $i \in B$, announce their Bayesian posteriors conditioning on their private information which depends on the announcements they observe. The analysis is based on the assump-

---

10 In the following $d_X()$ denotes the metric in metric space $X$ and $d_G()$ the distance in graph $G$.

11 $\mathcal{F}$ is generated by the join of partitions refined by $C$. See appendix B for details. The join of a set of partitions is the coarsest common refinement of the partitions.

12 $|S|$ denotes the cardinality of $S$. 

7
tions that Bayesian agents commonly know the (i) structure of the network, (ii) set of partitions \( \{ \mathcal{P}_i \}_{i \in V} \), (iii) type of each agent in the network, Bayesian or not-Bayesian, and (iv) the weight vector \( \theta^j \) of each non Bayesian agent \( j \). Moreover, it is assumed that the Bayesian agents truthfully report their posterior in each period even though there might be an incentive to lie in order to have an information gain in later periods. This myopia assumption is standard in the literature on knowledge and consensus and has been shown not to be restrictive generically. To be more specific, Mueller-Frank [29] has established that in Bayesian communication structures, generically in the space of priors, there is no gain from lying as all agents reveal their private information in each round of communication. Theorem 2 extends the result to mixed communication structures by establishing that generically Bayesian agents have no incentive to lie in mixed communication structures. For a detailed discussion of the generalization to a fully strategic setting see section 6.

As mentioned above, each non-Bayesian agent is described by the weights he assigns to his neighbors and himself. For a given network, the set of feasible weights of a non-Bayesian agent \( j \), or in other words the set of feasible updating functions, is denoted by \( \Theta_j \),

\[
\Theta_j = \left\{ \theta^j \in (0, 1)^{|N_j|+1} : \sum_{k \in N_j \cup \{j\}} \theta^j_k = 1 \right\} = \text{ri} \left( \Delta \left( |N_j| + 1 \right) \right).
\]

The set \( \Theta_j \) equals the relative interior of the unit simplex in \( \mathbb{R}^{|N_j|+1} \), an open, convex subset of an affine hyperplane in \( \mathbb{R}^{|N_j|+1} \) with dimension \( |N_j| \). Let \( \mathcal{L}_{|N_j|+1}(\Theta_j) \) denote the \( \sigma \)-algebra on \( \Theta_j \) and let \( \lambda_{\Theta_j} \) denote the corresponding measure which is derived from the Lebesgue measure in \( \mathbb{R}^{|N_j|} \).\(^{13}\)

I denote the product space of the sets of feasible updating functions of all non-Bayesian agents, \( \prod_{j \in \neg B} \Theta_j \), as \( \Theta \), the product \( \sigma \)-algebra \( \prod_{j \in \neg B} \mathcal{L}_{|N_j|+1}(\Theta_j) \) as \( \mathcal{L}(\Theta) \), and the corresponding product measure \( \prod_{j \in \neg B} \lambda_{\Theta_j} \) as \( \lambda_{\Theta} \). Since the realized weight state \( \theta \in \Theta \) is common knowledge among all Bayesian agent it is encoded in the fundamental state space. Each fundamental state includes a complete description of the state of the world. Therefore, it includes the weight state which is identical across all fundamental states.

In the following the tuple \( (\Omega, \mathcal{F}, p), \{ \mathcal{P}_i \}_{i \in V}, G \) is denoted as the information setting. For a given information setting, the communication structure is described by the set of non-Bayesian agents \( \neg B \). If the set of non-Bayesian agents equals the set of agents, \( \neg B = V \), I denote the communication structure as non-Bayesian. If there are Bayesian and non-Bayesian agents in the network, the communication structure is denoted as mixed. If the set of non-Bayesian agents is

\[^{13}\text{The unit simplex in } \mathbb{R}^n \text{ with the induced } \sigma \text{-algebra is isomorphic to a full dimensional subset of } \mathbb{R}^{n-1} \text{ endowed with its Lebesgue } \sigma \text{-algebra.}\]
empty, \(-B = \emptyset\), the communication structure is denoted as Bayesian.\(^{14}\)

The main objective of the paper is to evaluate the quality of information aggregation of non-Bayesian and mixed communication structures. The broad idea of this evaluation is to compare the limit belief of an agent to the Bayesian posterior conditioning on subsets of the private information of all agents. Let \(q_i^\infty(\omega \mid (p, \theta))\) denote the limit opinion of agent \(i\) in fundamental state \(\omega\).\(^{15}\) The limit opinions of each agent generate a partition \(\mathcal{P}_i^\infty(\cdot \mid (p, \theta))\) of the fundamental state space (for a given weight state and prior): states \(\omega\) and \(\omega'\) are in the same cell of \(\mathcal{P}_i^\infty(\cdot \mid (p, \theta))\) if and only if their corresponding limit beliefs are the same,

\[
\mathcal{P}_i^\infty(\omega \mid (p, \theta)) = \{ \omega' \in \Omega : q_i^\infty(\omega \mid (p, \theta)) = q_i^\infty(\omega' \mid (p, \theta)) \}.
\]

The join \(\bigvee_{i \in V} P_i(\omega)\) of a set of partitions \(\{P_i\}_{i \in V}\) is the coarsest common refinement of the partitions.
Two different concepts of information aggregation are considered.

**Definition 1** A communication structure achieves **perfect information aggregation** for prior \(p\) and weight state \(\theta\) if for every agent \(i \in V\) and every fundamental state \(\omega \in \Omega\)

\[
q_i^\infty(\omega \mid (p, \theta)) = \frac{p(C \cap \bigvee_{i \in V} P_i(\omega))}{p\left(\bigvee_{i \in V} P_i(\omega)\right)}.
\]

In a communication structure that satisfies perfect information aggregation the limit opinions of all agents are equal to the pooled information posterior probability of event \(C\).

**Definition 2** A communication structure is **Bayesian equivalent** for prior \(p\) and weight state \(\theta\) if for every agent \(i \in V\) and every fundamental state \(\omega \in \Omega\)

\[
q_i^\infty(\omega \mid (p, \theta)) = \frac{p(\mathcal{P}_i^\infty(\omega \mid (p, \theta)) \cap C)}{p(\mathcal{P}_i^\infty(\omega \mid (p, \theta))))}.
\]

Note that perfect information aggregation implies Bayesian equivalence but not vice versa. A communication structure is Bayesian equivalent if the limit belief of every agent in each fundamental

\(^{14}\)For Bayesian communication structures the set of feasible weight states and the weight state itself is simply equal to \(\emptyset\).

\(^{15}\)We have not yet shown that beliefs indeed individually converge in mixed communication structures but will do so in Theorem 2. Existing results establish that opinions individually converge to the same limit in Bayesian and non-Bayesian communication structures.
state of the world is consistent with some subset of the private information of agents. Consistency here refers to the limit belief equalling the Bayesian posterior conditioning on some subset of the private information of all agents. In the partitional information framework, a subset of the private information of all agents corresponds to a partition that is generated from (a subset of) the partitions of all agents, like for example the join of all partitions. The definition implies that if a communication structure is not Bayesian equivalent, then there exists no partition of the state space that generates the limit beliefs as conditional Bayesian posteriors. Therefore, a communication structure that is not Bayesian equivalent, fails to represent any subset of the private information of agents.

But what exactly does Bayesian equivalence mean in terms of information aggregation? If a communication structure is Bayesian equivalent, then the degree of fineness of the limit partition $P_i^\infty(\cdot|\{p, \Theta\})$ is a measure for the amount of private information being incorporated in the limit belief. The finer the limit partition, the more information is aggregated. In the best case the limit partition $P_i^\infty(\cdot|\{p, \Theta\})$ equals the join of partitions of all agents in which case the private information of all agents is perfectly aggregated in each state of the world.

3 Non-Bayesian Communication Structures

Non-Bayesian communication structures have been extensively analyzed in the literature, see for example DeMarzo, Vayanos and Zwiebel [11], and Golub and Jackson [18, 19]. It has been established that for weighted average updating functions, where all agents assign positive weights to themselves, the opinions of all agents converge to the same limit if the network is strongly connected. However, little attention has been given to the properties of the consensus limit beliefs for fixed networks and/or information structures that go beyond independent private signals. In this section I analyze how the limit beliefs relate to the underlying private information of agents. This analysis uses the concept of Bayesian equivalence as its main tool. Note that in some cases Bayesian equivalence is trivially satisfied. For example, in an information setting where the first period posteriors of all agents agree in every state of the world, Bayesian equivalence naturally holds as the limit belief equals the first period consensus belief. Such an information setting is denoted as trivial, that is an information setting is trivial if in every fundamental state $\omega \in \Omega$ we have

$$\frac{p(C \cap P_i(\omega))}{p(P_i(\omega))} = \frac{p(C \cap P_j(\omega))}{p(P_j(\omega))}$$

for all $i, j \in V$. The following theorem establishes a negative result for information aggregation in non-Bayesian communication structures.$^{16}$

$^{16}$Note that while the a.e. result of the theorem is established for a uniform measure derived from the Lebesgue measure, the result equally holds for all measures that are absolutely continuous with respect to the uniform measure.
Theorem 1 For a non-trivial information setting consider the corresponding non-Bayesian communication structure, $\neg B = V$. Bayesian equivalence fails for $[\lambda_{\Theta}]$ almost every weight state $\theta \in \Theta$.

Theorem 1 establishes that non-Bayesian communication structures yield, from a Bayesian perspective, incorrect limit beliefs for almost every weight state $\theta$. When considering a fixed network, Bayesian equivalence, and hence perfect information aggregation, fails.

The fact that non-Bayesian communication structures asymptotically achieve consensus among all agents might be useful and important from a pure coordination perspective. However, as established in Theorem 1, the consensus belief does not represent any subset of the private information of agents.

Whether the statement of the theorem holds for all as opposed to almost all weight states is unclear for the general information structures considered here. However, for some typical information settings Bayesian equivalence fails for all weight states. Consider the following example which is derived from the sequential social learning literature. Agents face uncertainty represented by a binary state space, $\Omega = \{C, \neg C\}$, and receive conditionally independent and identically distributed signals with finite support $S$. Consider the signal $s^{-}$ that minimizes the posterior probability of state $C$. Now consider the vector of such signals $s^{-}$ (across agents). Compare this vector of signals with any other vector $s$ of signal realization in regards to the vector of first round announcements they generate. Note that the limit opinion equals a weighted average of the first period announcements of all agents. For any weighted average that assigns positive weight to each agents posterior probability, the weighted average of the individual posterior probabilities corresponding to $s^{-}$ is strictly smaller than that of any $s \neq s^{-}$. Bayesian equivalence then requires the weighted average (corresponding to the stationary distribution) of the posteriors corresponding to $s^{-}$ to equal the posterior probability of event $C$ conditioning on $s^{-}$. But this is not the case as the posterior probability of event $C$ given $s^{-}$ is strictly smaller than the posterior probability conditioning on $s^{-}$.

Theorem 1 provides a result concerning information aggregation taking Bayesian equivalence as the tool to evaluate the limit opinions. But Bayesian equivalence is just one way to evaluate information aggregation in networks. It is a natural approach, in my opinion, as it asks whether the limit belief of an agent can be rationalized in a Bayesian manner. An alternative way of evaluating information aggregation is to check whether an agents limit belief is closer to the pooled information belief than his initial belief. For a given fundamental state let the pooled information posterior denote the posterior probability of event $C$ conditioning on the cell of the join containing the fundamental state. If in a given fundamental state an agents limit belief is closer to the

The same applies to all theorems in the paper.
true belief than his initial belief then one could say that the agent improves by participating in communication. While there are simple examples showing that some agents might improve in the sense just described in all fundamental states, on the cost of others, it can be shown that a non-Bayesian communication structure does not lead to a Pareto improvement of all agents, for almost all weight states. Whenever some agents improve in a given state, others are worse off. The concepts of Bayesian equivalence and posterior improvement have in common that under neither participating in a non-Bayesian communication structure is beneficial for all agents. The main point of this section is that non-Bayesian communication structures are (generically) strictly inferior to Bayesian communication structures from an information aggregation perspective under either evaluation concept.

3.1 Proof of Theorem 1

To preface the proof note that the weight state $\theta$ can be represented by a Markov transition matrix $T_\theta$ with $v$ rows and columns. Since the network is strongly connected and all agents assign positive weight to themselves, the transition matrix is irreducible and aperiodic and hence converges to its unique stationary distribution $\mu(\theta)$. This implies that the limit opinion of each agent is equal to $\mu(\theta) \times q^1(\omega | p)$ and hence the limit partitions $P^\infty_i(. | (p, \theta))$ coincide for all agents. Please note that in the following the relative interior of the unit simplex in $\mathbb{R}^k$ is denoted as $\Delta(k)$. The zero set of a function $f : X \to \mathbb{R}$ is equal to the preimage of 0 under $f$, i.e. the zero set of $f$ is defined as $\{x \in X : f(x) = 0\}$. The first lemma presents a result that is used in the proof of each of the three main theorems.

**Lemma 1** Let $f : \Delta(k) \to \mathbb{R}$ be a non-trivial polynomial function. Then the zero set of $f$ has $|\lambda_{\Delta(k)}|$ measure zero in $\Delta(k)$.

The proof of Theorem 1 makes use of the following two lemmas.

**Lemma 2** Consider $f_{q,q'} : \Theta \to \mathbb{R}$ such that

$$f_{q,q}(\theta) = \sum_{i=1}^v [\mu(\theta)]_i \times (q_i - q'_i).$$

and $q \neq q'$. Then the zero set of $f_{q,q'}$ has $|\lambda_\Theta|$ measure zero in $\Theta$.

**Lemma 3** There exists a $|\lambda_\Theta|$ full measure set $\Theta_{p1, p} \subset \Theta$ such that $P^\infty_i(. | (p, \theta)) = P^1_i(. | p)$ for all $\theta \in \Theta_{p1, p}$ and $i \in V$, where

$$P^1(\omega | p) = \{\omega' \in \Omega : q^1(\omega' | p) = q^1(\omega | p)\}.$$
See below for the proof of Theorem 1.

Proof. $\Theta_{P_1,p}$ denotes the set of weight states where $P_i^\infty(|p, \theta)) = P_i^1(|p)$ holds for all $i$, and denote its complement as $(\Theta_{P_1,p})^c$. By Lemma 3 we have $\lambda_{\theta}(\Theta_{P_1,p}) = \lambda_{\theta}(\Theta)$. I prove the statement of the theorem via contradiction. Suppose there exists a subset $\hat{\theta}$ of $\Theta$ such that all its elements are Bayesian equivalent and $\hat{\theta}$ contains a weight state $\theta \in \Theta_{P_1,p}$.

Let $q^*(\omega|p)$ denote the Bayesian equivalent limit belief for weight states $(\Theta_{P_1,p})^c$. By Lemma 3 we have $P_i^1(\omega|p) \cap C = P_i^1(\omega|p)$. I prove the statement of the theorem via contradiction. Suppose there exists a subset $\hat{\theta}$ of $\Theta$ such that all its elements are Bayesian equivalent and $\hat{\theta}$ contains a weight state $\theta \in \Theta_{P_1,p}$, then

$$q^*(\omega|p) = \frac{P_i^1(\omega|p) \cap C}{P_i^1(\omega|p)}.$$

For a Bayesian equivalent weight state $\theta \in \Theta_{P_1,p}$, the following holds by definition

$$\sum_{i \in V} [\mu(\theta)]_i = q^*(\omega|p).$$

As the information setting is not trivial there exists a fundamental state $\omega \in \Omega$ such that $q_1^1(\omega|p) \neq q_1^*(\omega|p)$, where $q_1^*(\omega|p)$ equals a vector with each component equaling $q^*(\omega|p)$. Consider the set of weight states $\Theta_{q_1^1(\omega),q_1^*(\omega)}$ such that average of the first period announcement vector $q_1^1(\omega|p)$ (for weights derived from the stationary distribution) is equal to the Bayesian equivalent limit belief $q^*(\omega|p)$. As all elements of $\Theta_{P_1,p}$ are Bayesian equivalent we have $\Theta_{P_1,p} \subset \Theta_{q_1^1(\omega),q_1^*(\omega)}$. But Lemma 2 implies that $\Theta_{q_1^1(\omega),q_1^*(\omega)}$ has measure zero and hence $\Theta_{P_1,p}$ has measure zero establishing a contradiction.

4 Perfect Information Aggregation in Mixed Communication Structures

The previous section established that non-Bayesian communication structures lead to incorrect consensus beliefs for almost all weight states. From the existing literature, see Mueller-Frank [29], we know that, for almost all priors, Bayesian communication structures lead to consensus beliefs that perfectly aggregate the private information of all agents.

In this section I consider the questions whether mixed communication networks, that consists of only some but not all Bayesian agents, lead to consensus, and under which conditions such mixed communication structures satisfy perfect information aggregation. For this part of the paper redefine information setting as the tuple $((\Omega, F), C, \{P_i\}_{i \in V}, G)$. This differs from the previous definition through the exclusion of the common prior $p$. Consider the space $\Delta$ of common priors on $(\Omega, F)$. The space of common priors $\Delta$ can be represented by the relative interior of the unit
simplex in $\mathbb{R}^k$.\footnote{Remember that only priors that assign positive probability to all non-empty elements of $\mathcal{F}$ are considered. For details on the space of common priors see appendix B.} Let $\mathcal{L}(\Delta)$ denote the $\sigma$-algebra on $\Delta$ and $\lambda_\Delta$ denote the corresponding measure which is derived from the Lebesgue measure in $\mathbb{R}^{k-1}$. Let $\Pi$ denote the product space $\Delta \times \Theta$ with the corresponding product $\sigma$-algebra and product measure $\lambda_\Pi = \lambda_\Delta \times \lambda_\Theta$. In order to gain some intuition for the analysis and the result to come, consider the following example.

**Example:** Consider an undirected star network consisting of four agents $V = \{1, 2, 3, 4\}$. Let 1 be the agent in the center of the star who is linked to every other agent. Agents 2, 3 and 4 are linked only to agent 1. Suppose that agent 2 is Bayesian and the remaining agents are non-Bayesian assigning equal weights to each of their neighbors and themselves, i.e.

$$\theta^1 = \left( \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4} \right)$$

$$\theta^2 = \left( \frac{1}{2}, 0, \frac{1}{2}, 0 \right)$$

$$\theta^3 = \left( \frac{1}{2}, 0, 0, \frac{1}{2} \right).$$

The state space is represented by the rectangle in figure 1 below.

![State space of example](image)

**Figure 1:** State space of example

Suppose that the agents share a uniform prior on the state space. The event $C$ is represented by the shaded subset of the state space. Agent 1 and agent 2 do not receive any private information, $\mathcal{P}_1(\omega) = \mathcal{P}_2(\omega) = \Omega$ for all $\omega \in \Omega$. The partition cells of agent 3 are given by the columns of the matrix,

$$\mathcal{P}_3 = \{ J_1 \cup J_3 \cup J_5; J_2 \cup J_4 \cup J_6 \}$$

while agent 4 learns the cells of the matrix

$$\mathcal{P}_4 = \{ J_1, J_2, J_3, J_4, J_5, J_6 \}.$$
Note that the join of the partitions equals the partition of agent 4. Consider the first period announcement vectors as a function of the realized state

\[ q^1(\omega) = \begin{cases} 
(1, 1, 1, 1) & \text{if } \omega \in J_1 \\
(1, 2, 2, 2) & \text{if } \omega \in J_2 \\
(1, 1, 1, 0) & \text{if } \omega \in J_3 \\
(1, 2, 2, 1) & \text{if } \omega \in J_4 \\
(1, 1, 2, 2) & \text{if } \omega \in J_5 \\
(1, 2, 2, 1) & \text{if } \omega \in J_6
\end{cases} \]

Since the join of the partitions equals the partition of agent 4, perfect learning occurs if the beliefs of all agents converge to the first period announcement of agent 4. Note that this occurs only if the Bayesian agent can infer the first period announcement of agent 4 through the announcements of the center agent. We have

\[ q^2(\omega) = \begin{cases} 
5 \frac{1}{12} & \text{if } \omega \in J_1 \\
7 \frac{1}{12} & \text{if } \omega \in J_2 \\
1 \frac{1}{3} & \text{if } \omega \in J_3 \\
2 \frac{1}{3} & \text{if } \omega \in J_4 \\
1 \frac{1}{2} & \text{if } \omega \in J_5 \\
1 \frac{1}{2} & \text{if } \omega \in J_6
\end{cases} \]

Suppose that a state \( \omega \) in \( J_5 \cup J_6 \) is realized which leads to

\[ q^3(\omega) = \begin{cases} 
\left( \frac{1}{2}, \frac{1}{2}, \frac{5}{12}, \frac{7}{12} \right) & \text{if } \omega \in J_5 \\
\left( \frac{1}{2}, \frac{1}{2}, \frac{7}{12}, \frac{5}{12} \right) & \text{if } \omega \in J_6
\end{cases} \]

If agent 2 observes agent 1 announcing \( \frac{1}{2} \) in period two, she concludes that the realized state of the world lies in either cell \( J_5 \) or \( J_6 \) as all other states would have lead to a different second round announcement of agent 1. Agent 2's third period announcement is hence equal to \( q^3_2(\omega) = \frac{1}{2} \). Due to the symmetric way agent 1 treats agents 3 and 4, and (the non-Bayesian) agents 3 and 4 treat agent 1, observing the announcement of agent 1 in periods \( t \geq 3 \) allows for no further rational inferences. Learning ends for agent 2 at the beginning of period three. Her announcement remains \( \frac{1}{2} \) for all periods \( t \geq 3 \). Straightforward arguments from the theory of absorbing Markov chains then establish that the opinions of the remaining agents converge to \( \frac{1}{2} \). Perfect learning fails for any state \( \omega \in J_5 \cup J_6 \) as the limit consensus belief \( \frac{1}{2} \) is unequal to the Bayesian posterior conditioning on the realized cell of the join \( J_5 \) respectively \( J_6 \) which is equal to \( \frac{2}{3} \) respectively \( \frac{1}{3} \). Note that for any pair of distinct first period announcement vectors \( q^1(\omega) \), \( q^1(\omega') \) every feasible weight vector \( \bar{\theta}^1 \) assigning different weights to agents 3 and 4 leads to different second period announcements of agent 1 implying that agent 2 can infer the first period announcement vector from the second period announcement of agent 1 and therefore the realized cell of the join of partitions. □
The example highlights that perfect learning can fail in mixed communication structures, for measure zero weight states, and suggest that perfect learning might hold generically. The following theorem establishes a positive result for information aggregation in mixed communication structures.

**Theorem 2** For a given information setting consider a mixed communication structure, \( \sim \mathcal{B} \notin \{\emptyset, V\} \). Perfect information aggregation occurs for \( \lambda, \lambda_\Theta \) almost every \( (p, \theta) \in \Delta \times \Theta \).

Theorem 2 establishes that asymptotic consensus holds in mixed communication structures for almost all priors and weight states. The Bayesian agents achieve consensus in finite time and determine the long run opinions of all agents. Moreover, despite the noise that is introduced through the presence of non-Bayesian agents, the Bayesian agents are able to, over time, infer the private information of all agents from the announcements of their neighbors. Therefore, in a strongly connected social network, Bayesian and non-Bayesian agents are informationally equivalent in the long run. While Bayesian and non-Bayesian agents are asymptotically informationally equivalent, the Bayesian agents might have a (vanishing) informational advantage for infinite periods. In appendix B, a setting is considered where, besides announcing their posterior, every agent myopically selects an expected utility maximizing action in each period. If the set of actions is finite, then Bayesian and non-Bayesian agents turn out to be generically equivalent in expected utility terms within finite time. To summarize the theorem, the presence of at least one Bayesian agent is sufficient for perfect information aggregation in social networks.

It is worthwhile to highlight a feature of the result of Theorem 2; perfect information aggregation occurs generically in any strongly connected network with some Bayesian agents, independent of their proportion and location in the network. Intuition might tempt one to believe that for perfect learning to occur the Bayesian agents require a prominent position in the network and by acting as hubs enable the result. However, the Bayesian agents can be located in a corner of the network, there might even be only one, nevertheless this one Bayesian has a striking impact on the long run opinion, and its quality, of all agents in the network.

Taking a big picture view, Theorem 2 highlights a positive consequence of social interaction; (i) the long run prevalence of informationally optimal opinions, and (ii) the tendency of non-Bayesian agents to obliviously converge to the consensus opinion of the Bayesian agents.

Theorem 3 relates the information aggregation properties of Bayesian and mixed communication structures.
Theorem 3 For a given information setting consider the Bayesian and a mixed communication structure, \( \neg B \notin \{\emptyset, V\} \). The Bayesian and mixed communication structure yield identical limit beliefs for \( \lambda_{\Delta} \times \lambda_\Theta \) almost every \((p, \theta) \in \Delta \times \Theta\), perfectly aggregating the private information of all agents.

Bayesian and mixed communication structures are equivalent in regards to the limit consensus belief. Both yield perfect information aggregation. As long as there exists one Bayesian agent, transforming more agents from non-Bayesian to Bayesian has no effect on the quality of information aggregation. The theorem follows directly from Theorem 2 above and Theorem 4 in Mueller-Frank [29] which establishes generic perfect information aggregation in Bayesian communication structures.

In their influential paper, Golub and Jackson [18] consider the information aggregation properties of non-Bayesian communication structures and show that perfect (or naive) learning occurs in the limit for a growing sequence of networks under a condition on the sequence of weight states, and an independence and additional distributional assumption on the private signals. While Theorem 1 implies that non-Bayesian communication structures fail to aggregate any subset of the private information of agents in finite networks, and hence naive learning fails, Theorem 2 reconciles our results with Golub and Jackson [18]. As long as there is one Bayesian agent in the network perfect information aggregation occurs generically, and hence all non-Bayesian agents naively learn, in any finite network. Moreover, no conditions on the weight state and the private information structure are required.

While the consensus beliefs do not depend on the structure of the network, nor the number and location of the Bayesian agents, the speed of convergence varies with these three parameters. A formal treatment of the speed of convergence of mixed communication structures requires the analysis of the speed of convergence of the infinite product of stochastic matrices to their limit.\(^{18}\) This is omitted here as it would distract from the main focus of the paper, the qualitative properties of information aggregation in mixed communication structures. Note however, that there exists a fundamental difference between the speed of convergence of Bayesian and mixed communication structures; in Bayesian communication structures consensus is achieved in finite time along the shortest path while in mixed communication structures consensus is generally only achieved in the limit.

There is one important point to mention in regards to the implications of the results stated so far. The proof of Theorem 2 heavily exploits the assumption that there is common knowledge among the Bayesian agents of the updating function each non-Bayesian agent uses. This assumption

\(^{18}\)For an analysis of speed of convergence in non-Bayesian learning models see for example Jadbabaie and Tahbaz-Salehi [25], Nedic and Ozdaglar [32], and Olshevsky and Tsitsiklis [33].
clearly fails in reality. No matter how sophisticated an agent is he cannot know the precise updating functions of all agents in a large network. So the results serve as a benchmark stating what would occur in a perfect world. In section 5 the effect of relaxing the assumption of common knowledge of the updating functions is analyzed.

4.1 Structure of the Proof of Theorem 2

This subsection not only provides an overview of the structure of the proof but the first part of the proof itself. The proof of the second part can be found in appendix A. The proof of the theorem is based on the following reasoning. Let \( T \) denote a \( v \times v \) row stochastic matrix derived from the weight state \( \theta \) and the communication structure such that

\[
T_{kl} = \begin{cases} 
1 & \text{if } k = l \in B \\
\theta^k_l & \text{if } k \in \neg B, \ l \in N_k \cup \{k\} \\
0 & \text{otherwise}
\end{cases}
\]

The announcement of agent \( i \in B \) in period \( t \geq 1 \) is given by his posterior conditioning on his information, while the announcement of a non-Bayesian agent \( j \) in period \( t \geq 2 \) is given by

\[
q^t_j(\omega | (p, \theta)) = [T \times q^{t-1}(\omega | (p, \theta))]_j.
\]

As the partitions of all agents are finite, for all Bayesian agents learning ends in finite time. Suppose that in period \( t^* \) learning ended for all Bayesian agents. We have

\[
q^{t^*+n}(\omega | (p, \theta)) = T^n \times q^{t^*}(\omega | (p, \theta))
\]

for all \( n \in \mathbb{N} \). Simple arguments from the theory on absorbing Markov chains then show that the limit announcement of every non-Bayesian agent is a weighted average of the learning-end announcements of a subset of the Bayesian agents.

The diameter, \( \text{diam}(G) \), of network \( G \) equals the greatest distance between any two agents in network \( G \). The proof establishes that for generic \( (p, \theta) \) (i) learning ends by period \( t^* = \text{diam}(G) + 1 \) for each Bayesian agent, and (ii) in \( t^* \) the announcement of every Bayesian agent equals the posterior probability of event \( C \) conditioning on the realized cell of the join of the partitions of all agents.

In the following I provide the first part of the proof of the theorem. First note that common knowledge of the (i) graph structure, (ii) weight state \( \theta \), (iii) partitions \( \mathcal{P}_t \) for all \( i \in V \), and (iv) the identity of the Bayesian and non-Bayesian agents, implies that the period-\( t \)-partition \( \mathcal{P}^t_i \) of each Bayesian agent is common knowledge among all Bayesian agents. For agent \( i \) let \( \mathcal{P}^{PL}_{i,t} \) denote the period-\( t \) perfect learning partition, the join of the partitions of agents within distance of \( t - 1 \) from
Bayesian learning of agent $i$ implies that agent $i$'s partition in period $t$, $P_{i,t}^t$, is at most as fine as $P_{i,t}^{PL}$, i.e. $P_{i,t}^{PL} \leq P_{i,t}^t$. The perfect learning announcement of agent $i$ in period $t$, $q_{i,t}^*(\omega|p)$, is defined as

$$q_{i,t}^*(\omega|p) = \frac{p\left(C \cap P_{i,t}^{PL}(\omega)\right)}{p\left(P_{i,t}^{PL}(\omega)\right)}.$$

From the perfect learning announcement in period $t$, $q_{i,t}^*$, construct the following partition $P_{i,t}^*$:

$$P_{i,t}^*(\omega|p) = \{\omega' \in \Omega : q_{i,t}^*(\omega'|p) = q_{i,t}^*(\omega|p)\}.$$

Note that $P_{i,t}^*$ is at most as fine as the perfect learning partition, $P_{i,t}^{PL} \leq P_{i,t}^*$. To establish the statement of the theorem it is sufficient to prove the following statement which I denote as statement A:

**Statement A:** Consider a Bayesian agent $i$. For every period $t$ there exists a generic set of priors $\Delta^* \subset \Delta$ such that for each $p \in \Delta^*$ there exists a generic set of weight states $\Theta^*_p \subset \Theta$ and for all weight states $\theta \in \Theta^*_p$ we have $q_{i,t}^t(\omega|(p, \theta)) = q_{i,t}^*(\omega|p)$ for every $\omega \in \Omega$.

If statement A is correct, then there exists a generic set of priors $\Delta^*$ and $a$, to each such prior, corresponding generic set of weight states $\Theta^*_p$, where $t^* = \text{diam}(G) + 1$, such that for all $\theta \in \Theta^*_p$ the announcement of every Bayesian agent $i$ in periods $t \geq t^*$ satisfies $q_{i,t}^t(\omega|(p, \theta)) = q_{i,t}^*(\omega|p)$ for all $\omega \in \Omega$. In order to show sufficiency of statement A for the statement of the theorem we require the following lemma.\footnote{Lemma 4 is used both in the proof of Theorem 2 and Theorem 4. Hence the more general formulation for $\Theta^k$ rather than $\Theta$.}

**Lemma 4** Consider the measure spaces $(\Delta, \mathcal{F}_\Delta, \lambda_\Delta)$, $(\Theta^k, \mathcal{F}_{\Theta^k}, \lambda_{\Theta^k})$, for $k \in \mathbb{N}$, and a generic set $\Delta^* \subset \Delta$ such that for each $p \in \Delta^*$ there is a corresponding generic subset $\Theta^*_p$ of $\Theta^k$, $\Theta^*_p \in \mathcal{F}_{\Theta^k}$. The set $\mathbf{F}$

$$\mathbf{F} = \{(p, \theta) : p \in \Delta^*, \theta \in \Theta^*_p\}$$

is generic in $\Delta \times \Theta^k$.

Statement A together with Lemma 4 then implies that there exists a generic subset $\Pi^*$ of the product space $\Delta \times \Theta$ in which perfect information aggregation occurs.
For the proof of statement A we require the following lemma.\(^{20}\) Let \(\phi_p\) be the announcement function which, for a given prior \(p\) on \(\Omega\), assigns the posterior probability of event \(C\) to each possible information set,

\[
\phi_p : 2^{c \in V} \setminus \{\emptyset\} \to [0,1].
\]

**Lemma 5** Define \(I \subset \mathcal{F}\) as

\[
I = \left\{ I \in 2^{c \in V} : I \cap C \neq \emptyset \text{ and } I \cap C^C \neq \emptyset \right\}
\]

There exists a \([\lambda_\Delta]\) full measure set of priors \(\Delta^* \subset \Delta\) such that \(\phi_p : I \to [0,1]\) is injective for every prior \(p \in \Delta^*\).

Statement A is proven by induction in the following. To establish the base case for \(t = 2\) consider a prior \(p \in \Delta^*\), a Bayesian agent \(i\) and his neighbor \(i_1\). Note that every agent’s first period announcement is based upon Bayesian updating conditioning on his realized partition cell. The realized weight state \(\theta\) does not affect the first period announcement of any agent. Suppose that \(\mathcal{P}_{i_1}(\omega) \notin I\). Since all priors \(p \in \Delta\) assign positive probability to every non-empty element of \(\mathcal{F}\) and \(p \in \Delta^*\), Lemma 5 implies that agent \(i\) learns the realized cell \(\mathcal{P}_{i_1}(\omega)\) through \(i_1\)’s first period announcement. Suppose that \(\mathcal{P}_{i_1}(\omega) = I\). By the same reasoning, agent \(i\) learns that \(\mathcal{P}_{i_1}(\omega)\) is a subset of \(C\) (or \(C^C\)) through agent \(i_1\)’s first period announcement. Therefore in each fundamental state \(\omega\) the following is satisfied for the second period partition of agent \(i\):

\[
\mathcal{P}^2_i(\omega) \begin{cases} 
\mathcal{P}^{i,l}_i(\omega) \\
\subset C \\
\subset C^C
\end{cases}
\]

In each of the three cases the second period announcement of \(i\) satisfies \(q^2_t(\omega | p) = q^*_t,2(\omega | p)\). By Lemma 5 \(\Delta^*\) is a full measure set. Setting \(\Theta^*_p = \Theta\) establishes the base case.

For the inductive step suppose that statement A is true for period \(t\), i.e. for each \(p \in \Delta^*\) there exists a full measure set \(\Theta^*_p\) such that \(q^*_t(\omega | (p, \theta)) = q^*_t,2(\omega | p)\) holds for every \(i \in B\), \(\omega \in \Omega\) and \(\theta \in \Theta^*_p\). To conclude the proof of statement A, I need to show that for each \(p \in \Delta^*\) there exists a full measure set \(\Theta^*_{p,t+1}\) such that \(q^*_{t+1}(\omega | (p, \theta)) = q^*_{t+1,2}(\omega | p)\) holds for every \(i \in B\), \(\omega \in \Omega\) and \(\theta \in \Theta^*_{p,t+1}\). Note that if the partition of agent \(i\) in period \(t + 1\), \(\mathcal{P}^t_{i,t+1}\) is finer than \(\mathcal{P}^*_t\), then

\(^{20}\)The same lemma is also established in the proof of Theorem 4 in Mueller-Frank [29]. However, this paper provides an alternative proof.
\[ q_{i,t+1}^{\ell}(\omega | (p, \theta)) = q_{i,t+1}^*(\omega | p) \] for all \( \omega \in \Omega \). Therefore, if for each \( p \in \Delta^* \) there exists a generic set \( \Theta_p^{i,t+1} \) such that \( \mathcal{P}_i^{l+1}(\cdot | (p, \theta)) \leq \mathcal{P}_i^{*}(\cdot | p) \) holds for every \( i \in B \) and \( \theta \in \Theta_{p,t+1}^i \), then statement A holds. This is established in the appendix. The proof requires some more detail on the non-Bayesian announcement functions in mixed communication structures and one additional lemma.

**Lemma 6** Let \( g : \Theta \to \mathbb{R} \) be a non-trivial polynomial function. Then the zero set of \( g \) has \([\lambda_\Theta]\) measure zero in \( \Theta \).

### 5 Lack of Common Knowledge of the Updating Functions and Perfect Information Aggregation

So far all Bayesian agents were assumed to commonly know the realized weight state \( \theta \) which was encoded in each fundamental state \( \omega \). This is a strong and unrealistic assumption. In this section we check the robustness of the perfect learning result to relaxing the assumption of common knowledge of \( \theta \), or in other words of common knowledge of the updating functions of the non-Bayesian agents.

Suppose that there exists a countable set of weight states \( \tilde{\Theta} \) which are commonly considered possible among the Bayesian agents. More precisely assume that for each non-Bayesian agent \( j \) there exists a set of countably many updating functions \( \tilde{\Theta}_j \subset \Theta_j \) that are commonly considered possible among the Bayesian agents. We have

\[ \tilde{\Theta} = \prod_{j \notin B} \tilde{\Theta}_j. \]

This generalization can be formally represented through an extended state space \( \Gamma \) consisting of the product of the fundamental state space and the set of possible weight states, \( \Gamma = \Omega \times \tilde{\Theta} \).

Suppose that each agent \( i \) is endowed with a probability measure \( p_i^\theta \) over \( \tilde{\Theta} \) such that \( p_i^\theta(\theta) > 0 \) for all \( \theta \in \tilde{\Theta} \). This allows for heterogeneous priors over the set of possible weight states. Let \( \mathcal{P}_i^\theta \) denote agent \( i \)'s partition of \( \tilde{\Theta} \), the set of possible weight states. I assume that each non-Bayesian agent \( j \) knows his own weight state \( \theta^j \) while each Bayesian agent \( i \) has no information regarding the realized weight state,

\[ \mathcal{P}_j^\theta(\theta) = \{ \tilde{\theta} \in \tilde{\Theta} : \tilde{\theta}^j = \theta^j \} \quad \mathcal{P}_i^\theta(\theta) = \tilde{\Theta}. \]

This implies the following extended state space partition for every agent \( i \in V \)

\[ \mathcal{G}_i(\omega, \theta) = \{ (\omega', \tilde{\theta}) : \omega' \in \mathcal{P}_i(\omega), \tilde{\theta} \in \mathcal{P}_i^\theta(\theta) \}. \]

The probability measure on the extended state space is the product of probability measures on \( \Omega \).
and $\Theta$. For all measurable sets $D \in \mathcal{F}$ and $S \subset \Theta$ we have

$$p_i^T(D, S) = p(D) \times p_i^T(S).$$

Let $N$ be the cardinality of the set of possible weight states $\Theta$.$^21$ Note that if the goal is to establish a result on information aggregation for generic sets $\Theta \in \Theta^N$ the statement has to be made for elements of the product space $\Theta^N$. Here $\Theta^N$ is the product measure space with the corresponding sigma algebra $\mathcal{F}^N$ and product measure $\lambda_{\Theta^N}$.$^22$ Let $\Xi$ denote the product space $\Delta \times \Theta^N$ with the corresponding product $\sigma$-algebra and product measure $\lambda_{\Xi} = \lambda_{\Delta \times \Theta^N}$.

In order to simplify the analysis let us assume that under no circumstances, even if pooling the private information of all agents, does an agent ever know with certainty whether the event $C$ or its complement is realized and denote this property of the information setting as property $I$.

**Property I:** For all cells $P \in \bigvee_{i \in V} \mathcal{P}_i$ we have $P \cap C \neq \emptyset$ and $P \cap C^C \neq \emptyset$.

The following theorem establishes the robustness of perfect learning in mixed communication structures to lack of common knowledge of the non-Bayesian updating functions.

**Theorem 4** For a given information setting satisfying property $I$ consider a mixed communication structure, $B \notin \{\emptyset, V\}$. For $[\lambda_{\Delta \times \Theta^N}]$ almost every $(p, \Theta) \in \Delta \times \Theta^N$ perfect information occurs in all $\theta \in \Theta$ and for all priors $\{p_i^\theta\}_{i \in V}$.

Perfect information aggregation occurs generically in any mixed communication structure as long as the pooled private information does not fully reveal the realization of the event, and, for each non-Bayesian agent, only countably many different updating functions are commonly considered possible among the Bayesian agents. Property $I$ is not necessary for the perfect information aggregation result.$^23$ It does however allow for an important simplification of the proof and is made for expositional purposes.

The perfect information aggregation result of Theorem 2 and the informational equivalence of Bayesian and mixed communication structures are based on the assumption that there is common knowledge among all Bayesian agents which updating rule every non-Bayesian agent uses. This is a very strong assumption and limits the ability to draw practical implications from the results.

$^{21}$ $N$ may be finite or countably infinite.

$^{22}$ Note that for infinite $N$ the measure $\lambda_{\Theta^N}$ exists by Kolgomorov’s Extension Theorem. For a simpler argument see Halmos [20] Theorem B, page 157.

$^{23}$ Similar to the proof of Theorem 2, the proof would otherwise be based upon establishing statement $A$.  

22
But the qualitative nature of the results does not depend on this strong assumption as Theorem 4 establishes. The assumption that there are only countably many possible updating functions for each agent is not particularly strong and the fact that the Bayesian agents do not require a common prior over the set of possible updating functions even less so. Theorem 4 shows that the perfect information aggregation property of mixed communication structures and the long run equivalence between Bayesian and non-Bayesian agents has a certain robustness to it. However, one still has to be careful when taking the normative implications of the model to the real world as the set of non-Bayesian updating functions is restricted to weighted averages. The natural next step following the current analysis would be to check the robustness of the results in regards to a more general class of non-Bayesian updating functions. This is left for future work.

5.1 Structure of the Proof of Theorem 4

This subsection does not only provide an overview of the structure of the proof but presents the first part of the proof itself. The second part can be found in appendix A. First note that common knowledge of the (i) network $G$, (ii) set $\Theta$, (iii) partitions $G_i$ for all $i$, and (iv) the identity of the Bayesian and non-Bayesian agents, imply that the period-$t$-partition $G^t_i$ of each Bayesian agent $i$ is common knowledge among all Bayesian agents. Let $G_i = (V_i, E_i)$ denote the subgraph of $G = (V, E)$ where all Bayesian agents but $i$ are removed,

$$V_i = \{i\} \cup V \setminus B$$
$$E_i = \{jk \in [V_i]^2 : jk \in E\}.$$

For agent $i$ let $G^{PL}_{i,t}$ denote the perfect learning partition in period $t$. For each $(\omega, \theta)$ the partition $G^{PL}_{i,t}$ satisfies

$$G^{PL}_{i,t}(\omega, \theta) = \left\{ (\omega', \bar{\theta}) \in \Gamma : \omega' \in \bigvee_{k \text{ s.t. } d_{G(i,k)} \leq t-1} P_k(\omega), \bar{\theta} \in \bigvee_{k \text{ s.t. } d_{G(i,k)} \leq t-2} \mathcal{P}_{\tilde{\Theta}}^k(\theta) \right\}.$$

where $d_{G(0)}(i,k) \in \{-1,0\}$ implies $k = i$. To establish the statement of the theorem it is sufficient to establish the following statement which I denote as statement B.

**Statement B:** For every period $t$ there exists a generic set of priors $\Delta^* \subset \Delta$ such that for each $p \in \Delta^*$ there exists a generic set $\Theta^N_{p,t} \subset \Theta^N$ such that for all $\tilde{\Theta} \in \Theta^N_{p,t}$ the period-$t$-partition partition $G^t_i(\cdot | (p, \tilde{\Theta}))$ of every Bayesian agent equals $G^{PL}_{i,t}$.

If statement B is correct, then there exists a generic set of priors $\Delta^*$ and a, to each such prior, corresponding generic set $\Theta^N_{p,t^*}$, where $t^* = \text{diam}(G) + 2$, such that $G^t_i(\cdot | (p, \tilde{\Theta})) = G^{PL}_{i,t^*}(\omega, \theta)$ for
all \( t \geq t^* \). Note that for \( G_i^t = G_{i,t}^{PL} \) simple equivalent transformations establish that the posterior probability of event \( C \in \mathcal{F} \) of agent \( i \) in period \( t \) is given by

\[
q_t^i(\omega, \theta \mid (p, \Theta)) = \frac{p_t^i((C, \Theta) \cap G_{i,t}^{PL}(\omega, \theta))}{p_t^i(G_{i,t}^{PL}(\omega, \theta))} = \frac{p(C \cap \bigvee_{k \in V} \mathcal{P}_k(\omega))}{p \left( \bigvee_{k \in V} \mathcal{P}_k(\omega) \right)}.
\]

due to the assumption that \( p_t \) is the product of probability measures \( p \) and \( p_t^i \). Therefore the announcement is independent of the prior \( p_t^i \). Lemma 4 then implies that there exists a generic subset \( \Xi^* \) of the product space \( \Delta \times \Theta^N \) in which perfect information aggregation occurs.

Statement B is proven by induction. Note that property I implies

\[
I = 2^k \setminus \{ \emptyset \}.
\]

The base case for \( t = 2 \) then follows from the generic injectivity of the Bayesian announcement function \( \phi_p \) by Lemma 5. For all \( p \in \Delta^* \) we have \( \Theta_{p,2}^N = \Theta^N \) which is a generic set.

For the inductive step suppose that the statement is true for period \( t \), i.e. for each \( p \in \Delta^* \) there exists a generic set \( \Theta_{p,t}^N \) such that the period-\( t \) partition of every Bayesian agent equals \( G_{i,t}^{PL} \). Consider a Bayesian agent \( i \) and the inferences he draws from the period-\( t \) announcement of each of his neighbors. Consider first a Bayesian neighbor \( i_1 \) who knows the realized cell of the join of fundamental partitions of all agents within distance of \( t - 1 \) (and hence distance \( t \) of agent \( i \)) due to the induction hypothesis. Due to the injectivity of \( \phi_p \) for \( p \in \Delta^* \) agent \( i \) learns the realized cell of the fundamental partition of all agents with distance of \( t - 1 \) of \( i_1 \) through the announcement of \( i_1 \) in period \( t \) for all \( \Theta \in \Theta_{p,t}^N \). While \( i \) might refine his fundamental partition based upon the period \( t \) announcement of his Bayesian neighbor, he cannot draw any further inferences in regards to the realized weight state.

Next consider a non-Bayesian neighbor \( j \) of agent \( i \). If for each prior \( p \in \Delta^* \) there exists a generic set \( \Theta_{p,t+1}^N(j) \subset \Theta^N \) such that \( i \) can infer, through the period-\( t \) announcement of \( j \), (i) the fundamental partition of all agents with distance of \( t - 1 \) of \( j \) in \( G \), and (ii) the realized weight state of all agents with distance of \( t - 2 \) of \( j \) in \( G_i \), for all \( \Theta \in \Theta_{p,t+1}^N(j) \), then the induction step is established as there are only finitely many agents. Consider a prior \( p \in \Delta^* \), a set \( \Theta \in \Theta^N \) and
compare the announcement of \( j \) in period \( t \) for two different extended states \((\omega', \theta') \notin \mathcal{G}_{j,t}(\omega, \theta)\)

\[
q_j^*_{j}(\omega, \theta | (p, \bar{\Theta})) = f_j^*_{j}(\theta) \times q_{Y_j}(\omega, \theta | (p, \bar{\Theta}))
\]

and

\[
q_j(\omega', \theta' | (p, \bar{\Theta})) = f_j^*_{j}(\theta') \times q_{Y_j}(\omega', \theta' | (p, \bar{\Theta})).
\]

If \( q_j^*_{j}(\omega, \theta) \) is not equal to \( q_j(\omega', \theta') \) then agent \( i \) can distinguish \((\omega', \theta')\) and \((\omega, \theta)\) through the announcement of \( j \) in period \( t \). We have to show that for a fixed prior \( p \in \Delta^* \) and generic sets \( \bar{\Theta} \) the announcements are unequal for all pairs \((\omega', \theta') \notin \mathcal{G}_{j,t}(\omega, \theta)\). Let us start by ordering the elements of \( \bar{\Theta} \) by their position from 1 to \( N \). The weight state \( \theta \in \bar{\Theta} \) on position \( \alpha \) is denoted by \( \bar{\Theta}_{\alpha} \). For a prior \( p \in \Delta^* \) the induction hypothesis implies that there exists a generic set \( \Theta_{p,t}^N \) such that the Bayesian vectors \( q_{Y_j} \) depend only on \( \omega \),

\[
q_{Y_j}(\omega, \bar{\Theta}_{\alpha} | (p, \bar{\Theta})) = q_{Y_j}(\omega | p)
\]

for all \( \omega \in \Omega \), all \( \bar{\Theta} \in \Theta_{p,t}^N \) and all \( \alpha \in \{1, ..., N\} \). As \( q_{Y_j}^* \) is a vector of Bayesian announcements of agents within distance of \( t - 1 \) of \( j \) for \( \omega' \in \mathcal{P}_{j,t}^L(\omega) \) we have

\[
q_{Y_j}^*(\omega | p) = q_{Y_j}^*(\omega' | p).
\]

The injectivity of \( \phi_p \) and the induction hypothesis further imply

\[
q_{Y_j}^*(\omega | p) \neq q_{Y_j}^*(\omega' | p)
\]

for \( \omega' \notin \mathcal{P}_{j,t}^L(\omega) \). The proof of Theorem 4 now constructs a countable set of polynomial functions such that the union of their zero sets in \( \Theta^N \) is equal to the subset of \( \Theta^N \) in which agent \( j \) does not distinguish among some pair \((\omega', \theta') \notin \mathcal{G}_{j,t}(\omega, \theta)\).

6 Truthful Communication versus Strategic Interaction

In the communication model analyzed in this paper agents are myopic. They truthfully announce their posteriors, not taking into account the externalities of their announcements on future announcements of other agents. For example, lying in one period might lead to superior information in a later period.

Myopic behavior is a common feature in the learning literature. The literature on repeated interaction, see for example Geanakoplos and Polemarchakis [17], Parikh and Krasucki [35], and Gale and Kariv [15] explicitly imposes myopic behavior on the agents. That is, agents are assumed
to truthfully communicate their posteriors or to select stage expected utility maximizing actions in each round.

In the sequential social learning literature, see for example Banerjee [5], Bikhchandani, Hirshleifer and Welch [7], and Smith and Sorensen [37], agents also select actions under absence of strategic consideration. In this literature however, myopia is a consequence of the modelling approach where each agent acts exactly once. Hence agents have no incentive to account for the externalities of their choice and the myopic action is indeed strategically optimal.

The main question addressed in this section is whether the perfect information aggregation result carries forward to a fully strategic setting, or, whether it is an artefact of agents being myopic.

Let us consider the following network interaction game. The informational setting

$((\Omega, \mathcal{F}, p), C, \{P_i\}_{i \in V}, G)$

is as in the communication model analyzed in sections 2 till 5. All agents share a common prior $p$. In each of countable rounds every agent (simultaneously) selects an action $a \in [0, 1]$. The stage utility of each agent is given by a strict proper scoring rule.\textsuperscript{24} Under a strict proper scoring rule the unique expected utility maximizing action is to announce one’s true distribution. In our case the relevant state space is given by $\{C, C^c\}$. One example of such a utility function is a squared loss function,\textsuperscript{25}

$$u(a, \omega) = - (a - 1 [\omega \in C])^2.$$ 

Under such a utility function an agent maximizes his expected stage utility by setting the action equal to the posterior probability of event $C$. Squared loss utility functions are common in the literature on information aggregation (for examples, see Acemoglu, Bimpikis and Ozdaglar [1], Galeotti, Ghiglino and Squintani [16], and Ostrovsky [34] who all use a variant of this utility function).

The players do not observe their stage utility realizations, but, as in the communication model, the history of actions of their neighbors. Suppose that the Bayesian agents are fully rational and have an unrestricted strategy space while the non-Bayesian agents use a weighted average updating function when revising their actions hence acting mechanically. The utility of a player is given by the discounted sum of stage utilities with a discount factor $\delta \in (0, 1)$.

Consider a Bayesian agent $i$ and suppose that all other Bayesian agents act myopically, setting

\textsuperscript{24}Scoring rules serve to elicit distributions over a state space. A scoring rule assigns payoffs according to the predicted distribution and the realized state.

\textsuperscript{25}$1 [\omega \in C]$ denotes the indicator function, which takes the value 1 if $\omega \in C$ and 0 otherwise.
their action equal to their posterior probability of event $C$. For some information settings and discount factors close to one, agent $i$ might have an incentive to deviate from his stage utility maximizing action in order to achieve an information gain in later periods. However, for almost all priors and non-Bayesian updating functions agent $i$’s best response is to act myopically in each round. That is, behaving myopically is a perfect Bayesian equilibrium of the network interaction game. This follows from the proof of Theorem 2, more precisely from statement A which establishes that learning occurs along the shortest path in the network. Therefore, behaving myopically is the only best response of agent $i$ to all other Bayesian agents acting myopically as it maximizes the expected stage utility in each round and learning cannot be improved upon by deviating.

Perfect information aggregation is therefore satisfied as a perfect Bayesian equilibrium property generically in Bayesian and mixed communication structures, and is hence not an artefact of agents acting myopically.

7 Applications: Trading and Prediction Markets

In this section I consider two specific applications that fit within the general framework. Both rely on the market scoring rule model which was introduced by Hanson [21, 22]. In a market scoring rule model agents repeatedly predict the realization of some uncertain event. Payments of agents are derived from strict proper scoring rules. In particular, the utility (or payoff) of each prediction equals the difference of the score of the prediction and the score of the prediction it replaced. The history of predictions is commonly observed by all market participants.

In a market scoring rule model, the myopically optimal action of each agent in a given round is to set the prediction equal to the conditional probability of the event. The market scoring rule model is therefore generically equivalent to our communication model which implies that the predictions of all agents converge to the pooled information posterior as long as there exists one Bayesian agent.\footnote{Note that in the communication model all agents simultaneously select an action, while in the market scoring rule model agents act sequentially. Nevertheless, our results carry forward to a strict sequential communication model. Likewise, the communication model is generically equivalent to a discounted market scoring rule model where agents act simultaneously in each round.}

Market scoring rule models are used in the theoretical literature and in practice. For example, prediction markets can be implemented via market scoring rules. In fact, so does Microsoft for its internal prediction market with the purpose of aggregating private information of its employees.\footnote{See Snowberg, Wolters and Zitzewitz [38].}

In the theoretical literature, Ostrovsky [34] models a financial market according to a market...
scoring rule framework.\textsuperscript{28} A finite set of strategic, privately informed traders sequentially predict the value of a security which generalizes the standard market scoring rule model where agents predict the realization of an event. Ostrovsky shows that for a large class of securities, the price of the security, that is the predictions, converges to the pooled information value of the security.

The analysis of this paper has the following implication for financial markets that are modeled based upon the market scoring rule framework. Consider the class of Arrow-Debreu securities. Suppose that a finite number of privately informed traders, some Bayesian and some non-Bayesian, who revise their predictions based upon a (commonly known) weighted average updating function, trade on a finite number of dynamic markets. The history of predictions within a market is common knowledge among the market participants. The market structures generate a trading network; two traders are neighbors in the network if they are trading on the same market. Theorems 2 then implies that the prices, that is predictions, on all markets converge to the pooled information value of the Arrow-Debreu security generically, as long as there is at least one Bayesian trader in some market and markets overlap, that is the underlying network is strongly connected. Therefore, when restricting attention to a smaller class of securities than Ostrovsky [34], neither all traders being rational nor common knowledge of predictions is necessary in order for predictions to perfectly aggregate the private information in the long run.

8 Related Literature

This paper contributes to the literature on opinion formation in social networks. The most related papers in the Bayesian learning literature are Gale and Kariv [15], and Mueller-Frank [29]. The non-Bayesian agents are modeled according to the standard DeGroot [10] approach by using weighted averages. The most prominent recent non-Bayesian learning papers that apply the DeGroot model are DeMarzo, Vayanos and Zwiebel [11], and Golub and Jackson [18, 19]. The approach of modeling the updating function of non-Bayesian agents as weighted averages is also common in other papers focusing on non-Bayesian learning as for example Acemoglu, Ozdaglar and ParandehGheibi [2], and Jadbabaie, Molavi, Sandroni and Tahbaz-Salehi [24].

An important distinction to DeMarzo, Vayanos and Zwiebel [11] and Golub and Jackson [18] lies (i) in the assumptions on the private information of agents, and (ii) the content or object of communication in the first round of interaction. In DeMarzo, Vayanos and Zwiebel [11] the private information agents receive is restricted to conditional independent, normally distributed random variables with mean \( \mu \). The announcement of an agent in the first period is a linear estimator of the mean \( \mu \) based on his signal. This paper utilizes the partitional information approach as

\textsuperscript{28}He also considers a dynamic trading model, based on Kyle [28].
common in the literature on knowledge and consensus. While I consider only finite partitions, no restrictions are imposed on the common prior distribution over the state space. In regards to the content of communication, I consider the case where agents announce their posterior probability of an uncertain event as in Geanakoplos and Polemarchakis [17] and Mueller-Frank [29]. These differences in approach yield different results. For example, Theorem 2 in DeMarzo, Vayanos and Zwiebel [11] establishes that in their setting a balanced non-Bayesian communication structure is necessary and sufficient for the limit belief to perfectly aggregate private information while there are simple examples within the setting considered in this paper where perfect information aggregation fails for balanced communication structures.

Golub and Jackson [18] consider the information aggregation properties of non-Bayesian communication structures by analyzing a sequence of growing networks in which agents receive conditional independent private signals with identical mean and variance. Golub and Jackson [18] show that the sequence of consensus opinions converges to an informationally perfect opinion if the influence of the most influential agent vanishes with the growing sequence of networks. Theorem 1 on the other hand establishes that finite non-Bayesian communication structures generically fail to aggregate any subset of the private information of agents. This difference in results derives from the fact that Golub and Jackson’s naive learning property is established only for the limit of the growing sequence of networks, relying on a law of large number argument.

Acemoglu, Ozdaglar, and ParandehGheibi [2] are concerned with information aggregation in social networks. The difference to this paper lies in the modelling approach and their focus only on non-Bayesian opinion formation. They analyze the spread of misinformation in a purely non-Bayesian opinion formation model under presence of forceful agents who influence the opinions of those they interact with while keeping their own opinions constant. The degree of misinformation is understood as the gap between the expected value of the limit opinion of agents and the true state of the world. They focus on a restricted set of signal generating distributions, show that stochastically asymptotic consensus occurs and provide results in regards to the degree of misinformation depending on the quantity and location of the forceful agents.

---

29 See for example Aumann [4], Geanakoplos and Polemarchakis [17], Krasucki [27], and Parikh and Krasucki [35].
30 However, the structure of the underlying utility function inducing the announcements is identical. As in this paper, De Marzo, Vayanos and Zwiebel [11] use a squared loss utility function.
31 An example of a balanced communication structure is a network where all agents have the same number of neighbors and assign the same weight to each of their neighbors and themselves. Note that balanced communication structures are have measure zero in $\Theta$. 
9 Conclusion

This paper analyzes the information aggregation properties of repeated interaction among agents with differing degrees of sophistication in social networks. I consider a communication model based upon Geanakoplos and Polemarchakis [17]. Each agent initially receives private information regarding the realized state of the world. In each of countable rounds of interaction all agents truthfully announce their (subjective) posterior probability of some relevant event to their neighbors in the network. Two types of agents are considered, fully rational or Bayesian agents and non-Bayesian agents that are modeled according to the standard model of non-Bayesian learning going back to DeGroot [10]. That is, they revise their opinions by averaging over the previous period’s opinions.

Two main results are established. First, the ensuing consensus opinion in networks of only non-Bayesian agents generically fails to represent any subset of the private information of agents. Unsurprisingly, the information aggregation properties of non-Bayesian networks are undesirable. Second, in networks consisting of Bayesian and non-Bayesian agents, all agents generically achieve consensus in the long run, and moreover, the consensus opinion perfectly aggregates the private information of all agents. Two summarize the two main results, the presence of at least one Bayesian agent is necessary and sufficient for perfect information aggregation in strongly connected social networks.

The positive result is quite robust. It carries forward as a perfect Bayesian equilibrium property to a fully strategic setting in which agents have squared loss stage utility functions and discount future utilities. Moreover, common knowledge of the updating functions of the non-Bayesian agents is not necessary. Perfect information aggregation is generically satisfied if there is a countable set of updating functions for each non-Bayesian agent commonly known to contain his true updating function.

While the paper is motivated by general theoretical questions, there are direct implications for specific economic settings. The perfect information aggregation result carries forward to market scoring rule models (see Hanson [21, 22]) which are used in practice to implement prediction markets. Moreover, our positive result implies that in a market scoring rule financial market as in Ostrovsky [34] the price of Arrow-Debreu securities converges to their pooled information value, even under presence of both Bayesian and non-Bayesian traders.

One interesting avenue to extend the analysis of the paper is the following. In this paper I adopt the standard approach in the literature to model the non-Bayesian agents via weighted average updating functions. A companion paper, Mueller-Frank [30], considers a general class of non-Bayesian updating functions and provides a characterization of the class of functions that achieve consensus in non-Bayesian communication structures. A natural extension of the analysis
undertaken here would allow for updating functions within this general class as opposed to the restricted class of weighted average updating functions. This is left for future work.

App\xa0pendix A

Proof of Lemma 1

Proof. As \( f \) is a polynomial function on a subset of \( \mathbb{R}^k \) it is defined on \( \mathbb{R}^k \). Denote the hyperplane in \( \mathbb{R}^k \) that includes \( \Delta(k) \) by \( H^k \) and the restriction of \( f \) to \( H^k \) and \( \Delta(k) \) as \( f_{H^k} \) respectively \( f_{\Delta(k)} \). We need to establish that \( \lambda_{\Delta(k)}(f_{\Delta(k)}^{-1}(0)) = 0 \).

1. Consider \( p \in H^k \). Note that each \( p \in H^k \) satisfies

\[
p = \left( p_1, \ldots, p_{k-1}, 1 - \sum_{i=1}^{k-1} p_i \right).
\]

Denote by \( \pi : H^k \to \mathbb{R}^{k-1} \) the following projection

\[
\pi(p) = (p_1, \ldots, p_{k-1}).
\]

Note that \( \pi \) is a continuous bijection from \( H^k \) to \( \mathbb{R}^{k-1} \). Let \( \lambda_{\mathbb{R}^{k-1}} \) denote the Lebesgue measure in \( \mathbb{R}^{k-1} \) and \( \lambda_{\Delta(k)} \) be a measure on \( \Delta(k) \), the relative interior of the unit simplex in \( \mathbb{R}^k \). For a measurable subset \( S \) of \( \Delta(k) \) we have

\[
\lambda_{\Delta(k)}(S) = \lambda_{\mathbb{R}^{k-1}}(\pi(S)).
\]

2. Consider the function \( g : \mathbb{R}^{k-1} \to \mathbb{R} \) which is derived from \( f \) in the following way

\[
g(p_1, \ldots, p_{k-1}) = f(\pi^{-1}(p_1, \ldots, p_{k-1})).
\]

As \( f \) is a polynomial function, so is \( g \).

3. By a result in geometric measure theory the zero set of a polynomial \( g : \mathbb{R}^{k-1} \to \mathbb{R} \) is either the domain \( \mathbb{R}^{k-1} \) or the zero set of the polynomial has Lebesgue measure zero in \( \mathbb{R}^{k-1} \).\(^{32}\) As \( g \) is not equal to zero we have

\[
\lambda_{\mathbb{R}^{k-1}}(g^{-1}(0)) = 0.
\]

\(^{32}\)See Federer [12] page 240.
4. By definition of $g$ we have
\[ \pi^{-1}(g^{-1}(0)) = f^{-1}(0). \]

5. Combining step 1 and step 4 gives
\[ \lambda_{\Delta(k)}(f^{-1}(0)) = \lambda_{\Delta(k)}(\pi^{-1}(g^{-1}(0))) = \lambda_{\mathbb{R}^{k-1}}(g^{-1}(0)). \]

Hence by step 3 we have
\[ \lambda_{\Delta(k)}(f^{-1}(0)) = 0 \]
concluding the proof. ■

Proof of Lemma 2

Proof. Consider the set $\Theta_{q,q'} \subseteq \Theta$,
\[ \Theta_{q,q'} = \{ \theta \in \Theta : f_{q,q'}(\theta) = 0 \}. \]

We need to show that $\lambda_{\Theta}(\Theta_{q,q'}) = 0$. The Freidlin and Wentzell Theorem implies the following functional form of the stationary distribution $\mu : \Theta \to \Delta(v)$,
\[ \mu(\Theta)_{i} = \frac{\sum_{S_i \in \mathcal{S}_i} \left( \prod_{ij \in E_{S_i}} [T_{\theta}]_{ij} \right)}{\sum_{i \in V} \left( \sum_{S_i \in \mathcal{S}_i} \left( \prod_{ij \in E_{S_i}} [T_{\theta}]_{ij} \right) \right)} \]
where $\mathcal{S}_i$ denotes the set of $i$-trees.\(^{33}\) Therefore $\Theta_{q,q'}$ is equal to the zero set of $g_{q,q'} : \Theta \to \mathbb{R}$ where
\[ g_{q,q'}(\theta) = \sum_{i=1}^{v} \left( \left( \sum_{S_i \in \mathcal{S}_i} \left( \prod_{ij \in E_{S_i}} [T_{\theta}]_{ij} \right) \right) \times (q_i - q'_i) \right). \]

Denote $\prod_{i=2}^{v} \Theta_i$ by $\Theta_{-1}$ and consider the product space $\Theta_1 \times \Theta_{-1} = \Theta$ with respective measures $\lambda_{\Theta_1}$ and $\lambda_{\Theta_{-1}}$. By Fubini’s Theorem $\Theta_{q,q'}$ has $\lambda_{\Theta}$ measure zero if and only if almost every $\Theta_{-1}$-section has measure zero.\(^{34}\) For a fixed $\theta^{-1} \in \Theta_{-1}$ consider the $\Theta_{-1}$-section of $\Theta_{q,q'}$,
\[ (\Theta_{q,q'})_{\theta^{-1}} = \{ \theta \in \Theta : g_{q,q}(\theta, \theta^{-1}) = 0 \}. \]

\(^{33}\)For details see Appendix B.

\(^{34}\)For details see Appendix B.
Note that the function $g_{q,q}(\cdot, \theta^{-1}) : \Delta([N_1] + 1) \to \mathbb{R}$ is a non-trivial polynomial function for every $\theta^{-1}$ since $q \neq q'$. By Lemma 1 the zero set of such a polynomial has $\lambda_{\Delta([N_1] + 1)}$ measure zero. Therefore, every $\Theta_{-1}$-section has measure zero. Applying Fubini’s Theorem then directly implies that $\Theta_{q,q'}$ has $\lambda_\Theta$ measure zero.

Proof of Lemma 3

Proof. First note that $P^1(.|p)$ is finite since the join of partitions is finite and $P^1(.|p)$ is (weakly) coarser than the join of partitions. The objective is to show that generically in $\Theta$ any pair of two different first period announcement vectors $q^1(\omega|p)$, $q^1(\omega'|p)$ leads to different limit consensus opinions $q^\infty(\omega|(p, \theta))$, $q^\infty(\omega'| (p, \theta))$ as the limit partition $P^\infty_i(.|(p, \theta))$ then generically coincides with $P^1(.|p)$. The limit opinion is determined by the first period announcement vector and the unique stationary distribution $\mu(\theta)$ for weight state $\theta$

$$q^\infty(\omega|(p, \theta)) = \sum_{i \in V} [\mu(\theta)]_i \times q^1(\omega|p)$$

and similarly for $q^\infty(\omega'| (p, \theta))$. For each pair $q^1(\omega|p)$, $q^1(\omega'|p)$ consider the corresponding set of weight states $\Theta_{q^1(\omega),q^1(\omega')}$, i.e. the set of weight states in which the limit opinions $q^\infty(\omega|(p, \theta))$ and $q^\infty(\omega'| (p, \theta))$ are identical. By Lemma 2, $\Theta_{q^1(\omega),q^1(\omega')}$ has measure zero in $\Theta$. Construct the set $\Theta_{P^1,p}$ as follows

$$\Theta_{P^1,p} = \Theta \setminus \left( \bigcup_{(q^1(\omega),q^1(\omega')):q^1(\omega) \neq q^1(\omega')} \Theta_{q^1(\omega),q^1(\omega')} \right).$$

Since the join of partitions is finite there are only finitely many distinct sets $\Theta_{q^1(\omega),q^1(\omega')}$ each of which has measure zero. Therefore, $\Theta_{P^1,p}$ is a generic set in $\Theta$.

Non-Bayesian Announcements in Mixed Communication Structures

Let $T$ denote a $v \times v$ row stochastic matrix derived from the weight state $\theta$ and the communication structure such that

$$T_{kl} = \begin{cases} 1 & \text{if } k = l \in B \\ \theta_i^k & \text{if } k \in -B, l \in N_k \cup \{k\} \\ 0 & \text{otherwise} \end{cases}.$$

The announcement of agent $i \in B$ in period $t \geq 1$ is given by the posterior probability of event $C$ conditioning on his information, while the announcement of a non-Bayesian agent $j$ in period $t \geq 2$ is given by

$$q^t_j(\omega|(p, \theta)) = [T \times q^{t-1}(\omega|(p, \theta))]_j.$$
The fact that the first period announcement of all agents is based upon Bayesian updating together with the recursive nature of the announcements of the non-Bayesian agents implies that the announcement of each non-Bayesian agent in period $t$ is a weighted average of only Bayesian announcements. Note that by Bayesian announcements I do not refer only to the announcements of Bayesian agents but all announcements that are based on Bayesian updating, i.e. announcements of Bayesian agents and the first period announcements of all agents. The dimension of the vector of Bayesian announcements a non-Bayesian agent $j$ averages over in period $t$, and the corresponding weights are determined by $T$. For a non-Bayesian agent $j$, let $f^t_j$ be the function that assigns the corresponding weights to the Bayesian announcements agent $j$ averages over in period $t$. To be more precise, consider a non-Bayesian agent $j$ and the network $G^* = (V, E^*)$ which is a subgraph of $G$. The only difference between the two networks is that the outgoing edges of the Bayesian agents are removed in $G^*$, $E^* = E \setminus \{ik : i \in B, k \in V\}$.

Let $Y^t_j(\omega | (p, \theta))$ be the set of announcements that have a strictly positive weight assigned to by $f^t_j : \Theta \rightarrow \text{ri} \left( \Delta \left( |Y^t_j| \right) \right)$. The set $Y^t_j(\omega | (p, \theta))$ has to be defined recursively. We have

$$Y^1_j(\omega | (p, \theta)) = \{q^1_j(\omega | (p, \theta)) : d_{G^*}(j, i) = 0\} = q^1_j(\omega | (p, \theta)).$$

For $Y^t_j(\omega | (p, \theta))$ we have

$$Y^t_j(\omega | (p, \theta)) = Y^{t-1}_j(\omega | (p, \theta)) \cup Z^t_j(\omega | (p, \theta))$$

where the announcements in $Z^t_j(\omega | (p, \theta))$ are those that are first averaged in period $t$ and consists of (i) the first period announcements of agents whose distance from $j$ is exactly $t - 1$, and (ii) one announcement of each Bayesian agent whose distance from $j$ is smaller than $t - 1$,

$$Z^t_j(\omega | (p, \theta)) = \{q^t_i(\omega | (p, \theta)) : d_{G^*}(j, i) = t - 1, i \in V \} \cup \bigcup_{m=2}^{t-1} \{q^m_i(\omega | (p, \theta)) : d_{G^*}(j, i) = t - m, i \in B\}.$$

Denote the set of agents with a corresponding announcement in $Z^t_j(\omega | (p, \theta))$ and $Y^t_j(\omega | (p, \theta))$ by $V_{Z^t_j}$ and $V_{Y^t_j}$ respectively. In each period $t$ agent $j$ has a corresponding weight function

$$f^t_j : \Theta \rightarrow \text{ri} \left( \Delta \left( |Y^t_j| \right) \right)$$

assigning a strictly positive weight to each announcement in $Y^t_j(\omega | (p, \theta))$. Note that $f^t_j$ is a
polynomial function. The announcement of agent $j$ in period $t$ is weighted average of $q_{Y_j}$,

$$
q_j^t(\omega | (p, \theta)) = \sum_{l=1}^{\lvert Y_j \rvert} [f_j^t(\theta)]_l \times [q_{Y_j}(\omega | (p, \theta))]_l.
$$

Proof of Lemma 4

*Proof.* We prove the statement by establishing that the complement $F^C$ of $F$ has measure zero. The complement of $F$ is given by

$$
F^C = \left\{ (p, \theta) : p \in \Delta, \theta \in \left( \Theta^k \right)^C \right\} \cup \left\{ (p, \theta) : p \in (\Delta^*)^C, \theta \in \Theta^k \right\}
$$

$$
= F^C_1 \cup F^C_2.
$$

In order to establish $\lambda_{\Delta \times \Theta^k}(F^C) = 0$ we apply Fubini’s Theorem to the sets $F^C_1$ and $F^C_2$. Note that both sets $F^C_1$ and $F^C_2$ are elements of the product $\sigma$-algebra $F_{\Delta \times \Theta^k}$. Moreover, both $\lambda_{\Delta}$ and $\lambda_{\Theta^k}$ are $\sigma$-finite. Consider a $\Delta$-section $(F^C_1)_{p'}$ of $F^C_1$

$$
(F^C_1)_{p'} = \left\{ \theta \in \Theta^k : (p', \theta) \in F^C_1 \right\} = \left( \Theta^k_{p'} \right)^C.
$$

Since $\lambda_{\Theta^k} \left( \left( \Theta^k_{p'} \right)^C \right) = 0$ for all $p'$, every $\Delta$-section of $F^C_1$ has measure zero implying by Fubini’s Theorem that the product measure of $F^C_1$ is equal to zero. Next consider a $\Delta$-section $(F^C_2)_{p'}$ of $F^C_2$

$$
(F^C_2)_{p'} = \left\{ \theta \in \Theta^k : (p', \theta) \in F^C_2 \right\} = \left\{ \begin{array}{ll}
\Theta^k & \text{if } p' \in (\Delta^*)^C \\
\emptyset & \text{otherwise}
\end{array} \right\}.
$$

As $\lambda_{\Delta}(\Delta^*) = \lambda_{\Delta}(\Delta)$ almost every $\Delta$-section of $F^C_2$ has measure zero implying by Fubini’s Theorem that the product measure of $F^C_2$ is equal to zero. 

Proof of Lemma 5

*Proof.* Consider the following partition $P$ which is generated by the join of $\{C, C^C\}$ and the join of partitions of all agents,

$$
P = \{C, C^C\} \vee \bigvee_{i \in V} P_i.
$$

Let $m$ denote the cardinality of $P$. The partition $P$ can be divided into partitions $P_C$ and $P_{CC}$
where

\[
P_C = \left\{ P \cap C : P \in \bigvee_{i \in V} P_i \right\}
\]

\[
P_{CC} = \left\{ P \cap C^C : P \in \bigvee_{i \in V} P_i \right\}.
\]

Each element \( p \) of the unit simplex \( \Delta(m) \) in \( \mathbb{R}^m \) is a probability measure over the cells of partition \( P \).

For a given information set \( I \in \mathcal{I} \) let \( f_I(p) \) denote the posterior probability of event \( C \) conditioning on \( I \) for a given probability measure \( p \). We have \( f_I : \Delta(m) \to [0,1] \) and

\[
f_I(p) = \frac{\sum_{P \in P_C \text{ s.t. } P \subset I} p(P)}{\sum_{P \in P_C \text{ s.t. } P \subset I} p(P) + \sum_{P \in P_{CC} \text{ s.t. } P \subset I} p(P)}.
\]

For two different information sets \( I \neq I' \) consider the set of probability measures \( S_{II'0} \) that lead to the same conditional probability,

\[S_{II'} = \{ p \in \Delta(m) : f_I(p) = f_{I'}(p) \}.
\]

Equivalent transformations yield

\[S_{II'} = \left\{ p \in \Delta(m) : \sum_{P_C \subset I} p(P_C) \times \left( \sum_{P_{CC} \subset I'} p(P_{CC}) \right) = \sum_{P_C \subset I'} p(P_C) \times \left( \sum_{P_{CC} \subset I} p(P_{CC}) \right) \right\}
\]

where \( P_C \) denotes a typical element of \( P_C \) and \( P_{CC} \) denotes a typical element of \( P_{CC} \). Define the function \( f_{II'} : \Delta(m) \to \mathbb{R} \) as

\[
f_{II'}(p) = \sum_{P_C \subset I} p(P_C) \times \left( \sum_{P_{CC} \subset I'} p(P_{CC}) \right) - \sum_{P_C \subset I'} p(P_C) \times \left( \sum_{P_{CC} \subset I} p(P_{CC}) \right).
\]

Therefore \( f_{II'}(p) \) equals zero if and only if \( p \in S_{II'} \). It is easy to see that \( f_{II'} \) is a non-trivial polynomial function. Hence by Lemma 1 we have \( \lambda_{\Delta(m)}(f_{II'}^{-1}(0)) = 0 \) and therefore \( \lambda_{\Delta(m)}(S_{II}) = 0 \).

As the powerset of the join is finite, there are finitely many pairs \( I, I' \) involving subsets of the powerset of the join. For each of these finitely many pairs the corresponding set of probability measures that lead to the same posterior announcement has measure zero. Therefore, the union over all sets \( S_{II'} \) has probability zero as well which implies that \( \phi_p \) is injective for [\( \lambda_{\Delta(m)} \)] almost every probability measure \( p \).
Proof of Lemma 6

Proof. Let \( \Theta_g \) denote the zero set of \( g \),

\[ \Theta_g = \{ \theta \in \Theta : g(\theta) = 0 \} \]

Denote \( \prod_{i \in -B - \{j\}} \Theta_i \) by \( \Theta_{-j} \) and consider the product space \( \Theta_j \times \Theta_{-j} = \Theta \) with respective measures \( \lambda_{\Theta_j} \) and \( \lambda_{\Theta_{-1}} \). By Fubini’s Theorem the zero set \( \Theta_g \) has \( \lambda_{\Theta} \) measure zero if and only if almost every of its \( \Theta_{-j} \)-section has measure zero. \(^{35}\) For a fixed \( \theta^{-j} \in \Theta_{-1} \) consider the \( \Theta_{-1} \)-section of \( \Theta_g \). As \( g \) is a non-trivial polynomial on \( \Theta \) there exists \( j \) such that the function

\[ g(\cdot, \theta^{-j}) : \Delta(|N_j| + 1) \to \mathbb{R} \]

is a non-trivial polynomial function on \( \Delta(|N_j| + 1) \) for every \( \theta^{-j} \). By Lemma 1 the zero set of such a polynomial has \( \lambda_{\Delta(|N_j| + 1)} \) measure zero and hence every \( \Theta_{-j} \)-section has measure zero. Therefore, the zero set of \( g \) has measure zero in \( \Theta \).

Proof of Theorem 2

Proof. In order to conclude the proof of Theorem 2 we need to show that if for each \( p \in \Delta^* \) there exists a generic set \( \Theta^t_p \) such that

\[ q^t_i(\omega | (p, \theta)) = q^*_{i,t}(\omega | p) \]

holds for every \( i \in B, \omega \in \Omega \) and \( \theta \in \Theta^t_p \), then for each \( p \in \Delta^* \) there exists a generic set \( \Theta^t_{p+1} \) such that \( \mathcal{P}^*_{i,t+1}(\cdot | p) \) is coarser than \( \mathcal{P}^t_{i+1}(\cdot | (p, \theta)) \), formally that \( \mathcal{P}^t_{i+1}(\cdot | (p, \theta)) \leq \mathcal{P}^*_{i,t+1}(\cdot | p) \) holds for every \( i \in B \) and \( \theta \in \Theta^t_{p+1} \).

Consider a prior \( p \in \Delta^* \), a weight state \( \theta \in \Theta^t_p \) and fundamental states \( \omega, \omega' \) such that \( \omega' \in \mathcal{P}^t_{i}(\omega | (p, \theta)) \) and \( \omega' \notin \mathcal{P}^*_{i,t+1}(\omega | p) \).

**Step 1.** \( \omega' \in \mathcal{P}^t_{i}(\omega | (p, \theta)) \) implies \( q^t_i(\omega | (p, \theta)) = q^*_{i,t}(\omega' | (p, \theta)) \). The induction hypothesis implies

\[ q^t_i(\omega | (p, \theta)) = \frac{p \left( C \cap \mathcal{P}^{PL}_{i,t}(\omega) \right)}{p \left( \mathcal{P}^{PL}_{i,t}(\omega) \right)}. \]

\(^{35}\) For details see Appendix B.
By the definition of \( P_{i,t+1}^* \), for \( \omega' \notin P_{i,t+1}^*(\omega|p) \) we have

\[
\frac{p \left( C \cap P_{i,t}^{PL}(\omega) \right)}{p \left( P_{i,t}^{PL}(\omega) \right)} \neq \frac{p \left( C \cap P_{i,t+1}^{PL}(\omega') \right)}{p \left( P_{i,t+1}^{PL}(\omega') \right)}.
\]

\( \omega' \in P_i^t(\omega|p, \theta) \) then implies \( P_i^t(\omega|p, \theta) ) \in I \). \( P_{i,t}^{PL} \leq P_i^t \), \( p \in \Delta^* \) and the injectivity of \( \phi_p \) on \( I \) imply

\[
P_i^t(\omega|p, \theta) = P_{i,t}^{PL}(\omega).
\]

**Step 2.** For \( \omega' \in P_i^t(\omega|p, \theta) \) the following holds: \( \omega' \notin P_{i,t+1}^*(\omega|p, \theta) \) if and only if there exists a neighbor \( i_1 \) of \( i \) such that \( q_i^t(\omega|p, \theta) \neq q_i^t(\omega'|p, \theta) \).

**Step 3.** \( \omega' \notin P_{i,t+1}^*(\omega|p) \) implies \( P_{i,t+1}^{PL}(\omega) \neq P_{i,t+1}^{PL}(\omega') \). Step 1 and \( \omega' \in P_i^t(\omega|p, \theta) \) then imply that there exists an agent \( i_t \) with \( d_G(i,i_t) = t \) such that \( \omega' \notin P_{i,t}(\omega) \). Note that \( P_{i,t+1}^{PL} \) satisfies

\[
P_{i,t+1}^{PL} = P_{i,t}^{PL} \lor \bigvee_{l:d_G(i,l) = t} P_l.
\]

Suppose that

\[
q_i^t(\omega|p, \theta) = q_i^t(\omega'|p, \theta)
\]

for all agents \( i_t \) with \( d_G(i,i_t) = t \). By Lemma 5 and \( p \in \Delta^* \) it then follows that for each \( i_t \)

\[
P_{i,t}(\omega) \cup P_{i,t}(\omega') \subset C
\]

or

\[
P_{i,t}(\omega) \cup P_{i,t}(\omega') \subset C^C
\]

which implies

\[
\bigvee_{l:d_G(i,l) = t} P_l(\omega) \cup \bigvee_{l:d_G(i,l) = t} P_l(\omega') \subset C
\]

or

\[
\bigvee_{l:d_G(i,l) = t} P_l(\omega) \cup \bigvee_{l:d_G(i,l) = t} P_l(\omega') \subset C^C.
\]

But then we have \( P_{i,t+1}^{PL}(\omega) \cup P_{i,t+1}^{PL}(\omega') \subset C \) respectively \( P_{i,t+1}^{PL}(\omega) \cup P_{i,t+1}^{PL}(\omega') \subset C^C \) contradicting \( \omega' \notin P_{i,t+1}^*(\omega|p) \). Therefore \( \omega' \notin P_{i,t+1}^*(\omega|p) \) implies that there exists an agent \( i_t \) with \( d_G(i,i_t) = t \) such that \( \omega' \notin P_{i,t}(\omega) \) and \( q_i^t(\omega|p, \theta) \neq q_i^t(\omega'|p, \theta) \).

**Step 4.** Consider a neighbor \( i_1 \) of \( i \) that lies on a shortest path to \( i_t \), i.e. \( d_G(i_1,i_t) = t - 1 \). If we can show that \( q_i^t(\omega|p, \theta) \neq q_i^t(\omega'|p, \theta) \) holds, by step 2 we have \( \omega' \notin P_{i,t+1}^*(\omega|p, \theta) \). Note that for two partitions \( P, Q \) if \( \omega' \notin P(\omega) \) implies \( \omega' \notin Q(\omega) \) for all pairs \( \omega, \omega' \) such that
\( \omega' \notin \mathcal{P}(\omega) \) then \( Q \leq P \), i.e. \( P \) is coarser than \( Q \). Therefore \( \omega' \notin \mathcal{P}^{t+1}(\omega \mid (p, \theta)) \) for all \( \omega, \omega' \) such that \( \omega' \notin \mathcal{P}^{t+1}_{i,t+1}(\omega \mid p) \) implies

\[
\mathcal{P}^{t+1}_i(. \mid (p, \theta)) \leq \mathcal{P}^{t+1}_{i,t+1}(. \mid p)
\]

concluding the proof. Two cases have to be distinguished: \( i_1 \) being Bayesian and \( i_1 \) being non-Bayesian.

Suppose first that \( i_1 \) is Bayesian. Step 3 establishes \( \omega' \notin \mathcal{P}^{t+1}_{i_1}(\omega) \) which implies \( \mathcal{P}^{PL}_{i_1,t}(\omega) \neq \mathcal{P}^{PL}_{i_1,t}(\omega') \). The joint use of the induction hypothesis, the injectivity of \( \phi_p \) on \( \mathcal{I} \) and an argument similar to the one applied in step 3.) imply that

\[
q_{i_1}^t(\omega \mid (p, \theta)) \neq q_{i_1}^t(\omega' \mid (p, \theta)).
\]

Next consider the case where \( i_1 \) is non-Bayesian. We need to establish that

\[
q_{i_1}^t(\omega \mid (p, \theta)) \neq q_{i_1}^t(\omega' \mid (p, \theta))
\]

for \( i_t \) with \( d_G(i_1, i_t) = t - 1 \) implies \( q_{i_t}^t(\omega \mid (p, \theta)) \neq q_{i_t}^t(\omega' \mid (p, \theta)) \) for all \( \theta \) in a full measure set

\[
\Theta_{p}^{t+1} \left( (q_{Y_{i_1}}^t(\omega \mid p), q_{Y_{i_1}}^t(\omega' \mid p) \right).
\]

Note that common knowledge of \( \theta \) implies that (i) the set of possible Bayesian announcement vectors \( \left\{ q_{Y_{i_1}}^t(\omega \mid (p, \theta)) \right\}_{\omega \in \Omega} \) every non-Bayesian agent \( j \) averages over in period \( t \), and (ii) the corresponding weight functions \( \{ f_j^t(\theta) \}_{j \in -B} \), are commonly known among all Bayesian agents. The proof follows seven steps. Consider a prior \( p \in \Delta^* \) and weight state \( \theta \in \Theta_{p}^t \).

1. By the induction hypothesis we have for all \( \theta \in \Theta_{p}^t \)

\[
q_{Y_{i_1}}^t(\omega \mid (p, \theta)) = q_{Y_{i_1}}^t(\omega \mid p).
\]

2. Suppose that there exists a shortest path from \( i_1 \) to \( i_t \) such that all agents (but \( i_t \) on the path are non-Bayesian. \( q_{i_1}^t(\omega \mid (p, \theta)) \neq q_{i_1}^t(\omega' \mid (p, \theta)) \) then immediately implies

\[
q_{Y_{i_1}}^t(\omega \mid p) \neq q_{Y_{i_1}}^t(\omega' \mid p).
\]

Suppose on the other hand that there exists no such exclusively non-Bayesian shortest path from \( i_1 \) to \( i_t \). This implies that there exists a Bayesian agent \( i_k \) on the path between \( i_1 \) and \( i_t \).

39
Let $d_{kl}$ denote the distance from $i_k$ to $i_l$. The new coming announcement of $i_k$ in $\mathbf{q}_{Y_{i_1}^*}$ satisfies

$$\left[ q^*_{Y_{i_1}^*} (\omega | p) \right]_{i_k} = \frac{p \left( C \cap \mathcal{P}_{i_k, d_{kl+1}}^{PL} (\omega) \right)}{p \left( \mathcal{P}_{i_k, d_{kl+1}}^{PL} (\omega) \right)}.$$ 

$q^1_i (\omega | (p, \theta)) \neq q^1_i (\omega' | (p, \theta))$ implies $\omega' \notin \mathcal{P}_{i_k, d_{kl+1}}^{PL} (\omega)$. Suppose

$$\left[ q^*_{Y_{i_1}^*} (\omega | p) \right]_{i_k} = \left[ q^*_{Y_{i_1}^*} (\omega' | p) \right]_{i_k}.$$ 

The injectivity of $\phi_p$ on $\mathcal{I}$ for $p \in \Delta^*$ then implies

$$\mathcal{P}_{i_k, d_{kl+1}}^{PL} (\omega) \cup \mathcal{P}_{i_k, d_{kl+1}}^{PL} (\omega') \subset C$$

or

$$\mathcal{P}_{i_k, d_{kl+1}}^{PL} (\omega) \cup \mathcal{P}_{i_k, d_{kl+1}}^{PL} (\omega') \subset C^C.$$ 

But $\mathcal{P}_{i_1, t+1} \leq \mathcal{P}_{i_k, d_{kl+1}}^{PL}$ then implies $q^*_{Y_{i_1}^*} (\omega | p) = q^*_{Y_{i_1}^*} (\omega' | p)$ contradicting $\omega' \notin \mathcal{P}_{i_1, t+1} (\omega | p)$. Therefore we have

$$q^*_{Y_{i_1}^*} (\omega | p) \neq q^*_{Y_{i_1}^*} (\omega' | p).$$

3. The announcement of agent $i_1$ in period $t$ is given by

$$q^t_{i_1} (\omega | (p, \theta)) = \sum_{l=1}^{\left| Y_{i_1}^* \right|} \left[ f^t_{i_1} (\theta) \right]_l \times \left[ q^*_{Y_{i_1}^*} (\omega | p) \right]_l.$$ 

4. Let $\Theta_p^{t+1} \left( q^*_{Y_{i_1}^*} (\omega | p), q^*_{Y_{i_1}^*} (\omega' | p) \right)$ denote the set of weight states such that

$$\sum_{l=1}^{\left| Y_{i_1}^* \right|} \left[ f^t_{i_1} (\theta) \right]_l \times \left[ q^*_{Y_{i_1}^*} (\omega | p) \right]_l \neq \sum_{l=1}^{\left| Y_{i_1}^* \right|} \left[ f^t_{i_1} (\theta) \right]_l \times \left[ q^*_{Y_{i_1}^*} (\omega' | p) \right]_l.$$ 

holds. Consider the following function $g : \Theta \to \mathbb{R},$

$$g(\theta) = \sum_{l=1}^{\left| Y_{i_1}^* \right|} \left[ f^t_{i_1} (\theta) \right]_l \times \left( \left[ q^*_{Y_{i_1}^*} (\omega | p) \right]_l - \left[ q^*_{Y_{i_1}^*} (\omega' | p) \right]_l \right).$$ 

Note that $\Theta_p^{t+1} \left( q^*_{Y_{i_1}^*} (\omega | p), q^*_{Y_{i_1}^*} (\omega' | p) \right)$ equals the complement of the zero set of the function $g.36$ As $g$ is a non-trivial polynomial function (since $q^*_{Y_{i_1}^*} (\omega | p) \neq q^*_{Y_{i_1}^*} (\omega' | p)$) its zero

---

36 The zero set of a function $g : X \to \mathbb{R}$ is equal to $\{ x \in X : g(x) = 0 \}$. 

40
set has measure zero in $\Theta$ by Lemma 6. Therefore $\Theta^{t+1}_p \left( q^{*}_{Y^t_j} (\omega | p), q^{*}_{Y^t_j} (\omega' | p) \right)$ is generic.

5. Consider the following set

$$\Theta^{t+1}_p = \Theta^t_p \cap \bigcap_{j \in \Lambda^t_p, \Lambda^t_p \neq \emptyset} \bigcap_{\omega \in \Omega, \omega' \notin P^t_j} (q^{*}_{Y^t_j} (\omega, q^{*}_{Y^t_j} (\omega'))).$$

As the join of partitions of all agents is finite, there exist finitely many distinct pairs $q^{*}_{Y^t_j} (\omega | p), q^{*}_{Y^t_j} (\omega' | p)$. By step 5 each of the sets $\Theta^{t+1}_p (q^{*}_{Y^t_j} (\omega | p), q^{*}_{Y^t_j} (\omega' | p))$ is generic. By the induction hypothesis $\Theta^t_p$ is generic. As the intersection of finitely many generic sets is generic, $\Theta^{t+1}_p$ is a generic set.\]

**Proof of Theorem 4**

*Proof.* Consider the following function $g^{t,j}_{(\omega, \omega'), (\alpha, \beta)} : \Theta^N \to \mathbb{R}$ where

$$g^{t,j}_{(\omega, \omega'), (\alpha, \beta)} (\Theta) = f^t_j (\Theta_\alpha) \times q^{*}_{Y^t_j} (\omega | p) - f^t_j (\Theta_\beta) \times q^{*}_{Y^t_j} (\omega' | p).$$

Note that for $\bar{\Theta} \in \Theta^N_p$ the function $g^{t,j}_{(\omega, \omega'), (\alpha, \beta)} (\bar{\Theta})$ satisfies

$$g^{t,j}_{(\omega, \omega'), (\alpha, \beta)} (\bar{\Theta}) = q^{*}_{Y^t_j} (\omega, \Theta_\alpha \mid (p, \bar{\Theta})) - q^{*}_{Y^t_j} (\omega', \Theta_\beta \mid (p, \bar{\Theta}))$$

which implies that for $g^{t,j}_{(\omega, \omega'), (\alpha, \beta)} (\bar{\Theta}) \neq 0$ agent $i$ can distinguish $(\omega, \Theta_\alpha)$ from $(\omega', \Theta_\beta)$ through the announcement of $j$ in period $t$. Consider the set of functions

$$S^{t,j}_p = \left\{ g^{t,j}_{(\omega, \omega'), (\alpha, \beta)} \right\}_{(\omega, \omega'), (\alpha, \beta) \in \Omega^2 \times \{1, \ldots, N\}^2}$$

such that $g^{t,j}_{(\omega, \omega'), (\alpha, \beta)} (\bar{\Theta}) \neq 0$ for some $\bar{\Theta} \in \Theta$. Note that $S^{t,j}_p$ is countable as the join of fundamental partitions is finite and $\bar{\Theta}$ is countable. Further note that for $g \in S^{t,j}_p$ we have either (i) $\omega' \notin P_{j,t} (\omega)$, (ii) $\alpha \neq \beta$ such that

$$\Theta_\alpha \notin \bigvee_{k \text{ s.t. } d_{C_j, i, k} \leq t-2} P^\Theta_k (\Theta_\beta),$$

or both. If we can show that for each $g \in S^{t,j}_p$ its zero set (preimage of 0) has measure zero in $\Theta^N$, then we can construct a generic set $\Theta^N_{p,t+1} (j)$ satisfying that the announcement of $j$ corresponding to $(\omega', \Theta') \notin G_{j,t} (\omega, \Theta)$ differs from the announcement corresponding to $(\omega, \Theta)$ for all $\Theta \in \Theta^N_{p,t+1} (j)$.
which concludes the proof of the theorem. The sets $\Theta^N_{p,t+1}$, $\Theta^N_{p,t+1}(j)$ equal

$$
\Theta^N_{p,t+1}(j) = \Theta^N_{p,t} \cap \bigcap_{g \in S^j_p} \left( \Theta^N \setminus g^{-1}(0) \right).
$$

$$
\Theta^N_{p,t+1} = \bigcap_{j \in \sim B : I \cap \{j\} \neq \emptyset \text{ for some } i \in B} \Theta^N_{p,t+1}(j)
$$

I conclude the proof by establishing that for every $g \in S^{t,j}_p$ the following holds

$$
\lambda_{\Theta^N}(g^{-1}(0)) = 0.
$$

Consider $g \in S^{t,j}_p$ for $p \in \Delta^*$. The proof that the preimage of 0 under $g$ has measure zero in $\Theta^N$ is established in two steps.

1. Consider a function $g \in S^{t,j}_p$ and its corresponding positions $\alpha$ and $\beta$. Note that for any pair $\bar{\Theta}, \bar{\Theta}' \in \Theta^N$ such that

$$
\bar{\Theta}_i = \bar{\Theta}'_i \text{ for } i = \alpha, \beta
$$

implies

$$
g(\bar{\Theta}) = g(\bar{\Theta}').
$$

Therefore there exists a function $\bar{g} : \Theta^2 \rightarrow \mathbb{R}$ that is derived by projecting the domain on positions $\alpha$ and $\beta$,

$$
\bar{g}(\bar{\Theta}_\alpha, \bar{\Theta}_\beta) = g(\bar{\Theta}).
$$

We have

$$
g^{-1}(0) = \bar{g}^{-1}(0) \times \Theta^N_{-\alpha, \beta}
$$

This implies that it is sufficient to show that the zero set of $\bar{g}$ has measure zero in $\Theta^2$.

2. Again we employ a technique based upon Fubini’s Theorem. For a given $\bar{g} \in S^{t,j}_p$ with corresponding states $\omega, \omega'$ and positions $\alpha, \beta$ consider its zero set

$$
\Theta^2_{\bar{g}} = \left\{ (\theta^1, \theta^2) \in \Theta^2 : \bar{g}(\theta^1, \theta^2) = 0 \right\}.
$$

By Fubini’s Theorem $\Theta^2_{\bar{g}}$ has $\lambda_{\Theta^2}$ measure zero in $\Theta^2$ if and only if almost every $\Theta^2$-section has measure zero in $\Theta^2_1$.\footnote{For details see Appendix B. Also note that there is a slight abuse of notation as denote $\Theta_i$ denotes a weight state space as opposed to the set of feasible updating functions of agent $i$. Similarly $\theta^i$ denotes a weight state as opposed to an updating function of agent $i$.}

For a fixed $\theta^2 \in \Theta_2$ consider the $\Theta_2$-section of $\Theta^2_{\bar{g}}$,

$$
(\Theta^2_{\bar{g}})_{\theta^2} = \left\{ \theta^1 \in \Theta_1 : \bar{g}(\theta^1, \theta^2) = 0 \right\}.
$$

37 For details see Appendix B. Also note that there is a slight abuse of notation as denote $\Theta_i$ denotes a weight state space as opposed to the set of feasible updating functions of agent $i$. Similarly $\theta^i$ denotes a weight state as opposed to an updating function of agent $i$.}
Note that the function $\tilde{g}(\cdot, \theta^2) : \Theta \to \mathbb{R}$ is a non-trivial polynomial for every $\theta^2 \in \Theta_2$. Hence by Lemma 6 the $\lambda_{\Theta}$ measure of its zero set is equal to zero. Therefore, by Fubini’s Theorem the $[\lambda_{\Theta^2}]$ measure of the zero set of $\tilde{g}$ is equal to zero which by step 1 implies that the $[\lambda_{\Theta^N}]$ measure of the zero set of $g$ is equal to zero.

Appendix B

The Space of Common Priors and the $\sigma$-Algebra $\mathcal{F}$

Let the partitions of all agents be finite and consider the following partition of the state space

$$
\mathcal{P} = \left\{ P \cap C : P \in \bigvee_{i \in V} \mathcal{P}_i \right\} \cup \left\{ P \cap C^C : P \in \bigvee_{i \in V} \mathcal{P}_i \right\}
$$

where $C^C$ denotes the complement of $C$. The partition $\mathcal{P}$ further refines the join $\bigvee_{i \in V} \mathcal{P}_i$ of partitions of all agents, by partitioning each cell of the join into states that lie in event $C$ and states that do not. The $\sigma$-algebra $\mathcal{F}$ is generated by $\mathcal{P}$. Let $m$ denote the cardinality of the partition $\mathcal{P}$. The relative interior of the unit simplex $\Delta(m)$ in $\mathbb{R}^m$, $\text{ri}\Delta(m)$, endowed with a uniform probability measure $\lambda_\Delta$, describes the set of all probability measures to be considered. Abusing notation the relative interior of the unit simplex is denoted as $\Delta$ and equals the set of all possible priors on the fundamental state space.

Freidlin and Wentzell Theorem

Consider a transition matrix $T_\theta$ and its corresponding stationary distribution $\mu(\theta)$. An $i$-tree, $S_i = (V, E_{S_i})$ of $T_\theta$ is a directed subgraph of $G$ with $v - 1$ edges such that from every node $j$ different from $i$ there exists a unique directed path from $j$ to $i$. Let $S_i$ denote the set of $i$-trees. Let the likelihood $P$ of an $i$-tree be defined as the product of weights $\theta^j_i$, where $j$ is $i$’s neighbor in $S_i$,

$$
P(S_i) = \prod_{ij \in E_{S_i}} [T_\theta]_{ij}.
$$

A result by Freidlin and Wentzell establishes the following relation between the stationary distribution and the $i$-trees of a transition matrix $T_\theta$. 

43
Theorem 5 (Freidlin and Wentzell) The stationary distribution \( \mu(\theta) \) has the property that the weight of each agent \( i \) is proportional to the sum of likelihoods of its \( i \)-trees,

\[
\mu_i(\theta) = \frac{g_\theta(i)}{\sum_{j \in V} g_\theta(j)},
\]

where \( g_\theta(i) = \sum_{S_i \in S_i} P(S_i) \).

Fubini’s Theorem

The following is taken from chapter 7 in Halmos [20]. Consider the measure spaces \( (X, \mathcal{F}_X, \lambda_X) \), \( (Y, \mathcal{F}_Y, \lambda_Y) \) and the corresponding product measure space \( (X \times Y, \mathcal{F}_{X \times Y}, \lambda_{X \times Y}) \). For \( E \subset X \times Y \) and \( x \in X \) denote the set \( E_x \)

\[
E_x = \{ y \in Y : (x, y) \in E \}
\]
as an \( X \)-section of \( E \).

Theorem 6 (Fubini) \( E \in \mathcal{F}_{X \times Y} \) has measure zero if and only if almost every \( X \)-section of \( E \) has measure zero in \( Y \).

Example: Non-Bayesian Communication

There are two agents, 1 and 2, that are mutual neighbors of each other. Let the set \( \Omega \) of all states of the world be represented by the rectangle in figure 2 and suppose that the agents share a uniform common prior probability distribution over the rectangle.

![Figure 2: State space \( \Omega \)](image-url)
The initial private information of agents is represented via partitions of the state space. Agent 1 learns whether the true state of the world lies in the first row or in the second row,

\[ \mathcal{P}_1 = \{J_1 \cup J_2; J_3 \cup J_4\}. \]

Agent 2 learns whether the true state of the world lies in the first column or in the second column,

\[ \mathcal{P}_2 = \{J_1 \cup J_3; J_2 \cup J_4\}. \]

Suppose that event \( C \) represented by the dark area is of interest to both agents, i.e. the agents communicate their posterior belief of event \( C \) in the first round of communication.

In the first round agent 1 announces a posterior probability of \( \frac{1}{2} \) if he observes the first row or a posterior of \( \frac{1}{4} \) if he observes the second row. Agent 2 announces a first round posterior of \( \frac{1}{2} \) if he observes the first column and a posterior of \( \frac{1}{4} \) otherwise. Hence both agents perfectly reveal their private information through their first round announcement.

As shown in section 3, the limit belief of both agents equals a weighted average of the first period announcement with the weights stemming from the stationary distribution \( \mu(\theta) \) of the transition matrix \( T_\theta \) derived from the weight state. Depending on the realized state the following holds for the limit consensus belief of the non-Bayesian agents,

\[ q^\infty_1(\omega | \theta) = q^\infty_2(\omega | \theta) = \begin{cases} \frac{1}{2} & \text{if } \omega \in J_1 \\ \frac{1}{2} [\mu(\theta)]_1 + \frac{1}{4} (1 - [\mu(\theta)]_1) & \text{if } \omega \in J_2 \\ \frac{1}{2} [\mu(\theta)]_1 + \frac{1}{4} (1 - [\mu(\theta)]_1) & \text{if } \omega \in J_3 \\ \frac{1}{4} & \text{if } \omega \in J_4 \end{cases}. \]

Depending on the realized weight state, the limit partitions \( \mathcal{P}^\infty_i(\cdot | \theta) \) are given by

\[ \mathcal{P}^\infty_i(\cdot | \theta) = \begin{cases} \{J_1; J_2 \cup J_3; J_4\} & \text{if } \theta^1 = \theta^2 = \theta^3 = \theta^4 = \frac{1}{2} \\ \{J_1; J_2; J_3; J_4\} & \text{otherwise} \end{cases}. \]

Bayesian equivalence requires that, in every state \( \omega \), the limit belief \( q^\infty_i(\omega | \theta) \) equals the posterior probability of event \( C \) conditioning on the realized cell of \( \mathcal{P}^\infty_i(\cdot | \theta) \). For every \( \theta \) we have

\[
p(\mathcal{P}^\infty_i(\omega | \theta) \cap C) \quad \frac{p(\mathcal{P}^\infty_i(\omega | \theta) \cap C)}{p(\mathcal{P}^\infty_i(\omega | \theta))} = \begin{cases} 1 & \text{if } \omega \in J_1 \\ 0 & \text{if } \omega \in J_2 \cup J_3 \\ \frac{1}{2} & \text{if } \omega \in J_4 \end{cases}.
\]

Since for every \( \omega \) and \( \theta \) the limit belief is bounded away from the Bayesian posterior conditioning on \( \mathcal{P}^\infty_i(\cdot | \theta) \) by at least \( \frac{1}{4} \), Bayesian equivalence fails for every non-Bayesian communication structure.
Communication and Choice

Suppose now that in each round of interaction individuals do not only communicate their opinion but choose an action as well. Let $A_t^i$ denote the set of actions of agent $i$ in period $t$. I assume that the set of actions does not vary with time nor across agents, i.e. $A_t^i = A$ for all $i \in V$ and $t \in \mathbb{N}$. All agents share a common stage utility function $u : A \times \Omega \to \mathbb{R}$, implying that there are no payoff externalities. The uncertain event $C$, whose likelihood is the object of communication, plays a crucial role in the utility function. In particular, I assume that under any action $a$, for utilities it only matters whether a state $\omega$ lies in $C$ or in its complement $C^C$. For all $a \in A$ we have

$$u(a, \omega) = \begin{cases} u_a^C & \text{if } \omega \in C \\ u_a^{C^C} & \text{if } \omega \in C^C \end{cases}.$$ 

In every period $t$ each agent $i$ (i) announces his (subjective) posterior probability of event $C$, $q^i_t(\omega | (p, \theta))$, and (ii) selects the action that maximizes his expected stage utility based on $q^i_t(\omega | (p, \theta))$. As the action choices of agents are a coarse representation of their posterior beliefs, the sequence of opinions and actions is independent of the actions being private or commonly observed. If the set of actions is finite, then Theorem 2 implies the following corollary result.

**Corollary 1** If $A$ is finite, then the following result holds for almost all $(p, \theta) \in \Delta \times \Theta$: There exists a finite time $t_\theta$ such that for all $t \geq t_\theta$ and $\omega \in \Omega$ the action $a^i_j$ chosen by any non-Bayesian agent $j$ maximizes the expected utility of all Bayesian agents conditioning on their Bayesian posterior.

**Proof.** By Theorem 2, all agents reach consensus asymptotically for generic priors and weight states. As the partitions of all agents are finite, for a fixed prior $p$, there exists a finite number of such consensus opinions. For each action $a$ the corresponding expected utility is a function $f_a : [0, 1] \to \mathbb{R}$ linear in the probability $q$ of event $C$,

$$f_a(q) = q u_a^C + (1 - q) u_a^{C^C}.$$ 

For a given probability $q^*$ consider the set of expected utility maximizing actions $A_{q^*}$ with typical element $a_{q^*}$. I establish that there exists a positive $\delta_{q^*} \in \mathbb{R}$ such that for a $q'$ within $\delta_{q^*}$ distance of $q^*$ the expected utility maximizing action is an element of $A_{q^*}$.

Consider the second best action given probability $q^*$,

$$a^2_{q^*} = \min_{a \in A \setminus A_{q^*}^*} f_a(q^*) - f_a(q^*).$$
Denote by $\epsilon_{q^*}$ the difference in expected utility between the best and second best action,

$$\epsilon_{q^*} = f_{a_{q^*}}(q^*) - f_{a_{q^2}}(q^*).$$

As $A$ is finite we have $\epsilon_{q^*} > 0$. Continuity of all functions $f_a$ implies that for $\frac{\epsilon_{q^*}}{2}$ there exists a $\partial_{q^*}$ such that

$$|q^* - q| < \partial_{q^*} \Rightarrow |f_a(q^*) - f_a(q)| < \frac{\epsilon_{q^*}}{2},$$

which then implies that for all $q$ within $\min_{a \in A} \partial_{q^*}$, $\min_{a \in A} \partial_{q^*} > 0$ since $A$ is finite, distance of $q^*$ the expected utility maximizing action lies in $A_{q^*}$. As there are finitely many consensus beliefs $q^*$, finitely many non-Bayesian agents and the beliefs of all non-Bayesian agents converge to $q^*(\omega | (p, \theta))$ in state $\omega$, there exists a $t_\theta$ such that for all $t \geq t_\theta$ we have

$$|q_j^t(\omega | (p, \theta)) - q^*(\omega | (p, \theta))| < \min_{q^*} \left( \min_{a \in A} \partial_{q^*} \right).$$

Hence all actions chosen by non-Bayesian agents from $t_\theta$ onward lie in the set of actions that maximize the expected utility of the Bayesian agents.

The corollary states that generically Bayesian and non-Bayesian agents become equivalent in finite time from an expected utility point of view. Differences in sophistication become irrelevant within a finite time frame as long as the underlying social network is strongly connected. Despite the fact that Bayesian agents might hold more accurate beliefs for infinitely many rounds this translates into an advantage from an expected utility point of view for only finitely many periods. Within a finite horizon all agents either select the same action, or if different actions are chosen infinitely often, then all actions that are chosen past a finite time threshold are expected utility maximizing for the Bayesian agents conditioning on their consensus belief.

References


