UNU-MERIT Working Paper Series

#2014-034

Beliefs dynamics in communication networks

Théophile T. Azomahou and Daniel C. Opolot

Maastricht Economic and social Research institute on Innovation and Technology (UNU-MERIT)
email: info@merit.unu.edu | website: http://www.merit.unu.edu

Maastricht Graduate School of Governance (MGSoG)
email: info-governance@maastrichtuniversity.nl | website: http://mgsog.merit.unu.edu

Keizer Karelplein 19, 6211 TC Maastricht, The Netherlands
Tel: (31) (43) 388 4400, Fax: (31) (43) 388 4499

UNU-MERIT Working Papers

ISSN 1871-9872

Maastricht Economic and social Research Institute on Innovation and Technology, UNU-MERIT

Maastricht Graduate School of Governance MGSoG

UNU-MERIT Working Papers intend to disseminate preliminary results of research carried out at UNU-MERIT and MGSoG to stimulate discussion on the issues raised.

Beliefs dynamics in communication networks

Théophile T. Azomahou

Maastricht University—UNU-MERIT, 6211 TC Maastricht, The Netherlands

Daniel C. Opolot*

Maastricht University—UNU-MERIT, 6211 TC Maastricht, The Netherlands

This Version—March 2014

Abstract: We study the dynamics of individual beliefs and information aggregation when agents communicate via a social network. We provide a general framework of social learning that captures the interactive effects of three main factors on the structure of individual beliefs resulting from such a dynamic process; that is historical factors—*prior beliefs*, learning mechanisms—*rational and bounded-rational learning*, and the topology of communication structure governing information exchange. More specifically, we provide conditions under which heterogeneity and consensus prevail. We then establish conditions on the structures of the communication network, prior beliefs and private information for public beliefs to correctly aggregate decentralized information. The speed of learning is also established, but most importantly, its implications on efficient information aggregation.

Keywords: Learning, social networks, public beliefs, speed of learning, information aggregation.

JEL classification: C70, D83, D85.

1. Introduction

Individual beliefs play a significant role in determining public opinions and decisions made under uncertainty, both of which in turn shape social welfare. For example the level of heterogeneity in beliefs about government policies such as public health and social integration initiatives affects their implementation. Decision making under uncertainty is a ubiquitous problem in economics and social settings. Examples include consumer decisions on a brand choice; adoption of agricultural products and information technologies; investment as well as legislative decisions among others. Our goal in this paper is to provide a comprehensive yet fundamental framework for characterizing the evolution of individual beliefs through social learning, and to establish conditions under which the resulting public beliefs correctly and efficiently aggregate decentralized information.

There are mainly three factors that influence the evolution of individual beliefs through social learning; historical factors—*prior beliefs*, the learning mechanism (the manner in which individuals incorporate new information into their beliefs)—-*rational or bounded-rational learning*, and the

*Corresponding author: Opolot D. C., Maastricht University—UNU-MERIT, Keizer Karelplein 19, 6211 TC Maastricht, The Netherlands, (opolot@merit.unu.edu), Tel. +31 433884440, Fax +31 433884499.

topology of communication structure governing information exchange. To this end we construct a comprehensive theoretical model that is fundamental enough to capture the main properties and general enough to enable us compare and contrast the properties of beliefs resulting from rational and bounded-rational learning. There exists a true state of nature unknown to all agents about which they update their beliefs. Agents possess heterogeneous prior beliefs that are not necessarily correlated to the true state of nature. Each agent receives private information that is informative about the true state of nature, and in turn updates their prior belief. The resulting *private beliefs* are then simultaneously communicated (or simply announced) to the immediate neighbors.[1] After observing the neighbors' announcements, each agent incorporates the communicated beliefs into their private beliefs either by deducing the associated private signal in the case of rational learning, or by taking the weighted average of the announcements in the case of bounded-rational learning. The communication and learning process continues until none of the agents has new information to learn from the announcements of his neighbors, in which case their beliefs become "local public information". To differentiate private beliefs from beliefs that result at the end of the learning process, we refer to the later as *public beliefs* since they become public information at least to the first-order neighbors.

Unlike previous studies that focused on each of the three factors separately (e.g. Geanakoplos and Polemarchakis (1982), Parikh and Krasucki (1990), Ellison and Fudenberg (1995), Gale and Kariv (2003), Rosenberg et al. (2009), Demarzo et al. (2003) and Golub and Jackson (2010)), the generalized framework we provide directly establishes an explicit characterization of how the three factors interactively shape individual beliefs. We also provide a stylized model for an *exit game* in which economic agents partially rely on their level of confidence in their beliefs in deciding whether or not to take an action (e.g. investing in a given project) or wait to collect more information. The waiting process is however costly, such that each agent has to choose the optimal waiting time to take an action. The purpose of this stylized model is to provide a characterization of the effect of the learning mechanisms and the topology of the communication structure on the efficiency of information aggregation. We start by making generalizations of some of the existing results in the literature. More specifically, we establish the structure of public beliefs under rational learning when agents are either certain or uncertain about others' prior beliefs. We then extend existing results on bounded-rational learning to dynamic communication networks. The main contribution of the paper is the theorems providing the conditions on prior beliefs, private information and communication network structures for public beliefs to correctly and efficiently aggregate decentralized information. Public beliefs are said to be *correct* or to correctly aggregate private information if they fully incorporate private information of all agents, and *asymptotically correct* if in the limit of the population size, they converge in probability to the true state of nature.

The main results that emerge from this paper are the following. Proposition 1 provides the expressions for the structure of public beliefs under rational learning for finite population. When the communication network is common knowledge and connected, then a consensus in public beliefs arises only if the prior beliefs are identical and observable to the neighbors. It is not necessary for prior beliefs to be common knowledge among all agents for a consensus to emerge provided they

[1] A private belief is what results after an agent incorporates his private information into his prior belief.

are observable to the neighbors. Proposition 1 also shows that when the communication network is complete, a consensus in public beliefs obtains under uncertainty of prior beliefs provided the realized prior beliefs are identical and correlated. Heterogeneity in public beliefs arises under two cases. The first case is when the realized prior beliefs of neighbors are observable but heterogeneous, and the second case is when the realized prior beliefs are identical but not observable to the neighbors.

In Proposition 2, we generalize the bounded-rational learning models to dynamic communication networks. We show that provided the *switching strategy* is such that there exists a positive probability of realizing a connected network, then a consensus will obtain in a long-run.[2] Public beliefs will be heterogeneous otherwise.

In Theorems 1 and 2, we establish conditions for public beliefs to be asymptotically correct. That is conditions under which public beliefs converge in probability for a large population size to the true state of nature. Under rational learning, public beliefs will be asymptotically correct even when agents are uncertain of others' prior beliefs provided the prior beliefs and signals are independently distributed and of finite spaces. The topology of the communication network does not affect correctness of asymptotic public beliefs provided that it is common knowledge and connected. Contrary to rational learning, under bounded-rational learning, the topology of the communication network plays a significant role in determining correctness of public beliefs. We show that the network must be perfectly balanced and asymptotically balanced for public beliefs and asymptotic public beliefs respectively to be correct. We then characterize classes of networks that satisfy perfect and asymptotic balancedness conditions. In addition to restrictions on the network topology, we also find that under bounded-rational learning, public beliefs are asymptotically correct if and only if the prior beliefs are correlated to the true state of nature. Specifically, agents prior beliefs must be normally distributed with mean equal to the true state of nature and with finite variance.

In section 6 we compare the efficiency of private information aggregation by public beliefs under rational and bounded-rational learning. We show that the *price of rationality*, which we define as the ratio of expected social welfare under rational learning to the expected social welfare under bounded-rational learning, is an inverse function of the population size. Implying that the larger the population, the higher is the relative benefit of having an economy made up of rational agents. We show that the speed at which private information is aggregated is faster under rational learning than under bounded-rational learning.

2. Contribution to the literature

This paper contributes to the existing literature in several ways. First, it is closely related to the literature on knowledge and consensus (e.g Geanakoplos and Polemarchakis (1982), Parikh and Krasucki (1990) and Krasucki (1996)). Geanakoplos and Polemarchakis (1982) showed that repeated communication of posterior beliefs between two agents who start with a common prior will eventually lead to a consensus in their public beliefs. Parikh and Krasucki (1990) generalize the framework of Geanakoplos and Polemarchakis (1982) to the case in which agents communicate a general family of functions that map information sets to messages, of which posterior beliefs are

[2] A switching strategy is a function that maps an agent's current position in the network (or simply the current neighborhood) to another position in the next period.

a special case. They then establish conditions on such functions under which a consensus in public beliefs obtains. Krasucki (1996) extends the framework of Parikh and Krasucki (1990) to multiple agents who communicate sequentially through a protocol that determines the sender and receiver. They show that if the communication protocol does not consist of *cycles*, then a consensus in public beliefs obtains.[3] In all the above models, communication is sequential. That is it is defined by a protocol that selects a pair of agents (a sender and a receiver) and in each period only one of them is active. This is contrary to the case in this paper in which agents act simultaneously. The second general difference is that in the above papers, private information is represented by an information set which is defined by a partition of the state space. Such representation of private information typically leads to multidimensional information structure, which implies that communication can take multiple rounds even in the case of two agents before a consensus in beliefs is finally reached. On the contrary we adopt a simpler information structure that allows us to focus on the main questions concerning learning in general networks and properties of public beliefs. Indeed, as in Geanakoplos and Polemarchakis (1982) and Parikh and Krasucki (1990) we find that when the communication network is complete or rather when posterior beliefs announcements are public, a consensus emerges provided prior beliefs are common knowledge and identical. But as an extension to this result, we show that it is not necessary for agents to be certain of others' prior beliefs for a consensus to obtain. All that matters is that prior beliefs are correlated and that the realizations are identical. When the communication network is not complete, we find as in Krasucki (1996) that the network must be connected if a consensus is to obtain. We then characterize conditions for heterogeneity and correctness in public beliefs to obtain.

Secondly, this paper is related to the literature on Bayesian learning with rational agents in social networks (e.g. Gale and Kariv (2003), Rosenberg et al. (2009) and Mueller-Frank (2013)). Just as we do in this paper, these papers also consider simultaneous communication among agents but with the difference that agents communicate their actions rather than the posterior beliefs.[4] These papers also consider a multidimensional information structure as in the case of the models on knowledge and consensus above. The primary focus of this literature is on the uniformity and local indifference in the actions chosen by the agents at the end of the learning process, which is contrary to our focus on public beliefs. Nevertheless, some of our findings extend directly to situations where agents communicate actions rather than posterior beliefs.[5] Rosenberg et al. (2009) and Mueller-Frank (2013) show that under the assumptions of connected network, common prior, common knowledge of strategies and network topology, heterogeneity in agents' actions at the end of learning process arises from. Here, we show that heterogeneity in public beliefs hence actions

[3] A cycle is a closed path, where a path from agent i to agent j is a connected set of links starting from i and ending in j.

[4] There also exists a literature on sequential Bayesian learning in which agents make a decision once in a lifetime in an exogenously predefined order. When it is an agent's turn to act, he observes the history of actions of all agents that acted before him. The primary concern of this literature is establishing conditions under which informational cascades and herds behavior occurs. The main contributions are Banerjee (1992), Bikhchandani et al. (1992), Smith and Sorensen (2000) and Acemoglu et al. (2011).

[5] This is particularly true because communicating posterior beliefs is equivalent to communicating actions when the action space is rich (for example the continuous action space) and/or when strategies are common knowledge as in the case partly considered by Mueller-Frank (2013).

(see section 6.2) can also arise from heterogeneity and uncertainty of prior beliefs.

Thirdly, this paper is related to the literature on agreement and disagreement under Bayesian rational learning (e.g Dixit and Weibull (2007), Cripps et al. (2008), Acemoglu et al. (2009) and Sethi and Yildiz (2012)). Cripps et al. (2008) and Acemoglu et al. (2009) study the validity of observational learning as a justification for the common prior assumption generally employed in most economics models. That is, individuals that share their experiences with each other will eventually have a shared history of events that are informative about the state of nature, and this will in turn lead to an agreement on their beliefs about the true state of nature. The framework employed by Cripps et al. (2008) and Acemoglu et al. (2009) to model this claim consists of two agents who observe a sequence of either private or public signals that are necessarily informative about the true state of nature. Cripps et al. (2008) show that when both agents observe a sequence of correlated private signals then their beliefs converge to a common public belief provided that the signal space is finite. Acemoglu et al. (2009) show that if agents start with heterogeneous prior beliefs, then observe a sequence of public signals and that they are uncertain about signal interpretation, then they do not necessarily converge to a common public belief. In our framework, rather than agents having to observe and learn from exogenously "communicated" signals, they in stead learn endogenously through direct communication with their immediate neighbors in the network. Our framework for rational learning is thus similar to that of Cripps et al. (2008) and Acemoglu et al. (2009), whereby an infinite sequence of signals corresponds to an infinite set of agents each of whose private information is realized independently of the others. There is a difference in terms of the "time factor" though. In our framework, the number of signals an agent receives (by observing the neighbors announcements) at any given period t depends on the number of t-order neighbors. As a consequence the speed of learning is generally faster (depending on the network topology) in our framework. Indeed, as in Cripps et al. (2008) we find that provided that the distribution from which the signals are drawn has bounded variance, a consensus obtains in the limit of the number of agents. In relation to Acemoglu et al. (2009), the uncertainty on how to interpret the signal is related to the case in which agents do not observe the prior beliefs of their neighbors. The uncertainty of neighbors' priors leads to uncertainty on how to interpret their announcements, which in turn leads to uncertainty on the signal interpretation. Contrary to Acemoglu et al. (2009) we show that asymptotic consensus obtains provided the signal space is finite and that the distribution of prior beliefs, hence of expected signals is bounded. Heterogeneity obtains otherwise. Additionally we show that under such conditions, a consensus also implies correct asymptotic beliefs.

Dixit and Weibull (2007) and Sethi and Yildiz (2012) study the polarization in public beliefs between two groups of agents when agents do not observe the prior beliefs of other agents outside of their group. They show how uncertainty of prior beliefs reinforces disagreement in opinions of the members of different groups. The findings in Dixit and Weibull (2007) and Sethi and Yildiz (2012) are special cases of this paper when the communication network is complete.

The closely related papers in the literature of bounded-rational learning are Demarzo et al. (2003) and Golub and Jackson (2010). The main difference with this paper is our assumption that agents start with private beliefs that are of finite precision. This assumption enables us to analyze the efficiency of information aggregation by public beliefs under bounded-rational learning and compare it with that under rational learning. In Proposition 2 we generalize the findings in

Demarzo et al. (2003), Golub and Jackson (2010) concerning the convergent beliefs to dynamic networks. We show that the communication network does not necessarily have to be connected at all periods for a consensus to obtain. Specifically, a consensus in public beliefs will always emerge in a long-run provide that there exists a positive probability of realizing a connected network. Golub and Jackson (2010) study conditions on the communication network for wisdom of crowds to obtain. Their definition of a wise crowd is closely related to our definition of correct asymptotic public beliefs in a weak sense (see Definition 2). They show that a society will be wise if for any finite-size subgroup of agents, the sum of weights connecting such a subgroup to the rest of the society is sufficiently large, and vise versa. That is, there should not be a subgroup of agents that stays prominent in the limit of population size. There are two major differences between our characterization of correct public beliefs from that of Golub and Jackson (2010). The first involves the initial conditions, whereby in Golub and Jackson (2010) it is assumed that agents' prior beliefs are drawn from a normal distribution with mean equal to the true state of nature. On the contrary, we model a general case in which prior beliefs are distributed heterogeneously across the population. We then show that a necessary condition for public beliefs to correctly aggregate private information is for prior beliefs to be correlated with the true state of nature. In other words, the initial condition of Golub and Jackson (2010) is derived as a necessary condition in our case. The second difference is that Golub and Jackson (2010) characterize conditions on the network topology under which wisdom does not occur. On the contrary, we characterize conditions under which wisdom occurs and in Theorem 2 we show that a network must be asymptotically balanced for a society to be wise. To demonstrate the generality of our approach, we provide an example in which public beliefs are not correct under Golub and Jackson (2010) but are correct in our case. We then further show as a corollary that correct public beliefs obtain in Erdös-Rényi family of random networks but not in networks formed through preferential attachment, like scale-free networks.

There is also a related literature on bounded-rational learning in physics and computer science which we shall not review in detail here. It focuses on conditions for a consensus to be attained contrary to this paper in which the primary goal is on the correctness of public beliefs and how it relates to rational learning. See Jackson (2008) for more references on related models.

The rest of the paper is organized as follows. In Section 3 we outline the framework of communication and learning, providing the informational and communication structure. Section 4 presents the characterization of the general structure of public beliefs for a finite population. Section 5 characterizes the conditions on the informational and communication network structures under which public beliefs correctly aggregate decentralize information. Section 6 deals with convergence rates and efficient information aggregation. All proofs are gathered in the Appendix.

3. The model

The set of agents is denoted by $N = \{1, \cdots, i, \cdots, n\}$. There is a state of nature X that is unknown and unobservable to all agents. The true value of X is $\bar{\mu}$ or generally a Dirac delta function centered at $\bar{\mu}$. Agents form and update beliefs about X. We assume without loss of generality that the prior belief of each $i \in N$ is a normal distribution with mean μ_i and unit variance. That is each $i \in N$ initially believes that X is normally distributed with mean $\mu_{i,0}$ and unit variance. The assumption of unit variance is for simplicity and does not affect the main conclusions of the paper.

We model agents' uncertainty of others' prior beliefs by assuming that they are normally distributed with mean ν_i for each $i \in N$ and variance-covariance matrix M with entries $m_{ij} \geq 0$ for all $(i,j) \in N$. That is

$$\boldsymbol{\mu}_0 \sim \mathcal{N}(\boldsymbol{\nu}, M) \quad \text{and} \quad X \sim_i \mathcal{N}(\mu_{i,0}, 1)$$

where $\boldsymbol{\mu}_0$ and $\boldsymbol{\nu}$ are column vectors of $\mu_{i,0}$ and ν_i for all $i \in N$ respectively, and $\sim_i$ means "distribution according to i". For a clearer exposition, we refer to the pair $(\mu_{i,0}, 1)$ as prior belief of $i \in N$ and the pair (ν_i, M_i), where M_i is the ith row of M, as i's prior belief distribution. When the prior beliefs are independently distributed then i's prior belief distribution will simply be the pair (ν_i, η^2), where η^2 is the associated variance. The assumption that X is normally distributed is also for simplicity and does not affect our main results.

Given prior beliefs, each agent observes a private signal s_i that is informative about X, and takes the form $s_i = X + \varepsilon_i$. Conditional on $\mu_{i,0}$, each i believes that X and $\varepsilon_1, \cdots, \varepsilon_n$ are independently distributed, and it is common knowledge that $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$ for all $i \in N$ Agents therefore differ with respect to their prior beliefs and private information, and our goal is to study the public beliefs resulting from deliberation under general communication networks. In what follows, we denote by $\mu_{i,t}$ for the mean associated with the posterior belief of i at period t, and by $var_{i,t}$ for the associated variance.

Lemma 1: *Given that X is normally distributed with mean $\mu_{i,0}$ and unit variance, after observing the signal s_i it follows by Bayes rule that i's posterior belief becomes $X \sim_i \mathcal{N}(\mu_{i,1}, var_{i,1})$, where $\mu_{i,1} = \frac{\sigma^2}{1+\sigma^2}\mu_{i,0} + \frac{1}{1+\sigma^2}s_i$ and $var_{i,1} = \frac{\sigma^2}{1+\sigma^2}$.*

Lemma 1 follows from the Bayesian relation that given $X \sim \mathcal{N}(\mu, \pi^2)$ and $\varepsilon \sim \mathcal{N}(0, \sigma^2)$, if $s = X + \varepsilon$ then

$$\mathbb{E}[X|s] = \frac{\sigma^2}{\pi^2 + \sigma^2}\mu + \frac{1}{\pi^2 + \sigma^2}s. \tag{1}$$

3.1. Communication network

The communication among agents is modeled through an associated network or graph. Let $G(n, E)$ be a graph with n vertices representing the number of agents and E is the set of edges linking different pairs of agents such that a graph g_{ij} defines a communication link between i and j. In particular if $g_{ij} > 0$, then j communicates to i, or simply that i observes j's posterior beliefs, and $g_{ij} = 0$ implies the absence of communication between i and j. No strict restrictions are imposed on g_{ij}'s except that $0 \leq g_{ij} \leq 1$. In particular it is not necessary that $g_{ij} = g_{ji}$. The corresponding adjacency matrix of interactions is denoted by G.[6]

The *first-order neighborhood* $N_{i,1}$ is the set of agents that directly communicate with i. That is, $N_{i,1} = \{j \in N; g_{ij} > 0\}$. The corresponding cardinality of $N_{i,1}$, $k_{i,1} = \#N_{i,1}$, is the *first-order degree* of i. A *path* between i and j is a connected set of links $\mathscr{P}_{ij} = \{g_{i1}, g_{12}, \cdots, g_{(j-1)j}\}$ such that $g_{ij} > 0$ for each $g_{ij} \in \mathscr{P}_{ij}$. The length of the path between i and j is denoted by $|\mathscr{P}_{ij}|$, which is simply the L^1 norm of $\mathscr{P}_{ij}$. We can then define the *second-order neighborhood* of i, $N_{i,2}$ as the

[6]We use G to denote both the underlying network and the corresponding matrix unless otherwise specified.

set of agents, indexed by j_2, such that for each $j_2 \in N_{i,2}$ there exists a path of length two between i and j. That is $N_{i,2} = \{j \in N;\ |\mathscr{P}_{ij}| = 2\}$. The *second-order degree* of i is $k_{i,2} = \#N_{i,2}$. In a similar logic we can define the *t-order neighborhood* of i as a set, $N_{i,t} = \{j \in N;\ |\mathscr{P}_{ij}| = t\}$. The corresponding *t-order degree* is denoted by $k_{i,t} = \#N_{i,t}$.

PROPERTY 1: *A communication network with the corresponding matrix G is said to be connected if for any two agents $(i,j) \in N$, there exists a path of at least one step from i to j and vice versa.*

DEFINITION 1: *(a) A geodesic $d_{ij}(G)$ between two agents $(i,j) \in N$ is the shortest path between them. That is $d_{ij}(G) = \min\{|\mathscr{P}_{ij}|;$ for a pair of agents $\{i,j\} \in N\}$.*

(b) The diameter of a network G, $D(G)$ is its longest geodesic. That is

$$D(G) = \max_{\{i,j\} \in N} \{|\mathscr{P}_{ij}|; \forall i,j \in N\}.$$

A network is said to be *finite* if its diameter is finite. Note that it is possible for the network to be finite when the population size is infinite.

3.2. Dynamic networks

We also consider dynamic communication structure in which the neighborhood of an agent changes over time. We do not define the specific mechanism by which agents switch their neighborhood but rather consider a generic and an arbitrary switching mechanism. Such mechanism can be strategic or simply random. The strategy can be a function of the current position in the network and/or other agents observable characteristics. See for example Sethi and Yildiz (2013) for a switching mechanism that depends on the precision of opponents' private information, and König et al. (2009) for a switching mechanism that depends on network related properties such as the centrality of an agent.

To be precise, let γ be the switching strategy or simply the switching signal, defined as $\gamma : N \to G$. We then denote by $\mathscr{G} = \{G_{\gamma(1)}, \cdots, G_{\gamma(t)}, \cdots\}$ as a class/set of all possible networks or graphs that can be defined on the set of agents N as a result of the switching strategy γ. $G_{\gamma(t)}$ is therefore the resulting network at time t from switching strategy γ. Under such dynamics, it is then possible that at certain periods some agents do not have any neighbors, that it $k_{i,0} = 0$.

A special case of this dynamic interaction structure includes that in which the switching strategy maps into the same network structure over time. That is $G_{\gamma(1)} = \cdots = G_{\gamma(t)} = \cdots$. We denote such a switching strategy by γ_0. Another special case is when the switching strategy induces a connected network structure at each period. That is each $G_{\gamma(t)}$ for all t is connected. Let such a switching strategy be denoted by γ_c.

3.3. Rational learning

Under rational learning, given the prior beliefs and after observing the signal s_i, each $i \in N$ computes their posterior belief. The posterior beliefs are then truthfully announced to the corresponding neighbors. After observing their neighbors' announcements, each $i \in N$ updates their belief and the resulting posterior beliefs are again simultaneously announced to the neighbors at the end of that period, and so forth. The sequential process continues until each agent's posterior announcements are constant, in which case each agent has no new information to learn from their neighbors. The limit belief of each $i \in N$ becomes common knowledge to his first-order neighbors, and we refer it as a *local public belief* or simply *public belief.* The crucial assumptions in the rational learning mechanism are common knowledge of rationality, and that agents have memory of the history of their neighbors announcements.

Example 1: *Consider two agents $N = \{a, b\}$ such that $g_{ab} = g_{ba} = 1$; an undirected link exists between a and b. Let the prior beliefs be normally distributed such that $X \sim_a \mathcal{N}(\mu_{a,0}, 1)$ and $X \sim_b \mathcal{N}(\mu_{b,0}, 1)$, and the private signals be $s_a = X + \varepsilon_a$ and $s_b = X + \varepsilon_b$ respectively. As defined above, assume that $X, \varepsilon_a, \varepsilon_b$ are independently distributed and that ε_a and ε_b are independent and identically distributed with $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$ for all $i = a, b$.*

In the first round, after observing signals, it follows from Lemma 1 that the posterior beliefs of both agents become $X \sim_i \mathcal{N}(\mu_{i,1}, var_{i,1})$ for $i = a, b$. Where$\mu_{i,1} = \frac{\sigma^2}{1+\sigma^2}\mu_{i,0} + \frac{1}{1+\sigma^2}s_i$ and $var_{i,1} = \frac{\sigma^2}{1+\sigma^2}$. At the end of the first round each agent announces their posterior beliefs. From b's announcement, a knows that

$$(1 + \sigma^2)\mu_{b,1} = \sigma^2\mu_{b,0} + s_b \tag{2}$$

If we assume that a knows the prior belief of b, that is a knows that b's prior belief is normally distributed with mean $\mu_{b,0}$ and unit variance, then a can deduce s_b from the first period announcement of b. The similar argument follows for b since $g_{ba} = 1$. Combining their posterior beliefs at the end of the first period together with the deduced signals s_a and s_b, both agents update their beliefs to $X \sim_i \mathcal{N}(\mu_{i,2}, var_{i,2})$ for $i = a, b$, where

$$\mu_{i,2} = \frac{\sigma^2}{2+\sigma^2}\mu_{i,0} + \frac{1}{2+\sigma^2}(s_a + s_b), \qquad var_{i,2} = \frac{\sigma^2}{2+\sigma^2}$$

From the announcements at the end of the second period, both a and b do not have anymore information to learn from each others' announcements. Learning stops, and their posterior beliefs become public information.

3.4. *Bounded-rational learning*

Under bounded-rational learning, agents update their beliefs sequentially just like in the case rational learning except that they are not able to disentangle between old and new information from announcements of their first-order neighbors. After receiving private signals, each agent updates their prior belief in accordance to Bayes rule as in Lemma 1. The resulting private beliefs are then communicated to the first-order neighbors. From the second period onwards, agents incorporate information from their first-order neighbors by simply taking the weighted average of their posterior beliefs. That is, at period t i's posterior mean is given by

$$\mu_{i,t+1} = \sum_{j=1}^{N} g_{ij}(t)\mu_{j,t} \quad i = 1, \cdots, n \tag{3}$$

where $0 \le g_{ij}(t) \le 1$ is the weight that i attaches to j's announcement at period t. Since each agents revises their belief in every period, it follows that $g_{ii}(t) > 0$ for all $i \in N$ and $t \ge 0$. If $G_{\gamma(t)}$

is the associated matrix of interactions in the t-th period, then (3) can be written as

$$\boldsymbol{\mu}_{t+1} = G_{\gamma(t)} \boldsymbol{\mu}_t \tag{4}$$

where $\boldsymbol{\mu}_t$ is a vector of posterior means in the t-th period. We assume that the same updating rule applies to even the variance of precision of agents beliefs.

4. Public beliefs

In this section, we characterize the general structure of public beliefs under the informational structure described above. We focus of the case of a finite population and provide conditions for heterogeneity and consensus in public beliefs to obtain.

4.1. *Rational learning*

The following theorem establishes the structure of public beliefs under rational learning.

Proposition 1: *Let the communication network be common knowledge and connected. Let also $\mu_{i,\infty}$ and $var_{i,\infty}$ denote the mean and variance of i's public belief respectively. Under rational learning the public belief of each $i \in N$ is normally distributed with mean and variance as follows.*

(i) If for each $i \in N$, $\mu_{j,0}$ for all $j \in N_{i,1}$ are observable, then

$$\mu_{i,\infty} = \frac{\sigma^2}{n+\sigma^2}\mu_{i,0} + \frac{1}{n+\sigma^2}\sum_{j=0}^{n} s_j \quad \text{and} \quad var_{i,\infty} = \frac{\sigma^2}{n+\sigma^2} \tag{5}$$

for all $i \in N$, where $s_0 = s_i$

(ii) If for each $i \in N$, $\mu_{j,0}$ for all $j \in N_{i,1}$ are unobservable but it is common knowledge that $\boldsymbol{\mu}_0 \sim \mathcal{N}(\boldsymbol{\nu}, \eta^2 \mathbf{I})$, where $\boldsymbol{\mu}_0$ is a column vector of all $\mu_{i,0}$ and $\mathbf{I}$ is an $n \times n$ identity matrix, then

$$\mu_{i,\infty} = \frac{(1+\sigma^2)(1+\eta^2\sigma^2)}{(1+\sigma^2)(1+\eta^2\sigma^2)+n-1}\mu_{i,1} + \frac{1}{(1+\sigma^2)(1+\eta^2\sigma^2)+n-1}\sum_{l\in N\setminus\{i\}} \mathbb{E}_i[s_l|\boldsymbol{\nu}]$$

and

$$var_{i,\infty} = \frac{\sigma^2(1+\eta^2\sigma^2)}{(1+\sigma^2)(1+\eta^2\sigma^2)+n-1}$$

where $\mathbb{E}_i[s_l|\boldsymbol{\nu}] = \sigma^2(\mu_{l,0} - \nu_l) + s_l$ for all $i \in N$ and all $l \in N \setminus \{i\}$ is the expected signal of l according to i given $\boldsymbol{\nu}$.

(iii) If the communication network is complete and it is common knowledge that $\boldsymbol{\mu}_0 \sim \mathcal{N}(\boldsymbol{\nu}, M)$, where $m_{ij} > 0$ for all $(i,j) \in N$, then $\mu_{i,\infty}$ and $var_{i,\infty}$ are as in (5) above.

Proof. See Appendix A.1 □

The first implication of Proposition 1 (i) is that when agents observe their neighbors' prior beliefs and that the communication network is common knowledge and connected, public beliefs fully incorporate all private information. The second implication is that under similar conditions on the network and prior beliefs, if the realized prior beliefs are identical, then a consensus in public beliefs obtain. This result is consistent with that of the literature on common knowledge and consensus, for example Geanakoplos and Polemarchakis (1982) and Parikh and Krasucki (1990) in which they find that a consensus arises under a common prior assumption. Similarly in the literature of Bayesian rational learning on network for example Gale and Kariv (2003) and Mueller-Frank (2013) it is established that a consensus (in actions) arises under common prior assumption and connectedness of the communication network. The only supplement to these papers in this regard is that Proposition 1 (i) also emphasizes the fact that prior beliefs do not have to be common knowledge for consensus to obtain, but rather they are observable to the first-order neighbor.

The third implication of Proposition 1 (i) is that heterogeneity in prior beliefs leads to heterogeneity in public beliefs. That is, let $\mu^o_{i,\infty}$ denote the public belief of i when neighbor's prior are observable, then for any pair of agents $(i, j) \in N$ it follows from (5) that

$$\mu^o_{i,\infty} - \mu^o_{j,\infty} = \frac{\sigma^2}{n + \sigma^2} (\mu_{i,0} - \mu_{j,0}) \tag{6}$$

in which case only the heterogeneity in prior beliefs leads to heterogeneity in public beliefs. This result also complements the literature on rational learning which has focused mainly on learning under common prior beliefs.

Proposition 1 (ii) establishes the structure of public beliefs when agents are uncertain about others' prior beliefs. Generally, public beliefs exhibit three sources of heterogeneity under uncertainty of prior beliefs. Heterogeneity could result from a difference in the realized priors and/or signals, as well as a difference in prior beliefs distributions if it exists. To see this notice that under assumptions of Proposition 1 (ii), for any pair of agents $(i, j) \in N$, $\mathbb{E}_i[s_l|\nu_k] = \mathbb{E}_j[s_l|\nu_r] = \sigma^2 (\mu_{l,0} - \nu_l) + s_l$ for each $k \in N_{i,1}$, $r \in N_{j,1}$ and all $l \in N$. Let $\mu^u_{i,\infty}$ denote the public belief of i when neighbor's prior are unobservable, then for any pair of agents $(i, j) \in N$

$$\mu^u_{i,\infty} - \mu^u_{j,\infty} = \frac{\sigma^2 \left(\eta^2\sigma^2(\mu_{i,0} - \mu_{j,0}) + \eta^2(s_i - s_j) + (\nu_i - \nu_j)\right)}{(1 + \sigma^2)(1 + \eta^2\sigma^2) + n - 1} \tag{7}$$

Implying that if the distributions of prior beliefs are heterogeneous, then heterogeneity in public beliefs can still arise even under identical realized priors and signals.

In comparison to the case in which priors are observable, the precision of public beliefs when priors are unobservable is always lower. That is let $var^o_{i,\infty}$ and $var^u_{i,\infty}$ be the variance of i's public beliefs when prior beliefs are observable and unobservable respectively, then

$$var^o_{i,\infty} - var^u_{i,\infty} = \frac{-\sigma^2(n-1)(\eta^2\sigma^2)}{(\sigma^2 + n)((1 + \sigma^2)(1 + \eta^2\sigma^2) + n - 1)} < 0 \tag{8}$$

This implies that agents' confidence in their beliefs at the end of the learning process is always lower when they are uncertain of their neighbors' prior beliefs compared to when they observe their neighbors' prior beliefs.

Proposition 1 (iii) entails two main implications. Since a pair of agents with an undirected link between them is a simplest form of a complete network, Proposition 1 (iii) implies that the findings in Geanakoplos and Polemarchakis (1982) extend to the case in which agents are uncertain of other's prior beliefs provided that their prior beliefs are correlated. This is a particularly strong result given that the literature on knowledge and consensus has always emphasized the condition of common knowledge of prior beliefs as a prerequisite for a consensus to obtain. Note the distinction we make between common knowledge of prior beliefs and common knowledge of prior beliefs' distributions. Proposition 1 (*iii*) shows that the necessary condition for consensus to emerge under rational learning when the network is complete, is prior beliefs' to be correlated and their distributions to be common knowledge. The second implication of Proposition 1 (*iii*) is that provided the network is complete and that prior beliefs are correlated, then public beliefs fully incorporate dispersed private information.

We do not go into the detailed analysis of the outcomes of rational learning when the communication network is dynamic because the requirements on agents' knowledge of the network switching strategies amounts to assuming common knowledge of the network structure, and therefore does not affect the final outcome. To see this, notice that the assumption that agents recall the history of their neighbors' announcements is indispensable if they are to be able to deduce the private information (or the expected private information) of their t-order neighbors. This then requires agents to have knowledge of the past and current positions of all agents in the network, which is equivalent to assuming common knowledge of the network. On the other hand, when the communication network is common knowledge, its topology does not affect the general properties of public beliefs. Hence, under rational learning, whether or not the network is dynamic does not affect the general properties of public beliefs.

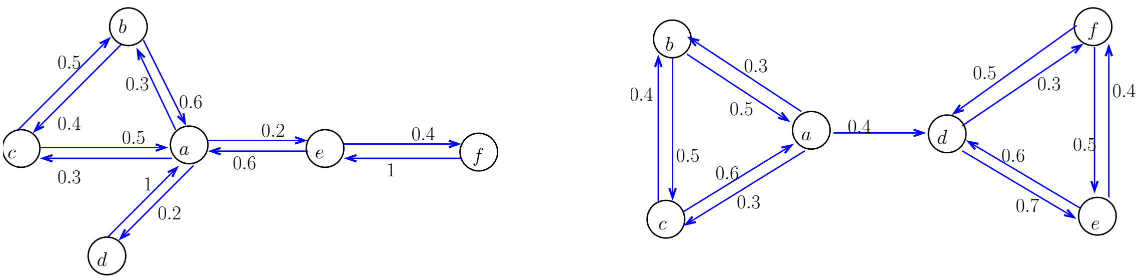
4.2. Bounded-rational learning

The dynamic system (4) can generally be treated as a non-homogeneous Markov chain, and the special case in which $\gamma = \gamma_c$ is a homogeneous Markov chain. The convergence properties of non-homogeneous Markov chains has been well established in the literature, and it particularly depends on the *irreducibility* and *aperiodicity* properties of the transition matrices $G_{\gamma(t)}$ for all $t \geq 0$. A Markov chain (transition matrix) is said to be irreducible if it is possible to make a transition from any one state to every other state, not necessarily in one time step. In the context of communication networks described in section 3.1, irreducibility of the network or the matrix induced by the network implies that a path $\mathscr{P}_{ij}$ between any pair of agents $(i, j) \in N$ exists. Aperiodicity on the hand implies that there does not exists two or more groups of agents for which communication is possible only among groups and not within groups. For the communication structure considered in this paper, aperiodicity is guaranteed since self loops exists. A more general property of Markov chains that guarantees convergence to well defined characteristics is that of *ergodicity*. Ergodic Markov chains are irreducible, aperiodic and *recurrent*. A Markov chain is recurrent if every state is revisited infinitely many times. The following lemma summarizes the properties of ergodic Markov chains and at the same time acts as a definition for an ergodic chain.

LEMMA 2: *Let $\mathscr{G} = \{G_{\gamma(t)}\}_{t\geq 0}$ be (an infinite) sequence of finite matrices. If a Markov chain formed as a product of matrices in $\mathscr{G}$ is ergodic, then*

$$\lim_{T\to\infty} \prod_{t=0}^{T} G_{\gamma(t)} = \mathbf{e}\mathbf{v}' \tag{9}$$

where $\mathbf{e}$ is a column vector of ones, $\mathbf{v}$ is a column vector of real values and $\mathbf{v}'$ is the transpose of $\mathbf{v}$.



(a) A connected network: G_1

(b) Network with prominent "family": G_2

Figure 1

Proof. The proof can be found in Wolfowitz (1963). □

Note that $\mathbf{ev}'$ is an $n \times n$ matrix with identical rows, hence it is of unit rank. The analysis of the convergence of (4) then reduces to determining the conditions under which a sequence of communication networks induced by a switching strategy γ lead to an ergodic Markov chain. The following proposition provides such conditions.

PROPOSITION 2: *Let $\tau \geq 0$ be a sufficiently large integer. If the switching strategy γ is such that in every time interval $[t, t+\tau)$ a connected communication network obtains, then*

$$\mu_{i,\infty} = \left[\lim_{T\to\infty} G^T \boldsymbol{\mu}_1\right]_i = \mathbf{v}'\boldsymbol{\mu}_1 \quad \textit{for every } i \in N \tag{10}$$

where $G^T = \prod_{t=0}^{T} G_{\gamma(t)}$.

Proof. See Appendix A.2. □

Proposition 2 states that under some restrictions on the switching strategy, specifically that there be a positive probability of realizing a connected network, a consensus in public beliefs obtains. The public beliefs are a weighted average of the initial private beliefs. The weights vector $\mathbf{v}$ generally depends on the network structure. To understand the nature of the weights vector let us consider the special case in which $\gamma = \gamma_c$, that is in which γ maps into the same network topology. This special case is that of static networks studied in Demarzo et al. (2003) and Golub and Jackson (2010), in which case the weight vector $\mathbf{v}$ is related to the left eigenvector of G_{γ_c} associated with the leading eigenvalue. The left eigenvector is normally associated with the measure of centrality; the *eigenvector centrality* (see Bonacich and Lloyd (2001) and the references therein). The centrality measures quantifies the level of influence of each agent. That is for each $i \in N$, the ith value of $\mathbf{v}$, v_i is the measure of how influential i is in shaping public beliefs. The composition of the vector $\mathbf{v}$ is thus central in determining the *correctness* and *quality* of public beliefs. For a detailed exposition of the nature of the weight vector $\mathbf{v}$ we refer the reader to Demarzo et al. (2003) and Golub and Jackson (2010), we only provide a simple example below.

EXAMPLE 2: *Let G_1 denote the associated transition matrix of the communication network in Figure 1a and G_2 for the that in Figure 1b, and let the corresponding weights vectors be $\mathbf{v}_1$ and $\mathbf{v}_2$ respectively. A power iteration of each transition matrix results into*

$$\mathbf{v}_1 = (0.363, 0.204, 0.191, 0.073, 0.121, 0.048) \quad \textit{and} \quad \mathbf{v}_2 = (0, 0, 0, 0.359, 0.381, 0.260).$$

Clearly, the influence of each agent depends on their first-order connectivity and the their closeness to other agents who are also have high first order-connectivity. Take for example the network G_1 *of Figure 1a in which agents* d *and* f *both observe announcements of only one other agent and both communicate to one other agent. Though the first-order neighbor of agent* f *attaches more weight to* f*'s announcements than does the first-order neighbor of agent* d*, agent* d *is more influential than* f *in a long-run. This is precisely the effect of being connected to other agents who are themselves have higher first-order connectivity, as is the case of* a *and* e *who are the first-order neighbors of* d *and* f *respectively.*

In the case of communication network in Figure 1b, there are two subgroups (that is $\{a, b, c\}$ *and* $\{d, e, f\}$*) each of whom form a complete subgroup. The inter-subgroup communication on the other hand is unidirectional, that is members of subgroup* $\{a, b, c\}$ *observe and learn from announcements of those in subgroup* $\{d, e, f\}$ *and not vise versa. As a consequence, a consensus emerges in a long-run in which member of subgroup* $\{a, b, c\}$ *adopt public beliefs of subgroup* $\{d, e, f\}$*. This example highlights the effect of the presence of prominent "families" discussed in Bala and Goyal (1998) in a different model but related in the sense that they also assume bounded-rationality of among as in this case.*

In general, under bounded-rational learning public beliefs are greatly influenced by the topology of the communication network and the distribution of private information. That is public belief of i will be normally distributed with mean

$$\mu_{i,\infty} = \sum_{j=1}^{n} v_j \left(\frac{\sigma^2}{1+\sigma^2} \mu_{i,0} + \frac{1}{1+\sigma^2} s_j \right). \tag{11}$$

and variance

$$\text{var}_{i,1} = \sum_{j=1}^{n} v_j \frac{\sigma^2}{1+\sigma^2} = \frac{\sigma^2}{1+\sigma^2}$$

The implication is that, under bounded-rational learning, if agents start with an identical level of precision or confidence in their beliefs, they stay so even after the learning process has ended. This result highlights the weakness associated with bounded-rational learning in correct aggregation of private information as will be elaborated in the following subsection.

5. Correct public beliefs

Correctness of public beliefs generally depends on whether or not they fully incorporate private information of all other agents and not just of their first-order neighbors. If private information of all agents is fully incorporated then we would expect the public beliefs to converge jointly in probability for large n to the true state of nature. We differentiate correctness in public beliefs for finite population from that when the population size is infinite. For a clear distinction, we refer to beliefs resulting from learning as the population size goes to infinity as *asymptotic public beliefs* as opposed to just public beliefs for finite population. Public beliefs are said to be correct if they fully incorporate private information of all other agents. In the case of asymptotic public beliefs we differentiate between correctness in a weak and strong sense. That is, public beliefs are correct in a *strong sense* if for each agent it converges in probability to a Dirac delta function centered at the true state of nature $\bar{\mu}$, and in the *weak sense* if it converges in probability to a normal distribution centered at the true state of nature and not necessarily zero variance. Based on these definitions it is easy to see that the basic requirement for both public beliefs and asymptotic public beliefs to be correct is the communication network to be connected. We therefore will not repeatedly emphasize this fact in most of the analysis that follows.

In the case of rational learning we have already provided conditions for public beliefs to fully incorporate private information of other agents, so we will now focus on correctness of asymptotic public beliefs below. To precisely define convergence in probability, we need to define a sequence of networks for a growing population size n. We let $G(n)$ denote the network or equivalently the corresponding matrix when the population size is n. This then implies that there are two limits in the learning processes above; the limit for time t and for population size n. To study such dynamic processes, we have to assume that one of the two limits is reached faster than the other. Alternatively, and perhaps even more intuitive, we first fix the network topology and derive the corresponding nature of public beliefs, then study the evolution of public beliefs for a growing population size. The population size should grow in such a way that the topology (or at least the main characteristics) of the communication network are preserved. Given the nature of public beliefs, we then study the sequence of networks $\{G(n)\}_{n\geq 2}$ of fixed topology and show conditions under which convergence in probability of public beliefs to the true state of nature occurs. We formally define strong and weak correctness of asymptotic public beliefs as follows.

Definition 2: *(a) Asymptotic public beliefs are said to be correct in a strong sense if for each* $i \in N$

$$\lim_{n\to\infty} \mathbb{P}\left(|\mu_{i,\infty}(n) - \bar{\mu}| > \epsilon\right) = 0 \quad \textit{and} \quad \lim_{n\to\infty} var_{i,\infty}(n) = 0 \tag{12}$$

for all possible realizations of $\mu_{i,0}$.

(b) Asymptotic public beliefs are said to be correct in a weak sense if for each $i \in N$, *condition* (12) *is satisfied but*

$$\lim_{n\to\infty} var_{i,\infty}(n) < \infty \tag{13}$$

Under strong correctness of asymptotic public beliefs we require that each agent's public belief converges in probability *precisely* to the true state of nature for all possible realizations of $\mu_{i,0}$. For weak correctness, the asymptotic public belief of each agent must place the most weight on the true state of nature for all possible realizations of $\mu_{i,0}$. With these definitions, if public beliefs are correct in a strong sense then they are also correct in a weak sense.

5.1. Rational learning

As shown in Example 1 and Proposition 1, full incorporation of private information by public beliefs in the case of rational learning depends on the observability of prior beliefs. We show below that in addition to observability of prior beliefs, correctness of asymptotic public beliefs also depends on the structure of prior beliefs and of private information, and in some cases on the structure of the communication network. The following theorem establishes necessary conditions for asymptotic public beliefs to be correct.

THEOREM 1: *As a general condition, let the sequence* $\{G(n)\}_{n\geq 2}$ *be common knowledge.*

(i) When it is common knowledge that $\boldsymbol{\mu}_0 \sim \mathcal{N}(\boldsymbol{\nu}, \eta^2\mathbf{I})$, *then* $\mu_{i,\infty}(n) \xrightarrow{p} \bar{\mu}$ *and* $\lim_{n\to\infty} var_{i,\infty}(n) = 0\ \forall i \in N$ *if* σ^2 *and* η^2 *are finite.*

(ii) When it is common knowledge that $\boldsymbol{\mu}_0 \sim \mathcal{N}(\boldsymbol{\nu}, M)$, *where* $m_{ij} > 0$ *for all* $(i,j) \in N$, *then* $\mu_{i,\infty}(n) \xrightarrow{p} \bar{\mu}$ *and* $\lim_{n\to\infty} var_{i,\infty}(n) = 0\ \forall i \in N$ *if the communication network is complete and that* $m_{ii} < \infty$.

Proof. See Appendix A.3. □

The general implication of Theorem 1 is that whenever asymptotic public beliefs are correct, they are correct in a strong sense. Depending on the observability of prior beliefs though, there exists a difference on the speed of convergence to correct asymptotic public beliefs, which is determined by the rate at which the variance of an agent's belief goes to zero. As indicated in the discussion of Proposition 1, the variance associated with an agent's public belief for a given population size depends on the observability of prior beliefs.

Theorem 1 (i) shows that despite the difference in the speed of convergence, correct asymptotic public beliefs obtain irrespective of uncertainty of others' prior beliefs provided it is common knowledge that prior beliefs are independently distributed, and that the signal space and prior beliefs space are finite. That is $\sigma^2 < \infty$ and $\eta^2 < \infty$. Note that in Theorem 1, correct asymptotic public beliefs also implies consensus in asymptotic public beliefs. The condition that the signal space must be finite for correct learning (public beliefs in this case) to occur has been pointed out in other models of learning in the literature. For example Cripps et al. (2008), in which they show that a finite signal space is necessary for a consensus to obtain in a learning model where two agents observe and learn from a sequence of correlated private signals. Similarly in the literature of sequential Bayesian learning, it is shown that boundedness of private beliefs plays a role in determining whether or not wrong herds will emerge (Smith and Sorensen, 2000). Recall that a private belief is what results after agents incorporate their realized signal into their prior belief, so in essence if the signal space is finite then private beliefs are also finite/bounded.

Our claim that correct asymptotic public beliefs obtain irrespective of uncertainty of prior beliefs provided that prior beliefs are independently distributed and that the signal and prior beliefs spaces are finite, is a particularly strong result. Since uncertainty in prior beliefs in turn leads to uncertainty in the signal interpretation, one would expect that the uncertainty in signal interpretation should derail correctness in asymptotic public beliefs. Although our framework is different, this finding is contrary to Acemoglu et al. (2009) who find that when two agents learn from a sequence of signals and that they are uncertain of signal interpretation, then their beliefs do not necessarily converge to a consensus (which is equivalent to correct public beliefs in our case). We find that asymptotic consensus fails to arise only if the distribution of prior beliefs, hence of expected signals is not bounded.

Theorem 1 (ii) establishes the basic condition for asymptotic public beliefs to be correct when the distributions of prior beliefs are correlated and that the correlation coefficients are not necessarily finite. In particular, the variance of the prior belief of each agent must be finite and that the communication network is complete.

5.2. *Bounded rational learning*

Under bounded-rational learning, the network topology almost solely determines the nature of public beliefs and should thus equally influence their correctness. Since a consensus exists in public beliefs, we can focus on the correctness of the public belief of a single agent for all possible realizations of prior beliefs. Under bounded-rational learning, since agents incorporate others' beliefs by taking weighted averages, it follows that the public belief of an agent is said to be correct if it attaches equal weight to the private beliefs of all other agents. The following definition formalizes this observation.

DEFINITION 3: *Under bounded-rational learning, public beliefs are said to be correct or fully aggregate private information if* $\mu_{i,\infty}(n) = \text{Ave}[\boldsymbol{\mu}_1]$ *and that* $var_{i,\infty}(n) < \infty$ *for each* $i \in N$*, where*

$$\text{Ave}[\boldsymbol{\mu}_1] = \frac{1}{n}\sum_{i=1}^{n} \mu_{i,1}.$$

The following additional definitions are useful for a complete characterization of the results below.

DEFINITION 4: *A matrix* $G(n)$ *is said to be doubly stochastic if* $\sum_{j=1}^{n} g_{ij}(n) = \sum_{i=1}^{n} g_{ij}(n) = 1$ *for all* $(i,j) \in N$

DEFINITION 5: *(a) A communication network is said to be perfectly balanced if the corresponding matrix* $G(n)$ *is doubly stochastic.*

(b) A sequence of networks $\{G(n)\}_{n\geq 2}$ *is said to be asymptotically balanced if* $\lim_{n\to\infty} G(n) = S$*, where* S *is an arbitrary doubly stochastic matrix.*

THEOREM 2: *Under bounded-rational learning, if* $\{\mu_{i,\infty}(n)\}_{n\geq 2}$ *for each* $i \in N$ *is the sequence of public beliefs corresponding to the sequence* $\{G(n)\}_{n\geq 2}$ *of networks, then*

(i) $\mu_{i,\infty}(n) = \text{Ave}[\boldsymbol{\mu}_1]$ *and* $var_{i,\infty}(n) < \infty \; \forall i \in N$ *if and only if* $\sum_{j=1}^{n} g_{ij}(n) = \sum_{i=1}^{n} g_{ij}(n)$ *for all* $(i,j) \in N$*.*

(ii) $\mu_{i,\infty}(n) \xrightarrow{p} \bar{\mu}$ *and* $\lim_{n\to\infty} var_{i,\infty}(n) = \frac{\sigma}{1+\sigma} \; \forall i \in N$ *if and only if* $\lim_{n\to\infty} G(n) = S$*,* $\boldsymbol{\mu}_0 \sim \mathcal{N}(\bar{\boldsymbol{\mu}}, \eta^2\mathbf{I})$*, and that* σ^2 *and* η^2 *are finite. Where* $\bar{\boldsymbol{\mu}}$ *is an* n *dimensional vector of* $\bar{\mu}$*.*

Proof. See Appendix A.4. □

Theorem 2 (i) implies that under bounded-rational learning, correct public beliefs obtain if and only if the network is balanced. The extensiveness of the class of networks that satisfy the balancedness condition depends on whether or not self-loops are permitted to be heterogeneous. If the communication structure is such that all agents must place the same weight on their beliefs (that is $g_{11} = \cdots = g_{nn}$), then there are limited network topologies that satisfy balancedness. If on the other hand the weights g_{ii} for all agents are permitted to be heterogeneous then there exists a wide range of networks that satisfy the balancedness condition. This can be checked by comparing the first-order influences using the balancedness condition, that is $\sum_{i=1}^{n} g_{ij} = \sum_{i=1}^{n} g_{ik}$ for $j \neq k$. Which is equivalent to

$$g_{jj} + \sum_{i\neq j} g_{ij} = g_{kk} + \sum_{i\neq k} g_{ik}$$

If $g_{jj} = g_{kk}$ for all $(j,k) \in N$ then it must be that $\sum_{i \neq j} g_{ij} = \sum_{i \neq k} g_{ik}$, which implies that all $(j,k) \in N$ must assign weights in an identical manner though not necessarily to the same first-order neighbors. If on the contrary $g_{jj} \neq g_{kk}$ for all $(j,k) \in N$ such that $\sum_{i \neq j} g_{ij} \neq \sum_{i \neq k} g_{ik}$ for all $(j,k) \in N$, then agents can assign weights in different and various ways.

Theorem 2 (*ii*) has three main implications. First, under bounded-rational learning asymptotic public beliefs can only be correct in a weak sense. Once agents start with beliefs that are not completely precise (in the sense that the variance of prior beliefs is greater than zero), they will not be able to learn the true state of nature with complete precision as in the case of rational learning. Secondly, unlike in the case of rational learning where the distribution of prior beliefs can be independent of the true state of nature, under bounded-rational learning asymptotic public beliefs can only be correct if both the signals and prior beliefs are informative about the true state of nature. That is the prior belief of each agent must be normally distributed with mean equal to the true state of nature and finite variance. This restrictive condition highlights the superiority of rational learning in aggregating privately information compared to bounded rational learning. Third, correct asymptotic public beliefs obtain if and only if the network is asymptotically balanced. The range of networks that are asymptotically balanced is wide and we explore their characterization in the following subsection.

5.3. *Asymptotically balanced networks*

To characterize the class of communication networks that are asymptotically balanced, we need to construct a measure of balancedness of a network/matrix. We denote by $\phi(n)$ for a network $G(n)$ as a measure of its balancedness defined as follows.

Definition 6: *Let $S(n)$ be closest doubly stochastic matrix in terms of the Frobenius norm to the matrix $G(n)$. Then $G(n)$ is said to be $\phi(n)$-balanced if given $S(n)$,*

$$\|S(n) - G(n)\|_F = \phi(n)$$

where $\|.\|_F$ is the Frobenius norm.

We can then rephrase definition 5 (b) as saying that a sequence $\{G(n)\}_{n \geq 2}$ is asymptotically balanced if

$$\lim_{n\to\infty} \phi(n) = \lim_{n\to\infty} \|S(n) - G(n)\|_F = 0$$

The following proposition provides an expression for $\phi(n)$ for any stochastic matrix of size n.

Proposition 3: *Let $G(n)$ with entries $g_{ij}(n)$ for $(i,j) \in N$ be a stochastic matrix corresponding to a communication network whose measure of balancedness $\phi(n) > 0$. Then*

$$\phi^2(n) = \frac{1}{n} \sum_{j=1}^{n} \left(1 - \sum_{i=1}^{n} g_{ij}(n)\right)^2 \tag{14}$$

Proof. See Appendix A.5. □

Equation (14) gives an expression for the measure of balancedness of any arbitrary stochastic matrix. The quantity $\sum_{i=1}^{N} g_{ij}(n) = d_j^{in}(n)$ is the *in-degree* of agent j. It specifies the influence of j

on her *first-order neighbors'* beliefs. Unlike the out-degree $d_i^{out}(n) = \sum_{j=1}^{N} g_{ij}(n)$ that is normalized to unity for all agents, the in-degree can generally be equal to, less or greater than unity. The higher the value of $d_j^{in}(n)$, the more influential j in terms of first-order influence. For various values of n, $d_j^{in}(n)$ for each j is a random variable whose distribution depends on how the communication network scales with n. For a sequence of networks $\{G(n)\}_{n\geq 2}$ there exists a corresponding sequence $\{d_i^{in}(n)\}_{n\geq 2}$ of in-degree for each $i \in N$. If we define $d_{max}^{in}(n) = \max_{i\in N}\{d_i^{in}(n); i \in N\}_{n\geq 2}$, then for each sequence $\{G(n)\}_{n\geq 2}$ there exists a corresponding sequence $\{d_{max}^{in}(n)\}_{n\geq 2}$.

The following theorem employs Proposition 3 to characterize conditions for a sequence of communication networks to be asymptotically balanced.

THEOREM 3: *A sequence of networks $\{G(n)\}_{n\geq 2}$ with the corresponding sequence of maximum in-degrees $\{d_{max}^{in}(n)\}_{n\geq 2}$ is asymptotically balanced if and only if*

(i) $\lim_{n\to\infty} d_{max}^{in}(n) = 1$

(ii) there exists an integer $n' > 0$ such that for all $n > n'$ the quantity

$$h(n) = \sum_{j=1}^{n} \left(1 - \sum_{i=1}^{n} g_{ij}(n)\right)^2$$

stays constant.

Proof. See Appendix A.6. □

There exists a wide range of network topologies that satisfy condition (i) in Theorem 3, including random networks (graphs). To see this, notice that Theorem 3 (i) also implies that the sequence of the influence of the most influential agent $\{v_{max}(n)\}_{n\geq 2}$, where $v_{max}(n) = \max_{i\in N}\{v_i(n); i \in N\}$ must converge to zero for large n if the sequence $\{G(n)\}_{n\geq 2}$ is to be asymptotically balanced. This argument follows directly from the fact that the most influential agents are either those that are the most connected or those that are connected to the most connected agents (see example 2). Now consider the Erdös-Rényi family of random networks with parameters n and p (that is a random network $G(n, p)$ of n agents in which each link is included in the network with probability p independently of the other links), it is shown that for $p \geq \frac{\log^6 n}{n}$ and for all $i \in N$ (Mitra, 2009)

$$c\frac{1}{\sqrt{n}}\frac{\log n}{\log(np)}\sqrt{\frac{\log n}{np}} - \frac{1}{\sqrt{n}} \leq v_1(i) \leq c\frac{1}{\sqrt{n}}\frac{\log n}{\log(np)}\sqrt{\frac{\log n}{np}} + \frac{1}{\sqrt{n}} \qquad (15)$$

with probability $1 - o(1)$, where $c > 0$ is some constant. Clearly, for a sequence of such graphs it follows that $\lim_{n\to\infty} v_{max}(n) = 0$.

The class of communication networks that is ruled out by Theorem 3 (i) is those that are formed through preferential attachment, such as scale-free and generally high-clustering networks. In such networks, $d_{max}^{in}(n) := f(n)$, where f is an increasing function in its argument. That is $d_{max}^{in}(n)$ increases with n.

The following examples illustrate more other general network topologies that satisfy the two conditions for asymptotic balancedness in Theorem 3 and one which does not.

Example 3: *(a) Consider a communication network that assumes a one dimensional lattice structure in Figure 2. This example is adapted from Golub and Jackson (2010) to demonstrate the generality of our Theorem 3 compared to the characterizations of wisdom of crowds in Golub and Jackson (2010). They show that a society assuming a topology in Figure 2 is not wise (that is public beliefs do not correctly aggregate private information), but we demonstrate the contrary using Theorem 3.*

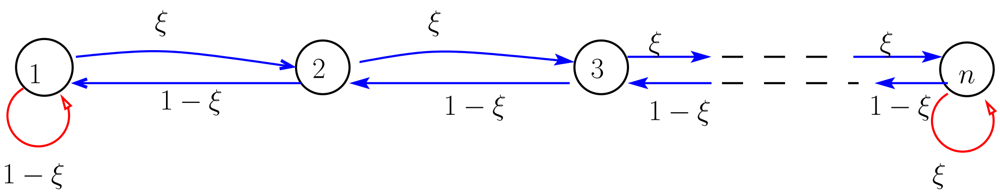


Figure 2: A network satisfying condition (ii) of Theorem 3

It is easy to see that in such a network, the quantity $h(n)$ is constant and that

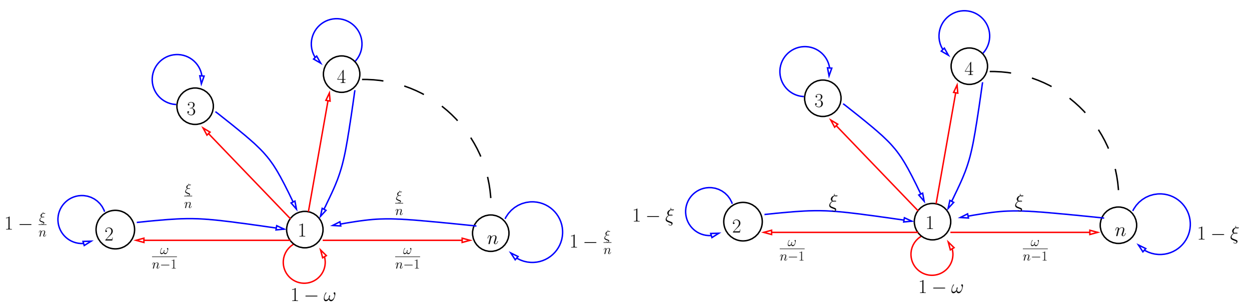
$$\phi^2(n) = \frac{2(1-2\xi)^2}{n} \tag{16}$$

Implying that $\lim_{n\to\infty}\phi(n) = 0$, hence public beliefs correctly aggregate private information.

(b) Consider the following two cases in which $h(n)$ is not independent of n. In the communication network topology of Figure 3a, each agent's first-order influence decays with n. The corresponding $\phi^2(n)$ is

$$\phi^2(n) = \frac{1}{n}\left[\left(\omega - \frac{n-1}{n}\xi\right)^2 + (n-1)\left(\frac{\xi}{n} - \frac{\omega}{n-1}\right)^2\right] \tag{17}$$

Consequently, $\lim_{n\to\infty}\phi(n) = 0$, in which case asymptotic balancedness obtains.



(a) Each agents influence decays with n

(b) Player 1's influence grows with n

Figure 3

For the communication network of Figure 3b

$$\phi^2(n) = \frac{1}{n}\left[(\omega-(n-1)\xi)^2 + (n-1)\left(\xi - \frac{\omega}{n-1}\right)^2\right], \tag{18}$$

in which case $\lim_{n\to\infty}\phi(n) = \infty$. *This is a general situation in which an agent or a subgroup of agents posses unbounded influence. In such communication networks asymptotic balancedness fails.*

6. Convergence rate and efficient information aggregation

6.1. Convergence rate

In both learning mechanisms, the convergence rate depends solely on the topology of the communication network. The following proposition summarizes the convergence rates in terms of network properties.

Proposition 4: *Let* $G_{\gamma(t)} = G_{\gamma_c}$ *for all* $t \geq 0$, *where* G_{γ_c} *is the transition matrix resulting from a communication network. Let* $\lambda_2(G_{\gamma_c})$ *and* $D(G_{\gamma_c})$ *be the second largest eigenvalue and the diameter of* G_{γ_c} *respectively. Denote by* r_R *and* r_{BR} *for the convergence rates of rational and bounded-rational learning respectively, then*

$$r_R = \frac{1}{D(G_{\gamma_c})} \geq \frac{\ln(1/\lambda_2(G_{\gamma_c}))}{\ln(c)} > \lambda_2(G_{\gamma_c}) = r_{BR}$$

Proof. See Appendix A.7. □

Proposition 4 shows that rational learning generally converges faster than bounded-rational learning. Specifically, under rational learning, it is the diameter of the communication network that matters and all other properties such as clustering and skewness of degree distributions do not matter as much. In the case of bounded-rational learning, clustering and the nature of degree distributions matter since the second eigenvalue depends on them. For example, a network with a given number of links in which agents are clustered into subgroups with few (or weak) connections between subgroups generally has a higher second eigenvalue than a network with equal number of agents and edges but in which edges are randomly distributed. Similarly, a network formed by removing links from a "parent" network generally has a higher second eigenvalue than its "parent" network. This argument follows from the well known concept of *interlacing eigenvalues* according to Fiedler (1973). Finally if agents are highly "inward-looking" when updating their beliefs, that is for each agent the weight of a self-loop is higher compared to the total weight of out-going links, then such a network also has a higher second eigenvalue, hence slow convergence rate.

The relevance of the network topologies in determining the convergence rate becomes more apparent in decision environments with discounted payoffs as will be demonstrated in the subsection below.

6.2. Efficient information aggregation

To establish the efficiency of the two learning mechanisms and that of the network topologies in aggregating information, we introduce a simple decision problem in which agents' objective is to minimize the expected loss from mismatching their action and the true state of nature. Specifically, each agent has two choices in each period; either to take an action $a \in A_i$, where A_i is a continuous action space for i, or "wait" for the next period(s). The joint action space (which we assume to be homogeneous for all agents, $A_1 = \cdots = A_n$) for each agent is thus $\mathscr{A} = \{A, w\}$, where w stands for "wait". If an agent takes an action within the action space A, he "exits" the game, meaning that he no longer learns and transmits new information but only transmits the same information he possesses at the time of exit. If on the other hand the agent decides to wait, then his expected loss for the next period is increased by a factor of $\frac{1}{\delta}$, where $0 \leq \delta \leq 1$. That is

$$U_{it}(\tilde{a}_{it}, X) = \begin{cases} -\frac{1}{\delta^t}\mathbb{E}_{i,t}\left[(\tilde{a}_{it} - X)^2\right] & \text{if } \tilde{a}_{it} = a \in A \\ 0 & \text{if } \tilde{a}_{it} = w \end{cases}$$

where $\mathbb{E}_{i,t}$ stands for expectation according to agent i at period t and $\tilde{a} \in \mathscr{A}$. We assume for simplicity that agents are homogeneous in terms of payoff structure. The strategy of each agent thus entails choosing the optimal period to exit the game.

Under this setup, it is easy to see that the optimal action $a \in A$ for each agent at any given period is

$$a^*_{it} = \mathbb{E}_{i,t}[X] = \mu_{i,t}$$

and the payoff corresponding to the optimal action is

$$U_{it}(a^*_{it}, X) = -\frac{1}{\delta^t}\mathbb{E}_{i,t}\left[(\mu_{i,t} - X)^2\right] = -\frac{1}{\delta^t} var_{i,t}$$

where $var_{i,t}$ is the variance of X according to i at period t. This implies that an agent's exit time depends on his confidence in his beliefs.

We can also define the associated expected social welfare as the average of individual optimal expected utilities as follows.

$$W_t(X) = \frac{1}{n}\sum_{i=1}^{n} U_{it}(a^*_{it}, X) = -\frac{1}{n\delta^t}\sum_{i=1}^{n}\mathbb{E}_{i,t}\left[(\mu_{i,t} - X)^2\right]$$

Consider first the case in which there is no cost on the payoffs associated with waiting, that is when $\delta = 1$. It is clear that it is optimal for agents to wait until all information has been exchanged among all agents, which in the case of bounded-rational learning implies until a consensus is reached. Now, consider a balanced network (which supports correct information aggregation under bounded-rational learning) of size n. The optimal action (taken at the end of information exchange process) under rational learning when priors are observable is

$$a^*_{i\infty} = \frac{\sigma^2}{n+\sigma^2}\mu_{i,0} + \frac{1}{n+\sigma^2}\sum_{j=0}^{n} s_j \tag{19}$$

and the corresponding expected social welfare $W_\infty(X)$ is

$$W_\infty(X) = -\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_{i,t}\left[(a^*_{i\infty} - X)^2\right] = -\frac{1}{n}\sum_{i=1}^{n} var_{i,\infty} = -\frac{\sigma^2}{n+\sigma^2} \tag{20}$$

where the last equality follows from Proposition 1 (i). Similarly for bounded-rational learning, the optimal action is

$$a^*_{i\infty} = \frac{1}{n}\sum_{j=1}^{n}\left(\frac{\sigma^2}{1+\sigma^2}\mu_{i,0} + \frac{1}{1+\sigma^2}s_j\right). \tag{21}$$

where we have used the fact that for a balanced network of size n, the weights or influence vector is $\mathbf{v} = (\frac{1}{n}, \cdots, \frac{1}{n})$ (see the proof of Theorem 2 in Appendix A.4 for details). And the corresponding expected social welfare is

$$W_\infty(X) = -\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_{i,t}\left[(a_{i\infty}^{*} - X)^2\right] = -\frac{1}{n}\sum_{i=1}^{n} var_{i,\infty} = -\frac{\sigma^2}{1+\sigma^2} \tag{22}$$

The *Price of Rationality (PoR)* defined as the ratio of expected social welfare under rational learning to the expected social welfare under bounded-rational learning, can then be used to analyze how the efficiency of information aggregation (or learning in general) degrades due to bounded rationality of the agents. From (20) and (22) we thus have,

$$PoR = \frac{1+\sigma^2}{n+\sigma^2} \tag{23}$$

Note that since the expected social welfare is the sum of the expected loss or cost, its possible maximum value is zero. It follows that the minimum possible value for the PoR is zero and its maximum is one. When PoR is one, the two learning mechanisms measure equally and rationality plays no role in the individual optimal decisions. If on the other hand PoR is zero, then rational learning is infinitely more efficient in aggregating information than bounded rational learning. From (23), PoR tends to zero with large n, implying that rationality becomes more and more important in efficient aggregation of information as the population size increases. So long as $\delta = 1$ and agents do not find it costly to wait until all private information has been exchanged before taking action, then the communication network does not matter and only the population size does matter. If on the other hand $\delta < 1$, the network topology starts to matter through its role in determining the convergence rate. The convergence rate becomes critical since agents have to optimize when to exit the game and if the convergence rate is high such that some or all agents exit the game before all the private information has been exchanged, then the actions they take at the end of the learning process will lead to higher expected loss hence lower expected social welfare. To formalize this argument, we need to formally define *correct asymptotic learning* in terms of actions taken at the end of the learning process. In this particular decision problem, correct asymptotic learning coincide with correct asymptotic public beliefs, but this does not always have to be the case.

Definition 7: *Given a sequence of networks $\{G(n)\}_{n\geq 2}$, correct asymptotic learning is said to occur if*

$$\lim_{n\to\infty}\lim_{t\to\infty}\mathbb{P}\left(|a_{i,t}(n) - \bar{\mu}| > \epsilon\right) = 0 \quad \textit{for all } i \in N \tag{24}$$

where $\bar{\mu} = a_{i,\infty}(\infty)$ is the possible optimal action for all agents.

When $\delta = 1$ the conditions for correct asymptotic learning is precisely those described in Theorems 1 and 2 for rational and bounded-rational learning respectively, and the convergence rate does not play much of a role in achieving correct asymptotic learning or public beliefs. If on the other hand $\delta < 1$, the convergence rate plays a big role in determining whether or not correct asymptotic leaning actually occurs. This follows from the fact that when $\delta < 1$ agents have to choose an optimal exit time, and this has an impact on how much of the private information gets exchanged before the learning process ends. It also implies that the topologies of the communication network that support correct asymptotic learning (and correct public beliefs) will no longer be those illustrated in Theorems 1 and 2. The following proposition states conditions on the communication network for correct asymptotic learning to occur when $\delta < 1$.

Proposition 5: *Let $\{G(n)\}_{n\geq 2}$ be a sequence of connected communication networks, and that let $\delta < 1$. If $\{D(G(n))\}_{n\geq 2}$ is the corresponding sequence of networks diameters, then correct asymptotic learning obtains under rational learning if*

$$\lim_{n\to\infty} \frac{1}{n} D(G(n)) = 0 \tag{25}$$

Under bounded-rational learning, the network must be complete.

Proof. See Appendix A.8 □

Proposition 5 basically emphasizes the importance of the topology of the communication network in determining the convergence rate hence efficient information aggregation, and can be summarized as follows. When $\delta < 1$, though waiting reduces the expected loss associated with a mismatch between ones action and the true state of nature, waiting for too long on the other hand becomes costly as prescribed by the factor $1/\delta$. It is therefore important that the convergence rate be high if correct learning is to ever be achieved. Condition (25) is a necessary condition for both fast and correct aggregation of information, hence correct asymptotic learning.

The network topology depicted in Figure 3 for example in which $D(G(n)) = 2$ for all n, satisfies condition (25). The network topology in Figure 2 on the other hand does not satisfy condition (25), since $D(G(n)) = n$. In the case of random graphs the diameter is generally an increasing function of n (see for example Bollobás (1981) for analysis on the diameter of random graphs), hence correct asymptotic learning does not obtain in random networks when $\delta < 1$.

The failure by agents to account for informational externalities under bounded rational learning is even more pronounced when $\delta < 1$. All agents would prefer to take an action in the action space A in the first or second periods since waiting longer does not improve their confidence in their beliefs.

We conclude this section by noting the indirect implication of Proposition 5. When the communication network has a large diameter and that agents find it costly to wait before taking an action, even if agents start with a common prior, heterogeneity in public beliefs and actions will obtain.

7. Conclusion

Individual beliefs play a significant in determining the success of policies, initiatives and theories, and play a significant role in determining outcomes in decisions under uncertainty. Contrary to prediction from models of rationality and common prior assumptions that learning through deliberation leads to a consensus in beliefs, heterogeneity in beliefs is more of a rule than an exception in most economic and social environments. In this paper, we investigated the properties of beliefs resulting from rational and bounded-rational learning in social networks. We established conditions under which heterogeneity prevails, and under which such beliefs correctly aggregate private information. Under bounded-rational learning, correctness in public beliefs is determined mainly by the topology of the communication network. Specifically, for a finite population the network must be balanced and for large population it must be asymptotically balanced. In addition to restrictions on the network topology, we also find that under bounded-rational learning, public beliefs are asymptotically correct if and only if the prior beliefs are correlated to the true state of nature.

In section 6 we provide a "toy" model of decision making under uncertainty, in which agents find it costly to wait much longer before taking an action. It highlights another possible source of heterogeneity in public beliefs that is not necessarily a result of historical factors or of observability of prior beliefs, but rather resulting from the topology of the communication network. That is, under such conditions heterogeneity in public beliefs will arise if the communication network has a large diameter.

Through out the paper we assumed that agents start with homogeneous confidence in their prior beliefs. Relaxing this assumption would give a richer understanding of the nature of public beliefs and distribution of actions in exit games or observed levels of diversity in public opinions across the population even when agents are exposed to the same sources of information (e.g Chamley and Gale (1994) and Murto and Välimäki (2011)). But this analysis requires a more detailed model (than the one provided in section 6) with specific preference structure, and we postpone this question for future research.

APPENDIX

A.1. Proof of Proposition 1

(i) If $X, \varepsilon_1, \cdots, \varepsilon_n$ are independent and it is common knowledge that $\varepsilon_1 \sim \mathcal{N}(0, \sigma^2)$, then from Lemma 1

$$\mu_{i,1} = \frac{\sigma^2}{1+\sigma^2}\mu_{i,0} + \frac{1}{1+\sigma^2}s_i, \quad var_{i,1} = \frac{\sigma^2}{1+\sigma^2} \quad \text{for all } i \in N$$

We assume that under rational learning $g_{ij} = 1$ if j communicates to i and zero otherwise. Let G be static and common knowledge, and let j_t be the index for the t-order neighbors. Consider any $i \in N$ with first-order degree $k_{i,1}$. After the first round of announcements, such an i updates his beliefs to a normal distribution mean

$$\mu_{i,2} = \frac{\sigma^2}{1+k_{i,1}+\sigma^2}\mu_{i,0} + \frac{1}{1+k_{i,1}+\sigma^2}\left(s_i + \sum_{j_1 \in N_{i,1}} s_{j_1}\right) \quad var_{i,2} = \frac{\sigma^2}{1+k_{i,1}+\sigma^2}$$

where $s_{j_1} = (1+\sigma^2)\mu_{j_1,1} - \sigma^2\mu_{j_1,0}$ for all $j_1 \in N_{i,1}$. If $\mu_{j_1,0}$ for all $j_1 \in N_{i,1}$ are observable to i then s_{j_1}'s are accurately deduced, otherwise $\mu_{j_1,0}$ is simply i's ex-ante belief about the prior distribution of j_1.

Similarly, for all $j_1 \in N_{i,1}$ and all $l_1 \in N_{j_1,1}$ we have

$$\mu_{j_1,2} = \frac{\sigma^2}{1+k_{j_1,1}+\sigma^2}\mu_{j_1,0} + \frac{1}{1+k_{j_1,1}+\sigma^2}\left(s_{j_1} + \sum_{l_1 \in N_{j_1,1}} s_{l_1}\right)$$

and $var_{j_1,2} = \frac{\sigma^2}{1+k_{j_1,1}+\sigma^2}$.

Since i knows $k_{j_1,1}$ for all $j_1 \in N_{i,1}$, hence $N_{i,2}$ and $k_{i,2}$, he can deduce each $\sum_{l_1 \in N_{j_1,1}} s_{l_1}$ from the first and second period announcements of all $j_1 \in N_{i,1}$. The posterior belief of i in the third period is thus normally distributed with mean

$$\mu_{i,3} = \frac{\sigma^2}{1+k_{i,1}+k_{i,2}+\sigma^2}\mu_{i,0} + \frac{1}{1+k_{i,1}+k_{i,2}+\sigma^2}\left(s_i + \sum_{j_1\in N_{i,1}} s_{j_1} + \sum_{j_2\in N_{i,2}} s_{j_2}\right)$$

and variance

$$var_{i,3} = \frac{\sigma^2}{1+k_{i,1}+k_{i,2}+\sigma^2}$$

For a finite network, there exists a bound on the order of the neighborhood for all agents. That is there exists a T_i for all $i \in N$ such that for any $t > T_i$, $N_{i,t} = 0$, hence i stops updating his beliefs after $t = T_i$. It then follows that $T_m = \max\{T_i; i \in N\}$, which is also the diameter of the communication network, is the period after which learning stops for all agents. After T_m no agents has new information to learn from his neighbors. The iteration of posterior beliefs in a manner described above leads to public belief of each $i \in N$ to be normally distributed with the mean and variance of the form

$$\mu_{i,\infty} = \frac{\sigma^2}{\sum_{t=0}^{T_m} k_{i,t}+\sigma^2}\mu_{i,0} + \frac{1}{\sum_{t=0}^{T_m} k_{i,t}+\sigma^2}\sum_{t=0}^{T_m}\sum_{j_t\in N_{i,t}} s_{j_t} \qquad var_{i,\infty} = \frac{\sigma^2}{\sum_{t=0}^{T_m} k_{i,t}+\sigma^2} \tag{A.1}$$

where $s_{j_0} = s_i$ and $k_{i,0} = 1$, and that $\sum_{t=0}^{T_m} k_{i,t} = n$, $\quad \sum_{t=0}^{T_m}\sum_{j_t\in N_{i,t}} s_{j_t} = \sum_{j=1}^{n} s_j$.

(ii) Let agents' prior beliefs be uncorrelated and that $\boldsymbol{\mu}_0 \sim \mathcal{N}(\boldsymbol{\nu}, \eta^2\mathbf{I})$, where $\boldsymbol{\mu}_0$ is a column vector of all $\mu_{i,0}$ and $\mathbf{I}$ is an $n \times n$ identity matrix. Assume also that agents know the distribution of their neighbors prior beliefs but do not observe the realized value. Let $\mathbb{E}_i[s_k|\nu_j]$ for each $j \in N_{i,1}$ denote the expected signal of k according to i given the distribution of his neighbor j's prior belief. It follows from Lemma 1 that, after observing his neighbors first period announcements, i deduces the expected signals of each $j \in N_{i,1}$ to be

$$\begin{aligned}\mathbb{E}_i[s_j|\nu_j] &= \left(1+\sigma^2\right)\mu_{j,1} - \sigma^2\nu_j\\ &= \sigma^2\left(\mu_{j,0}-\nu_j\right) + s_j\end{aligned} \tag{A.2}$$

where the second equality result from substituting for $\mu_{j,1}$. The corresponding variance of the expected signal according to i is $\mathrm{var}[\mathbb{E}_i[s_j|\nu_j]] = \sigma^4\eta^2 + \sigma^2$.

After incorporating the expected signals from his neighbors, each $i \in N$ updates his beliefs to a normal distribution with mean and variance

$$\mu_{i,2} = \frac{(1+\sigma^2)(1+\eta^2\sigma^2)}{k_i + (1+\sigma^2)(1+\eta^2\sigma^2)}\mu_{i,1} + \frac{1}{k_i+(1+\sigma^2)(1+\eta^2\sigma^2)}\sum_{j\in N_{i,1}}\mathbb{E}_i[s_j|\nu_j]$$

and

$$var_{i,2} = \frac{\sigma^2(1+\eta^2\sigma^2)}{k_i+(1+\sigma^2)(1+\eta^2\sigma^2)}$$

Let l be the index for the neighbors of $j \in N_{i,1}$. We maintain the assumption that agents have memory of past announcements of their neighbors and that the network is common knowledge. From the second period announcements of his neighbors, i deduces the new information from each $j \in N_{i,1}$ to be

$$\sum_{l\in N_{j,1}}\mathbb{E}_i[s_l|\nu_j] = \left(k_j + \left(1+\sigma^2\right)\left(1+\eta^2\sigma^2\right)\right)\mu_{j,2} - \left(1+\sigma^2\right)\left(1+\eta^2\sigma^2\right)\mu_{j,1} \tag{A.3}$$

where each $\mathbb{E}_i[s_l|\nu_j]$ is of the form $\mathbb{E}_i[s_l|\nu_j] = \sigma^2(\mu_{l,0} - \nu_l) + s_l$. At the end of the third period, i updates his beliefs to a normal distribution with mean and variance

$$\mu_{i,3} = \frac{(1+\sigma^2)(1+\eta^2\sigma^2)}{k_i + k_{i,2} + (1+\sigma^2)(1+\eta^2\sigma^2)}\mu_{i,1}$$

$$(A.4) \qquad + \frac{1}{k_i + k_{i,2} + (1+\sigma^2)(1+\eta^2\sigma^2)}\left(\sum_{j\in N_{i,1}} \mathbb{E}_i[s_j|\nu_j] + \sum_{j\in N_{i,1}}\sum_{l\in N_{j,1}} \mathbb{E}_i[s_l|\nu_j]\right)$$

and

$$var_{i,3} = \frac{\sigma^2(1+\eta^2\sigma^2)}{k_i + \sum_{j\in N_{i,1}} k_j + (1+\sigma^2)(1+\eta^2\sigma^2)}$$

By iterating the posterior beliefs until the end of the learning process and noting that $\sum_{t=0}^{T_m} k_{i,t} = n$, the public belief of each $i \in N$ is then normally distributed with mean and variance

$$\mu_{i,\infty} = \frac{(1+\sigma^2)(1+\eta^2\sigma^2)}{(1+\sigma^2)(1+\eta^2\sigma^2)+n-1}\mu_{i,1} + \frac{1}{(1+\sigma^2)(1+\eta^2\sigma^2)+n-1}\sum_{l\in N\setminus\{i\}} \mathbb{E}_i[s_l|\boldsymbol{\nu}]$$

and

$$var_{i,\infty} = \frac{\sigma^2(1+\eta^2\sigma^2)}{(1+\sigma^2)(1+\eta^2\sigma^2)+n-1}$$

(iii) In Proposition 1 (*iii*) we assume that $\boldsymbol{\mu}_0 \sim \mathcal{N}(\boldsymbol{\nu}, M)$, where each element of M, $m_{ij} > 0$ for all $(i,j) \in N$.

The following notations will be used in the proof. We write $\boldsymbol{\mu}$ and $\boldsymbol{\mu}_{-i}$ for the column vectors of all μ_i for all $i \in N$ and for all $i \in N \setminus \{i\}$ respectively. Similarly, $\boldsymbol{\mu}$ and $\boldsymbol{\nu}_{-i}$ denote the column vectors of all ν_i for all $i \in N$ and for all $i \in N \setminus \{i\}$ respectively. $M_{-i,-i}$ denotes and $n-1 \times n-1$ variance-covariance matrix of all agents excluding i, and $M_{-i,i}$ is the ith column of M with the ith row excluded. We also denote by $\mathbf{1}_{k\times l}$ for a $k \times l$ dimensional matrix of ones, and by $\mathbf{I}$ the identity matrix.

We employ the following well known concepts for normally distributed random variables and specifically adopted to the distributions of prior beliefs of agents. Let $\boldsymbol{\mu}_0$ be the column vector of prior beliefs of all agents, $\mu_{i,0}$ $\forall i \in N$ and $\boldsymbol{\mu}_{-i,0}$ $\forall i \in N \setminus \{i\}$. If $\boldsymbol{\mu}_0 \sim \mathcal{N}(\boldsymbol{\nu}, M)$, then

$$(A.5) \qquad \mathbb{E}\left[\boldsymbol{\mu}_{-i,0}|\mu_{i,0}\right] = \boldsymbol{\nu}_{-i} + m_{ii}^{-1}M_{-i,i}(\mu_{i,0} - \nu_i)$$

$$(A.6) \qquad \text{var}\left[\boldsymbol{\mu}_{-i,0}|\mu_{i,0}\right] = M_{-i,-i} - m_{ii}^{-1}M_{-i,i}M'_{-i,i}$$

When agents priors are correlated, the expected signal of each $j \in N_{i,1}$ according to i given the realization $\mu_{i,0}$ and j's first period announcement $\mu_{j,1}$ is given by

$$(A.7) \qquad \mathbb{E}_i\left[s_j|\mu_{i,0}, \mu_{j,1}\right] = (1+\sigma^2)\mu_{j,1} - \sigma^2\mathbb{E}_i\left[\mu_{j,0}|\mu_{i,0}\right]$$

In a similar manner, we can write the joint expected signals for all $j \in N_{i,1}$ according to an agent i who observes the announcements of all other agents as follows

$$(A.8) \qquad \mathbb{E}_i\left[\mathbf{s}_{-i}|\mu_{i,0}, \boldsymbol{\mu}_{-i,1}\right] = (1+\sigma^2)\boldsymbol{\mu}_{-i,1} - \sigma^2\mathbb{E}_i\left[\boldsymbol{\mu}_{-i,0}|\mu_{i,0}\right]$$

which by substituting for $\boldsymbol{\mu}_{-i,1}$ is equivalent to

$$(A.9) \qquad \mathbb{E}_i\left[\mathbf{s}_{-i}|\mu_{i,0}, \boldsymbol{\mu}_{-i,1}\right] = \sigma^2\left(\boldsymbol{\mu}_{-i,0} - \mathbb{E}_i\left[\boldsymbol{\mu}_{-i,0}|\mu_{i,0}\right]\right) + \mathbf{s}_{-i}$$

The variance associated with the expected signal is then

$$\begin{aligned}\text{var}\left[\mathbf{s}_{-i}|\mu_{i,0},\boldsymbol{\mu}_{-i,1}\right] &= \sigma^4\text{var}\left[\boldsymbol{\mu}_{-i,0}|\mu_{i,0}\right]+\sigma^2\mathbf{I}\\ &= \sigma^4\left(M_{-i,-i}-m_{ii}^{-1}M_{-i,i}M'_{-i,i}\right)+\sigma^2\mathbf{I}\end{aligned} \tag{A.10}$$

We now derive the coefficient of $\mathbb{E}_i\left[\mathbf{s}_{-i}|\mu_{i,0},\boldsymbol{\mu}_{-i,1}\right]$ in the second period announcement of i. Note that the variance associated with $\mu_{i,1}$ is $\frac{\sigma^2}{1+\sigma^2}$, and together with var $\left[\mathbf{s}_{-i}|\mu_{i,0},\boldsymbol{\mu}_{-i,1}\right]$ we can define an $n-1\times n-1$ matrix C as follows

$$\begin{aligned}C &= \frac{\mathbf{1}_{n-1\times n-1}}{\left(\frac{\sigma^2}{1+\sigma^2}\right)\mathbf{1}_{n-1\times n-1}+\text{var}\left[\mathbf{s}_{-i}|\mu_{i,0},\boldsymbol{\mu}_{-i,1}\right]}\\ &= \frac{(1+\sigma^2)\mathbf{1}_{n-1\times n-1}}{\sigma^2\mathbf{1}_{n-1\times n-1}+\sigma^2(1+\sigma^2)\left(\sigma^2\left(M_{-i,-i}-m_{ii}^{-1}M_{-i,i}M'_{-i,i}\right)+\mathbf{I}\right)}\end{aligned} \tag{A.11}$$

Let C_i denote the ith row of C, that is

$$C_i=(1+\sigma^2)\mathbf{1}_{1\times n-1}\left[\sigma^2\mathbf{1}_{n-1\times n-1}+\sigma^2(1+\sigma^2)\left(\sigma^2\left(M_{-i,-i}-m_{ii}^{-1}M_{-i,i}M'_{-i,i}\right)+\mathbf{I}\right)\right]^{-1}$$

The elements of C_i are the coefficients associated with the expected signals of all $j\in N\setminus\{i\}$ in the second period announcement of i. We thus have

$$\begin{aligned}\mu_{i,2} &= \left(1-C_i\mathbf{1}_{1\times n-1}\right)\mu_{i,1}+C_i\mathbb{E}_i\left[\mathbf{s}_{-i}|\mu_{i,0},\boldsymbol{\mu}_{-i,1}\right]\\ &= \left(1-C_i\mathbf{1}_{1\times n-1}\right)\mu_{i,1}+(1+\sigma^2)C_i\boldsymbol{\mu}_{-i,1}-\sigma^2C_i\boldsymbol{\nu}_{-i}-\sigma^2m_{ii}^{-1}C_iM_{-i,i}\left(\mu_{i,0}-\nu_i\right)\end{aligned} \tag{A.12}$$

From the second period announcement of i, all agents who observe i's announcements know that

$$\mu_{i,0}=\nu_i+\frac{\left(1-C_i\mathbf{1}_{1\times n-1}\right)\mu_{i,1}+(1+\sigma^2)C_i\boldsymbol{\mu}_{-i,1}-\sigma^2C_i\boldsymbol{\nu}_{-i}-\mu_{i,2}}{\sigma^2m_{ii}^{-1}C_iM_{-i,i}} \tag{A.13}$$

From (A.13), each agent who communicates to i and i communicates to, will correctly deduce $\mu_{i,0}$ at the end of the second period announcements if and only if that agent also observes the announcements of all other agents that communicate to i and i communicates to. This follows from the fact that for any $j\in N_{i,1}$ to correctly deduce $\mu_{i,0}$, then j must also observe all $\mu_{k,1}\in\boldsymbol{\mu}_{-i,1}$ for $k\neq j$. The conclusion is that the communication network must be complete if agents are to correctly deduce the neighbors' realized prior beliefs and hence private information. That is, from the second period announcements, all agents (who observes i's announcements) deduces that

$$s_i=(1+\sigma^2)\mu_{i,1}-\sigma^2\left(\nu_i+\frac{\left(1-C_i\mathbf{1}_{1\times n-1}\right)\mu_{i,1}+(1+\sigma^2)C_i\boldsymbol{\mu}_{-i,1}-\sigma^2C_i\boldsymbol{\nu}_{-i}-\mu_{i,2}}{\sigma^2m_{ii}^{-1}C_iM_{-i,i}}\right) \tag{A.14}$$

Since the network is complete, the third period announcement of each $i\in N$ a normal distribution with mean

$$\mu_{i,\infty}(n)=\frac{\sigma^2}{n+\sigma^2}\mu_{i,0}+\frac{1}{n+\sigma^2}\sum_{j=0}^{n}s_j \tag{A.15}$$

A.2. *Proof of Proposition 2*

To proof the proposition, we first define the notion *coefficient of ergodicity* as a measure for ergodicity (Seneta, 1979).

Definition 8: *Given a matrix G with entries g_{ij}, the coefficient of ergodicity $\rho(G)$ defined on the L^1-norm of G is*

$$\rho(G) = \frac{1}{2}\left\{\max_{i,j}\sum_{k=1}^{n}(|g_{ik} - g_{jk}|)\right\}, \tag{A.16}$$

with the following properties;

(i) $0 \leq \rho(G) \leq 1$.

(ii) For two matrices G_1 and G_2, $\rho(G_1G_2) \leq \rho(G_2)\rho(G_2)$.

(iii) $\rho(G) = 0$ if and only if rank(G)=1; that is $G = \mathbf{ev}'$

Let $G^T = \prod_{t=0}^{T} G_{\gamma(t)}$, then property (ii) of Definition 2 implies that

$$\rho(G^T) \leq \rho(G_{\gamma(T)})\rho(G_{\gamma(T-1)})\cdots\rho(G_{\gamma(0)}) \tag{A.17}$$

Property (iii) together with Lemma 2 imply that ergodic Markov chains are those in which

$$\lim_{T\to\infty}\rho(G^T) = 0 \tag{A.18}$$

We can thus deduce from properties (ii) and (iii) of Definition 2, specially equations (A.17) and (A.18) that a chain is ergodic if there exists a $\tau \geq 0$ such that for every time interval $[t, t+\tau)$ there exists a sequence of matrices $\{G_{\gamma(t)}, \cdots, G_{\gamma(t+\tau)}\}$, for which the coefficient of ergodicity of their product is less than unity. That is $\rho(G_{\gamma(t)}\cdots G_{\gamma(t+\tau)}) < 1$. Matrices or a product of matrices for which the coefficient of ergodicity is less than unity are known as *scrambling matrices*. The question we then ask is, what switching strategy induces a sequence of networks whose product on intervals of time result into scrambling matrices? The following lemmas establish the necessary conditions.

Lemma 3: *Let G_1 and G_2 be two network induced transition matrices, and that both matrices are aperiodic. If G_1 is connected, then the products G_1G_2 and G_2G_1 are also connected and aperiodic.*

Proof. Write G_2in the form $G_2 = \text{Diag}(G_2) + G_2'$, where $\text{Diag}(G_2)$ is a diagonal matrix whose elements are the diagonal of G_2, and G_2' is the residual $G_2 - \text{Diag}(G_2)$. Since G_2 is non-negative then so is G_2'. We thus have that

$$G_2G_1 = \text{Diag}(G_2)G_1 + G_2'G_1$$

Denote by g_{ii}^2 and g_{ij}^1 for $1 \leq i \leq n$ as the elements of $\text{Diag}(G_2)$ and G_1 respectively. Then

$$\text{Diag}(G_2)G_1 = \begin{pmatrix} g_{11}^2 g_{11}^1 & \cdots & g_{11}^2 g_{1n}^1 \\ \vdots & & \vdots \\ g_{nn}^2 g_{n1}^1 & \cdots & g_{nn}^2 g_{nn}^1 \end{pmatrix}$$

Clearly if G_1 is connected and aperiodic, then so is $\text{Diag}(G_2)G_1$. The matrix product is non-negative since G_1 is also non-negative, such that when added to $\text{Diag}(G_2)G_1$ the properties of connectedness aperiodicity are preserved. A similar argument follows for G_1G_2. □

Lemma 4: *Let $\{G_{\gamma(t)}\}_{t\geq 0}$ be a sequence of connected aperiodic transition matrices, then there exists a sufficiently large $\tau > 0$ such that*

$$G^{\tau} = G_{\gamma(t)} \cdots G_{\gamma(t+\tau)}$$

is a scrambling matrix.

Proof. The proof follows from the fact that a network that is connected and aperiodic induces a primitive transition matrix. That is let G be such a matrix, then there exists a sufficiently large t such that G^t is a positive matrix. Since positive stochastic matrices are scrambling, it follows that the product of a sequence of connected aperiodic matrices is scrambling. □

The following lemma then directly follows from the above two

Lemma 5: *Let $T_2 < T_1 < T$ be sufficiently large integers. If γ is such that there exist a $T_2 \geq 0$ and $T_1 \geq 0$ where by for every time interval $[t, t+T_2)$ a connected communication network obtains, such that within the time interval $[t, t+T_1)$ a scrambling connected network obtains, then*

$$\lim_{T\to\infty} \prod_{t=0}^{T} G_{\gamma(t)} = \mathbf{e}\mathbf{v}'$$

The expression on the right hand side of (10) then follows directly from Lemma 5

A.3. Proof of Theorem 1

The general assumption in all the proofs below is that the sequence $\{G(n)\}_{n\geq 2}$ is common knowledge

(i) There are two parts to the proof of Theorem 1 (i); when the prior beliefs are observable and when they are not. We first prove for the case in which the realized prior beliefs are observable, that is for all $i \in N$, each $\mu_{j,0}$ for all $j \in N_{i,1}$ is observable to i. Under this assumption, the public belief of each i is normally distributed with mean

(A.19)
$$\mu_{i,\infty}(n) = \frac{\sigma^2}{n+\sigma^2}\mu_{i,0} + \frac{1}{n+\sigma^2}\sum_{j=0}^{n} s_j$$

The variance of $\mu_{i,\infty}(n)$ is then

$$\text{var}[\mu_{i,\infty}(n)] = \frac{\sigma^4\eta^2}{(n+\sigma^2)^2} + \frac{n\sigma^2}{(n+\sigma^2)^2}$$

From Chernoff bound, it follows that

(A.20)
$$\lim_{n\to\infty} \mathbb{P}\left(|\mu_{i,\infty}(n) - \bar{\mu}| > \epsilon\right) \leq \lim_{n\to\infty} \left(\frac{\text{var}[\mu_{i,\infty}(n)]}{\epsilon}\right) = \lim_{n\to\infty} \left(\frac{\sigma^4\eta^2}{(n+\sigma^2)^2} + \frac{n\sigma^2}{(n+\sigma^2)^2}\right) = 0$$

It is easy to see that the right hand side of (A.20) will be zero if and only if both σ and η are finite such that $\sigma^4\eta^2 < \infty$. This conditions applies to all $i \in N$

The limit of the variance of X is

$$\lim_{n\to\infty} var_{i,\infty}(n) = \lim_{n\to\infty} \frac{\sigma^2}{n+\sigma^2} = 0.$$

Now we prove for the case in which the realized priors are unobservable. From Proposition 1, if for each $i \in N$, $\mu_{j,0}$ for all $j \in N_{i,1}$ are unobservable but it is common knowledge that $\boldsymbol{\mu}_0 \sim \mathcal{N}(\boldsymbol{\nu}, \eta^2 \mathbf{I})$, where $\boldsymbol{\mu}_0$ is a column vector of all $\mu_{i,0}$, then the public belief of each i is normally distributed with mean

(A.21)

$$\mu_{i,\infty}(n) = \frac{(1+\sigma^2)(1+\eta^2\sigma^2)}{(1+\sigma^2)(1+\eta^2\sigma^2)+n-1}\mu_{i,1} + \frac{1}{(1+\sigma^2)(1+\eta^2\sigma^2)+n-1}\sum_{l\in N\setminus\{i\}} \mathbb{E}_i[s_l|\boldsymbol{\nu}]$$

For the sake of notational convenience, let $\varphi = (1+\sigma^2)(1+\eta^2\sigma^2)$, and recall that $\mathbb{E}_i[s_l|\boldsymbol{\nu}] = \sigma^2(\mu_{l,0}-\nu_l)+s_l$ for all $i \in N$ and all $l \in N \setminus \{i\}$. We can then rewrite (A.21) as

$$\text{(A.22)} \qquad \mu_{i,\infty}(n) = \frac{\varphi}{\varphi+n-1}\mu_{i,1} + \frac{\sigma^2}{\varphi+n-1}\sum_{l\in N\setminus\{i\}}(\mu_{l,0}-\nu_l) + \frac{1}{\varphi+n-1}\sum_{l\in N\setminus\{i\}} s_l$$

The variance associated with $\mu_{i,\infty}(n)$ in (A.22) is

$$\text{(A.23)} \qquad \text{var}[\mu_{i,\infty}(n)] = \frac{\varphi^2\sigma^4\eta^2+\varphi^2\sigma^2}{(\varphi+n-1)^2(1+\sigma^2)^2} + \frac{\sigma^4\eta^2}{(\varphi+n-1)^2} + \frac{(n-1)\sigma^2}{(\varphi+n-1)^2}$$

Similarly from (A.23), the $\lim_{n\to\infty} \text{var}[\mu_{i,\infty}(n)] = 0$ if and only if η and σ are finite. Note that the sequence of random variables $\{(\mu_{l,0}-\nu_l)\}_{l\in N\setminus\{i\}}$ is of mean zero and variance η^2. This implies that if each term in the sum on the right hand side of (A.22) converges in probability, then the first and the second term both converge to zero and the third term converges to $\bar{\mu}$. It then follows from (A.20) that $\mu_{i,\infty}(n) \xrightarrow{p} \bar{\mu}$.

If η and σ are finite, then limit of the variance of X when priors beliefs are not observable is

$$\lim_{n\to\infty} \frac{\sigma^2(1+\eta^2\sigma^2)}{(1+\sigma^2)(1+\eta^2\sigma^2)+n-1} = 0.$$

(ii) From (A.20) it follows that when the conditions of Theorem 1 *(ii)* are satisfied, $\mu_{i,\infty}(n) \xrightarrow{p} \bar{\mu}$.

A.4. Proof of Theorem 2

(i) Consider the case in which $G(n)$ is doubly stochastic such that $\sum_{j=1}^n g_{ij}(n) = \sum_{i=1}^n g_{ij}(n)$. Let $\mathbf{z}$ and $\mathbf{y}$ be the left and right eigenvalues of $G(n)$ associated with the leading eigenvalue. If $G(n)$ is doubly stochastic then $\mathbf{z} = \mathbf{y}$, such that

$$\mathbf{z}'G(n) = \mathbf{z}' \quad \text{and} \quad G(n)\mathbf{z} = \mathbf{z}$$

Without loss of generality we can assume that $\|\mathbf{z}\| = \mathbf{z}'\mathbf{z} = 1$. Specifically, $\mathbf{z} = \frac{1}{\sqrt{N}}\mathbf{e}$, where $\mathbf{e}$ is a length n vector of ones. From Lemma 2 it follows that $\mathbf{v} = \frac{1}{n}\mathbf{e}$, such that for each $i \in N$

$$\mu_{i,\infty}(n) = \frac{1}{n}\sum_{i=1}^n \mu_{i,1}$$

(ii) Recall that

$$\mu_{i,\infty}(n) = \sum_{j=1}^{n} v_j(n)\left(\frac{\sigma^2}{1+\sigma^2}\mu_{j,0} + \frac{1}{1+\sigma^2}s_j\right)$$

$$= \frac{1}{1+\sigma^2}\bar{\mu} + \frac{\sigma^2}{1+\sigma^2}\sum_{j=1}^{n} v_j(n)\mu_{j,0} + \frac{\sigma^2}{1+\sigma^2}\sum_{j=1}^{n} v_j(n)\varepsilon_j \tag{A.24}$$

where the second equality follows from the fact that $s_j = \bar{\mu} + \varepsilon_j$ and that $\sum_{j=1}^{n} v_j(n) = 1$. Let the summation components on the right hand side of A.24 be denoted as follows

$$V_\mu(n) = \sum_{j=1}^{n} v_j(n)\mu_{j,0}, \quad \text{and} \quad V_\varepsilon(n) = \sum_{j=1}^{n} v_j(n)\varepsilon_j.$$

Since $\bar{\mu}$ is a parameter, it follows that any variance in $\mu_{i,\infty}(n)$ is due to the variances in $V_\mu(n)$ and $V_\varepsilon(n)$. This in turn implies that for $\mu_{i,\infty}(n)$ to converge in probability, then both $V_\mu(n)$ and $V_\varepsilon(n)$ must also converge in probability to their respective limits. Consider first the in which the communication network is balanced, in which case (A.24) becomes.

$$\mu_{i,\infty}(n) = \frac{1}{1+\sigma^2}\bar{\mu} + \frac{\sigma^2}{1+\sigma^2}\frac{1}{n}\sum_{j=1}^{n} \mu_{j,0} + \frac{\sigma^2}{1+\sigma^2}\frac{1}{n}\sum_{j=1}^{n} \varepsilon_j \tag{A.25}$$

This implies that both $V_\mu(n)$ and $V_\varepsilon(n)$ are simply sample averages of the random variables $\mu_{i,0}$ and ε_i of all $i \in N$ respectively. It then follows from the law of large numbers that $V_\varepsilon(n)$ converge in probability to zero, and $V_\mu(n)$ converges in probability to $\bar{\mu}$ only if $\boldsymbol{\mu}_0 \sim \mathcal{N}(\bar{\boldsymbol{\mu}}, \eta^2\mathbf{I})$. This in turn implies that $\mu_{i,\infty}(n)$ also converges in probability to $\bar{\mu}$. That is

$$\mu_{i,\infty}(\infty) = \frac{1}{1+\sigma^2}\bar{\mu} + \frac{\sigma^2}{1+\sigma^2}\bar{\mu} = \bar{\mu} \tag{A.26}$$

Generally speaking, both $V_\mu(n)$ and $V_\varepsilon(n)$ are sample weighted averages of n random variable $\mu_{i,0}$ and ε_i respectively drawn with probabilities $v_i(n)$ for each $i \in N$, from normal distributions with means $\bar{\mu}$ and zero variance. These sample means converge in probability if and only if their variances converge to zero for large n. That is

$$\lim_{n\to\infty} \text{var}[V_\mu(n)] = 0 \quad \text{and} \quad \lim_{n\to\infty} \text{var}[V_\varepsilon(n)] = 0$$

The respective variances are

$$\text{var}[V_\mu(n)] = \eta^2 \sum_{j=1}^{n} v_j^2(n) \quad \text{and} \quad \text{var}[V_\varepsilon(n)] = \sigma^2 \sum_{j=1}^{n} v_j^2(n),$$

which converge to zero if and only if the summand $\sum_{j=1}^{n} v_j^2(n)$ converges to zero with n. This sum converges to zero if each square $v_j^2(n)$ converges to zero with large n. In fact it is enough to say that $\max_{i\in N}(v_i(n)) \to 0$ for large n. This condition is satisfied by asymptotically balanced networks in that if $G(n)$ is asymptotically balanced, then $G(n)$ converges to doubly stochastic matrix S. Where for any doubly stochastic matrix, the elements of the corresponding weight vector $v_i(n) = \frac{1}{n}$, converge to zero for large n.

In the case of the variance of public belief, it follows directly that

$$\begin{aligned} var_{i,\infty}(n) &= \sum_{j=1}^{n} v_j(n)\frac{\sigma}{1+\sigma} \\ &= \frac{\sigma}{1+\sigma}\sum_{j=1}^{n} v_j(n) = \frac{\sigma}{1+\sigma} \end{aligned} \tag{A.27}$$

A.5. *Proof of Proposition 3*

Given $G(n)$, the closest doubly stochastic matrix $S(n)$ to $G(n)$ is that which minimizes the quantity $\|S(n) - G(n)\|_F^2$. Our objective is then as follows:

$$\begin{array}{lll} \min_{S(n)\in\mathcal{S}(n)} & \|S(n) - G(n)\|_F^2 & \\ \text{subject to} & S(n)\mathbf{z} = \mathbf{z} & \mathbf{z}'S(n) = \mathbf{z}' \end{array}$$

where $\mathcal{S}(n)$ is the set of all doubly stochastic matrices of size n, and $\mathbf{z}$ is the left and right eigenvectors corresponding to the first (unit) eigenvalue of $S(n)$. We assume without loss of generality that $\|\mathbf{z}\| = \mathbf{z}'\mathbf{z} = 1$. The elements of $G(n)$ are denoted by $g_{ij}(n)$ and those for $S(n)$ by s_{ij}. It is easy to see that the objective function to be minimized is of the form

$$f(G(n)) = (s_{11} - g_{11})^2 + \cdots + (s_{1n} - g_{1n})^2 + \cdots + (s_{n1} - g_{n1})^2 + \cdots + (s_{nn} - g_{nn})^2$$

Note that the constraints together make a total of $2n$ linear equations, which leads to $2n$ Lagrange multipliers denoted (in vector form) by $2\alpha = (2\alpha_1, \cdots, 2\alpha_n)$ and $2\beta = (2\beta_1, \cdots, 2\beta_n)$. Note also that the factor of 2 in α and β is to account for the factor of 2 in the derivative of $f(G(n))$ below. The Kuhn-Tucker conditions yield the following matrix equation.

$$(S(n) - G(n)) + \alpha\mathbf{z}' + \mathbf{z}\beta' = 0 \tag{A.28}$$

By substituting for $S(n) = G(n) - \alpha\mathbf{z}' - \mathbf{z}\beta'$ into $S(n)\mathbf{z} = \mathbf{z}$ and $\mathbf{z}'S(n) = \mathbf{z}'$, together with the assumption that $\mathbf{z}'\mathbf{z} = 1$, one obtains the following set of simultaneous equations.

$$G(n)\mathbf{z} - \alpha - \mathbf{z}\beta'\mathbf{z} = \mathbf{z} \tag{A.29}$$

$$\mathbf{z}'G(n) - \mathbf{z}'\alpha\mathbf{z}' - \beta' = \mathbf{z}', \tag{A.30}$$

which can be written in matrix form as

$$\begin{bmatrix} I & \mathbf{z}\mathbf{z}' \\ \mathbf{z}\mathbf{z}' & I \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} G(n)\mathbf{z} - \mathbf{z} \\ G(n)'\mathbf{z} - \mathbf{z} \end{bmatrix} \tag{A.31}$$

We can then solve for α and β by multiplying both sides of (A.31) by the inverse of $\begin{bmatrix} I & \mathbf{z}\mathbf{z}' \\ \mathbf{z}\mathbf{z}' & I \end{bmatrix}$. That is

$$\begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \frac{1}{1 - \mathbf{z}\mathbf{z}'} \begin{bmatrix} I & -\mathbf{z}\mathbf{z}' \\ -\mathbf{z}\mathbf{z}' & I \end{bmatrix} \begin{bmatrix} G(n)\mathbf{z} - \mathbf{z} \\ G'(n)\mathbf{z} - \mathbf{z} \end{bmatrix} \tag{A.32}$$

which yields,

$$\alpha = G(n)\mathbf{z} - \mathbf{z} \quad \text{ and } \beta = G'(n)\mathbf{z} - \mathbf{z}$$

The assumption that $\mathbf{z}'\mathbf{z} = 1$ also implies that $\mathbf{z} = \frac{1}{\sqrt{n}}\mathbf{e}$. Substituting for α and β in (A.28) and noting that $G(n)\mathbf{e} = \mathbf{e}$, we obtain

$$S(n) = G(n) + \frac{1}{n}\mathbf{e}\mathbf{e}' - \frac{1}{n}\mathbf{e}\mathbf{e}'G(n) \tag{A.33}$$

It then follows that

$$\begin{aligned} \phi(n)^2 = \|S(n) - G(n)\|_F^2 &= \frac{1}{n^2}\|\mathbf{e}\mathbf{e}' - \mathbf{e}\mathbf{e}'G(n)\|_F^2 \\ &= \frac{1}{n}\sum_{j=1}^{n}\left(1 - \sum_{i=1}^{n} g_{ij}(n)\right)^2 . \end{aligned} \tag{A.34}$$

A.6. Proof of Theorem 3

The proof follows from Proposition 3, where a network is said to be asymptotically balanced if

$$\lim_{n\to\infty} \phi(n) = 0.$$

For Theorem 3 (i), we have that

$$\begin{aligned} \phi^2(n) &= \frac{1}{n}\sum_{j=1}^{n}\left(1 - \sum_{i=1}^{n} g_{ij}(n)\right)^2 \\ &= \frac{1}{n}\sum_{j=1}^{n}\left(1 - 2\sum_{i=1}^{n} g_{ij}(n) + \left(\sum_{i=1}^{n} g_{ij}(n)\right)^2\right) \\ &= \frac{1}{n}\left(n - 2\sum_{i=1}^{n}\sum_{j=1}^{n} g_{ij}(n) + \sum_{j=1}^{n}\left(\sum_{i=1}^{n} g_{ij}(n)\right)^2\right) \\ &\leq \frac{1}{n}\left(d_{max}^{in}(n)\left(\sum_{i=1}^{n}\sum_{j=1}^{n} g_{ij}(n)\right) - n\right) \\ &= d_{max}^{in}(n) - 1 \end{aligned} \tag{A.35}$$

such that $\lim_{n\to\infty} \phi(n) = 0$ if $\lim_{n\to\infty} d_{max}^{in}(n) = 1$.

For the case of Theorem 3 (ii), since for $n > n'$ the quantity $D(n)$ is constant say c, then for all $n > n'$

$$\phi^2(n) = \frac{1}{n}c$$

In which case $\lim_{n\to\infty} \frac{1}{n}c = 0$.

A.7. Proof of Proposition 4

Under Bayesian rational learning, the proof follows directly from the fact that learning stops after the two agents that form the longest geodesic (diameter of the network) have communicated their private informations (see proof of Proposition 1). The time it takes private beliefs to become public is thus $t_c = D(G_{\gamma_c})$, in which case the convergence rate is basically $\frac{1}{t_c}$. The relationship between the diameter and the second largest eigenvalue of a graph is a well studied concept in graph theory (e.g Chung (1989)). Generally it assumes the form,

$$D(G_{\gamma_c}) \leq \frac{\ln(c)}{\ln(1/\lambda_2(G_{\gamma_c}))}$$

where c is some constant.

In the case of Bayesian bounded-rational learning mechanism, it is well known that a homogeneous Markov chain with transition matrix G_{γ_c} whose second largest eigenvalue is $\lambda_2(G_{\gamma_c})$ converges at the equal to $\lambda_2(G_{\gamma_c})$. We write λ_i for $\lambda_i(G_{\gamma_c})$ for the sake of notational cumbersomeness. Let $\mathbf{z_i}$ and $\mathbf{y_i}$ be the right and left eigenvectors associated with the eigenvalue λ_i, and let λ_i in ordered as $\lambda_1 > \lambda_2 \geq \cdots$. Then the above argument follows from the eigendecomposition of G_{γ_c} (assuming that G_{γ_c} is connected such that $\lambda_2(G_{\gamma_c}) < 1$, and that G_{γ_c} is actually eigendecomposable), where the convergence rate is defined precisely as,

$$\begin{aligned} r_{BBR} &= \lim_{t\to\infty} \|G^t_{\gamma_c}\mu_\mathbf{1} - \bar{\mu}\|^{\frac{1}{t}} \\ &= \lim_{t\to\infty} \| \left(\mathbf{z_1}\mathbf{y_1}'\mu_1 - \bar{\mu}\right) + \sum_{i=2}^{n} \lambda_i^t \mathbf{z_i}\mathbf{y_i}'\mu_\mathbf{1}\|^{\frac{1}{t}} \\ &= |\lambda_2| \lim_{t\to\infty} \|\mathbf{z_2}\mathbf{y_2}'\mu_\mathbf{1} + \sum_{i=3}^{n} \left(\frac{\lambda_i}{\lambda_2}\right)^t \mathbf{z_i}\mathbf{y_i}'\mu_\mathbf{1}\|^{\frac{1}{t}} = |\lambda_2| \end{aligned} \tag{A.36}$$

A.8. Proof of Proposition 5

First consider the case of Bayesian rational learning mechanism where

$$U_{it}(a^*_{it}(n), X) = -\frac{1}{\delta^t}\mathbb{E}_{i,t}\left[(\mu_{i,t}(n) - X)^2\right] = -\frac{1}{\delta^t} var_{i,t}(n) \tag{A.37}$$

For a given δ and n, $U_{it}(a^*_{it}(n), X)$ is a non-monotone function of time. The variance $var_{i,t}(n)$ a generally non-increasing function of time for a given n, and $\frac{1}{\delta^t}$ is an increasing function of time. Note that the time until the learning process ends under Bayesian ration learning is equal to the diameter $D(G(n))$ of the network. This implies that for a given δ and n, there exists a positive integer $t_{i,e}$ such that at $t = t_{i,e} \leq D(G(n))$, $U_{it}(a^*_{it}(n), X)$ is minimum. The time $t = t_{i,e}$ is the optimal exit time for agent i and it generally depends on i's position in the network.

The exit time can be ordered with respect to $D(G(n))$ for groups of agents. Let $Q(G(n))$ be a set of agents in $G(n)$ whose optimal exit time is less than the diameter of the graph. That is

$$Q(G(n)) = \{i \in N : t_{i,e} < D(G(n))\}$$

and let $\#Q(G(n))$ denote the respective cardinality. Then for two networks $G_1(n)$ and $G_2(n)$ such that $G_1(n) > G_2(n)$, then it must be $Q(G_1(n)) \geq Q(G_2(n))$

Lemma 6: *Given a sequence of networks $\{G(n)\}_{n\geq 2}$ and the corresponding sequence $\{Q(G(n))\}_{n\geq 2}$, correct asymptotic learning obtains if and only if*

$$\lim_{n\to\infty} \#Q(G(n)) = 0 \tag{A.38}$$

If condition (A.38) is not fulfilled then even for a large n there exists a finite number of agents who exit the game after receiving signals only from a finite number of other agents. Lemma 6 can be proved by contradiction.

Consider the contrary case in which $\lim_{n\to\infty} Q(G(n)) = Q_e$, that is Q_e is a set of agents that exited the game after receiving the private information of a finite number of other agents. Denote by $a_{i,e}$ for the corresponding action taken by an agent $i \in Q_e$ at the end of the learning process, and by $n_{i,e}$ as the number of agents whose signals i received before exiting the game. Let $n_e = \max_{i\in Q_e}\{n_{i,e}\}$. The above definitions and notations imply that

$$a_{i,e}(n) = \lim_{n\to\infty} a_{i,t}(n) = \mu_{i,\infty}(n)$$

such that the variance of $a_{i,e}(n)$ is

$$\mathrm{var}[a_{i,e}(n)] \equiv \mathrm{var}[\mu_{i,\infty}(n)]$$

Given the sequence of networks $\{G(n)\}_{n\geq 2}$, then for an agent $i \in Q_e$ for whom $n_{i,e} = n_e$, we have

$$\begin{aligned}\lim_{n\to\infty} \mathrm{var}[a_{i,e}(n)] &= \mathrm{var}[\mu_{i,\infty}(\infty)] \\ &= \mathrm{var}\left[\frac{\sigma^2}{n_e+\sigma^2}\mu_{i,0} + \frac{1}{n_e+\sigma^2}\sum_{j=0}^{n_e} s_j\right] \\ &= \frac{\sigma^4\eta^2 + n_e\sigma^2}{(n_e+\sigma^2)^2}\end{aligned} \tag{A.39}$$

From Chernoff bound, it follows that

$$\begin{aligned}\lim_{n\to\infty}\lim_{t\to\infty} \mathbb{P}\left(|a_{i,t}(n) - \bar{\mu}| > \epsilon\right) &> \lim_{n\to\infty}\lim_{t\to\infty}\left(1 - \frac{\mathrm{var}[a_{i,t}(n)]}{\epsilon}\right) \\ &= 1 - \frac{\sigma^4\eta^2 + n_e\sigma^2}{(n_e+\sigma^2)^2} > 0\end{aligned} \tag{A.40}$$

In which case $\lim_{n\to\infty}\lim_{t\to\infty} a_{i,t}(n) \neq \bar{\mu}$. If this is true for an agent $i \in Q_e$ for whom $n_{i,e} = n_e$, and from the fact that for all $i \in Q_e$ $n_{i,e} \leq n_e$, then (A.40) must be true for all $i \in Q_e$.

Lemma 7: *Condition (A.38) in Lemma 6 is satisfied if and only if for a sequence $\{G(n)\}_{n\geq 2}$ of networks and the corresponding sequence $\{D(G(n))\}_{n\geq 2}$*

$$\lim_{n\to\infty} \frac{1}{n} D(G(n)) = 0 \tag{A.41}$$

The proof of Lemma 7 follows directly from the argument that for a sequence $\{G(n)\}_{n\geq 2}$, if by contradiction the corresponding sequence $\{D(G(n))\}_{n\geq 2}$ is such that $D(G(2)) < D(G(2)) < \cdots$, then similarly for the sequence $\{Q(G(n))\}_{n\geq 2}$ it must be that $Q(G(2)) \leq Q(G(2)) \leq \cdots$ with at least one strict inequality. In which case condition (A.38) in Lemma 6 will not be fulfilled.

In the case of bounded-rational learning, the variance of X is constant over time and n, which implies that the "discounted" expected loss is non-increasing, unlike the the case of rational learning in which it is non-monotonic in nature. Agents therefore do not have any incentive to wait until longer before taking an action $a \in A$. That is $\lim_{n\to\infty}\lim_{t\to\infty} a_{i,t}(n) = \bar{\mu}$ can occur only if the network is complete, such that agents do not have to wait for more than one or two periods before taking an action in A.

Acknowledgments

We thank Dimitri Vayanos for helpful comments, and the participants at The Royal Economic Society Postgraduate Presentation Meeting & Job Market 2014 for the fruitful discussions. This research was supported by UNU-MERIT. The usual disclaimer applies.

References

Acemoglu, D., V. Chernozhukov, and M. Yildiz (2009). Fragility of asymptotic agreement under bayesian learning. Technical report.

Acemoglu, D., M. A. Dahleh, I. Lobel, and A. Ozdaglar (2011). Bayesian learning in social networks. *Review of Economic Studies 78*(4), 1201–1236.

Bala, V. and S. Goyal (1998). Learning from neighbours. *Review of Economic Studies 65*(3), 595–621.

Banerjee, A. V. (1992). A simple model of herd behavior. *The Quarterly Journal of Economics 107*(3), 797–817.

Bikhchandani, S., D. Hirshleifer, and I. Welch (1992). A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of Political Economy 100*(5), 992–1026.

Bollobás, B. (1981). The diameter of random graphs. *Transactions of the American Mathematical Society 267*(1), 41–52.

Bonacich, P. and P. Lloyd (2001). Eigenvector-like measures of centrality for asymmetric relations. *Social Networks 23*(3), 191–201.

Chamley, C. and D. Gale (1994). Information revelation and strategic delay in a model of investment. *Econometrica 62*(5), 1065–85.

Chung, F. R. (1989). Diameters and eigenvalues. *Journal of the American Mathematical Society*, 187–196.

Cripps, M. W., J. C. Ely, G. J. Mailath, and L. Samuelson (2008). Common learning. *Econometrica 76*(4), 909–933.

Demarzo, P. M., D. Vayanos, and J. Zwiebel (2003). Persuasion bias, social influence, and unidimensional opinions. *The Quarterly Journal of Economics 118*(3), 909–968.

Dixit, A. K. and J. W. Weibull (2007). Political polarization. *Proceedings of the National Academy of Sciences 104*(18), 7351–7356.

Ellison, G. and D. Fudenberg (1995). Word-of-Mouth Communication and Social Learning. *The Quarterly Journal of Economics 110*(1), 93–125.

Fiedler, M. (1973). Algebraic connectivity of graphs. *Czechoslovak Mathematical Journal 23*(98), 298–305.

Gale, D. and S. Kariv (2003). Bayesian learning in social networks. *Games and Economic Behavior 45*(2), 329–346.

Geanakoplos, J. D. and H. M. Polemarchakis (1982). We can't disagree forever. *Journal of Economic Theory 28*(1), 192 – 200.

Golub, B. and M. O. Jackson (2010). Naïve learning in social networks and the wisdom of crowds. *American Economic Journal: Microeconomics 2*(1), 112–49.

Jackson, M. O. (2008). *Social and Economic Networks.* Princeton University Press. Princeton University Press.

König, M., C. J. Tessone, and Y. Zenou (2009). A dynamic model of network formation with strategic interactions. Technical report.

Krasucki, P. (1996). Protocols forcing consensus. *Journal of Economic Theory 70*(1), 266–272.

Mitra, P. (2009). Entrywise bounds for eigenvectors of random graphs. *The Electronic Journal of Combinatorics 16*(1), Research Paper R131, 18 p.

Mueller-Frank, M. (2013). A general framework for rational learning in social networks. *Theoretical Economics 8*(1), 1–40.

Murto, P. and J. Välimäki (2011). Learning and Information Aggregation in an Exit Game. *Review of Economic Studies 78*(4), 1426–1461.

Parikh, R. and P. Krasucki (1990). Communication, consensus, and knowledge. *Journal of Economic Theory 52*(1), 178–189.

Rosenberg, D., E. Solan, and N. Vieille (2009). Informational externalities and emergence of consensus. *Games and Economic Behavior 66*(2), 979–994.

Seneta, E. (1979). Coefficients of ergodicity: Structure and applications. *Advances in Applied Probability 11*(3), 576–590.

Sethi, R. and M. Yildiz (2012). Public disagreement. *American Economic Journal: Microeconomics 4*(3), 57–95.

Sethi, R. and M. Yildiz (2013). Perspectives, opinions, and information flows. *Working paper, Columbia University and the Santa Fe Institute.*

Smith, L. and P. Sorensen (2000). Pathological outcomes of observational learning. *Econometrica 68*(2), 371–398.

Wolfowitz, J. (1963). Products of indecomposable, aperiodic, stochastic matrices. *Proceedings of the American Mathematical Society 14*(5), 733–737.

The UNU-MERIT WORKING Paper Series

2014-01 *The medium-term effect of R&D on firm growth* by Marco Capasso, Tania Treibich and Bart Verspagen

2014-02 *Diverse and uneven pathways towards transition to low carbon development: The case of diffusion of solar PV technology in China*Michiko Iizuka

2014-03 *User innovators and their influence on innovation activities of firms in Finland* by JariKuusisto, MerviNiemi and Fred Gault

2014-04 *Migration, remittances and household welfare in Ethiopia* by Lisa Andersson

2014-05 *Path-breaking directions of nanotechnology-based chemotherapy and molecular cancer therapy* by Mario Coccia and Lili Wang

2014-06 *Poor trends - The pace of poverty reduction after the Millennium Development Agenda*Richard Bluhm, Denis de Crombrugghe, Adam Szirmai

2014-07 *Firms' adoption of international standards: Evidence from the Ethiopian floriculture sector* by MuluGebreeyesu

2014-08 *School choice, segregation, and forced school closure*by Cheng Boon Ong andKristof De Witte

2014-09 *Gender difference in support for democracy in Sub-Saharan Africa: Do social institutions matter?*by MatyKonte

2014-10 *Why are women less democratic than men? Evidence from Sub-Saharan African countries* by Cecilia García-Peñalosa and MatyKonte

2014-11 *Tipping points? Ethnic composition change in Dutch big city neighbourhoods* by Cheng Boon Ong

2014-12 *Technology life cycle and specialization patterns of latecomer countries. The case of the semiconductor industry* by Giorgio Triulzi

2014-13 *Patents as quality signals? The implications for financing constraints on R&D* by Dirk Czarnitzki, Bronwyn H. Hall and Hanna Hottenrott

2014-14 *Assessment of effectiveness of Chinese aid in competence building and financing development in Sudan* by SamiaSatti Osman Mohamed Nour

2014-15 *Education, training and skill development policies in Arab Gulf countries: Macro-micro overview* by SamiaSatti Osman Mohamed Nour

2014-16 *Structure of labour market and unemployment in Sudan* by SamiaSatti Osman Mohamed Nour

2014-17 *Overview of knowledge transfer in MENA countries - The case of Egypt* by SamiaSatti Osman Mohamed Nour

2014-18 *The impact of ICT in public and private universities in Sudan*by SamiaSatti Osman Mohamed Nour

2014-19 *End-user collaboration for process innovation in services: The role of internal resources* by Mona Ashok, Rajneesh Narula and Andrea Martinez-Noya

2014-20 *Public investment and regional politics: The case of Turkey* by Mehmet Guney Celbis, Denis de Crombrugghe and Joan Muysken

2014-21 *Infrastructure and the international export performance of Turkish regions* by Mehmet Guney Celbis, Peter Nijkamp and Jacques Poot

2014-22 *Discovering and explaining work-family strategies of parents in Luxembourg* by Nevena Zhelyazkova

2014-23 *Parental leave take up and return to work of mothers in Luxembourg: An application of the model of nested dichotomies* by Nevena Zhelyazkova

2014-24 *Millennium Development Goals: Tool or token of global social governance?* by Mueid Al Raee, Elvis Amoateng, Elvis Korku Avenyo, Youssef Beshay, Mira Bierbaum, Charlotte Keijser and Rashmi Sinha

2014-25 *One Europe or several? Causes and consequences of the European stagnation* by Jan Fagerberg and Bart Verspagen

2014-26 *The harmony of programs package: Quasi-experimental evidence on deworming and canteen interventions in rural Senegal* by Théophile Azomahou, Fatoumata Diallo and Wladimir Raymond

2014-27 *Country Terms of Trade 1960-2012: Trends, unit roots, over-differencing, endogeneity, time dummies, and heterogeneity* by Thomas Ziesemer

2014-28 *The structure and comparative advantages of China's scientific research - Quantitative and qualitative perspectives* by Lili Wang

2014-29 *Transition to knowledge-based economy in Saudi Arabia* by Samia Satti Osman Mohamed Nour

2014-30 *Challenges and opportunities for transition to knowledge-based economy in Arab Gulf countries* by Samia Satti Osman Mohamed Nour

2014-31 *Migration of international students and mobilizing skills in the MENA Region* by Samia Satti Osman Mohamed Nour

2014-32 *Beyond product innovation; improving innovation policy support for SMEs in traditional industries* by René Wintjes, David Douglas, Jon Fairburn, Hugo Hollanders and Geoffrey Pugh

2014-33 *The impact of innovation support programmes on SME innovation in traditional manufacturing industries: an evaluation for seven EU regions* by Dragana Radicic, Geoffrey Pugh, Hugo Hollanders and René Wintjes

2014-34 *Beliefs dynamics in communication networks* by Théophile T. Azomahou and Daniel C. Opolot