The Ronald O. Perelman Center for Political Science and Economics (PCPSE)
133 South 36th Street
Philadelphia, PA 19104-6297

pier@econ.upenn.edu
http://economics.sas.upenn.edu/pier

PIER Working Paper 19-007

Higher Order Information Complementarities and Polarization

CARLOS SEGURA-RODRIGUEZ
University of Pennsylvania

April 2019

Higher Order Information Complementarities and Polarization*

Carlos Segura-Rodriguez †

April 2019

Abstract

I study endogenous network formation in an environment in which individuals want to forecast a stochastic state and it is costly for them to communicate with others to exchange some exogenously observed information. Due to the existence of information complementarities, individuals' preferences for networks in which they have multiple neighbors cannot be characterized by a linear ranking of the pairwise correlations between their signals. Instead, these complementarities generate a counterintuitive result: for a fixed number of individuals, information structures exist in which all signals are conditionally positively correlated, and these are preferred to a structure in which all signals are conditionally independent. Therefore, it may be that the only strongly stable network consists of two cliques with signals that are highly positively correlated within each clique that generate different beliefs across cliques, even when there are opportunities to exchange information with individuals sharing less correlated signals. Thus, this model exemplifies how homophily and belief polarization can coexist in a rational environment.

Keywords: Information, Communication, Endogenous Networks, Homophily, Polarization.

JEL Classification: C71, D83, D85.

*I would like to express my thanks to the following people: Ashwin Kambhampati and Peng Shao for early conversations that helped me shape the question I consider in this project; Ben Golub, George Mailath, Andrew Postlewaite, Aislinn Bohren, and Annie Liang for their insightful comments and support throughout the project; and the participants at the UPenn Micro Theory Lunch and the UPenn Econometrics Lunch. Any errors are my own.

†University of Pennsylvania, Department of Economics, The Ronald O. Perelman Center, 133 South 36th Street; Philadelphia, PA 19104, cseg@sas.upenn.edu.

Contents

1 Introduction

Homophily and belief polarization are two of the most pervasive characteristics of social networks. Homophily is the tendency of individuals to relate to others with similar backgrounds in terms of characteristics such as ethnicity, gender, religion, and education,[1] while belief polarization is the fact that individuals' beliefs and conceptions about specific matters differ across groups.[2] In a frictionless environment in which individuals are completely rational and their objective is to learn, one would not expect homophily and belief polarization to arise since the agents within the various groups would communicate with those who have different beliefs from their own, and information would spread throughout the whole network (DeGroot (1974), DeMarzo, Vayanos, and Zwiebel (2003)). The literature has pointed out the distortions in an environment that might disrupt this logic, ranging from different meet-rates across and inside groups, along with the higher returns of being with individuals in the same group (Currarini, Jackson, and Pin, 2009), to psychological factors, such as feeling at ease among those with similar backgrounds (Kets and Sandroni, 2016).

I investigate a new motive that arises in a fully rational environment in which individuals are only able to communicate their private signals: due to complementarities in information it may be optimal for an individual to communicate with those who have information that is highly correlated with her signal. These preferences originate from the fact that some signal structures allow for forecasting a stochastic state when all signals are positively conditionally correlated with higher precision than when the signals are conditionally independent. As a result, I can find instances in which homophily and belief polarization simultaneously appear in the endogenous network. In my model a finite number of individuals want to forecast a stochastic state about which they are uncertain. Each individual receives a private signal and can decide whether to share his

[1]For a summary of the empirical evidence, see, for example, McPherson, Smith-Lovin, and Cook (2001).

[2]For example, despite strong scientific evidence that the planet is warming, in large part due to human activity, a partisan divide exists in public opinion as to whether or not this is true (Dunlap, McCright, and Yarosh (2016)).

signal with others. If two individuals agree to engage in costly communication,[3] a new social link is created. Communication decisions such as these generate a social network. For the sake of simplicity, I assume that individuals face a quadratic loss when forecasting, so their objective is to pick the right set of links to minimize the communication cost plus the conditional variance given the information they gather.

I prove that for generic correlation matrices, the only network that satisfies a strong definition of stability, due to Dutta and Mutuswami (1997), is a network of disconnected cliques. Focusing only on disconnected cliques, I characterize the conditions under which $n > 2$ signals generate a very precise forecast of the state. It turns out that there is a continuum of correlation matrices such that observing all the signals will allow individuals to learn the state perfectly, and many of these matrices contain only positive correlations.[4] These complementarities in information can be very powerful. For any number of signals K, there are information structures such that after observing all K signals, the agents learn the state perfectly, but if they observe any subset of the signals, they learn almost as little as a single individual would.

This approach leads to new economic insights. First, I present an example where the indirect preferences become reversed: an individual's preferences over who he wants to share information with depends on the number of people with whom he wants to communicate. Second, the size of the cliques in the stable network is limited only by the cost of forming links; if this cost is small enough, it is always possible to find a correlation matrix for which the unique strongly stable network consists of a clique containing all the individuals.

Finally, homophily and belief polarization can coexist in my model's environment. For each small $\delta \geq 0$, there is a two-block matrix, with all intra-clique correlations larger than a threshold $\tilde{\rho} > 0$, for which the unique strongly stable network consists of the two cliques and the ex-ante expected value of the squared difference of the forecasts between the two cliques is larger than δ. Due to the information complementarities, individuals with similar information optimally communicate between themselves, emulating the

[3] Time and resources have to be spent in order to communicate with other individuals.

[4] This result is robust in the sense that the variance reduction is continuous in the correlations.

homophily patterns that have been found in social networks. At the same time, as the individuals' forecasts are different across cliques, for an external observer it will appear as if the cliques' beliefs are polarized. An interesting applied insight is that these phenomena are easier to sustain when the cost of communicating is smaller, and certainly this cost has decreased in recent years due to the surge in online social networks; and during this time belief polarization has simultaneously increased.

I show that in my model these two phenomena are closely associated since higher levels of belief disagreement require higher intra-clique correlations. On one hand, an upper bound for the intra-clique posterior variance is the posterior variance between any pair, and this variance decreases with respect to the correlation. On the other hand, a low correlation between two individuals in a clique means that they can provide more information to the individuals in the other clique, making it more difficult to deter deviations.

1.1 Related Literature

As noted above, my result can be considered a new explanation for the homophily phenomenon (see Currarini, Jackson, and Pin (2009),Kets and Sandroni (2016)). Golub and Jackson (2012) show that homophily can reduce the speed of learning when individuals are only able to take averages of their neighbors' beliefs. My work focuses on what kinds of networks can be formed when only signals, and not beliefs, can be communicated, and I show how this may result in the presence of homophily in the endogenous stable network.

There is a growing literature that studies how individuals acquire signals in similar Gaussian environments. Sethi and Yildiz (2016)] study an environment where individuals do not know other priors and choose whom to observe, so they can simultaneously learn about the state and others' priors. Liang, Mu, and Syrgkanis (2017) study information acquisition patterns by two agents, a myopic one and a fully rational one, when the agents can decide sequentially which signals to acquire from a set of correlated signals that identify the state. Kambhampati, Segura-Rodriguez, and Shao (2018) study the inefficiency patterns that can arise in a stable matching, when individuals can decide with whom to match and how many signals to acquire from a correlated information structure after they are matched.

The paper whose results resemble mine the most is Sethi and Yildiz (2018). In their environment individuals have different priors and can decide whom to observe over every period. The prior realizations are both correlated between individuals in the same group and independent across groups. Their main result is that when the intra-group correlation is large enough, individuals will observe only others in the same group. In contrast to the complementarities of information I describe, the explanation for homophily in their environment is that individuals understand other biases better when the correlation between their priors is high. There is previous work (e.g., Meyer and Strulovici (2015)) that has focused on finding the right measure of interdependencies for non-Gaussian environments with more than two signals. I contribute by showing that in any information environment, the correlation structures of signals that do not provide new information or that lead to perfect learning cannot be summarized by simple linear rankings of the pairwise correlations.

2 Model

I study a society with a finite number of agents $\mathcal{N} = \{1, ..., N\}$, $N \geq 2$. Individuals are uncertain about a stochastic state, $\theta \in \mathbb{R}$, which they want to forecast. They evaluate their estimate using a quadratic loss function; that is, if their forecast is $a \in \mathbb{R}$ they receive a utility equal to $u(a, \theta) = -(a - \theta)^2$.

I assume that all the agents share a common prior belief that follows a Normal distribution, with mean μ_θ and precision τ_θ. Each agent privately observes an unbiased Normal signal, y_i, with precision τ. I allow the individuals' signals to be correlated in any feasible way. Its correlation matrix, which I denote by Σ, only needs to be a positive semidefinite matrix, and, in particular, it does not have to be invertible. Explicitly, the vector of signals is distributed according to:

$$\begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix} \sim N\left(\begin{pmatrix} \theta \\ \vdots \\ \theta \end{pmatrix}, \tau^{-1}\begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1N} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{N1} & \cdots & \cdots & 1 \end{bmatrix}\right).$$

Individuals can communicate with others and learn their signal realizations. Communication is a costly activity: if an agent communicates with K other individuals, he has to pay a cost $c(K)$, with c an increasing function.[5] I impose three restrictions in the communication protocol. First, individuals must decide with whom to communicate before they observe the realization of their own signals. Second, communication between two individuals actually occurs only if both of them agree to communicate. Finally, individuals can communicate only the realization of their signals, not their beliefs.[6]

The communication links that are created define a network in the set $\mathcal{N}$. The endogenous network is represented by the matrix g, where $g_{ij} = 1$ if agents i and j communicate and 0 otherwise.

The timing in the model is summarized in Figure 1. Each individual first decides with whom to communicate. Once the communication network has formed, Nature draws the individuals' signals. Upon observing their own signals, the agents communicate, and, with the signals gathered, each individual chooses a forecast a.

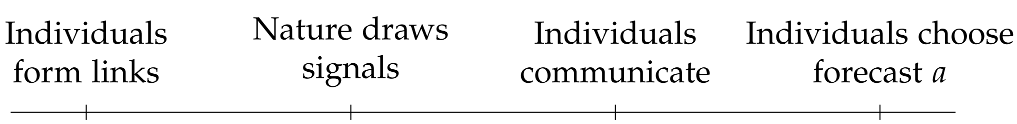

Figure 1: Timing in the model.

3 Optimal Action and Endogenous Networks

The objective of this section is twofold. First, I describe the optimal agents' forecast after observing any set of signals. Second, I characterize the shape of the communication

[5] If communication were without cost, individuals would communicate with everybody since observing more signals would allow them to make a better forecast.

[6] One possible interpretation for this assumption is that individuals can only communicate information they can support with evidence.

network that endogenously arises. I do not model how each communication link forms, but I require that the resulting network must satisfy a strong notion of stability.

Fix a network g. Agent i observes her own signal and her neighbors' signals—that is, she learns the signals received by $N_i(g) = \{j : g_{ij} = 1\} \cup \{i\}$. Since individual i suffers a quadratic loss, her optimal forecast is equal to the conditional expectation, and her expected utility, which I denote by $U_i(g)$, is equal to the negative of the posterior variance minus the communication cost she has to pay, $c(|N_i(g)| - 1)$. In the Normal environment the posterior variance is independent of the signals' realization and can be expressed in a simple closed form solution. Lemma 1 formalizes this discussion.[7]

Lemma 1 *Fix agent i and a network g. Let $\Sigma_{N_i(g)}$ be the correlation matrix between the signals i observes. For any signal realization $x_{N_i(g)}$, individual i's optimal action is $a = E[\theta \mid x_{N_i(g)}]$, and individual i's expected utility is*

$$U_i(g) = U_i(N_i(g)) = -Var\left(\theta \mid x_{N_i(g)}\right) - c(|N_i(g)| - 1).$$

Let B be a base of the null space of $\Sigma_{N_i(g)}$. If $1'_{N_i(g)}B = 0$ then

$$Var(\theta \mid x_{N_i(g)}) = (\tau 1'_{N_i(g)} \Sigma^+_{N_i(g)} 1_{N_i(g)} + \tau_\theta)^{-1},$$

where $\Sigma^+_{N_i(g)}$ is the Moore-Penrose pseudoinverse of $\Sigma_{N_i(g)}$, and if $1'_{N_i(g)}B \neq 0$ then $Var(\theta \mid x_{N_i(g)}) = 0$.

I use the Moore-Penrose pseudoinverse, a generalization of the inverse of a matrix, to account for the possibility that the correlation matrix of a group of signals is singular. When the correlation matrix is invertible this generalization coincides with the inverse of the matrix, and if it is singular it can be thought of as an approximation of the inverse by the inverse of nearby positive definite matrices.[8] The lemma provides two important insights. First, an agent, conditional on the number of signals, prefers the set of signals with the highest sum of the entries in $\hat{\Sigma}^+_{N_i(g)}$. As the inverse of a matrix is a highly nonlinear function, no simple ranking of the pairwise correlations corresponds to the in-

[7]The proof of this lemma and all others not in the main text can be found in Appendix A.

[8]Formally, for any symmetric matrix A, $A^+ = T(\lim_{\delta \to 0}(D^2 + \delta^2 I)^{-1} D)T'$, where D is a diagonal matrix and T is an orthogonal matrix such that $A = TDT'$. Its proof can be found on page 23 Albert (1972).

dividual's preferences over groups of signals. I analyze in Section 4 the relation between the pairwise correlations and the individuals' preferences. Second, an individual's utility depends only on the neighbors she has. In particular, an agent is indifferent as to the choice od networks as long as she has the same set of neighbors. This fact will be critical for my network characterization.

In my environment there are two ways in which an individual's utility can be equal in two different networks, g and g'. The first one occurs when the individual has the same number of neighbors in the two networks and $1'_{N_i(g)}\Sigma^{-1}_{N_i(g)}1_{N_i(g)} = 1'_{N_i(g')}\Sigma^{-1}_{N_i(g')}1_{N_i(g')}$. The second occurs if the difference in communication costs is offset by the difference in the posterior variances. While the first occurs only in a zero measure space, the second one can be destroyed by a small perturbation in the cost function. To avoid these peculiar cases I introduce the following assumption.

Assumption 1 *If individual i's neighbors differ in two networks g and g', then $U_i(g) \neq U_i(g')$.*

Since I know the payoff for any individual in any possible network, I can characterize which networks endogenously arise. I require that the endogenous network be strongly stable, a concept introduced by Dutta and Mutuswami (1997).[9] I say that a network is strongly stable (Dutta and Mutuswami, 1997) if there is no coalition that can strictly improve the utility of each of its individual members by merely creating extra links within the coalition or cutting an existing link that is held by someone in the coalition. Formally,[10]

Definition 1 *A network g is* strongly stable *with respect to U if there is no $S \subseteq \{1, \ldots, N\}$ or network g' that satisfies*

- *If $g'_{ij} = 1$ and $g_{ij} = 0$ then $i, j \in S$.*
- *If $g'_{ij} = 0$ and $g_{ij} = 1$ then either $i \in S$ or $j \in S$.*

[9] I could ask for a weaker solution concept such a pairwise stability, a notion introduced by Jackson and Wolinsky (1996). This alternative notion only requires that no pair has incentive to deviate, but this requirement is too weak for my interests.

[10] There is another notion of strong stability due to Jackson and Van den Nouweland (2005). The only difference is that the group will deviate if it can weakly increase the utility of all its members and strictly increase the utility of at least of one of them. Assumption 1 makes both definitions equivalent in my environment.

such that $\forall i \in S$, $U_i(g') > U_i(g)$.

Since in my environment all individuals attain the same level of utility when observing the same set of signals, it makes sense to look at networks where if individual i has as neighbors both j and k, then individuals j and k are neighbors as well. In other words, individuals may have incentive to form closed and disconnected cliques. Definition 2 formally defines this class of networks. Theorem 1 shows that a disconnected clique network is the only one that satisfies strong stability.

Definition 2 *The network* g *consists of* disconnected cliques $g_1, \ldots, g_M$ *if* $\cup_i g_i = \mathcal{N}$*, for all* $i, j \in g_l$, $g_{ij} = 1$ *and if* $i \in g_{l_i}$, $j \in g_{l_j}$, $l_i \neq l_j$ *then* $g_{ij} = 0$.

Theorem 1 *Suppose assumption 1 is satisfied.*[11] *There is a unique strongly stable network, and it is a disconnected clique network.*

Proof The set $2^{\mathcal{N}} \setminus \emptyset$ contains all the possible cliques that can be formed. By Lemma 1 each clique is associated with a unique utility level for each of its members. Since $2^{\mathcal{N}} \setminus \emptyset$ is a finite set, we can order the utility generated in each of the cliques, from the highest to the lowest.

Form the clique that gives the highest utility (if there are ties, pick one of the largest cliques) and call it S^1. From the cliques without individuals in S^1 pick the clique that gives the highest utility (if there are ties, pick one of the largest cliques) and call it S^2. Continue inductively. This process delivers a partition $S := \{S^1, S^2, \ldots, S^m\}$. I will show that this partition is strongly stable.

Consider the incentives to deviate for an individual in clique S^1. If individual i in clique S_1 decides to create links with individuals in A_1 and cut links with individuals in C_1, he will obtain utility as being in group $S^1 \cup A_1 \setminus C_1$. However, individual i weakly prefers group S^1 over $S^1 \cup A_1 \setminus C_1$. Therefore, no player in S^1 can be part of a deviating group. Continuing inductively allows us to conclude that the disconnected clique network defined by S is strongly stable.

[11]Without Assumption 1, the first part of the proof implies that one of the strongly stable networks is a disconnected clique network, but I cannot argue that all strongly stable networks belong to this class.

To prove uniqueness, I use assumption 1. Suppose a network g is strongly stable. Take the individuals in $S_1 = \{i_1^1, \ldots, i_1^m\}$. Let $D = \{1 \leq l \leq m : \exists k \neq l \in S^1 \text{ with } g_{i^k i^l} = 0 \text{ or } \exists j \notin S_1 \text{ with } g_{i^l j} = 1\}$ and suppose that D is non-empty. By definition all individuals in D are derive a lower utility than would be the case if they formed the clique S_1. If $D = S_1$ they have a strict incentive to create the network where clique S_1 forms. If $D \neq S_1$ for all $j \in S_1 \setminus D$, individual j is connected to everyone else in S_1 and has no connections outside S_1. Therefore, the individuals in D are able to create the clique S_1 by creating links between individuals in D and cutting links with individuals outside S_1, a network they strictly prefer. This contradicts the assumption that network g is strongly stable. Thus it must be that g contains the disconnected clique S_1. Inductively we conclude that g must be the disconnected clique network defined by S. ∎

This result implies that in an environment where individuals can only communicate their own experiences, they will create disjointed clusters of information.[12] From now on I assume that the endogenous network is a disconnected clique network. My objective is to characterize how individuals partition into cliques.

To the best of my knowledge, Theorem 1 is not a direct consequence of any of the results in the literature. Jackson and Van den Nouweland (2005) focuses on studying the set of strongly stable networks when the value function satisfies anonymity. My environment does not satisfy this property; instead, an individual's identity characterizes the complementarity of others' information with his own. Dutta and Mutuswami (1997) focuses on finding an allocation rule that delivers strongly stable outcomes. My allocation rule is fixed and does not coincide with the one they proposed.

Since Theorem 1 implies that the only strongly stable network consists of disconnected cliques, my result is related to the literature on coalition formation. In particular, if I restrict attention to coalitions alone, my environment satisfies the Top Coalition Property introduced by Banerjee, Konishi, and Sönmez (2001). Since by assumption 1 the preferences in my environment are strict, I obtain that the cliques in the network de-

[12]The disconnected clique network characterized in Theorem 1 satisfies other notions of stability. It satisfies pairwise stability, and under Assumption 1, it is the only strongly stable network as defined by Jackson and Van den Nouweland (2005) and the unique farsighted stable network as defined by Herings, Mauleon, and Vannetelbosch (2009).

scribed in the theorem correspond to the elements of the unique partition in the Core.[13] This means that the strongly stable network maximizes social welfare as well.

4 Higher Order Information Complementarities

This section studies two questions: How does the posterior variance relate to correlation structure inside any possible clique? And, for a given clique, what is the marginal value of adding a new individual? The answer to these questions will give us some interesting economic insights that I will present and discuss in Section 5.

I start by studying how the posterior variance relates to the correlation matrix. Consider first the simple case in which there are only two individuals in a clique, $\{i, j\}$. By Lemma 1 the posterior variance is given by

$$\frac{1}{\frac{2}{1+\rho_{ij}}\tau + \tau_\theta}.$$

This posterior variance has some important properties. First, when the correlation is -1, the posterior variance is 0. The reason is that the realized signals are going to be symmetrically located around the mean, and thus the individuals can perfectly forecast the state by taking the average of the two signals. Second, when the correlation is 1 the reduction of the variance after observing both signals is exactly the same as when observing only one of the signals since the realizations of the signals coincide. Third, the posterior variance is continuous and strictly increases with respect to the pairwise correlation.

I show through a series of results that most of these properties still hold when considering larger cliques. Theorem 2 establishes that, for any clique size, when a new member is added the resulting posterior variance ranges from 0 to the posterior variance in the original clique. However, the intuition is trickier than when the clique has two agents. The new challenge is that the variance reduction that results when a new signal is observed is not a monotonous function of the correlation coefficients or linear combinations of them, and thus I need to focus my analysis on the correlation matrix

[13] See Theorem 2 in Banerjee, Konishi, and Sönmez (2001).

itself. The key insight is that the signal observed by an individual provides direct information, and simultaneously it helps, in a nonlinear fashion, in the interpretation of the information provided by others. I call such complementaries *Higher Order Information Complementarities.*

Theorem 2 *Let A_{n-1} be a positive definite correlation matrix of dimension $(n-1) \times (n-1)$ and P a vector of dimension $(n-1) \times 1$ such that $P' A_{n-1}^{-1} P \leq 1$. Let*

$$A_n = \begin{pmatrix} A_{n-1} & P \\ P' & 1 \end{pmatrix}.$$

1. *A_n is positive semidefinite.*

2. *$Var(\theta \mid A_n) = Var(\theta \mid A_{n-1})$ if and only if $\exists \{a_1, \ldots, a_{n-1}\}$ such that $\sum_{k=1}^{n-1} a_n = 1$ and*

$$P = a_1 A^1 + \ldots + a_{n-1} A^{n-1},$$

where A^i is A_{n-1}'s i-th column. Further, A_n is invertible if and only if $\forall i,\ a_i \neq 1$.

3. *$Var(\theta \mid A_n) = 0$ if and only if A_n is singular and none of the entries of P are equal to 1.*[14]

The proof of Theorem 2 consists of two parts. First, by using the Cholesky decomposition of the correlation matrix, I can express the n-th signal as a linear combination between the state, the previous $n-1$ signals, and independent noise:

$$y_n = \left(1 - \sum_{k=1}^{n-1} a_{n,k}\right) \theta + a_{n,1} y_1 + a_{n,2} y_2 + \ldots + a_{n,n-1} y_{n-1} + a_{nn} \epsilon_n.$$

This clearly defines two extreme cases. On one hand, if $a_{nn} = 0$ and $1 \neq \sum_{k=1}^{n-1} a_{n,k}$, after observing the n signals the individuals in the clique can recover the state perfectly since the state can be expressed as a linear combination of the signals. On the other hand, if $1 = \sum_{k=1}^{n-1} a_{n,k}$, observing the first $n-1$ signals or observing all of them produces the same reduction in the variance since the last signal is just a linear combination of the previous signals and independent noise.

[14] I use the normality assumption to prove only the "only if" part of the argument. Therefore, in any setting with unbiased signals, it is possible to identify correlation matrices that generate the extreme values of the posterior variance.

The second part of the proof studies how these linear combinations relate to the correlation structure between the n signals. First, I show that observing all n signals allows individuals to learn the state perfectly if and only if the correlation matrix between the n signals is singular and all the correlation submatrices are invertible.[15] Second, I show that the $n-th$ signal does not provide any extra information if and only if the correlation vector between the n-th signal and the previous signals can be expressed as a linear combination of the columns of the correlation matrix between the first $n-1$ signals, where the sum of the weights in the linear combination is equal to 1.

Example 1 presents interesting insights that are embedded in the result. First, it may be that the individuals learn the state perfectly even when many of the correlations are positive. Second, the amount of information provided by the signals is not a monotonic function of the correlation coefficients. Finally, a correlation matrix does not need to be singular to contain a signal that is redundant.

Example 1 *Suppose the correlation matrix is given by*

$$\begin{pmatrix} 1 & 0 & \frac{\sqrt{2}}{2} \\ 0 & 1 & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 1 \end{pmatrix}.$$

According to Theorem 2, perfect learning occurs in a clique with signals that share this correlation pattern since this matrix is singular and all of the correlation submatrices are invertible.

Now, suppose the correlation matrix is given by

$$\begin{pmatrix} 1 & 0 & 0.5 \\ 0 & 1 & 0.5 \\ 0.5 & 0.5 & 1 \end{pmatrix}.$$

In this case the third column of the matrix is a convex combination of the first two columns of the matrix. According to Theorem 2 this implies that the third person signal is redundant in

[15]The condition that all the correlation submatrices must be invertible has two implications. First, none of the correlations can be 1. Second, any case in which the individuals can learn the state perfectly with fewer than n signals must be excluded: n must be construed as the smallest number of signals that allow the individuals to learn the state perfectly.

this clique.This happens even when this matrix is invertible.

Furthermore, some of the correlations in the second matrix are smaller than in the first matrix. This exemplifies the argument that the reduction in the variance does not need to be monotonic in the correlation coefficients.

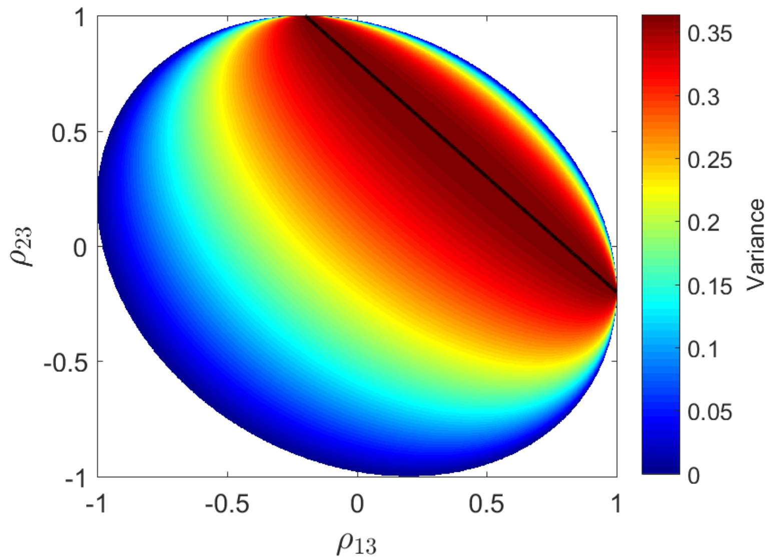

Figure 2: Ex-post variance for different correlations ρ_{13} and ρ_{23} when fixing $\rho_{12} = -0.2$, $\sigma = 1$ and $\sigma_\theta = 2$.

Figure 2 represents the posterior variance as a function of the correlation matrix when there are three signals. In the figure I fix the correlation between y_1 and y_2 and allow the other two correlations to vary. I highlight three properties. First, due to the complementarities in information that I discussed earlier, many correlation matrices that contain only positive correlations generate a small posterior variance. Second, there is a continuum of correlation matrices in which the third signal is redundant. This set corresponds to the hyperplane that connects the two extreme points with a correlation equal to 1, and except for the two extreme points, all matrices in this set are invertible. Third, the posterior variance is not monotone in the pairwise correlations: along the main diagonal the posterior variance is strictly concave, a result that I generalize in Proposition 2.

Example 2 shows how we can use Theorem 2 to built interesting extreme cases. The example shows that there are information structures for which the posterior variance of a clique with two or three signals is similar to the posterior variance after observing only one signal, but after observing four signals the clique learns the state perfectly. The key insight from the example is that having small groups that generate little information does not imply that there cannot be larger groups containing these smaller ones that generate plenty of information. I extend this intuition to the case with many signals in my analysis of the economic insights in Section 5.

Example 2 *Suppose $\tau = \tau_\theta = 1$. Suppose that there is a group with four individuals with correlation matrix*

$$\Sigma = \begin{pmatrix} 1 & 0.9 & 0.982 & 0.9876 \\ 0.9 & 1 & 0.918 & 0.9219 \\ 0.983 & 0.918 & 1 & 0.9994 \\ 0.9876 & 0.9219 & 0.9994 & 1 \end{pmatrix}$$

$(1,2)$ *is the size-two group with the smallest posterior variance:* 0.4872*. Upon the formation of group* $(1,2,4)$*, the posterior variance only reduces to* 0.4869*, and this is the lowest conditional variance in a group with three agents. However, the posterior variance in the clique with the four individuals reduces to* 0*; that is, the state is learned perfectly.*

I have discussed cases in which adding a new person to a clique either adds or does not add information; I have also looked a cases in which adding a new person allows the clique to learn the state perfectly. Although this is an interesting characterization, I am interested in whether the variance reduction is similar for nearby information structures. Proposition 1 states a particular form of continuity that holds everywhere except at a finite number of points. I cannot strengthen this result to uniform continuity since the variance reduction is more sensitive to small changes in the correlations when the correlations are positive.

Proposition 1 *Let A_n be a correlation matrix of dimension $n \times n$, A_{n-1} the correlation matrix of the first $n-1$ signals, and P the correlation vector between the n-th signal and all other $n-1$ signals. Suppose that either $P' A_{n-1}^{-1} P \neq 1$ or $\forall i \in \{1, \ldots, n-1\}$ $p_i \neq 1$. For all $\epsilon > 0$, there* *exists $\delta > 0$ such that if $\left\|P - \hat{P}\right\| < \delta$ and $\hat{P}' A_{n-1}^{-1} \hat{P} \leq 1$ then*

$$\left|Var(\theta \mid A_n) - Var(\theta \mid \hat{A}_n)\right| < \epsilon,$$

where $\hat{A}_n$ is obtained from A_{n-1} by adding $\hat{P}$.

The other interesting question is how much variance reduction the agents can obtain when their correlation matrix is far from the extreme cases described in Theorem 2. Proposition 2 shows that the posterior variance is concave when I consider a particular direction in the set of feasible correlation matrices. In the proposition I show that if the new signal has a uniform correlation with the previous $n-1$ signals, there exists a positive threshold such that the posterior variance increases for correlations lower than the threshold and it decreases in the opposite case. While the effect for correlations to the left of the threshold follows the intuition from the two signals case, an opposite effect appears for positive correlations when I consider many signals.

Proposition 2 *Let A_{n-1} be a positive definite correlation matrix of dimension $(n-1) \times (n-1)$ and let $P = (p, \ldots, p)$ be the correlation vector between the n-th signal and the previous $n-1$ signals such that $P' A_{n-1}^{-1} P \leq 1$. Let A_n be the resulting $n \times n$ correlation matrix. Then, $\frac{\partial Var(\theta|A_n)}{\partial p}$ is negative if $p > \frac{1}{1'_{n-1} A_{n-1}^{-1} 1_{n-1}}$ and it is positive if $p < \frac{1}{1'_{n-1} A_{n-1}^{-1} 1_{n-1}}$.*

5 Economic Implications

In this section I present the main economic insights from this approach. First, I introduce an example that shows how individuals' preferences concerning their interlocutors can change when clique size is taken into consideration. Second, I show that the fact that a correlation matrix is positive semidefinite does not restrict the maximum size of a clique. Finally, I show that certain information structures endogenously generate cliques of individuals where information is highly correlated and individuals' forecasts differ across cliques. When both of these effects occur simultaneously I say that in equilibrium *homophily* and *polarization* coexist.

5.1 Reversion of Preferences

An interesting property of this environment is that the reduced form preferences of individuals deciding which group to join depend critically on the size of the potential group. An individual's choice of a partner with whom to form a group of two depends only on partial correlations, while choosing fellow members of a group of three people depends on the inverse of a 3×3 matrix. Here I present an interesting example where individuals' preferences in terms of joining a group become reversed when clique size increases.

Suppose $\tau = \tau_\sigma = 1$ and the signals are correlated according to the network depicted in Figure 3. The number in each circle is the identity of the individual, and the numbers next to the edges are the correlations between their signals.

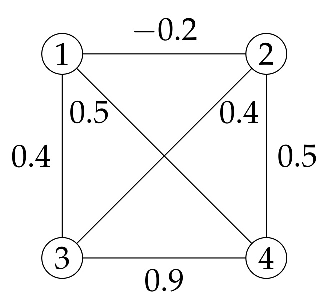

Figure 3: Preferences Reversion Examples.

By Proposition 2, an individual considering the formation of only one communication link prefers to join the person who receives the signal that is less correlated with hers. Thus, Player 1 (Player 2) prefers to form a pair with Player 3 rather than with Player 4; that is,

$$3 \succ_{1(2)} 4.$$

Theorem 2 implies that Player 3 does not add any extra information when he joins the pair $(1, 2)$ since $\rho_{13} + \rho_{23} = 1 + \rho_{12}$. But this condition does not hold for Player 4; that is, Player 4 will provide extra information beyond the information shared by Players 1 and 2. Hence, their preferences have reversed:

$$4 \succ_{(1,2)} 3.$$

Although in this section I introduced a very particular example, it exemplifies a pervasive property of the model. My analysis in Section 4 shows that a coalition may prefer to include new individuals with whom they have highly correlated information, and this preference depends as well on the identity of the coalition members.[16]

5.2 Which is the optimal size of a clique?

In this section I tackle the question of whether it is possible to characterize the size of the cliques in a strongly stable network. I show that the fact of the correlation matrix being positive semidefinite does not impose any limit in the size of an endogenous clique. The reason for this is that it is always possible to build a correlation matrix for which any clique of a size $K-1$ or less learns nearly the same amount as a single individual, whereas the K individuals coalition can figure out the state perfectly.

This matrix can be built in the following way. I start with two individuals with a very high pairwise correlation—say, $\rho_{12} = 0.99$, and they learn almost as much as a single individual. By Theorem 2, I can add a third individual that does not add any extra information. I can repeat the process any number of times—say, until I have $K-1$ individuals in a group. Finally, Theorem 2 allow us to add a K-th individual such that the K individuals coalition can learn the state perfectly. This procedure will fulfill my objective unless the last individual is part of coalition with fewer than $K-1$ individuals in which extra information is created. Lemma 2 shows that I can pick the last individual such that this is not the case. The critical observation is that I can pick the last signal close to the one that generates a singular matrix and does not provide any extra information. If this is done, the last signal is closed to the hyperplane in which no information is created for any coalition with fewer than K individuals.

Lemma 2 *For each number of individuals K there exists a correlation matrix Σ such that the state is learned perfectly when all signals are observed and, for any subset of signals, an individual can learn almost as much as he can when observing only one signal; that is, for each $\epsilon > 0$ and any*

[16]Deciding whom to spend free time with is another situation where this result may be applied. An agent may spend time with different groups of friends instead of simultaneously meeting all of them.

subgroup of k individuals with $k < K$, $Var(\theta \mid \Sigma_k) \geq \frac{1}{\tau+\tau_\theta} - \epsilon$.

This implies that it is always possible to find correlation structures for which the only strongly stable network consists of a clique containing all of the individuals in society. This happens whenever the cost of forming a clique containing all individuals is smaller than the difference between the value of learning the state perfectly and the value of the information an individual can learn by herself. In other words, the fact that the correlation matrix is positive semidefinite does not impose any limit in the optimal size of a clique.

Theorem 3 *Suppose $c(N-1) < \frac{1}{\tau+\tau_\theta}$. Then there exists a correlation matrix such that forming one clique with all individuals in it is the only strongly stable network.*

Proof Take $\epsilon = \frac{\frac{1}{\tau+\tau_\theta} - c(N-1)}{2} > 0$. By Lemma 2 there exist K signals with correlation matrix Σ such that $Var(\theta \mid \Sigma) = 0$ and for any subgroup of $k < K$ signals $Var(\theta \mid \Sigma_k) \geq \frac{1}{\tau+\tau_\theta} - \epsilon$. By forming a group with $k < K$ individuals, each of the individuals in this subgroup obtains a utility smaller than or equal to

$$-\frac{1}{\tau+\tau_\theta} + \epsilon - c(k-1) \leq -\frac{1}{\tau+\tau_\theta} + \epsilon < -c(K-1),$$

which is the utility each of them can obtain by forming the clique with all individuals. If Σ is the correlation matrix of the signals, then the only strongly stable network has a clique with all individuals in it. ■

5.3 Homophily and Belief Polarization

In this section I analyze how homophily and belief polarization can be rationalized by my model. I first introduce the definition will use for these phenomena.

Definition 3 *I say that a strongly stable network is*

a. $\rho-$homophilic *if the correlations between all individuals in a clique are larger than ρ; and*

b. $\delta-$polarized *if the ex-ante expected value of the squared difference of the forecasts between two cliques is at least equal to δ.*

For high ρ, saying that the stable network is $\rho-$homophilic means that individuals with similar information optimally communicate with each other, creating patterns that appear homophilic. For high δ, saying that the endogenous network is $\delta-$polarized means that we expect individuals in different cliques to have different beliefs about the value of the stochastic state.

There is a close relationship between $\delta-$polarization and the posterior variances in each of the cliques. When in both cliques the posterior variance is close to the prior variance, the difference in the forecasts must be small since in both cliques the forecasts are close to the prior mean. When in both cliques the posterior variance is close to zero, the difference in the forecasts has to be small since in both cliques the forecasts are close to the realization of θ. The difference between the forecasts has to be larger for intermediate values of the posterior variances: when forecasting, the agents assign similar weights to the prior mean and to the signal realization. Therefore, as we increase the posterior variance in both cliques, the difference between the forecasts first increases and then decreases as both posterior variances approach the prior variance. The next lemma formalizes this intuition.

Lemma 3 *Suppose that in a strong stable network two cliques form and all the signals across the cliques are independent. Let v_1 and v_2 be the posterior variance in each of the cliques. Then the ex-ante expected value of the squared difference of the forecasts is:*

$$v_1 + v_2 - 2\tau_\theta v_1 v_2.$$

Suppose $v_2 = kv_1$, $k \neq 0$. Then this expectation increases in v_1 iff $v_1 \leq \frac{1+k}{4k\tau_\theta}$ and decreases in v_1 otherwise.

Example 3 shows that the endogenous network that arises in my environment can exhibit both homophily and belief polarization. Due to the complementarities of information that I introduced in Section 4, individuals with high correlations form a clique even when there are opportunities to link with individuals that have signals that are independent from theirs. At the same time, since there is no information flowing across the two cliques, the agents' posterior beliefs across the two cliques are different.

Example 3 *Assume $\tau = \tau_\theta = 1$, $c(k) = 0.01k^2$, and $N = 6$. The correlation structure between the agents' signals is depicted on Figure 4.*

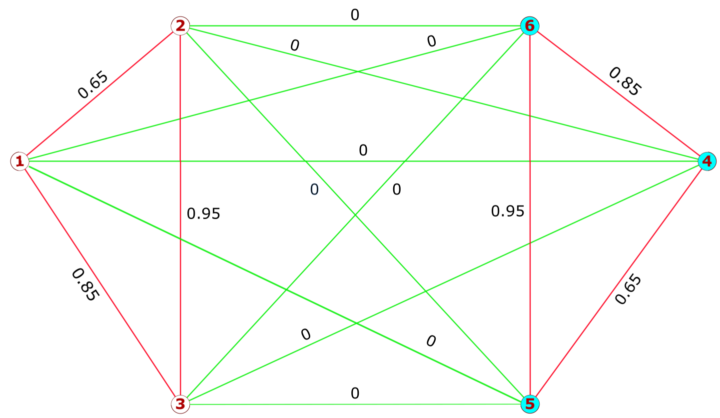

Figure 4: Correlation structure for example 3.

In this signal structure there are two identical cliques, $\{1, 2, 3\}$ and $\{4, 5, 6\}$, with positive intra-clique correlations, and all signals are independent across these cliques. I show that the unique strongly stable network only contains these two cliques.

First, it can be calculated that the utility in a coalition with two individuals who receive independent signals is -0.343, which is the largest utility of a two-agent coalition. Naively carrying over the intuition from the two-agent coalitions, I would conclude that the clique with two zero correlations and a correlation equal to 0.65 should have the smaller variance of any three-agent coalition, but this is not the case. The three-agent coalitions that minimize the variance are $\{1, 2, 3\}$ and $\{4, 5, 6\}$. The net utility in any of these coalitions is -0.2708. One of the four-agent coalitions that maximizes the net utility is $\{1, 2, 3, 4\}$. In this coalition the net utility is -0.2775. Groups with 5 and 6 individuals are too costly to be formed in equilibrium. Therefore, the only strongly stable network consists of the cliques $\{1, 2, 3\}$ and $\{4, 5, 6\}$.

Since the posterior variance in each of the cliques is equal to 0.23, Lemma 3 implies that the expected value of the squared difference between the forecasts is 0.41, which is almost half of the prior standard deviation. Therefore, this network is $0.65-$homophilic and $0.41-$polarized.

Theorem 4 generalizes the example. I assume that there are two cliques such that all inter-clique signals are independent. Under some mild conditions on the cost function, I show that for any δ small enough there exists ρ sufficiently large such that the only strongly stable network is $\rho-$homophilic and $\delta-$polarized. The last two conditions in the theorem are always satisfied for small enough δ since there is a strictly increasing mapping between δ and v. The first condition deserves some discussion. This condition implies that the level of inter-clique disagreement that can be sustained decreases as the cliques become large. The reason is that a clique with a large number of individuals is stable only if the agents in the clique can learn very well, a condition implying immediately that the variance inside the clique is small and, by Lemma 1, that there is little polarization. Another implication is that obtaining a polarized society is easier when the communication cost is smaller. This finding is especially relevant in light of recent technologies, such as online social networks, which have reduced the communication cost at the same time that political polarization has increased.

Theorem 4 *Suppose there are two cliques of individuals of sizes $3 \leq n \leq m$ such that all inter-clique correlations are 0. Let $\delta < \frac{4\tau}{(\tau_\theta+2\tau)^2}$ and $v = \frac{1-\sqrt{1-2\tau_\theta\delta}}{2\tau_\theta}$. Suppose that the following three conditions are satisfied:*

1. $c(m-1) < \frac{1}{\tau_\theta+2\tau} - v$,
2. $\min\{c(n)-c(n-1), c(m)-c(m-1)\} > \frac{v^2\tau}{v\tau+1}$, *and*
3. $c(n+m-1) - c(m-1) > \frac{v}{2}$.

There exist $\tilde{\rho} > 0$ and a correlation matrix with all intra-clique correlations larger than $\tilde{\rho}$ such that the only strongly stable network consists of the two cliques, and this network is $\tilde{\rho}-$homophilic and $\delta-$polarized.

I conclude this section by analyzing the relationship between level of polarization and homophily in a network. There are two effects. First, high levels of belief disagreement require large intra-clique correlations: if they are too low, then individuals will be able to learn the state very well just by forming pairs. However, there is a second effect that reinforces the first one: two individuals in a clique and a third one in another clique can learn more when the correlation between the signals of the two individuals in the same clique is lower,making a joint deviation between these three individuals more profitable.

Therefore, this outside option effect reduces the number of correlation structures that are consistent with high levels of disagreement.

Corollary 1 *Suppose $\delta < \frac{1}{\tau_\theta + 2\tau}$. Higher beliefs polarization requires that the endogenous network is more homophilic.*

6 Discussion

In this paper I have introduced an environment in which rational individuals want to forecast a stochastic state, and communicating the information they exogenously observed to others is costly. By studying the endogenous network that forms from the individuals' optimal communication decisions, I conclude that homophily and belief polarization can simultaneously emerge. Thus I offer a novel explanation for why homophily patterns are observed in the real world, using a model that does not require introducing behavioral biases or asymmetric frictions across groups.

In the process of working out this explanation, I discovered a counterintuitive statistical result: after observing a finite number of highly conditionally correlated signals, an individual can learn the state perfectly. This is an interesting result that might be important when studying other environments, and I point out two related situations that I believe are interesting for future work.

As a first example, consider a storyteller who wants to maximize the time a listener is exposed to his message and who wants to convey a narrative that is coherent and carries a clear message to the audience. My results suggest a way for the storyteller to achieve these objectives simultaneously: he can send a coherent message and keep the listener's attention for a long time by splitting the information into many positively correlated signals in a way that these signals communicate the whole message he wants to convey.

As another related example, consider a listener who needs to decide between two storytellers and pay attention to one of them. The listener will also have an opportunity to communicate with others who have listened to the storytellers. Assume that each storyteller sends a finite number of messages that are positively correlated and allow the listener to learn perfectly the state after processing all the messages, but these messages are independent across storytellers. A coordination effect emerges if the listener knows that he might miss some of the messages his storyteller sends: by communicating with individuals that have listened to the same storyteller, he is likely to recover information he has missed and recover the state perfectly, while if he communicates with individuals who have listened to the other storysteller, he may never learn the state perfectly.

References

ALBERT, A. (1972): *Regression and the Moore-Penrose pseudoinverse*. Elsevier.

BANERJEE, S., H. KONISHI, AND T. SÖNMEZ (2001): "Core in a simple coalition formation game," *Social Choice and Welfare*, 18(1), 135–153.

BOYD, S., AND L. VANDENBERGHE (2004): *Convex optimization*. Cambridge university press.

CURRARINI, S., M. O. JACKSON, AND P. PIN (2009): "An economic model of friendship: Homophily, minorities, and segregation," *Econometrica*, 77(4), 1003–1045.

DEGROOT, M. H. (1974): "Reaching a consensus," *Journal of the American Statistical Association*, 69(345), 118–121.

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.

DUNLAP, R. E., A. M. MCCRIGHT, AND J. H. YAROSH (2016): "The political divide on climate change: Partisan polarization widens in the US," *Environment: Science and Policy for Sustainable Development*, 58(5), 4–23.

DUTTA, B., AND S. MUTUSWAMI (1997): "Stable networks," *Journal of economic theory*, 76(2), 322–344.

GOLUB, B., AND M. O. JACKSON (2012): "How homophily affects the speed of learning and best-response dynamics," *The Quarterly Journal of Economics*, 127(3), 1287–1338.

Herings, P. J.-J., A. Mauleon, and V. Vannetelbosch (2009): "Farsightedly stable networks," *Games and Economic Behavior*, 67(2), 526–541.

Jackson, M. O., and A. Van den Nouweland (2005): "Strongly stable networks," *Games and Economic Behavior*, 51(2), 420–444.

Jackson, M. O., and A. Wolinsky (1996): "A strategic model of social and economic networks," *Journal of economic theory*, 71(1), 44–74.

Kambhampati, A., C. Segura-Rodriguez, and P. Shao (2018): "Matching to Produce Information," *Working paper*.

Kets, W., and A. Sandroni (2016): "A belief-based theory of homophily," .

Liang, A., X. Mu, and V. Syrgkanis (2017): "Dynamic Information Acquisition from Multiple Sources," Discussion paper, Working Paper.

McPherson, M., L. Smith-Lovin, and J. M. Cook (2001): "Birds of a feather: Homophily in social networks," *Annual review of sociology*, 27(1), 415–444.

Meyer, M., and B. Strulovici (2015): "Beyond correlation: Measuring interdependence through complementarities," Discussion paper, CSIO Working Paper.

Rao, C. R. (1973): *Linear statistical inference and its applications*vol. 2, pp. 527–528. Wiley New York.

Sethi, R., and M. Yildiz (2016): "Communication with unknown perspectives," *Econometrica*, 84(6), 2029–2069.

——— (2018): "Culture and Communication," *Available at SSRN 3263560*.

A Proofs

Proof Lemma 1

For any function $g : X_{N_i(g)} \to \mathbb{R}$, where $X_{N_i(g)}$ is the set of possible realizations of signals observed by individual i,

$$\begin{aligned}
-\mathbb{E}_{x_{N_i(g)},\theta}\left[(g(x_{N_i(g)}) - \theta)^2\right] &\leq -\mathbb{E}_{x_{N_i(g)},\theta}\left[(\mathbb{E}(\theta \mid x_{N_i(g)}) - \theta)^2\right] \\
&= -\mathbb{E}_{x_{N_i(g)}}\left[\mathbb{E}_\theta\left[(\mathbb{E}(\theta \mid x_{N_i(g)}) - \theta)^2 \mid x_{N_i(g)}\right]\right] \\
&= -Var(\theta \mid x_{N_i(g)}).
\end{aligned}$$

The inequality follows from $\mathbb{E}\left[(b-\theta)^2|x_{N_i(g)}\right]$ being minimized by setting $b = \mathbb{E}[\theta|x_{N_i(g)}]$. The first equality follows from the Law of Iterated Expectations, and the second equality follows from the definition of conditional variance.

Let Σ be the correlation matrix of joint signals $X_{N_i(g)}$, and $1_{|N_i(g)|}$ be a $|N_i(g)|$-column vector of 1s.

First, suppose that Σ is invertible. The likelihood function of joint signals is $p(x_{N_i(g)}|\theta) = \det(2\pi\sigma^{-2}\Sigma)^{-\frac{1}{2}} \exp\Big(-\frac{1}{2}\big[(\theta \cdot 1_{|N_i(g)|} - x_{N_i(g)})'\sigma^{-2}\Sigma^{-1}(\theta \cdot 1_{|N_i(g)|} - x_{N_i(g)})\big]\Big)$, and the prior density is $p(\theta) = (2\pi\sigma^{-2})^{-\frac{1}{2}} \exp\Big(-\frac{1}{2}\big[(\theta-\mu_\theta)^2\sigma_\theta^{-2}\big]\Big)$.

By Bayes rule, the posterior distribution of $\theta|x_{N_i(g)}$ is proportional to,

$$\begin{aligned}
p(x_{N_i(g)}|\theta)p(\theta) &\propto \exp\Big(-\frac{1}{2}\big[(\theta-\mu_\theta)^2\sigma_\theta^{-2} + (\theta \cdot 1_{|N_i(g)|} - x_{N_i(g)})'\sigma^{-2}\Sigma^{-1}(\theta \cdot 1_{|N_i(g)|} - x_{N_i(g)})\big]\Big) \\
&\propto \exp\Big(-\frac{1}{2}\big[\theta^2(\sigma_\theta^{-2} + \sigma^{-2}1'_{|N_i(g)|}\Sigma^{-1}1_{|N_i(g)|}) \\
&\qquad - \theta(2\mu_\theta\sigma_\theta^{-2} + \sigma^{-2}(x'_{N_i(g)}\Sigma^{-1}1_{|N_i(g)|} + 1'_{|N_i(g)|}\Sigma^{-1}x_{N_i(g)})\big]\Big) \\
&\propto \exp\Big(-\frac{1}{2}\big[\theta - A\big]'C\big[\theta - A\big]\Big),
\end{aligned}$$

where $C = (\sigma_\theta^{-2} + \sigma^{-2}1'_{|N_i(g)|}\Sigma^{-1}1_{|N_i(g)|})$, $A = C^{-1}(\mu_\theta\sigma_\theta^{-2} + \sigma^{-2}1'_{|N_i(g)|}\Sigma^{-1}x_{N_i(g)})$, and the proportionality operator eliminates positive constants. Since the derived expression is the kernel of a normal distribution, $Var(\theta \mid x) = C^{-1}$.

Now suppose that Σ is singular with rank $r < |N_i(g)|$. Let the vector B be a basis of the null space of Σ. With a singular correlation matrix the distribution assigns positive probability only in an affine subspace of $\mathbb{R}^r$. In this subspace the density can be expressed (see for example Rao (1973)) as

$$f(x_{N_i(g)} \mid \theta, \Sigma) = (2\pi)^{-\frac{1}{2}} det(\Lambda)^{-\frac{1}{2}} \exp\{-\tfrac{1}{2}(x_{N_i(g)} - \theta \cdot 1_{|N_i(g)|})' \Sigma^{+} (x_{N_i(g)} - \theta \cdot 1_{|N_i(g)|})\}$$

$$\text{such that } B' x_{N_i(g)} = \theta B' 1_{|N_i(g)|} \text{ with probability 1,}$$

where $\hat{\Sigma}^{+}$ represents the Moore-Penrose pseudoinverse of Σ, and Λ is the $r \times r$ diagonal matrix that contains the positive eigenvalues of Σ.

Suppose first that $B'1_{|N_i(g)|} = 0$. Then for any $x_{N_i(g)}$ such that $B'x_{N_i(g)} = 0$ I can follow the same procedure as before and obtain $Var(\theta \mid x_{N_i(g)}) = (\sigma_\theta^{-2} + \sigma^{-2} 1'_{|N_i(g)|} \Sigma^{+} 1_{|N_i(g)|})^{-1}$. If in the contrary, $B'1_{|N_i(g)|} = k \neq 0$ I can rewrite the condition that defines the subspace as $\theta = \frac{B' x_{N_i(g)}}{k}$, that is, after observing the realization $x_{N_i(g)}$, the individuals can perfectly recover θ. In such a case, $Var(\theta \mid x_{N_i(g)}) = 0$.

Proof Theorem 2

The Schur Complement of A_{n-1} in A_n is given by $1 - P' A_{n-1}^{-1} P$. The proof of 1. is immediate from the Schur Complement characterization of positive definiteness (semidefiniteness) (See for example Boyd and Vandenberghe (2004)).

Before proving parts 2. and 3. I present two lemmas that will be important for their proof. I start by showing that a generalization of the Cholesky Decomposition is valid in my environment.

Lemma 4 *Fix a correlation matrix A_n of dimension $n \times n$. There exist a vector $\epsilon = (\epsilon_1, \ldots, \epsilon_n)$ of independent random variables and a lower triangular matrix $C \in \mathbb{R}^{n \times n}$ such that*

$$\begin{matrix} y = & \theta & + & C & \epsilon \\ \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{pmatrix} = & \begin{pmatrix} \theta \\ \theta \\ \vdots \\ \theta \end{pmatrix} & + & \begin{pmatrix} 1 & 0 & 0 & \ldots & 0 \\ c_{21} & c_{22} & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ c_{M1} & c_{M2} & c_{M3} & \ldots & c_{MM} \end{pmatrix} & \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_M \end{pmatrix} \end{matrix}$$

Proof If A_n is positive definite, the result is immediate from the Cholesky Decomposition of A_n. If A_n is singular use the LDL decomposition and take $C = LD^{\frac{1}{2}}$, which is well defined since D is a non-negative diagonal matrix. ■

Let $\bar{C}_{n-1}$ be the $(n-1)\times(n-1)$ principal submatrix obtained by removing the last column and row from C and C_n the first $n-1$ entries of the n-th row of C. The matrix $\bar{C}_{m-1}$ is invertible as long as all of its diagonal components are non zero, which is implied by the correlation matrix between the $n-1$ signals being positive definite. From the definition of the matrix C, ρ_{ij} can be rewritten as $\rho_{ij} = \sum_{k=1}^{n} c_{ik}c_{jk}$. Then the correlation vector between the observation n-th and the first $n-1$ observations can be written as:

$$P' = C_n\bar{C}'_{n-1} = C_n\bar{C}^{-1}_{n-1}\bar{C}_{n-1}\bar{C}'_{n-1} = C_n\bar{C}^{-1}_{n-1}A_{n-1} \tag{1}$$

Let $a_{n,i}$ to be the $i-th$ entry of the vector $C_n\bar{C}^{-1}_{n-1}$. Therefore, $\rho_{ni} = \sum_{k=1}^{n-1} a_{n,i}A_{ki}$, or equivalently $P = a_{n,1}A_1 + \ldots + a_{n,n-1}A_{n-1}$.

Furthermore, if the correlation matrix between the first $n-1$ signals is positive definite I have:

$$Y_{n-1} = \theta_{n-1} + \bar{C}_{n-1}\mathcal{E}_{n-1} \Rightarrow \mathcal{E}_{n-1} = \bar{C}^{-1}_{n-1}(Y_{n-1} - \theta_{n-1})$$

Therefore, signal y_n can be rewritten as a function of only the first $n-1$ signals, the state and ϵ_n as follows:

$$\begin{aligned} y_n = \theta + (C_n, c_{nn})\mathcal{E}_n &= \theta(1 - C_n\bar{C}^{-1}_{n-1}1'_{(n-1)}) + C_n\bar{C}^{-1}_{n-1}\hat{Y}_{n-1} + c_{nn}\epsilon_n \\ &= \theta\left(1 - \sum_{k=1}^{n-1} a_{n,k}\right) + a_{n,1}y_1 + \ldots a_{n,n-1}y_{n-1} + c_{nn}\epsilon_n. \end{aligned}$$

The following lemma characterizes in which cases the coalition can learn perfectly the state and in which cases the last signal is redundant in terms of these linear combinations.

Lemma 5 *Suppose that $n-1$ signals have a positive definite correlation matrix. Then:*

1. *If $\left(1 - \sum_{k=1}^{n-1} a_{nk}\right) \neq 0$ and $c_{nn} = 0$, the n signals perfectly reveal the state.*
2. *If $\left(1 - \sum_{k=1}^{n-1} a_{nk}\right) = 0$, the n-th signal is redundant given the other $n-1$ signals.*

Proof The assumptions in case 1 imply that $y_n = a_{n,0}\theta + a_{n,1}y_1 + \ldots + a_{n,n-1}y_{n-1}$ with $a_{n,0} = 1 - \sum_{k=1}^{n-1} a_{n,k} \neq 0$. Then after observing the n signals the agents can invert the expression above to find θ as

$$\theta = \frac{1}{a_{n,0}} y_n - \frac{a_{n,1}}{a_{n,0}} y_1 - \ldots - \frac{a_{n,n-1}}{a_{n,0}} y_{n-1}.$$

The assumption in case 2 implies that $y_n = a_{n,1} y_1 + \ldots + a_{n,n-1} y_{n-1} + c_{nn}\epsilon_n$. Signal y_n does not contain any extra information about θ, so signal y_n is redundant given all the other $n-1$ signals. ■

Now I proceed to prove part 2. From Lemma 5 I conclude the first direction in part 2.: if $\sum_i a_{ni} = 1$ it has to be that $Var(\theta \mid A_n) = Var(\theta \mid A_{n-1})$.

Now suppose that $Var(\theta \mid A_n) = Var(\theta \mid A_{n-1})$. If A_n is singular then $c_{nn} = 0$ and $\sum_i a_{ni}$ has to be equal to 1; if not the coalition could learn the state perfectly according to Lemma 5. Suppose A_n is invertible. Then Lemma 1 implies that $1_n' A_n^{-1} 1_n = 1_{n-1}' A_{n-1}^{-1} 1_{n-1}$. A_n can be written as:

$$A_n = \begin{pmatrix} A_{n-1} & P \\ P' & 1 \end{pmatrix} = \begin{pmatrix} A_{n-1} & 0_{n\times 1} \\ 0_{1\times n} & 1 \end{pmatrix} + \begin{pmatrix} 0_{n\times 1} & P \\ 1 & 0 \end{pmatrix} \begin{pmatrix} P' & 0 \\ 0_{1\times p-1} & 1 \end{pmatrix} = \bar{A}_{n-1} + UV.$$

Clearly $\bar{A}_{n-1}$ is invertible since A_{n-1} is. The Woodbury Matrix Identity implies that:

$$A_n^{-1} = \bar{A}_{n-1}^{-1} - \bar{A}_{n-1}^{-1} U (I - V \bar{A}_{n-1}^{-1} U)^{-1} V \bar{A}_{n-1}^{-1}.$$

The sum of the entries of $\bar{A}_{n-1}^{-1}$ is equal to $1_{n-1}' A_{n-1} 1_{n-1} + 1$. From the definition of the matrices and using that the inverse of a block matrix is equal to the matrix formed by the inverse of each block, I obtain,

$$\begin{aligned} \bar{A}_{n-1}^{-1} U (I - V \bar{A}_{n-1}^{-1} U)^{-1} V \bar{A}_{n-1}^{-1} &= \begin{pmatrix} 0_{n\times 1} & A_{n-1}^{-1} P \\ 1 & 0 \end{pmatrix} \frac{1}{1 - P' A_{n-1}^{-1} P} \begin{pmatrix} 1 & -P' A_{n-1}^{-1} P \\ -1 & 1 \end{pmatrix} \begin{pmatrix} P' A_{n-1}^{-1} & 0 \\ 0_{n\times 1} & 1 \end{pmatrix} \\ &= \frac{1}{1 - P' A_{n-1}^{-1} P} \begin{pmatrix} -A_{n-1}^{-1} P P' A_{n-1}^{-1} & A_{n-1}^{-1} P \\ P' A_{n-1}^{-1} & -P' A_{n-1}^{-1} P \end{pmatrix} \end{aligned}$$

and the sum of its entries is $\frac{1}{1 - P' A_{n-1}^{-1} P} [1' A_{n-1}^{-1} P (2 - P' A_{n-1}^{-1} 1_{n-1}) - P' A_{n-1}^{-1} P]$. I conclude that,

$$1_n' A_n^{-1} 1_n = 1_{n-1}' A_{n-1}^{-1} 1_{n-1} + 1 - \frac{1}{1 - P' A_{n-1}^{-1} P} [1_{n-1}' A_{n-1}^{-1} P (2 - P' A_{n-1}^{-1} 1_{n-1}) - P' A_{n-1}^{-1} P]. \quad (2)$$

Since $1_n' A_n^{-1} 1_n = 1_{n-1}' A_{n-1}^{-1} 1_{n-1}$, it has to be that $1_{n-1}' A_{n-1}^{-1} P (2 - P' A_{n-1}^{-1} 1_{n-1}) = 1$. This is a quadratic equation in $P' A_{n-1}^{-1} 1_{n-1}$ with unique solution $P' A_{n-1}^{-1} 1_{n-1} = 1$. From Equation (1) I conclude that $C_n \bar{C}_{n-1}^{-1} 1_{n-1} = 1$, that is, $\sum_i a_{ni} = 1$.

The next lemma characterizes when is that the matrix A_n is invertible if $\sum_i a_{ni} = 1$, which concludes the proof of part 2.

Lemma 6 *Suppose that $\sum_i a_{ni} = 1$. The matrix A_n is singular iff $\exists i \in 1, \ldots, n-1$ such that $a_{ni} = 1$ and $a_{nj} = 0$ for $j \neq i$.*

Proof Remember that in the Cholesky decomposition $c_{nn} = 0$ if and only if A_n is singular. Since all the signals have the same variance it has to be that

$$
\begin{aligned}
c_{nn} &= 1 - \textstyle\sum_{k=1}^{n-1} a_{nk}^2 + 2 \sum_{k=1}^{n-1} \sum_{j>k} a_{nk} a_{nj} \rho_{kj} \\
&\geq 1 - \textstyle\sum_{k=1}^{n-1} a_{nk}^2 - 2 \sum_{k=1}^{n-1} \sum_{j>k} a_{nk} a_{nj} \\
&= 1 - (\textstyle\sum_{k=1}^{n-1} a_{nk})^2 = 0
\end{aligned}
$$

and the inequality is strict if two different a_{nk} and a_{nj} are non-zero, since by assumption $|\rho_{jk}| < 1$. Then $c_{nn} = 0$ if and only if $a_{ni} = 1$ for some i and $a_{nj} = 0$ for $j \neq i$. ∎

To prove 3. I will use the following lemma.

Lemma 7 *If $P' A_{n-1}^{-1} P = 1$, the conditions $\sum_i a_{ni} = 1$ and $\exists i \in 1, \ldots, n-1$ such that $p_i = 1$ are equivalent.*

Proof First, by lemma 6 I know that $P' A_{n-1}^{-1} P = 1$ and $\sum_i a_{ni} = 1$ imply that $\exists i \in 1, \ldots, n-1$ such that $a_{ni} = 1$ and $a_{nj} = 0$ for $j \neq i$. Then $p_i = \rho_{ni} = a_{ni} = 1$.

Now suppose that $\exists i \in 1, \ldots, n-1$ such that $p_i = 1$. WLOG, reorder the first $n-1$ observations such that $p_1 = 1$. As $p_1 = c_{n1}$, it has to be that $c_{n1} = 1$. By part 2 I have that $1 = P' A_{n-1}^{-1} P = C_n \bar{C}_{n-1}^{-1} A_{n-1} \bar{C}_{n-1}'^{-1} C_n' = C_n C_n'$. Then $c_{nj} = 0$ for $j \neq 1$. I conclude that $C_n \bar{C}_{n-1}^{-1} = (1, 0, \ldots, 0)$ since $c_{11} = 1$. Therefore, $a_{n1} = 1$ and $a_{nj} = 0$ for $j \neq 1$, that is, $\sum_i a_{ni} = 1$. ∎

From part 1. I know that A_n being singular implies that $P' A_{n-1}^{-1} P = 1$. By the last lemma I can replace the condition $p_i \neq 1$ by $\sum_i a_{ni} \neq 1$. As A_n is singular we have

$c_{nn} = 0$. Therefore, Lemma 5 implies the if direction of part 3.

For the other direction suppose that $Var(\theta \mid A_n) = 0$. By Lemma 1 it has to be that A_n is singular. Further, $\sum_i a_{ni}$ has to be different from 1; if not, Lemma 5 implies that $Var(\theta \mid A_n) = Var(\theta \mid A_{n-1})$ and this cannot be true since A_{n-1} being invertible implies that $Var(\theta \mid A_{n-1}) \neq 0$. By Lemma 7 it has to be that for all i $p_i \neq 1$.

Proof Proposition 1

Suppose $P' A_{n-1}^{-1} P \neq 1$. By Lemma 1 and Theorem 2, $Var(\theta \mid A_n) = (\tau_\theta + \tau 1_n' A_n 1_n)^{-1}$. Besides, by equation 2

$$1_n' A_n^{-1} 1_n = 1_{n-1}' A_{n-1}^{-1} 1_{n-1} + \frac{1}{1 - P' A_{n-1}^{-1} P} [1 - 1_{n-1}' A_{n-1}^{-1} P (2 - P' A_{n-1}^{-1} 1_{n-1})]. \tag{3}$$

Clearly, $1_n' A_n^{-1} 1_n$ is a continuous function of P, so $Var(\theta \mid A_n)$ is a continuous function of P.

Suppose $P' A_{n-1}^{-1} P = 1$ and $p_i \neq 1 \; \forall i \in \{1, \ldots, n-1\}$. Lemma 1 and Theorem 2 imply that $Var(\theta \mid A_n) = 0$. Take a sequence $P_k \to P$ such that the matrix A_n^k generated by adding P_k to A_{n-1} is positive semidefinite. As $p_i \neq 1 \; \forall i \in \{1, \ldots, n-1\}$ it has to be that $P' A_{n-1}^{-1} 1 \neq 1$. Therefore for large k, $P_k' A_{n-1}^{-1} 1 \neq 1$. Furthermore, $P_k' A_{n-1}^{-1} P_k \to 1$.

If the sequence $P_k' A_{n-1}^{-1} P_k$ is finally constant and equal to 1, I have $Var(\theta \mid A_n^k) = 0$ for large k, completing the proof.

Suppose for all k, there exists $k' > k$ such that $P_{k'}' A_{n-1}^{-1} P_{k'} \neq 1$. Take the subsequence generated by those indexes where the equality does not hold. Then I can use equation 3 to calculate $1_n' A_n^{k'^{-1}} 1_n$. The function $x(2-x)$ has a unique maximum at 1 and the maximum value is 1 which is reached when $x = 1$. Then, for large k the numerator of the last summand is strictly positive since $P_{k'}' A_{n-1}^{-1} 1_{n-1} \neq 1$. As $k' \to \infty$ the denominator approaches 0 from the right since $A_n^{k'}$ is positive definite and part 1 in Theorem 2. Therefore, as $k' \to \infty$, $1_n' A_n^{k'^{-1}} 1_n \to \infty$ and $Var(\theta \mid A_n^{k'}) \to 0$.

From both cases, if $\|P - \hat{P}\| < \delta$ and $\hat{A}_n$ is positive definite then the inequality $|Var(\theta \mid A_n) - Var(\theta \mid \hat{A}_n)| < \epsilon$ holds, where $\hat{A}_n$ is obtained from A_{n-1} by adding $\hat{P}$.

Proof Proposition 2

By equation 2, $1_n' A_n^{-1} 1_n = 1_{n-1}' A_{n-1}^{-1} 1_{n-1} + 1 - \frac{1_{n-1}' A_{n-1}^{-1} P(2 - P' A_{n-1}^{-1} P) - P' A_{n-1}^{-1} P}{1 - P' A_{n-1}^{-1} P}$. If $P = (p, \ldots, p)$, $1_{n-1}' A_{n-1}^{-1} P = p 1_n' A_{n-1}^{-1} 1_n$ and $P' A^{-1} P = p^2 1_{n-1}' A_{n-1}^{-1} 1_{n-1}$. Then from Lemma 1 I conclude that:

$$\frac{\partial Var(\theta \mid A_n)}{\partial p} = \frac{\frac{\partial 1_n' A_n^{-1} 1_n}{\partial p}}{Var(\theta \mid A_n)^2}$$

Solving for the derivative in the numerator I obtain

$$\begin{aligned} \frac{\partial 1_n' A_n^{-1} 1_n}{\partial p} &= \frac{2 1_{n-1}' A_{n-1}^{-1} 1_{n-1} (1 - p 1_{n-1}' A_{n-1}^{-1} 1_{n-1})(1 - p^2 1_{n-1}' A_{n-1}^{-1} 1_{n-1}) - 2p 1_{n-1}' A_{n-1}^{-1} 1_{n-1} (1 - p 1_{n-1}' A_{n-1}^{-1} 1_{n-1})^2}{(1 - p^2 1_{n-1}' A_{n-1}^{-1} 1_{n-1})^2} \\ &= \frac{2 1_{n-1}' A_{n-1}^{-1} 1_{n-1} (1 - p 1_{n-1}' A_{n-1}^{-1} 1_{n-1})(1 - p^2 1_{n-1}' A_{n-1}^{-1} 1_{n-1} - p + p^2 1_{n-1}' A_{n-1}^{-1} 1_{n-1})}{(1 - p^2 1_{n-1}' A_{n-1}^{-1} 1_{n-1})^2} \\ &= \frac{2 1_{n-1}' A_{n-1}^{-1} 1_{n-1} (1 - p 1_{n-1}' A_{n-1}^{-1} 1_{n-1})(1 - p)}{(1 - p^2 1_{n-1}' A^{-1} 1_{n-1})^2}, \end{aligned}$$

which is positive iff $p < \frac{1}{1_{n-1}' A_{n-1}^{-1} 1_{n-1}}$, since $1_{n-1}' A_{n-1}^{-1} 1_{n-1} > 0$.

Proof Lemma 2

Fix $\epsilon > 0$. Pick $\rho_{12} < 1$ such that $\frac{1}{\tau \frac{1}{1+\rho_{12}} + \tau_\theta} \geq \frac{1}{\tau + \tau_\theta} - \frac{\epsilon}{2}$. Such correlation always exists since the posterior variance continuously decreases with respect to ρ_{12} and when $\rho_{12} = 1$ it is given by $\frac{1}{\tau + \tau_\theta}$.

By Theorem 2 I can sequentially find vectors $P_2, \ldots, P_{K-1}$ such that the matrix with $K-1$ observations still have the same posterior variance and the matrix with the $K-1$ observations is positive definite. Using the same Theorem, I can find the vector $\hat{P}_K$ such that the posterior variance remains the same and the matrix $\hat{\Sigma}$ is singular. This implies that the posterior variance after observing any k signals, with $k < K$, is greater than or equal to $\frac{1}{\tau + \tau_\theta} - \frac{\epsilon}{2}$.

By Theorem 2 and Proposition 1, I can find a vector P_K that satisfies two properties. First, the correlation matrix between any $k < K$ signals that is generated by the correlation structure with vector P_K is invertible. Second, P_K is close enough to $\hat{P}_K$ such that the posterior variance after observing any k signals, with $k < K$, is greater than or equal to $\frac{1}{\tau + \tau_\theta} - \epsilon$. I only need to show that I can pick a vector P_K as close as I want to $\hat{P}_K$ that satisfies the first property and such that $P_K' \Sigma_{k-1} P_K = 1$ and $P_K' \Sigma_{k-1} 1_{K-1} \neq 1$.

When considering the Cholesky decomposition with $\hat{P}_K$ it has to be that $\hat{c}_{nn} = 0$ and

$\exists i \in \{1, \ldots, n\}$ such that $\hat{a}_{ni} = 1$ and $\hat{a}_{nj} = 0$ for $j \neq i$. Build a new signal such that $c_{nn} = 0$ and $a_{ni} = 1 - \delta$ and $a_{nj} = \gamma$ for $j \neq i$. Since none of the correlations between the first $K - 1$ signals is 1, it has to be that $\delta \neq (K-2)\gamma$. Then under the new signal with correlation vector P_K, $P_K' \Sigma_{k-1} P_K = 1$ and $P_K' \Sigma_{k-1} 1_{K-1} \neq 1$, that is, the state would be perfectly learned. When δ and γ are small, $\hat{P}_K$ and P_K are close and the correlation matrix between any $k < K$ signals that is generated by the correlation structure with vector P_K is invertible.

Proof Lemma 3

Let μ_1 and μ_2 to denote the expected value of the forecast. From Lemma 1, for any vector of signals x, the team forecast is given by, $\mu_i = \frac{\tau_\theta \mu_\theta + \tau 1' \Sigma_i^{-1} x}{\tau_\theta + \tau 1' \Sigma_i^{-1} 1}$. Therefore, $\mathbb{E}[\mu_i \mid \theta] = \frac{\tau_\theta \mu_\theta + \tau 1' \Sigma_i^{-1} 1 \theta}{\tau_\theta + \tau 1' \Sigma_i^{-1} 1} = (\delta_i \tau_\theta \mu_\theta + \theta(1 - \delta_i \tau_\theta))$, and $Var(\mu_i \mid \theta) = \frac{\tau 1' \Sigma_i^{-1} 1}{(\tau_\theta + \tau 1' \Sigma_i^{-1} 1)^2} = \delta_i (1 - \delta_i \tau_\theta)$.

As the signals across the two cliques are independent

$$z \mid \theta = \mu_1 - \mu_2 \mid \theta \sim N(\tau_\theta(\mu_\theta - \theta)(\delta_1 - \delta_2), \delta_1(1 - \delta_1 \tau_\theta) + \delta_2(1 - \delta_2 \tau_\theta)) = N(\mu_z, \sigma_z^2).$$

Therefore,

$$\begin{aligned}
\mathbb{E}[z^2] &= \mathbb{E}[\mathbb{E}[z^2 \mid \theta]] = \mathbb{E}[\sigma_z^2 + \mu_z^2] \\
&= \mathbb{E}[\tau_\theta^2 (\mu_\theta - \theta)^2 (\delta_1 - \delta_2)^2 + \delta_1(1 - \delta_1 \tau_\theta) + \delta_2(1 - \delta_2 \tau_\theta)] \\
&= -2\delta_1 \delta_2 \tau_\theta + \delta_1 + \delta_2,
\end{aligned}$$

where in the third line we use that $E[(\theta - \mu_\theta)^2] = \tau_\theta^{-1}$. $\mathbb{E}[z^2]$ value is always positive since $\delta_i < \tau_\theta^{-1}$.

If $\delta_2 = k\delta_1$ then $\frac{\partial \mathbb{E}[z^2]}{\partial \delta_1} = 1 + k - 4k\delta_i \tau_\theta$, so it increases with respect to δ_1 iff $\delta_1 \leq \frac{1+k}{4k\tau_\theta}$.

Proof Theorem 4

Fix $\delta < \frac{4\tau}{(\tau_\theta + 2\tau)^2}$ and let $v = \frac{1 - \sqrt{1 - 2\tau_\theta \delta}}{2\tau_\theta}$. Notice that v is one of the solutions to the equation $\delta = 2v - 2v\tau_\theta$. Therefore, by Lemma 1 if we can find a strongly stable network with two cliques in which the posterior variance is equal to v we are done. Further,

$v < \frac{1}{\tau_\theta + 2\tau}$. This is true since

$$\frac{1-\sqrt{1-2\tau_\theta\delta}}{2\tau_\theta} < \frac{1}{\tau_\theta+2\tau}$$

$$\Leftrightarrow \quad 2\tau - \tau_\theta < (\tau_\theta + 2\tau)\sqrt{1-2\tau_\theta\delta}.$$

If the expression in the left is negative we are done. Suppose it is positive then the inequality is equivalent to

$$4\tau^2 - 4\tau\tau_\theta + \tau_\theta^2 < (4\tau^2 - 4\tau\tau_\theta + \tau_\theta^2)(1-2\tau_\theta\delta)$$

$$\Leftrightarrow \quad \delta < \frac{4\tau}{(\tau_\theta+2\tau)^2}.$$

Let $\gamma = c(m-1) + v$. Find $\tilde{\rho}$ to be the minimum correlation that satisfies simultaneously that $\gamma \leq \frac{1}{\tau_\theta + \frac{4}{1+\rho}\tau}$, $v \leq \frac{1}{\tau_\theta + \frac{2}{1+\rho}\tau}$ and $\frac{v^2\tau}{v\tau + \frac{1+\rho}{2}} \leq c(n+1) - c(n-1)$. I can find always a correlation smaller than 1 that satisfies all the inequalities since $v < \frac{1}{\tau_\theta+2\tau}$, condition one in the 1. in the theorem implies that $\gamma < \frac{1}{\tau_\theta+2\tau}$, and as c is strictly increasing, condition 2. implies that $c(n+1) - c(n-1) > \frac{v^2\tau}{v\tau+1}$.

I build a correlation matrix inside the first clique such that the posterior variance inside the clique is equal to v and all intra-clique correlations are larger than $\tilde{\rho}$. First, pick the correlation between the first two individuals such that $\rho_{12}^1 > \tilde{\rho}$. Then by the definition of $\tilde{\rho}$, it satisfies $\frac{1}{\tau_\theta + \frac{2}{1+\rho_{12}^1}\tau} > v$, that is, the first two individuals have a posterior variance that is larger than the objective. Now for the individuals 3 to $n-1$, I can pick sequentially correlation vectors, as described in Theorem 2, such that after observing the first $n-1$ signals they have learned the same as if they had observed only the first two, and all the correlations in these vectors are larger than $\tilde{\rho}$. Let $\epsilon_1 = \frac{\frac{1}{\tau_\theta + \frac{4}{1+\rho_{12}^1}\tau} - \gamma}{2}$, which is larger than 0 since $\rho_{12}^1 > \tilde{\rho}$. A small variation of Lemma 2 shows that I can add the n-th individual such that with the n signals the variance reduces to v and for any smaller subgroup the integrated variance is at least $\frac{1}{\tau_\theta + \frac{2}{1+\rho_{12}^1}\tau} - \epsilon_1$.

I want to show that forming the two cliques is the only strongly stable network. The utility for each individual of forming clique 1 is $-v - c(n-1)$. The utility of forming any subgroup of individuals in 1 is at most

$$-\frac{1}{\tau_\theta + \frac{2}{1+\rho_{12}^1}\tau} + \epsilon_1 < -\frac{\frac{1}{\tau_\theta+\frac{2}{1+\rho_{12}^1}\tau} + \gamma}{2} < -\gamma \leq -v - c(n-1),$$

where the first inequality follows from the definition of ϵ_1 and the second one follows from the inequality $\rho_{12}^1 > \tilde{\rho}$ and the definition of $\tilde{\rho}$. Then deviations to smaller groups inside 1 are not profitable. An analogous argument holds for smaller groups inside 2.

Now consider a joint deviation for a proper subset S_1 of individuals in 1 and a proper subset S_2 of individuals in 2. The correlation matrix between the signals of this individuals is a block matrix. Then the posterior variance in this case is $\frac{1}{\tau_\theta+\tau\left(1'_{|S_1|}\Sigma_{S_1}^{-1}1_{|S_1|}+1'_{|S_2|}\Sigma_{S_2}^{-1}1_{|S_2|}\right)}$. Since after observing the signals in S_i the variance is at least $\frac{1}{\tau_\theta+\frac{2}{1+\rho_{12}^i}\tau} - \epsilon_i > \gamma > v$, we can find $\hat{\rho}^I > \tilde{\rho}$ such that $1'_{|S_i|}\Sigma_{S_i}^{-1}1_{|S_i|} = \frac{2}{1+\hat{\rho}^i}$. Therefore, the utility of such deviation is at most

$$-\frac{1}{\tau_\theta + \tau\left(1'_{|S_1|}\Sigma_{S_1}^{-1}1_{|S_1|} + 1'_{|S_2|}\Sigma_{S_2}^{-1}1_{|S_2|}\right)} = -\frac{1}{\tau_\theta + \tau\frac{2}{1+\hat{\rho}^1} + \tau\frac{2}{1+\hat{\rho}^2}} < -\frac{1}{\tau_\theta + \tau\frac{4}{1+\tilde{\rho}}} = -\gamma,$$

so this deviation is not profitable for any of the individuals.

Now consider the deviation where all individuals in group 1 and a proper subset S_2 of individuals in 2 are together. Since $v = \frac{1}{\tau_\theta+\tau 1'_n\Sigma_n^{-1}1_{|n|}}$, $1'_n\Sigma_n^{-1}1_n = \frac{1-v\tau_\theta}{v\tau}$. If in S_2 there is only one individual, every deviator's utility is

$$-\frac{1}{\tau_\theta+\tau+\tau 1'_n\Sigma_n^{-1}1_n} - c(n) = -\frac{v}{v\tau+1} - c(n) < -\frac{v}{v\tau+1} - \frac{v^2\tau}{v\tau+1} - c(n+1) = -v - c(n-1),$$

so this deviation is not profitable. Now, if there are more than one individual in S_2, I can find $\hat{\rho}^2 > \tilde{\rho}$ as before and the utility is at most

$$\begin{aligned}
-\frac{1}{\tau_\theta+\tau\frac{2}{1+\hat{\rho}^2}+\frac{1-v\tau_\theta}{v}} - c(n-1+|S_2|) &< -\frac{v}{\frac{2v\tau}{1+\tilde{\rho}}+1} - c(n-1+|S_2|) \\
&= \frac{v^2\tau}{v\tau+\frac{1+\tilde{\rho}}{2}} - v - c(n-1+|S_2|) \\
&< c(n+1) - c(n-1) - v - c(n+1) = -c(n-1) - v,
\end{aligned}$$

and this deviation is not profitable. A similar argument holds for the deviation where all individuals in clique 2 and some individuals in group 1 are together.

The last possible deviation is to create the grand coalition. Individuals in group 2, since this is the largest group, are the ones that can gain the most. The utility of any individual, when in the grand coalition, is

$$\begin{aligned}
-\frac{1}{\tau_\theta + \tau \frac{2(1-v\tau_\theta)}{v\tau}} - c(n+m-1) &= -\frac{v}{2-v\tau_\theta} - c(n+m-1) \\
&< \frac{v}{2} - v - c(n+m-1) \\
&< c(n+m-1) - c(m-1) - v - c(n+m-1) = -c(m-1) - v,
\end{aligned}$$

and forming the grand coalition is not optimal.

In an analogous way we can build the correlation matrix inside the second clique such that the posterior variance in the second clique is equal to v as well and all the intra-clique correlations are larger than $\tilde{\rho}$.

Proof Corollary 1

From the definition of $\tilde{\rho}$ in the proof of Theorem 4 it is immediate that higher posterior variance inside each clique requires a higher threshold $\tilde{\rho}$. By Lemma 1 this means that higher levels of beliefs' disagreement requires larges correlations inside each clique.