Communication and the emergence of a unidimensional world

Philippos Louis[1], Orestis Troumpounis[2], and Nikolas Tsakas[3]

[1]School of Economics, University of New South Wales and AGORA Centre for Market Design

[2]Department of Economics, Lancaster University Management School

[3]Department of Economics, Univesity of Cyprus

December 7, 2016

Abstract

We provide theoretical and experimental support on the emergence of a unidimensional world through communication. Both theoretical and experimental results suggest that when boundedly rational individuals communicate their opinions over multiple issues, disagreement can eventually be summarized on a unidimensional spectrum, even when imposing very little structure on the communication process. The presence of structured social networks is however crucial in determining whether an individual forms moderate or extreme views.

1 Introduction

Sociologists, political scientists and economists seem to often agree on one thing: the world is not flat; in fact it is unidimensional! The world we refer to is that of *opinions*. While individuals have opinions on myriads of issues, spanning domains such as politics, the economy or lifestyle, it is often the case that using a *unidimensional* spectrum one can describe an individual's opinions on all dimensions.

We encounter the best example in politics. Describing someone as a leftist or a rightist usually provides enough information about her opinions on an array of political issues (e.g., redistribution, protection of the environment, attitude towards immigration or gun possession). Indeed, representing political competition and analyzing voting behavior on a "left-right" unidimensional spectrum has dominated the political economy literature.[1] Empirical evidence supports this view, as individuals' opinions on different issues seem to be significantly correlated, and the underlying ideology can explain voting behavior of both legislators and individual voters (Poole and Rosenthal, 1997; Lee et al., 2004; Ansolabehere et al., 2008). Importantly, unidimensionality seems to extend even beyond the world of politics, and ideological cleavages spillover to preferences over leisure activities, consumption and art, as well as personal morality (see DellaPosta et al., 2015 and references therein, as well as Wilson and Haidt, 2014).

The prevalence of a unidimensional world, not only as a handy theoretical simplification, but often also as an accurate description of opinions across domains, raises the question regarding its origin. Some potential explanations rely on the philosophical underpinnings of ideologies (Bobbio and Cameron, 1996) and how those relate to personality traits (Carney et al., 2008; Gerber et al., 2010), cognitive and neural characteristics (Duckitt, 2001; Amodio et al., 2007) or even genetics (Alford et al., 2005). While all these explanations may be relevant, we note that all these rely on ideologies being stable across time and societal context, an assumption that seems rather strong (McDonald et al., 2007).[2,3]

We therefore approach the emergence of unidimensional worlds in a dynamic context. More specifically, we study a model where individuals communicate their opinions on an array of issues to others, repeatedly over several periods, and update their opinions on each issue in each period by taking a weighted average of their own prior opinion and those of others. In such a process, first introduced by DeGroot (1974), opinions eventually converge to a single

[1]From a theoretical perspective, formal political economy models relying on a unidimensional policy space dating back to Downs (1957) and Black (1958) still serve as the workhorse in the analysis of electoral competition. As Plott (1967) first pointed out extending to more dimensions can prove challenging, given that equilibria exist only in very particular cases.

[2]Recently, The Economist went as far as bidding farewell to 'left versus right', as it is replaced by 'open against closed' (*"The new political Divide"*, 2016).

[3]The definition of ideology itself has proven a non-trivial matter, see for example Gerring (1997) and Knight (2006) for a discussion of the relevant difficulties.

point.[4] But DeMarzo, Vayanos, and Zwiebel (2003) observed that, if opinions are communicated over a *fixed* communication structure, even if opinions eventually converge there is disagreement at any point in time, which can be summarized by a single dimension. In our theory, we show that this dynamic process of opinion formation can give rise to a unidimensional world even if the communication channels vary over time. That is, in contrast to DeMarzo et al. (2003), we show that individuals need not communicate in every period with the same individuals, nor assign the same subjective weight to one's opinion in all periods.

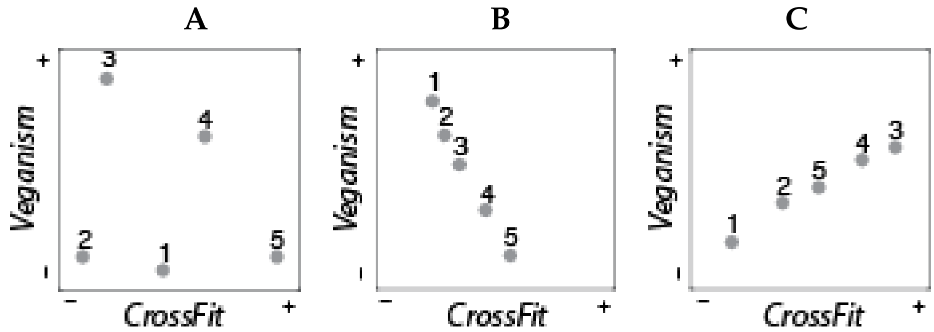


Figure 1: Possible distribution of opinions on two issues. Each point represents an individual's opinions on the two issues: CrossFit and veganism. Opinions can vary from extremely negative to extremely positive. Panel A shows an example of uncorrelated opinions. Panels B and C show examples of unidimensional opinions.

Of course permitting a dynamic communication structure comes at a cost compared to a static world with a fixed communication structure. While our approach is more general and possibly more realistic compared to DeMarzo et al. (2003) regarding the communication channels, it remains partly agnostic about the exact characteristics of the unidimensional world that may emerge. The following example illustrates how this is so. Figure 1 illustrates the opinions of five individuals across two issues: the practice of veganism and CrossFit training.[5] Panel A represents a multi-dimensional world, with no apparent

[4]A long literature focuses on the necessary and sufficient conditions for a society to reach consensus, as well as the speed at which this can be achieved (see for instance Golub and Jackson, 2010, 2012; Tahbaz-Salehi and Jadbabaie, 2008). For a broad overview see Jackson (2008) Chapter 8.3 and references therein.

[5]We pick two random issues so as to illustrate the concept of unidimensionality. Notice that in our setup opinions can be defined in a very broad sense including beliefs, judgements,

correlation between issues. On the other hand, panels B and C represent two different unidimensional worlds, where opinions on the two issues are strongly correlated. While we observe *unidimensionality* in both panels B and C, there are differences between the two. The first regards individuals' *relative positions* on the unidimensional spectrum. Individuals 1 and 5 have extreme positions in panel B on the opposite side of each other. While 1 is still an extremist in panel C, 5 has a more moderate position in that world. Another difference is the *direction of disagreement*, captured by the slope of the line indicating the importance of different issues in the overall disagreement as well as whether the correlation is positive or negative. While this direction is negative and steep in panel B, it is positive and flatter in panel C. In terms of our results, our model does predict that opinions will move from a multidimensional world (panel A) towards a unidimensional one (panels B and C). Nevertheless, stricter assumptions about the communication structure as in DeMarzo et al. (2003) also differentiate between panels *B* and *C*. Our results therefore distinguish between the exact role of fixed social networks and communication in the opinion formation process: While communication is enough to give rise to a unidimensional world, the presence of fixed channels is crucial so as to predict its characteristics.

Our study moves beyond theory to provide empirical validation in the form of a lab experiment. Subjects communicated their opinions over two issues across ten rounds in groups of five. In two treatments subjects are linked through a fixed network. Networks differed minimally between these treatments. In a third treatment subjects listen the opinions of different randomly picked subjects from their group in each round. In line with the theoretical predictions, our results support the emergence of a unidimensional world in all three treatments. We also find support for the role of networks in determining individuals' relative positions. Predictions regarding the direction of disagreement prove less robust, as we do not find support for the theory in this dimension.

Previous experiments provide some evidence to support the idea that individuals update their opinions by repeatedly averaging their network neighbors'

attitudes and all such fundamental drivers of behaviour that are amenable to social influence or the advent of new information.

opinions (see Corazzini et al. 2012; Brandts et al. 2015; Battiston and Stanca 2015). Communication in these experiments is over a single issue and the analysis focuses on the individual updating process. Our work therefore extends this knowledge by permitting communication of opinions over multiple issues and focusing on the emergence of a unidimensional world.

Before proceeding to our theoretical model and experimental results let us link further our work with relevant strands of the literature. In the political economy literature, McMurray (2014) shows how political competition can lead to a unidimensional policy space for parties. This result depends crucially on voters' ideal points across issues exhibiting a non-zero correlation. Our theory provides support for this assumption: if voters' ideal points are monotonic functions of opinions, communication will lead to them exhibiting perfect correlation across issues. The political economy literature has also recently provided theoretical links between the "correlation neglect" bias and polarization and the competitiveness of the electoral system (Ortoleva and Snowberg 2015; Levy and Razin 2015; Glaeser and Sunstein 2009; Levy and Razin 2016). Since our assumed updating process can be attributed to such bias, we complement these results with a link between "correlation neglect" and the emergence of a unidimensional world in the presence of dynamic communication channels.

In the communication literature, Spector (2000) shows how unidimensional beliefs can emerge in a model of sequential cheap-talk communication preceding a collective decision. Besides the more restrictive setting in his model, there is a significant qualitative difference in results: in his model individuals end up agreeing in all but one issue; in our unidimensional world this is not true. While individuals' opinions in the long run move arbitrarily close to each other, disagreement remains across all issues. In a very recent paper, Sethi and Yildiz (2016) take a different approach and study how the communication network is shaped by individuals' simultaneous and complementary efforts to learn about the state of the world and about others' perspectives. Our study focuses on the shape of disagreement. We view these two approaches as complementary. Assuming that the communication structure is exogenous, as we do, seems appropriate for the shorter run, where individuals update their opinions on a specific set of issues. In the longer run, it is natural to assume that individuals will try to optimize over their potential interlocutors, as they do in Sethi and Yildiz (2016). Finally, Klar and Shmargad (forthcoming) conduct an online experiment to study the effect of social network structure on the spread of information. They find that in more closely connected networks, opposing views spread equally and opinions are more balanced, even if one view is initially under-represented.

2 Theory

2.1 The opinion updating process

Our theoretical framework pertains to the family of average–based updating processes originated by DeGroot (1974). Our analysis extends on DeMarzo et al. (2003), who were the first to consider and observe the emergence of unidimensional worlds in such processes. In our framework we impose very little structure on the communication process. By contrasting the two cases we can highlight the role of the social network in the shape of the ensuing unidimensional world and formulate testable predictions to be explored experimentally.

Consider a population D consisting of N agents, forming opinions on K different issues.[6] They communicate in discrete time periods $t \in \{1, 2, \dots\}$ and update their opinions. Their initial opinions at time $t = 0$ are given exogenously. The opinion of agent i on issue k at time t, is $s_{i,k}(t) \in \mathbb{R}$ and the $N \times 1$ column vector $\mathbf{s}_k(t)$ denotes the opinions of all agents on issue i at period t. We summarize all agents' opinions in all dimensions at time t by the $N \times K$ matrix $\mathbf{s}(t)$, where $\mathbf{s}(0)$ is the matrix of initial opinions.

Communication occurs as follows: at every period $t \in \{1, 2, \dots\}$, an agent i observes the opinions across all K issues in period $t - 1$ of a subset of the population $D_i(t) \subseteq D$, which is called i's neighborhood. Communication may

[6]In the current context, it is perhaps best to think of opinions as agents' estimates about the state of nature, as this is the way we induce opinions in our experiment. Another alternative would be to think of them as preference parameters, like attitudes towards different choice alternatives. In any case, the internal consistency of the model of opinion dynamics is not affected by their exact meaning, as it is a purely mechanical process.

not be reciprocal, meaning that $j \in D_i$ need not imply that $i \in D_j$. The collection of neighborhoods $D_i(t)$ for all $i \in N$ defines the network of communication in period t. This network can be represented by a $N \times N$ adjacency matrix $\mathbf{G}(t)$, where $G_{i,j}(t) = 1$ if $j \in D_i(t)$ –defining j as i's neighbor–, $G_{i,i} = 1$ for all i as every agent is assumed to remember her opinion in $t-1$, and $G_{i,j}(t) = 0$ if $j \notin D_i(t)$. This network is assumed to be strongly connected, which is a necessary condition to ensure the flow of information in the population.[7] Crucially, this network may change in every period.

Opinion updating occurs as follows: at every period $t \in \{1, 2, \dots\}$, agent i assigns weight $T_{i,i}(t) \in (0, 1)$ to her own prior opinion, weight $T_{i,j}(t) \in (0, 1)$ to the observed opinion of each of her neighbors $j \in D_i(t)$, and weight $T_{i,j}(t) = 0$ to all $j \notin D_i(t)$ such that $\sum_{j=1}^{N} T_{i,j} = 1$. This weight can be considered as the relative precision agent i assigns to j's opinion, compared to the rest of her neighbors. The collection of all weights forms a $N \times N$ matrix $\mathbf{T}(t) = \big(T_{i,j}(t)\big)$, which will be called the *listening matrix*. It is useful to define a sequence of such listening matrices as $\mathcal{T}_t = \{\mathbf{T}(\tau)\}_{\tau=1}^{t}$ when finite and as $\mathcal{T}_\infty = \{\mathbf{T}(\tau)\}_{\tau=1}^{\infty}$ when infinite. Notice that if the network varies, then by definition the listening matrix varies as well. However, even if the network were to remain the same the listening matrix may still vary. We can now formalize the opinion updating process in its general form as:

$$\mathbf{s}(t+1) = \mathbf{T}(t+1) \cdot \mathbf{s}(t) \tag{1}$$

where (1) can be also written as follows:[8]

$$\mathbf{s}(t+1) = \prod_{\tau=1}^{t+1} \mathbf{T}(\tau) \cdot \mathbf{s}(0) \tag{2}$$

The distinction between the listening matrix and the underlying network reflects the two key ingredients of the opinion updating process. The latter

[7] A network is said to be strongly connected if there is a directed path from any agent i to any other agent j in the network. A directed path from i to j is a directed sequence of distinct agents $(i_1, i_2, \dots, i_l)$ such that $i_1 = i$, $i_l = j$ and $G_{i_h, i_{h+1}} = 1$ for all $h \in \{1, 2, \dots, l-1\}$.

[8] As matrix multiplication is in general non–commutative, it should be clarified that we consider *backwards matrix products*, i.e. $\prod_{\tau=1}^{t} \mathbf{T}(\tau) = \mathbf{T}_t \cdot \mathbf{T}_{t-1} \cdot \dots \cdot \mathbf{T}_1$.

captures the communication structure within the population: *who listens to whom in each point in time?* The listening matrix adds to that the behavioral elements of opinion updating: *how much weight one assigns on her neighbors' opinions in each point in time?* The distinction between these two becomes particularly important in our experimental setup. While in the lab it is possible to control the shape of the communication structure (the network), it is not possible to control the weight put by each subject to others (the listening matrix). Compared to DeMarzo et al. (2003) who assume that the listening matrix and network remained fixed, we permit a very general communication process were both are permitted to vary across rounds. A critical restriction we retain in common is that the agents are assumed to use the same listening matrix in all issues during the same period, which implies that they assign the same relative weight to the opinion of a given individual in all issues of discussion.

2.2 Properties of an updating process

2.2.1 Unidimensional opinions

We say that opinions are unidimensional when the points describing each agent's opinion on the K dimensions all fall on a straight line that traverses $\mathbb{R}^K$. To formalize this idea we introduce some notation related to *principal component analysis* (PCA). In general, principal component analysis allows the projection of multidimensional data in fewer dimensions, in a way that most of the total variance is still captured, despite the reduced number of dimensions. In our case, the data correspond to the multidimensional opinions of the agents and the parameter of interest is $\beta^P(t) \in \left[\frac{1}{K}, 1\right]$, which is the percentage of total variance explained by the 1^{st} principal component, $\mathbf{PC_1}(t)$.[9] The major advantage of

[9]The calculation of principal components is done as follows: Let $\mathbf{\hat{s}}(t) = \mathbf{s}(t) - \mathbb{1}\,\mathbf{\bar{s}}(t)$, where $\mathbf{\bar{s}}(t)$ is an 1 x K vector containing the mean opinion in each dimension at time t and $\mathbb{1}$ is a N x 1 vector of ones. Let $\mathbf{PC_n}(t)$ be the eigenvector corresponding to the n-highest eigenvalue of the K x K covariance matrix of $\mathbf{\hat{s}}(t)$. $\mathbf{PC_1}(t)$ is the 1^{st} principal component of opinions at time t. Then, $\mathbf{s^P}(t) = (\mathbf{\hat{s}}(t) \cdot \mathbf{PC_1}(t))^T$ is the projection of agents' opinions on this principal component. Finally, let $\beta^P(t) \in \left[\frac{1}{K}, 1\right]$ be the percentage of total variance explained by the 1^{st} principal component. To calculate $\beta^P(t)$ one needs to calculate the projection of $\mathbf{\hat{s}}(t)$ on all principal components and take the covariance matrix of that. This is a diagonal matrix and $\beta^P(t)$ is the ratio of the first element of the diagonal over the sum of all diagonal entries. For a thorough discussion on principal

PCA is that it allows us to consider the relation between multiple dimensions at once, instead of performing bilateral comparisons. Intuitively, when $\beta^P(t)$ is close to one, this means that there exists a dimension (which is a linear combination of the dimensions corresponding to different issues) that can capture most of the observed disagreement between agents. When $\beta^P(t) = 1$ opinions are unidimensional. Thus, we define the following property:

Property 1. (Unidimensionality) An opinion formation process that can be described by (1) has the *Unidimensionality* property when

$$\lim_{t\to\infty} \beta^P(t) = 1$$

The *unidimensionality* property formalizes the idea that communication can lead to a unidimensional world (e.g., panels *B* and *C* of Figure 1).

2.2.2 Relative positions

Notice that *Unidimensionality* ensures that the relative positions of each agent with respect to all others will be the same on all issues. This comes as a direct result of the linear relation between opinions. Thus, once *Unidimensionality* is achieved relative positions on the different issues can be summarized by an agent's position on the line. Referring again to Figure 1, when comparing the opinions of individual 3 and 5, one can say that while in Panel B individual 5 has more extreme opinions than individual 3 the opposite holds in Panel C.

To summarize formally a population's relative positions we construct the $N \times N$ *opinion comparison matrix* $\mathbf{C^{m,n}}$ whose element $C^{m,n}_{i,j}$ is equal to 1 whenever i's opinion on the line relative to j's is concordant to m's opinion relative to n's, and equal to 0 otherwise.[10] If *unidimensionality* holds, the choice of the reference

component analysis see Jolliffe (2014).

[10]A formal definition would be as follows: Consider the following *relative comparison function*

$$c(x,y;m,n) = \begin{cases} 1, & if\ x > y \,\&\, m > n \\ & or\ x < y \,\&\, m < n \\ 0, & otherwise \end{cases}$$

The opinion comparison matrix $\mathbf{C}^{m,n}(t)$ has elements $\mathbf{C}^{m,n}_{i,j}(t) = c(s^P_i(t), s^P_j(t); s^P_m(t), s^P_n(t))$, where the pair (m, n) is inconsequential, resulting in the same comparison matrix up to transposition.[11]

At any point in time t, the opinion comparison matrix is determined via (1) by $\mathbf{s}(0)$ and $\mathcal{T}_t$, a set of parameters that grows infinitely with t. We will later show that in some cases much fewer information is sufficient for determining the opinion comparison matrix of the process in the long–run. In those cases the process is said to have *Position determinacy* which is the following property.

Property 2. (Position determinacy) An opinion formation process that can be described by (1) has *position determinacy* if:

$$\lim_{t\to\infty} \mathbf{C}^{m,n}(t) = \hat{\mathbf{C}}\,(\mathcal{T})\ , \quad \text{for any } \mathbf{s}(0) \in \mathbb{R}^K$$

where $\hat{C}$ is an *opinion comparison matrix* that depends only on a finite set of parameters $\mathcal{T}$, which is independent of $\mathbf{s}(0)$.

Therefore, if *Position determinacy* holds, then relative positions in the long-run converge in a way that is captured by the long–run *opinion comparison matrix* and do not depend on the initial opinions.

2.2.3 Direction of disagreement

Finally, we turn to the determinacy of the line's direction. The way a line traverses the K-dimensional space represents how much of the disagreement in opinions can be attributed to each dimension. For instance, if $K = 2$, a line perpendicular to one of the axes, means that all agents agree on that dimension and all disagreement comes from the other dimension. It also informs about the sign of the correlation of opinions between pairs of issues. The direction of the line is given by the 1^{st} principal component of opinions at time t, $\mathbf{PC}_1(t)$.

arguments of function c are denoted as defined in principal component analysis. It follows from (2) that an opinion comparison matrix is a function of the sequence of listening matrices and initial opinions: $\mathbf{C}^{m,n}(t) = C(\mathcal{T}_t, \mathbf{s}(\mathbf{0}))$.

[11]Relative positions can be summarized by other, perhaps simpler, measures. The advantage of the opinion comparison matrix for our study will come to light in the analysis of the experimental data. There it allows for a direct comparison of relative positions of pairs of agents in different treatments.

It is easy to see that in a process like (1), if *Unidimensionality* holds then once opinions are unidimensional the direction of the line cannot change. Moreover, notice that at each point in time t $\mathbf{PC}_1(t)$ is determined via (1) by $\mathcal{T}_t$ and $\mathbf{s}(0)$. Similarly to the previous property, we will later show that in some cases fewer information will be sufficient to determine the direction of disagreement in the long–run. In those cases the process will be said to have *Direction determinacy* which is the following property.

Property 3. (Direction determinacy) An opinion formation process that can be described by (1) has *Direction determinacy* if:

$$\lim_{t\to\infty} \mathbf{PC_1}(t) = \mathbf{W}(\mathcal{T}, \mathbf{s}(0))$$

where $\mathbf{W}$ is a vector that depends on a finite set of parameters $\mathcal{T}$ and $\mathbf{s}(0)$.

Unlike for *Position determinacy*, in the case of *Direction determinacy* initial opinions still play a role, as these set constraints on what directions are achievable. For example if there is no difference in the initial opinions on one dimension, then this cannot change in the long run.

2.3 Theoretical results

The process described by (1) has been shown under a broad range of conditions to lead to consensus in all issues, i.e. in the long–run all agents end up having common opinions in each issue.[12] However, despite the fact that opinions converge, at each point in time there is disagreement. *Unidimensionality* implies that as opinions evolve and before they fully converge this disagreement can be summarized in a single dimension, which is a linear combination of all issues.

Before stating the results, we need to introduce some additional terminology and notation. Namely, for any listening matrix $\mathbf{T}$ we can calculate and rank its eigenvalues.[13] We denote by α_2 the second largest eigenvalue of this matrix.

[12]In our case, the conditions that the network is strongly connected with all agents putting strictly positive weights to themselves and to all their neighbors in each period is sufficient to obtain the result with a direct adaptation of the proof by Tahbaz-Salehi and Jadbabaie (2008).

[13]The eigenvalues are ranked according to their modulus, as they might be complex numbers. The modulus $\|\alpha\|$ of a complex number $\alpha = a + ib$ is $\|a + ib\| = \sqrt{a^2 + b^2}$.

Moreover, recall that the network is considered to be strongly connected, with each agent putting a strictly positive weight both to her own opinion and to that of each one of her neighbors.[14]

Based on these, DeMarzo et al. (2003) study a particular class of such updating processes where agents are only allowed to change the weight they put on their own opinion across periods. Namely:

$$\mathbf{T(t)} = (1 - \lambda(t))\,\mathbf{I} + \lambda(t)\mathbf{T} \tag{3}$$

where $\lambda(t) \in (0, 1]$ and $\mathbf{T}$ is a listening matrix that remains fixed. It turns out that as long as the agents do not become "too stubborn, too early"[15] the analysis can be concentrated on the properties of $\mathbf{T}$ and the following result is obtained:

Theorem 1 (restatement of Theorem 4, DeMarzo et al. (2003))**.** *Consider a generic listening matrix* $\mathbf{T}$ *with* $\alpha_2 \in \mathbb{R}$*. Then, the opinion process described by (3) satisfies Unidimensionality, Position determinacy and Direction determinacy.*

In the original article the authors provide the exact relationship between $\mathbf{T}$ and long run relative positions, as well as how $\mathbf{T}$ and $\mathbf{s}(0)$ determine the direction of disagreement. In fact, the result is proven by showing that relative opinions stabilize in the same way across all issues and therefore they must be unidimensional. Note that this argument cannot be extended to cases where the relative importance of neighbors' opinions change, since then even if opinions become unidimensional the relative positions may keep changing.

In addition to this, assuming T to be constant, apart from being very restrictive for the agents' behavior, it is also not easily testable. However, *Unidimensionality* turns out to be a more general property that emerges even for the general class of process described in expression (1), where the listening structure is allowed to change in a very general way. Namely:[16]

[14]This standard assumption is essential for the analysis, as it allows us to interpret the listening matrix $\mathbf{T}$ as the transition matrix of a finite, irreducible and aperiodic Markov chain.

[15]The formal necessary condition for the results is that $\sum_{t=1}^{\infty} \lambda(t) = +\infty$.

[16]The result is proven in terms of perfect correlation between two arbitrary issues, therefore it can be directly extended to K–dimensional opinions where perfect correlation is achieved between each pair of issues. Perfect correlation is conceptually identical to *Unidimensionality* as explained via PCA. Nevertheless, it should be noted that in general there is no one–to–one

Theorem 2. *Consider a sequence of generic listening matrices, $\mathcal{T}_\infty$, with $\alpha_2(t) \in \mathbb{R}$ for all t, and a matrix of initial opinions* $\mathbf{s}(0)$*. Then, the opinion process described by (1) satisfies Unidimensionality.*

Notice that this result does not specify the way the elements of the sequence are selected, but only their properties. This means that the matrices could be determined either randomly or deterministically, as long as they satisfy the necessary assumptions. It therefore suggests that unidimensional worlds should be the norm. As long as the repeated averaging updating process is an accurate description of opinion dynamics, we should expect to see such worlds arise. No further restrictions on the structure of communication are necessary.

For clarification, in both theorems the term *generic* means that the listening matrices are *diagonalizable* and that there are *no ties* in the ranking of eigenvalues, except for complex conjugates. Both of these properties are generically satisfied and made here for technical reasons. Diagonalizability allows us to rewrite opinions as a combination of eigenvalues and eigenvectors, whereas 'no ties', together with the second eigenvalue being real, ensure that the impact of eigenvectors corresponding to smaller eigenvalues on opinion differences will die out earlier, thus allowing the emergence of unidimensionality. The condition requiring the second eigenvalues to be real is more intuitive as it rules out cases where the listening matrix is dominated by a one–way cycle that can prevent the emergence of unidimensionality. In fact, our theorem would still hold if these two requirements hold for all but finitely many elements of the sequence.

Taken together, Theorems 1 and 2 can be summarized as follows: a) communication and social influence lead to the emergence of a unidimensional world, b) a stable underlying structure of social interactions determines the relative positions of agents in this world, c) combining this structure with the starting point of the process determine the direction of disagreement. These theoretical predictions are not only provokingly strong, but also surprisingly crisp. Yet, they depend crucially on the assumption that agents update their opinions through averaging. The literature already supports the idea that this is a good approx-

relation between correlation and the percentage of variance explained by PCA, despite the two measure being intuitively similar.

imation: individuals seem to update their opinions *as if* taking the average of their neighbors' opinions. What interests us is whether the predictions about the shape of disagreement in a population, stemming from the formal averaging model, carry over to an empirical setting. We use a lab experiment to answer this question.

3 Experimental Design

In the experiment we aim to recreate the stylized communication setup of the model. What we cannot control is the way subjects in the lab update their opinions after communication. As mentioned, this is a crucial assumption in the theory of the emergence of unidimensional worlds. Therefore, the experimental exercise tests the robustness of our theoretical predictions to the behavioral elements of the model.

We use three different experimental treatments. In all treatments, subjects repeatedly communicate their opinions about an unknown two-dimensional state. In the baseline treatment, communication takes place in a fixed network structure, i.e. subjects can listen to the same other subjects in each round. In a second treatment, the network remains fixed but is different than the one in the baseline treatment. In the third treatment, the network changes randomly in each round of communication. In what follows we describe the way we induce and elicit opinions, how subjects communicate, and give more detail about the network structures used in the different treatments.

3.1 The experimental task

The main task during the experiment is a non-competitive guessing game presented in the following form:

In a tank there are 100000 balls. These balls are either RED or BLUE. The number of balls of each colour is random and any combination is equally likely. You are asked to guess the number of RED balls in the tank. This number could be anywhere between 0 and 100000. Before making your guess, you observe a sample of 100 balls picked *randomly from the tank.*

The number of red balls represents an unknown state and the guess represents the subject's opinion about what this state is. For the main experiment we ask subjects to make guesses about two different tanks simultaneously, thus obtaining a two-dimensional state and respective opinions. The high number of balls is chosen so as to avoid the necessity of decimal numbers for subjects to give finer guesses.

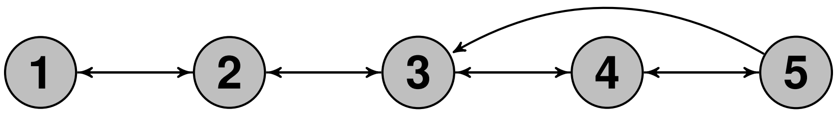
3.2 The experiment

Subjects play the guessing game for three phases, consisting of 10 rounds each.[17] In each phase they are playing two guessing games simultaneously: there are two tanks, one with red and blue balls, and one with green and orange balls. The state and private signals for each of the two parallel games are drawn independently. In each round subjects enter two guesses, one for each tank and observe both guesses for each of their neighbors.

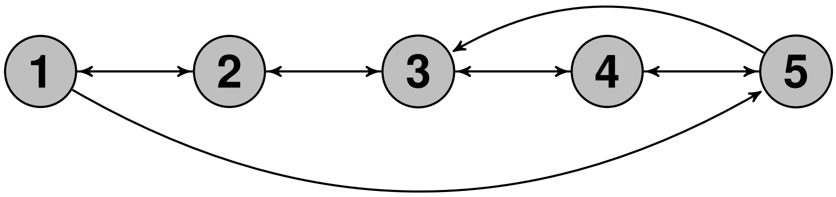
We use three **treatments:** ***Fixed 1*** **(F1),** ***Fixed 2*** **(F2)** and ***Random*** **(R)** as depicted in Figure 2. What varies across treatments is the network structure and its stability. In treatments *Fixed 1* and *Fixed 2* the network remains fixed but is different in each one (Figures 2a and 2b). *Fixed 1* serves as our baseline treatment. The network in *Fixed 2* is minimally different than the baseline: it is obtained by adding a single directed link to the baseline. In treatment *Random* the network structure changes randomly in each round of communication (Figure 2c). Each node observes the same number of neighbors as the corresponding node in the baseline. The identity of these neighbors is drawn randomly in each round. We explain the choice of the exact network structures at the end of this section, as it will be facilitated by the presentation of our research hypotheses.

In the two treatments where the network structure is fixed, the identity of each subjects' neighbors (group members whose guesses she observes) remains

[17] The main experiment was preceded by two parts that aimed at familiarizing subjects with the information and communication environments. In the first part subjects made guesses for a single tank without communication. In the second part subjects could communicate but again made guesses for only one tank. Instructions for all parts can be found in the appendix.



(a) *Fixed 1*



(b) *Fixed 2*

(c) *Random*

Figure 2: Treatments. The graphs represent the 5-node network structure used for communication in each treatment. An arrow from one node to another means that the latter listens to the former. In treatment *Random* each node has the same number of neighbors as the corresponding node in *Fixed 1*, but these change randomly in each round.

fixed throughout the experiment and subjects are informed that this is so. Together, the sets of neighbors for each subject in a group form a directed network. Subjects are not informed about the structure of their group's network. They are told that observing another group member's guess does not mean that that group member can observe their own guess. In the treatment with a random network subjects are informed that the number of neighbors they observe remains fixed, but a new set of neighbors is drawn randomly in each round. Neighbors are always drawn from within the same 5-member group.

Initial opinions in the experiment are induced by providing each subject at the beginning of each phase with a 100-ball sample from each tank. Each tank

k contained a number θ_k of target balls (red for tank 1, green for tank 2). Each θ_k is drawn from a uniform distribution over $\{0, 1, 2, ..., 100, 000\}$. The samples were i.i.d. draws from a binomial distribution with parameters $n = 100$ and $p = \frac{\theta_k}{100{,}000}$. For each group there was a set of $2 \times 3 = 6$ draws for θ_k (2 for each of the 3 phases) and $5 \times 2 \times 3 = 30$ samples (one 2 for each of the 5 group-members in each phase). Across the experiment we used 3 such sets in approximately equal proportions in each treatment.

3.3 Logistics

The experiment took place at the Lancaster Experimental Economics Laboratory hosted at the Department of Economics at the Lancaster University Management School (LUMS). A total of 180 subjects were recruited among LUMS students.[18] In total we had 12 groups for each treatment.

Final earnings were determined by selecting randomly the payoffs in one of the three phases in part 1, one of the five rounds in part 2 and one of ten rounds for each of the three phases in part 3. Subjects received an additional £3 participation fee. Average total payment was around £10 and the experiment lasted about 90 minutes.

3.4 Hypotheses

We start by forming preliminary hypotheses about the implicit assumptions of our model. Namely, whether subjects do update their opinions in a way that resembles averaging over one's neighbors' opinions and whether they do this in the same way for both dimensions. Notice that if the former holds then opinions need to come closer over time, something that would be reflected on a diminishing coefficient of variation. If the latter holds, then the decrease of the coefficient of Variation should be the same across the two dimensions. We will therefore first test the following two hypotheses:

[18]There were 15 sessions in total. Most sessions had 15 participants, except for 1 with 10 subjects and 3 with 5 subjects due to low turnout. Initially 190 students were recruited. During one of the sessions a technical issue affected the play of two groups (10 subjects) altering the intended treatment. We excluded this data from further analysis.

Hypothesis 0 (a). *The coefficient of variation of guesses across a single dimension is decreasing across all rounds.*

Hypothesis 0 (b). *The coefficient of variation of guesses across a single dimension moves the same way for both dimensions.*

Following Theorem 2, *Unidimensionality* should obtain in all treatments. That is, as long as subjects update their guesses for both dimensions by taking some form of average of their own and others' guesses of the previous round, these guesses should get aligned. Whether the network is fixed or not should not matter. We therefore consider the following hypothesis:

Hypothesis 1. *In all treatments the variance of guesses explained by the first principal component converges to 1.*

Following Theorem 1, it should be possible to predict relative positions if we knew the listening matrices used and these were fixed. But we cannot know the exact weight each individual puts on each of the others' guesses, neither whether such a weight – if in fact it exists! – remains fixed. What the lab environment allows us to do is to restrict the listening matrices each group can use by fixing a particular network structure – like in treatments *Fixed 1 & 2* – or force it to change in each round – like in treatment *Random*.

Without making any further assumptions on the specific listening matrices used by subjects in the experiment we can still formulate an hypothesis about the treatment effects on relative positions.

Hypothesis 2 (a). *In treatments* Fixed 1 *and* Fixed 2, *agents' relative positions as projected on to the long-run opinions' first principal component are determined by the agents' positions in the network and therefore converge to a specific opinion comparison matrix:* $\hat{\mathbf{C}}_{F1}^{m,n} \neq \hat{\mathbf{C}}_{F2}^{m,n}$.

In other words, we expect relative positions to become the same within each treatment and expect to see differences across treatments. In treatment *Random* Theorem 1 does not apply, as the listening matrix cannot remain the same. Relative positions can therefore not be determined by each subject's label.

Hypothesis 2 (b). *In treatment* Random *agents' relative positions as projected on to the long-run opinions' first principal component are random.*

To obtain more crisp predictions about subjects' relative positions after some rounds of communication we look at the theoretical predictions that follow the assumption that subjects put an equal weight to their own and every other neighbor's opinion. While strong, the assumption is also natural, since subjects in the experiment are not aware of the network structure and therefore all neighbors are identical to them. The listening matrices $\mathbf{T}_{F1}$ and $\mathbf{T}_{F2}$ would then be equal to:

$$\mathbf{T}_{F1} = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 & 0 \\ 1/3 & 1/3 & 1/3 & 0 & 0 \\ 0 & 1/4 & 1/4 & 1/4 & 1/4 \\ 0 & 0 & 1/3 & 1/3 & 1/3 \\ 0 & 0 & 0 & 1/2 & 1/2 \end{pmatrix} \qquad \mathbf{T}_{F2} = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 & 0 \\ 1/3 & 1/3 & 1/3 & 0 & 0 \\ 0 & 1/4 & 1/4 & 1/4 & 1/4 \\ 0 & 0 & 1/3 & 1/3 & 1/3 \\ 1/3 & 0 & 0 & 1/3 & 1/3 \end{pmatrix}$$

The difference between these two listening matrices is in the last row representing the weight that agent 5 assigns to his neighbors. While in *Fixed 1* she is only observing the guesses of agent 4, in *Fixed 2* she is also observing the guess of agent 1.

Relative positions are determined by the ranking of agents' corresponding element in the second row eigenvector of $\mathbf{T}$.[19] For $\mathbf{T}_{F1}$ it is equal to $V_2^c(T_1) = (-2.5, -1.667, 0, 0.667, 1)^T$, which means that the relative positions for the five agents is $(1, 2, 3, 4, 5)$ (or $(5, 4, 3, 2, 1)$). The second row eigenvector of $\mathbf{T}_{F2}$ is equal to $V_2^c(T_2) = (-4.5, -1.068, 3.585, 5.356, 1)^T$, which means that the relative positions of the five agents from extreme left to extreme right will be $(1, 2, 5, 3, 4)$ (or $(4, 3, 5, 2, 1)$).[20]

Intuitively, the more access an agent has to the opinions of others, the more moderate she becomes. Agents with extreme opinions tend to be those who do not have access to a lot of information. This intuition is supported by computer simulations. Therefore, while in *Random* it is not possible to predict on which

[19] For details see DeMarzo et al. (2003).

[20] Notice that these are also the relative positions of individuals in the examples of unidimensional worlds shown in Figure 1: in panel B for $\mathbf{T}_{F1}$ and in panel C for $\mathbf{T}_{F2}$.

side of the unidimensional spectrum one will end up, we can make predictions about the likelihood of a specific agent being more or less extreme than others. In particular, subjects labeled 1 and 5 are expected to be the most extreme, then 2 and 4, and finally 3 is expected to be the most moderate. Using computer simulations we can calculate a specific expected opinion comparison matrix $E_R\left[\hat{\mathbf{C}}^{m,n}\right]$.

We summarize our predictions from assuming equal weights in the following hypothesis:

Hypothesis 2 (c). *Agents' relative position as projected on to the long-run opinions' first principal component will converge to the following opinion comparison matrices:*

$$\hat{\mathbf{C}}^{1,3}_{fixed\ 1} = \begin{pmatrix} 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \qquad \hat{\mathbf{C}}^{1,3}_{fixed\ 2} = \begin{pmatrix} 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \end{pmatrix}$$

$$E_{random}\left[\hat{\mathbf{C}}^{1,3}\right] = \begin{pmatrix} 0 & 0.82 & 1 & 0.82 & 0.79 \\ 0.18 & 0 & 0.54 & 0.54 & 0.56 \\ 0 & 0.46 & 0 & 0.45 & 0.55 \\ 0.18 & 0.46 & 0.55 & 0 & 0.57 \\ 0.21 & 0.44 & 0.45 & 0.43 & 0 \end{pmatrix}$$

Notice that we use $m = 1$ and $n = 3$ for the opinion comparison matrices. The predicted matrices for *Fixed 1* and *Fixed 2* are not affected by this choice, but we would get slightly different numbers for the prediction for *Random*. We choose these for the construction of the opinion comparison matrices because for the case of fixed networks subjects at these nodes remain at the same distance in both networks and are the furthest apart from all such pairs of nodes. This makes the exercise less sensitive to noise.

Because of Direction determinacy we should expect to see groups within the same fixed network treatment that start off from the same initial constellation of guesses to converge to the same line. This is captured by the rotation of the first principal component with respect to the horizontal axis, measured in radians (rads). At zero rads all disagreement is with regard to guesses about tank 1. At $\pi/2$ rads the variance of guesses in the two dimensions is the same. Negative values reflect the fact that guesses are inversely correlated.

Hypothesis 3. *In fixed network opinion dynamics, the rotation of the first principal component of long-run guesses is determined by the network's structure and the initial constellation of guesses.*

It should now be clear why we chose these particular network structures for our experiment. For the baseline structure we had two desiderata: i) it should be simple so that theoretical predictions regarding subjects relative positions in the opinion space can be directly traced back to their position in the network; ii) it should be possible to minimally alter the baseline and obtain a substantially different prediction for relative opinions, in order to test our hypothesis of the direct link between the ordering and the network structure. A simple undirected linear network would satisfy the first condition but not the second: adding a single directed link to the line, like in *Fixed 1*, gives the same predicted ordering. By adding to *Fixed 1* a single directed link from node 1 to node 5 we obtain *Fixed 2*, where predicted relative positions are now different. This satisfies our second condition, providing the desired testbed for our ordering hypothesis.

4 Results

Figure 3 shows the evolution of opinions across rounds in a phase for one of the groups in the experiment. There is a high degree of heterogeneity in what we observe, but this example is shown to facilitate the understanding of the different measures we use to summarize the data. It also showcases some features that characterize the data set.

First of all, guesses become closer over time, as captured by the *normalized coefficient of variation (NCV)*.[21] Most of this convergence happens in the first five rounds. In this example we also see the group's guesses quickly align, captured

[21] The coefficient of variation of guesses for one tank is $CV(t) = \frac{st.dev.(guesses\ at\ t)}{mean(guesses\ at\ t)}$. We then report the normalized coefficient $NCV(t) = \frac{CV(t)}{CV(1)}$.

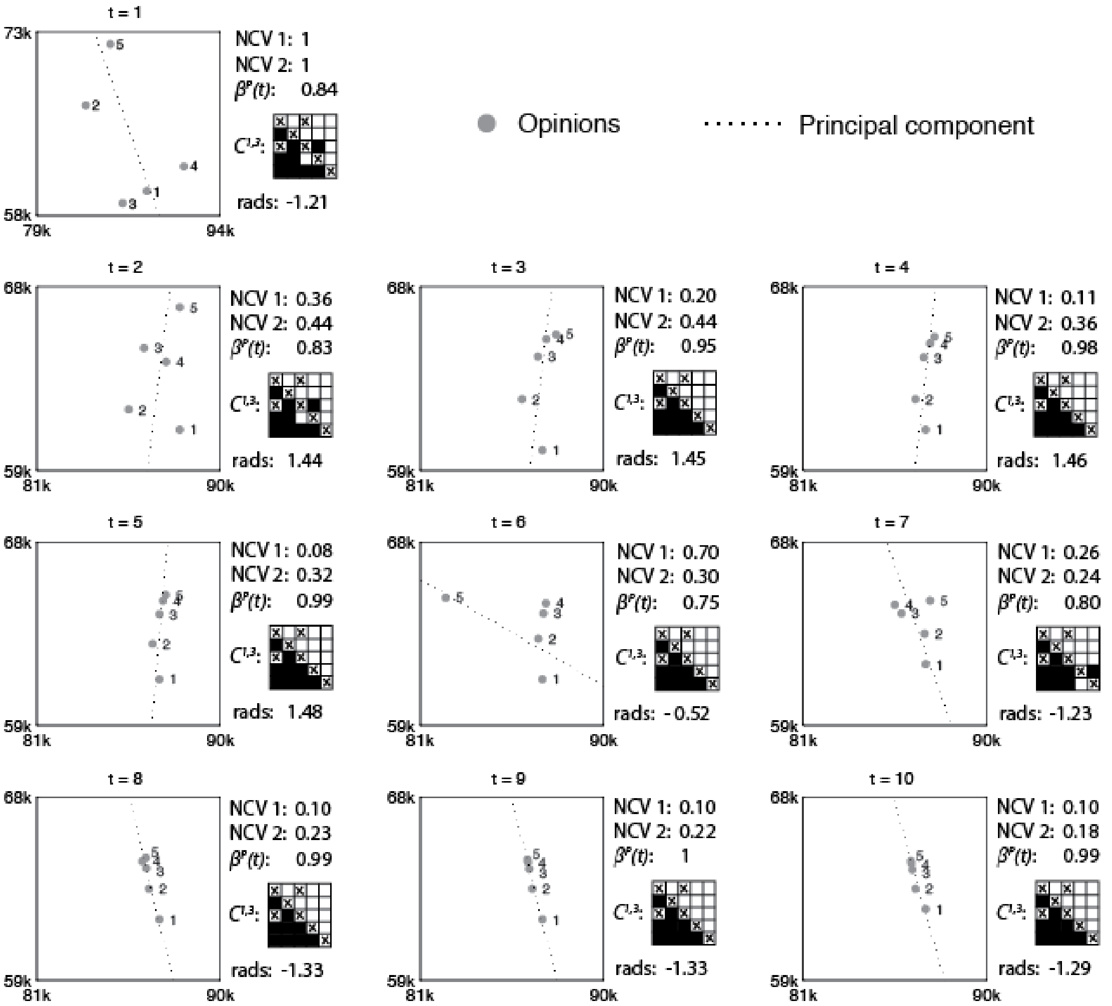


Figure 3: *The evolution of opinions.* Guesses across rounds in a phase for a particular group in treatment *Fixed 1* in the experiment (*session 6, group 2, phase 2).* Each graph represents guesses in each round t, from 1 to 10. Each point represents a subject's guesses (in thousands) for each tank. Tank 1 in the horizontal axis and tank 2 in the vertical axis. Labels 1 to 5 refer to subjects positions in the network. The dashed line traces the first principal component. NCV 1 and 2 is the coefficient of variation of guesses for the respective tank, normalized to be 1 in $t = 1$. $\beta^P(t)$ is the variance explained by the first principal component in t. The opinion comparison matrix $C^{1,3}$ is represented as a 'heat map', with white for entries equal to 1, black for entries equal to 0. An X is used for the main diagonal and elements $c_{1,3}$ and $c_{3,1}$, i.e. the elements that are determined by definition and remain the same across all rounds. *Rads* refer to the angle of rotation of the first principal component with respect to the horizontal axis, which measures the direction of disagreement.

by the high percentage of variance explained by the first principal component ($\beta^P(t)$). The relative positions of subjects' projection on this line also converges very quickly and follows the order of their labels: 1 next to 2, next to 3 etc. This is captured by the opinion comparison matrix $C^{1,3}$, represented here graphically for easier comparison. The process of convergence, as captured by all these different measures is interrupted in round 6, where subject 5 makes a guess for tank 1 that is far from her own and others' previous guesses. "Jumps" or perturbations like this one occur somewhat frequently in the data and are sometimes of much greater magnitude. Interestingly, we observe that the process of convergence picks up again immediately, only now on a line 'tilted' by the jump.

4.1 Convergence and Perturbations

The three upper panels of Figure 4 show the average NCV per round of guesses for each tank in each treatment. What immediately stands out is the jump in the value for tank 2 in round 3 of treatment *Fixed 1* that jumps up to 150%. This is mostly driven by a particular case where the group's NCV for tank 2 jumped up to 2653%. Jumps like that, although smaller in magnitude are common in the data. See the example in Figure 3 for such an instance in $t = 6$. Some of those, especially the biggest in magnitude, can be attributed to 'mistakes', such as mistyping one's guess. Others may be deliberate, although there does not seem to be some systematic pattern of behavior to explain them.[22] We do observe that such perturbations are less common in *Fixed 2*. This can be seen in Figure 4 by noticing that the mean and median for the NVC in *Fixed 2* are closer than in the other two, for both dimensions, which is evidence of a distribution with fewer extreme values.

Irrespectively of what causes them, these perturbations can be useful in our study. Individuals' opinions in real life may also be subject to shocks. Even if their causes are different from what makes subjects in the lab "jump",

[22] One reason for such perturbations could be that a subject tries to hedge by making guesses across a reasonable range of values, even if that is not optimal. Another reason could be boredom, which can affect subjects' behavior in repetitive experiments as this and may lead to arbitrary choices. The observation that boredom may lead to random responses in guessing games was first made by Siegel (1961).

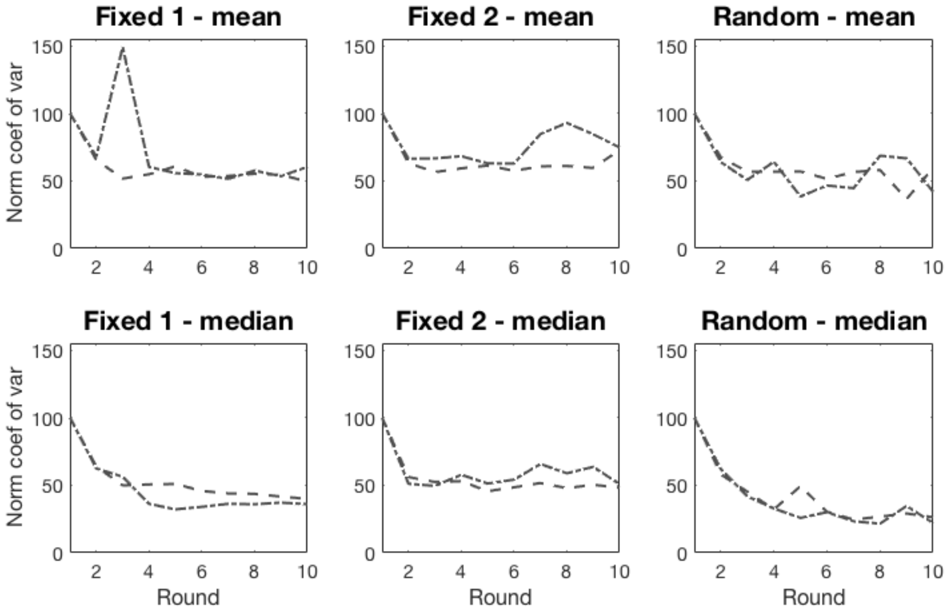


Figure 4: *Normalized coefficient of variation.* The graphs show the mean and median NCV values per round in each treatment. *Dashed lines* correspond to the NCV of guesses for *tank 1* and *dot-dashed lines* correspond to the same for *tank 2*.

having this feature in the lab can be informative about the robustness of the unidimensionality properties to similar "noise".

The lower panels of Figure 4 show again the median NCV in each round, for each tank, in all three treatments. The pattern of convergence can be seen much better here. We do not observe any systematic differences in the convergence pattern between the two dimensions in each treatment. Across treatments we observe that convergence appears to last longer in *Random*, where it also reaches higher levels (lower NCV). In all treatments, we see that NCV decreases mostly in the first five rounds and remains rather flat in the last five rounds.

At this point we can state the following results regarding Hypothesis 0 (a) and (b).

Result 0 (a)**.** *Hypothesis 0 (a), stating that the coefficient of variation of guesses across a single dimension is decreasing across all rounds, cannot be rejected.*

Support: Figure 4 provides some graphical support for this result based on aggregate data. Based on the non-parametric seasonal Mann-Kendall test for trend, we can reject the hypothesis that there is no trend in the coefficient of variation series for guesses for Tank 1 ($p < 0.001$) and Tank 2 ($p < 0.001$) in *Fixed 1*, for Tank 1 ($p < 0.001$) and Tank 2 ($p < 0.001$) in *Fixed 2*, and for Tank 1 ($p < 0.001$) and Tank 2 ($p < 0.001$) in *Random*.

Result 0 (b). *Hypothesis 0 (b), stating that the coefficient of variation of guesses across a single dimension moves the same way for both dimensions, cannot be rejected.*

Support: Figure 4 provides some graphical support for this result based on aggregate data. In all three treatments, the NCV decreases significantly over the first 5 rounds and then stabilizes in both dimensions. A Wilcoxon signed-rank test comparing pairs of groups' NCV for each dimension in each round of each treatment does not reject the null that these are different in any but two instances: in round 7 of treatment *Fixed 2* ($p = 0.053$) and in round 5 of treatment *Random* ($p = 0.04$). Given the high number of comparisons (3 treatments $\times$ 9 rounds = 27 comparisons) it is expected to obtain some false positives. Applying any of the standard corrections for multiple testing will render both these cases non-significant even at the 10% level.

4.2 Unidimensionality

We now turn our attention to the first of our main research questions: can communication lead to unidimensional opinions? Recall from Theorem 2 that this should be true for an arbitrary sequence of listening matrices and hence, as stated in Hypothesis 1, we should observe evidence of this in all three experimental treatments.

The left panel of Figure 5 shows the mean β^P, which is the percentage of variance in the guesses explained by the first principal component, in each round for each of the three treatments. The right panel of the figure shows the median for the same value.

Result 1. *Hypothesis 1, stating that in all treatments the variance of guesses explained by the first principal component converges to 1, cannot be rejected.*

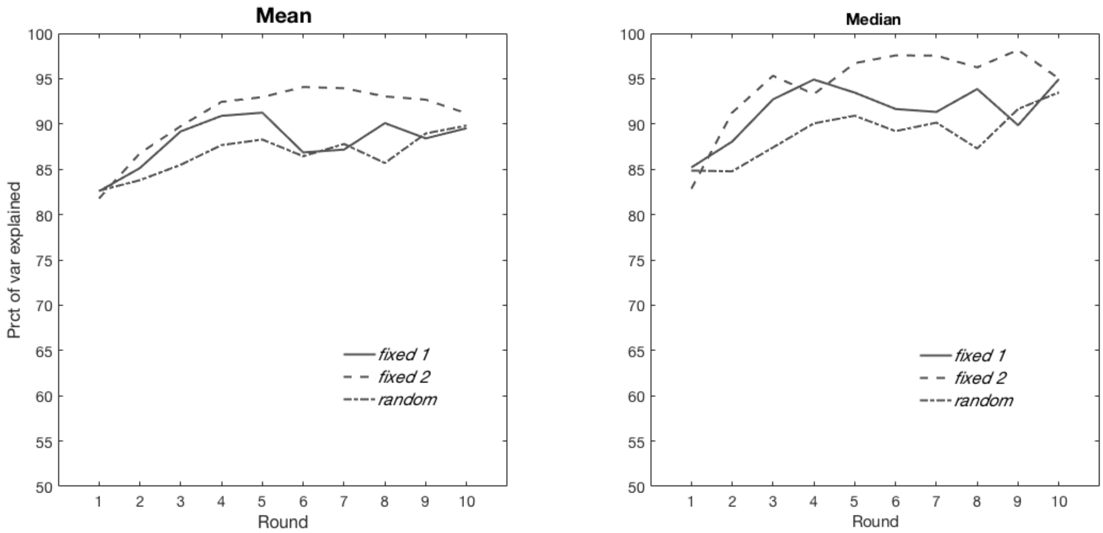


Figure 5: *Percentage of variance in the data explained in the data, across round and treatment.* The left panel shows the mean percentage of variance explained by the first principal component, β^P, across all group observations in a given treatment. The right panel does the same for the median. By definition β^P lies between 0.5 and 1.

Support: Figure 5 provides some graphical support for this result based on aggregate data. More conclusively, based on the non-parametric seasonal Mann-Kendall test for trend, we can reject the hypothesis that there is no positive trend in the series for β^P in *Fixed 1* ($p < 0.001$), in *Fixed 2* ($p < 0.001$) and in *Random* ($p = 0.002$).[23] Furthermore, the median group has a $\beta^P(t)$ of 96.9 on average in rounds 6 to 10 in *Fixed 1*, and the same value is 90.4 for *Fixed 2* and 92.3 for *Random*.

There are a few things to note with respect to this result. First of all, while it holds for all treatments, convergence to unidimensionality is clearly stronger in *Fixed 2*. As we mentioned above, in this treatment there is less noise due to perturbations, which could explain this difference. Second, it is clear from the graphs that the increase in β^P happens mostly in the first five rounds. This is confirmed by running the seasonal Mann-Kendall test for trend but restricting the data to the respective rounds. For rounds 1 to 5 we can strongly reject the null that there is no positive trend in all three treatments ($p < 0.001$ for all three

[23] Note that the seasonal Mann Kendall test uses information from individual group observations, not just the aggregate data shown in Figure 5.

tests). For rounds 6 to 10 we cannot reject the null of no trend in any of the treatments ($p = 0.484$ in *Fixed 1*, $p = 0.134$ in *Fixed 2*, $p = 0.388$ in *Random*). This seems to reflect what we observed for general convergence as measured by NCV, which also seems to happen mainly in the first 5 rounds.

Notice that the mean (and median) β^P in round 1 seems relatively high, which may lead to the concern that the high levels achieved in subsequent rounds are due to this. Recall that in this round subjects have not yet observed any of their neighbors guesses. The high β^P can therefore only be attributed to the random draw of private signals used in the experiment. Following standard experimental procedures, the draw was kept random, as explained in the instructions, and we did not make any selection of specific signal sets. More to the point, there is no significant correlation between a group's $\beta^P(0)$ and its average β^P for the last 6 rounds. This means that the increasing trend we observe is not a result of the (on average) high β^P induced by the initial draw of signals.

4.3 Relative Positions

To compare subjects relative positions we rely on the *opinion comparison matrices*, $C^{1,3}_{i,j}$. This method allows us to focus on the pairwise comparisons of subjects and is therefore less sensitive to the perturbations discussed earlier than methods based on rank correlation. It does require to make a choice of a reference pair (m, n) and as noted in the previous section we choose $m = 1$ and $n = 3$, as this choice seems the most sensible and robust to perturbations given our design. Still, in theory any choice should give the same results in the absence of noise, and indeed we do find qualitatively similar results when trying different pairs.

By definition, any matrix $C^{m,n}_{i,j}$ satisfies the condition $C_{i,j} + C_{j,i} = 1$ for all $i \neq j$. The elements of the diagonal are 0 and the elements (m, n) and (n, m) are 1 and 0 respectively, both again by definition. It therefore suffices to look at the remaining elements below the diagonal to have a complete picture of the relative positions in the group. For the groups in our experiments these are the following 9 elements: (2,1), (3,2), (4,1), (4,2), (4,3), (5,1), (5,2), (5,3), (5,4). Figure 6 shows the average value for each of those elements across rounds for all observations in each treatment.

If Property 2 holds, then these values should converge to either 0 or 1. Which elements go to 0 and which go to 1 will depend on the specific listening matrix. This means that, even without any assumptions on the shape of the listening matrix, we would expect in the fixed network treatments the graph for each element to converge to one or the other extreme, and no such convergence for *Random*. Furthermore, we would expect to see a different convergence pattern –different elements converging to different extremes– in *Fixed 1* and *Fixed 2*. The graphs on the left in figure 6 pretty much conform to these expectations.

In particular we see that all elements in *Fixed 1* converge to average values below 0.5. In *Fixed 2* the same is true for all elements except (5,3) and (5,4). Recall from Hypothesis 2 (c) that assuming equal weights in the listening matrix, it is precisely in these two elements where we should observe a treatment effect across the two fixed network treatments. This is depicted on the right side of figure 6. Alas, the exact prediction would be for these two elements to converge to 1, which we do not observe here. Still, their average values do seem to remain above 0.5. Notice that the difference between the treatments is the simple addition of one link from node 1 to node 5, allowing the latter to observe the former's guess. If the weight put by node 5 to node 1's guess is very low, then relative positions should not be different between treatments. Instances like that could explain why these elements' average value is far from 1.

In *Random* we do not observe any tendency for elements to converge to extreme values. In fact, on average we find that elements take values very close to the prediction obtained by assuming equal weights in the listening matrix.

Figure 7 shows heat-maps representing *average opinion comparison matrices* for each treatment in specific rounds. These allow for an easy visual comparison of subjects' "average" relative positions in each treatment across rounds. We aggregate data from rounds 6 to 10 to "average out" perturbations. While the differences between *Fixed 1* and *Fixed 2* are less clear in this figure, it does show the difference between relative position patterns in the fixed network treatments and in *Random*. It also shows how close to the theoretical prediction (under the assumption of equal weights) the relative positions in *Random* are.

Result 2. *We find support for Hypothesis 2 (a) and (b) and partial support for Hypoth-*

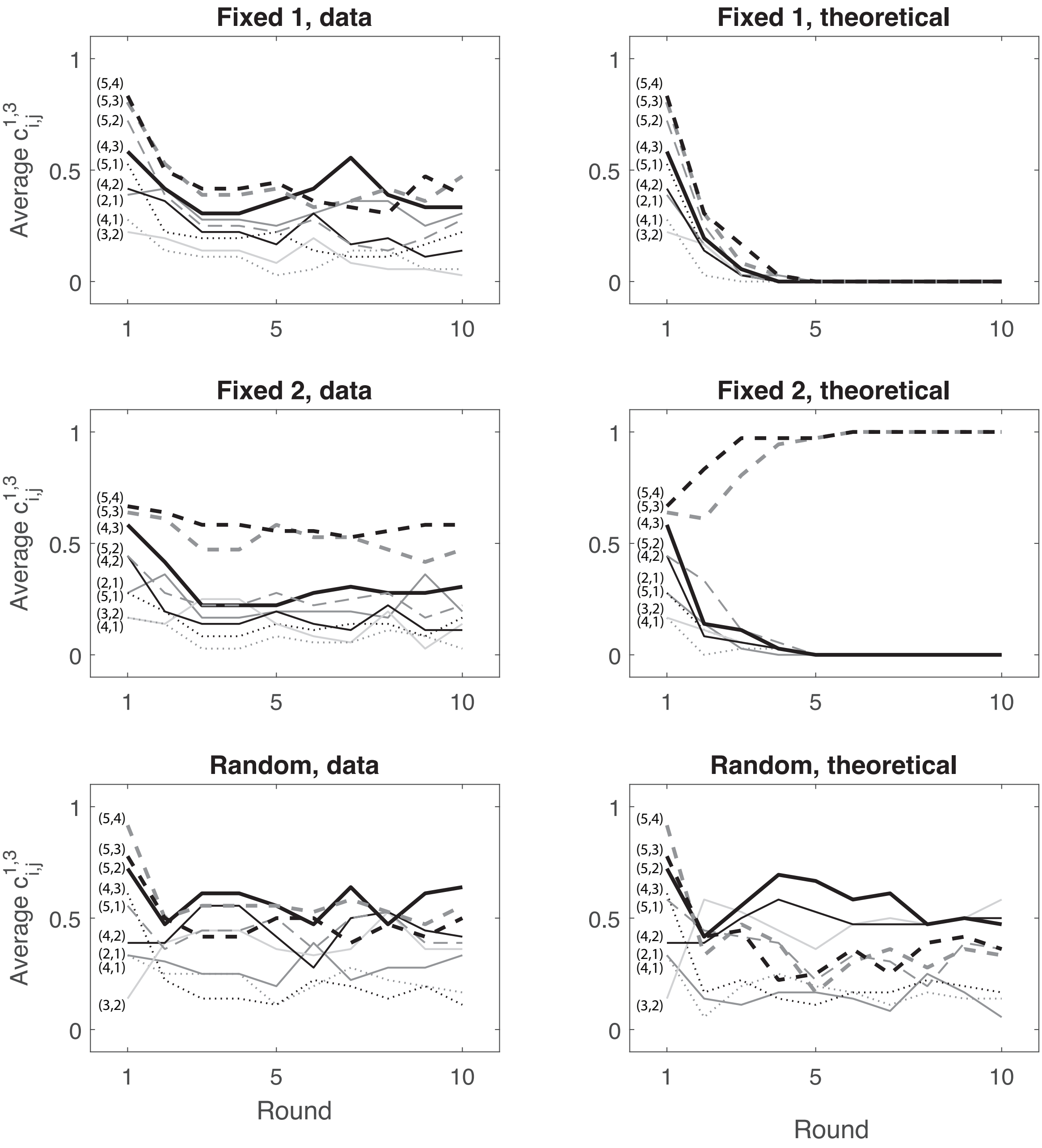


Figure 6: *The evolution of relative positions in each treatment.* Each line corresponds to an element below the diagonal of the opinion comparison matrix, specified by the pairs in parenthesis on the right of each graph. Values indicate the average value for that element across all groups in each treatment in a given round. The graphs on the left show the values obtained in the experiment. The graphs on the right show how these values would evolve if subjects put equal weights to all of their neighbors and their own previous guesses, and without any noise.

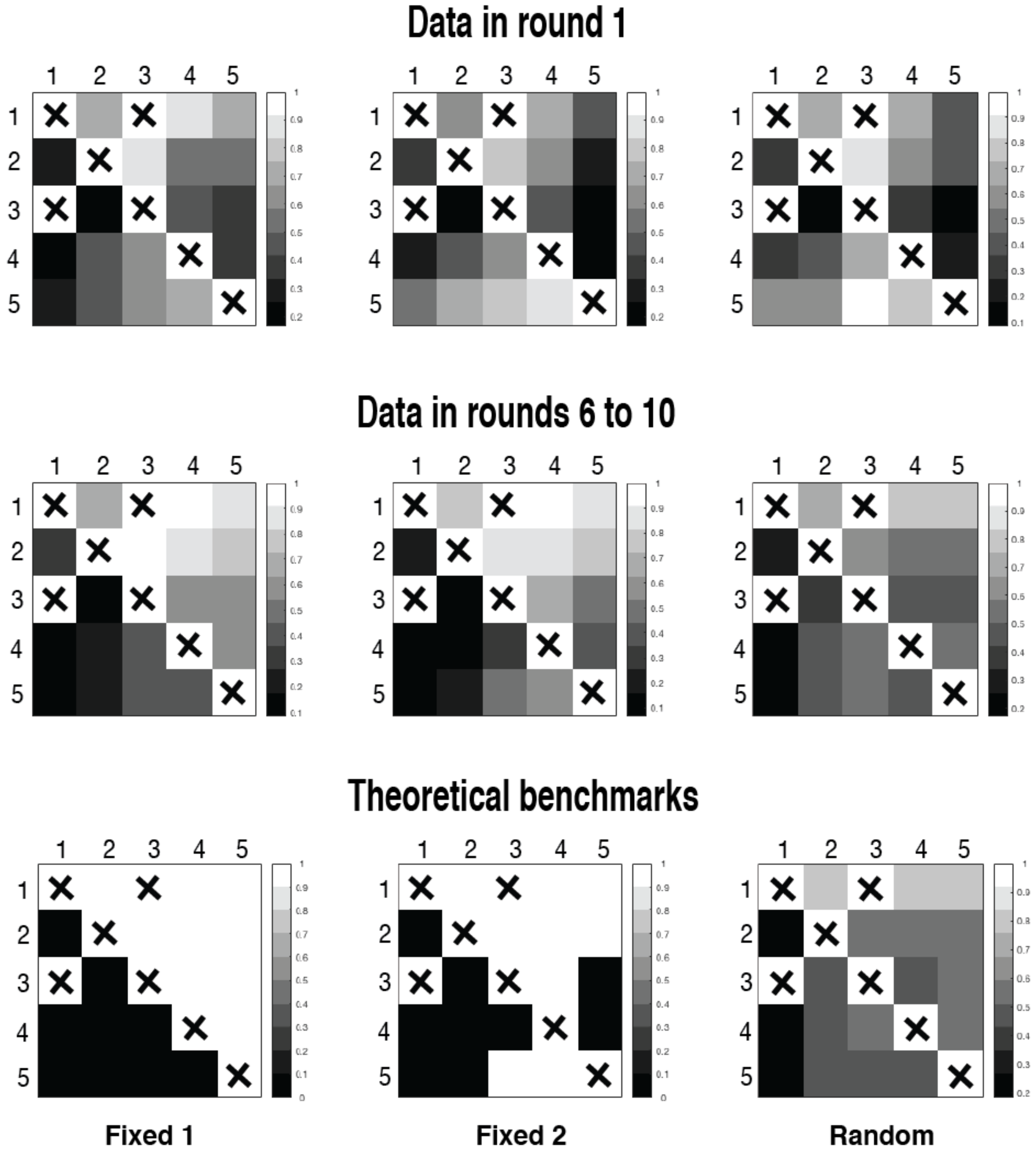


Figure 7: *Heat-maps representing average opinion comparison matrices.* Each cell in a heat-map represents the average value for the corresponding element of the opinions comparison matrices pertaining to the specific treatment for the given rounds. Cell shadings correspond to values, with darker cells corresponding to values closer to 0. Cells (1,3), (3,1) and the main diagonal are crossed out as their values do not depend on the data. The upper three heat-maps represent the data from round 1. The middle three represent average data from rounds 6 to 10. The lowest three represent theoretical benchmarks obtained by assuming that each subject assigns equal weight to all neighbors in the listening matrix (see Hypothesis 2 (c) for the corresponding matrices).

esis 2 (c).

Support: As discussed, Figures 6 and 7 provide graphical evidence of the differences in relative position patterns predicted in Hypotheses 2 (a) and (b). For the differences between the pattern in *Fixed 1* and *Fixed 2* we also perform the Barnard's exact test for the comparison of the average values between elements in $C^{1,3}_{Fixed\ 1}$ and $C^{1,3}_{Fixed\ 2}$, averaged across rounds 6 to 10. We find significant differences between elements (2,1) (p = 0.023), (4,3) (p = 0.011), (5,3) (p = 0.037) and (5,4) ($p < 0.001$). Concerning Hypothesis 2 (c) one can see from both figures that *Random* comes very close to the theoretical prediction obtained from assuming equal weights in the listening matrix. A Barnard's exact test between the theoretical prediction and the average values of elements in $C^{1,3}_{Random}$ averaged across rounds 6 to 10 shows that the only significant difference is in element (2,1) (p = 0.005): the value in the data is 0.30 when theory predicts 0.18. In both *Fixed 1* and *Fixed 2* the data does not coincide with the theoretical prediction. Still, qualitatively the patterns are similar: in *Fixed 1* where the theory predicts all elements to be zero, these are all below 0.5. In *Fixed 2* all values predicted to be 0 are below 0.5 and the two elements predicted to be 1 are on average close to 0.5 and significantly higher than the corresponding elements in *Fixed 1*.

4.4 Direction of Disagreement

We measure the direction of disagreement by the angle of rotation of the first principal component of guesses with respect to the horizontal axis (tank 1), measured in radians. In all treatments we used three sets of private signals, each one with a different draws for each phase. If Hypothesis 3 is true, then the difference in rotation between two groups starting off with the same set of private signals in the same fixed network should be zero. Furthermore, the difference in rotation between two groups in the same network starting with different private signals should be positive. Finally, the difference in rotation between two groups in different fixed networks, starting with the same set of private signals should also be positive.

Figure 8 shows the difference in rotation per round between pairs of groups with the same initial signal set (light grey boxplots) versus pairs of groups with

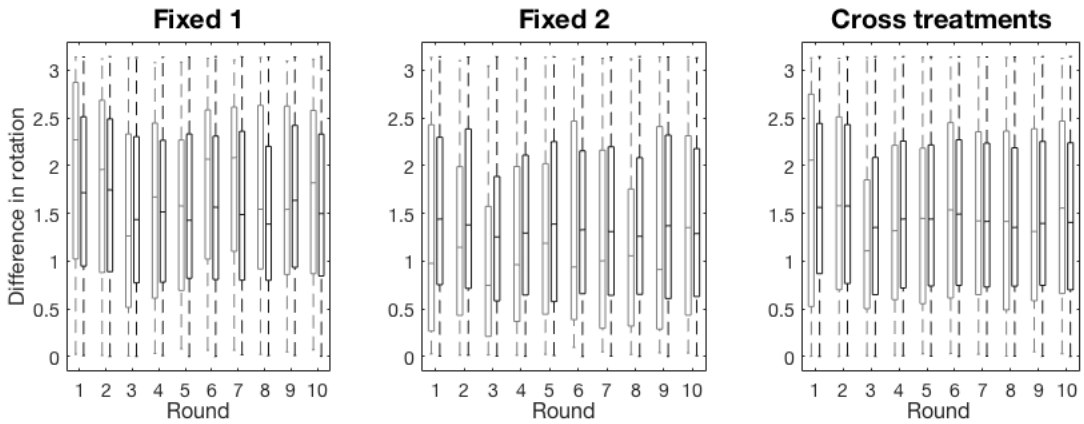


Figure 8: *Rotation differences between groups with the same vs different private signals.* Light grey boxplots show the distribution of pairwise differences in rotation for pairs with the same initial signal sets. Dark boxplots show the distribution of pairwise differences in rotation for pairs with different initial signal sets. Boxes represent the inter-quartile range (area between the 25th and 75th percentile) with horizontal lines representing medians. Whiskers extend 1.5 times the inter-quartile in each direction.

different initial signal sets (dark boxplots). The first and second panel depict these differences for *Fixed 1* and *Fixed 2* respectively. The third panel shows differences for pairs of groups where each group is in a different treatment. The main conclusion from this exercise is that differences in rotation are similar for any pair of groups compared, irrespectively of whether they start off with the same set of signals or not. While this is concordant with the theory regarding cross treatment comparisons, this should not be the case, if Property 3 holds, for within treatment comparisons. While for *Fixed 2* differences of rotation for pairs of groups with the same initial signals are on average lower than for pairs with different initial signals, these differences are not significant.[24]

Result 3. *We do not find support for Hypothesis 3.*

Support: See Figure 8 for graphical support and the discussion above.

Property 3 appears as the less robust to the perturbations in guesses we observe in the lab. To understand why this may be true it can be useful to look back at the example in Figure 3. The group there seems to converge to an axis of disagreement with a rotation of approximately 1.46 rads by round 5. A

[24] According to a non-parametric rank-sum test for differences (Mann-Whitney-Wilcoxon).

perturbation in the guess of subject 5 for tank 1 disturbs the process. By round 8 the group's guesses are aligned again, with the same relative positions, as predicted by properties 1 and 2. But the rotation of the axis of disagreement has now changed. As long as such perturbations are common and not correlated to initial signals it is not possible to use any information on the initial conditions to predict the group's direction of disagreement.

5 Discussion and Conclusions

It may be worth discussing a bit further one of our model's implicit assumptions, namely individuals using the same listening matrix to update opinions across all issues. This implies that in a given round they listen to the same set of individuals for all issues and assign the same weight to a given individual's opinion for all issues. In our experiment we restrict subjects to listen to the same set of others in each round, but they are of course free to update their opinion on each issue in any way they want. We do not find significant differences on how they do this in between the two issues, but this is perhaps not so surprising in the stylized guessing task they face. It is reasonable to think that in "real life" individuals may place different weights on a friend's opinions about political issues and sports. In fact, one may only discuss specific issues in specific social circles. It remains hence an open question how different the listening matrices for each dimension can be to still observe the emergence of unidimensional worlds. For now we can expect unidimensional worlds made up from issues that are discussed in the same social network.

Our results could be of interest to the ongoing discussion concerning privacy in online networks. Where one stands in a network determines her stance across an array of potentially sensitive issues, and this is something to be taken into account when designing regulation to protect privacy. However, we show that correlations across issues may exist even in the absence of specific network structures. Simply knowing one's opinion on a subset of even trivial issues can reveal information about their views on other more delicate matters. This poses further challenges for privacy regulation, although one might say it simply strengthens the view of those claiming that such attempts to protect privacy online are futile.

Finally, our results lead us to reflect upon some of the conventional wisdom about interventions aiming at influencing opinions in a population using knowledge of the underlying structure of social interactions, such as marketing and public awareness campaigns. A typical intervention of this sort would seed information to individuals holding key positions in the network, in a way that achieves the maximum effect while targeting a small number of influencers. In our model, the desired effect would play-out through direction determinacy. But our experiment shows that precisely this property is empirically not robust. This suggests that in some cases interventions may have a deeper effect when aiming to change the shape of the social network, rather than some of its members' opinions in a particular point in time. Nevertheless, considering that data regarding the exact network of social interactions are often hard to obtain, the general feature of correlated opinions across different issues provides a strong additional tool to the campaigner. Essentially, the campaigner could infer previously unobserved individual preferences by simply managing to identify patterns of correlation across issues.

References

Alford, J. R., C. L. Funk, and J. R. Hibbing (2005): "Are political orientations genetically transmitted?" *American Political Science Review*, 99, 153–167.

Amodio, D. M., J. T. Jost, S. L. Master, and C. M. Yee (2007): "Neurocognitive correlates of liberalism and conservatism," *Nature Neuroscience*, 10, 1246–1247.

Ansolabehere, S., J. Rodden, and J. M. Snyder (2008): "The strength of issues: Using multiple measures to gauge preference stability, ideological constraint, and issue voting," *American Political Science Review*, 102, 215–232.

Battiston, P. and L. Stanca (2015): "Boundedly rational opinion dynamics in directed social networks: Theory and experimental evidence," *Journal of Economic Behavior and Organization*, 119, 400–421.

Black, D. (1958): *The theory of committees and elections*, Cambridge: Cambridge University Press.

Bobbio, N. and A. Cameron (1996): *Left and right: The significance of a political distinction*, University of Chicago Press.

Brandts, J., A. E. Giritligil, and R. A. Weber (2015): "An experimental study of persuasion bias and social influence in networks," *European Economic Review*, 80, 214–229.

Carney, D. R., J. T. Jost, S. D. Gosling, and J. Potter (2008): "The secret lives of liberals and conservatives: Personality profiles, interaction styles, and the things they leave behind," *Political Psychology*, 29, 807–840.

Corazzini, L., F. Pavesi, B. Petrovich, and L. Stanca (2012): "Influential listeners: An experiment on persuasion bias in social networks," *European Economics Review*, 56, 1276–1288.

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

DellaPosta, D., Y. Shi, and M. Macy (2015): "Why do liberals drink lattes?" *American Journal of Sociology*, 120, 1473–1511.

DeMarzo, P. M., D. Vayanos, and J. Zwiebel (2003): "Persuasion bias, social influence, and unidimensional opinions," *Quarterly Journal of Economics*, 909–968.

Downs, A. (1957): *An Economic Theory of Democracy*, Harper, New York.

Duckitt, J. (2001): "A dual-process cognitive-motivational theory of ideology and prejudice," *Advances in Experimental Social Psychology*, 33, 41–113.

Gerber, A. S., G. A. Huber, D. Doherty, C. M. Dowling, and S. E. Ha (2010): "Personality and political attitudes: Relationships across issue domains and political contexts," *American Political Science Review*, 104, 111–133.

Gerring, J. (1997): "Ideology: A definitional analysis," *Political Research Quarterly*, 50, 957–994.

Glaeser, E. L. and C. R. Sunstein (2009): "Extremism and social learning," *Journal of Legal Analysis*, 1, 263–324.

Golub, B. and M. O. Jackson (2010): "Naive learning in social networks and the wisdom of crowds," *American Economic Journal: Microeconomics*, 2, 112–149.

——— (2012): "How homophily affects the speed of learning and best response dynamics," *Quarterly Journal of Economics*, 127, 1287–1338.

Jackson, M. O. (2008): *Social and economic networks*, Princeton University Press, Princeton, NJ.

Jolliffe, I. (2014): *Principal Component Analysis*, John Wiley and Sons, Ltd.

Karlin, S. and H. Taylor (1981): *A Second Course in Stochastic Processes*, Academic Press.

Klar, S. and Y. Shmargad (forthcoming): "The effect of network structure on preference formation," *Journal of Politics*.

Knight, K. (2006): "Transformations of the concept of ideology in the twentieth

century," *American Political Science Review*, 100, 619–626.
Lee, D. S., E. Moretti, and M. J. Butler (2004): "Do voters affect or elect policies? Evidence from the U.S. house," *Quarterly Journal of Economics*, 119, 807–859.
Levy, G. and R. Razin (2015): "Correlation neglect, voting behavior, and information aggregation," *American Economic Review*, 105, 1634–1645.
——— (2016): "Does polarization of opinions lead to polarization of platforms? The case of correlation neglect," *Quarterly Journal of Political Science*, 10, 321–355.
McDonald, M. D., S. M. Mendes, and M. Kim (2007): "Cross-temporal and cross-national comparisons of party left-right positions," *Electoral Studies*, 26, 62–75.
McMurray, J. (2014): "Why the political world is flat: An endogenous left-right spectrum in multidimensional political conflict," in *APSA 2014 Annual Meeting Paper*.
Ortoleva, P. and E. Snowberg (2015): "Overconfidence in political behavior," *American Economic Review*, 105, 504–535.
Plott, C. R. (1967): "A notion of equilibrium and its possibility under majority rule," *American Economic Review*, 787–806.
Poole, K. T. and H. Rosenthal (1997): *Congress: A political economic history of roll–call voting*, Oxford University Press, Oxford, UK.
Sethi, R. and M. Yildiz (2016): "Communication with unknown perspectives," *Econometrica*, 84, 2029–2069.
Siegel, S. (1961): "Decision making and learning under varying conditions of reinforcement," *Annals of the New York Academy of Sciences*, 89, 766–783.
Spector, D. (2000): "Rational debate and one–dimensional conflict," *Quarterly Journal of Economics*, 115, 181–200.
Tahbaz-Salehi, A. and A. Jadbabaie (2008): "A necessary and sufficient condition for consensus over random networks," *IEEE Transactions on Automatic Control*, 53, 791–795.
Wilson, C. and J. Haidt (2014): "It's true: Liberals like cats more than conservatives do," *TIME, published on Feb. 18th 2014*.

6 Appendix

6.1 Proofs

Proof of Theorem 2. First, notice that for a matrix of opinions $\mathbf{s}(t)$ the following equivalence relation

$$\lim_{t\to\infty} \beta^P(t) = 1 \Leftrightarrow \lim_{t\to\infty} \rho(\mathbf{s_i}(t), \mathbf{s_j}(t))^2 = 1 \text{ for all } i, j \in \{1, \ldots, K\} \text{ with } i \neq j$$

Hence, the Theorem holds if the following proposition is true:

Proposition 1. *Let $\rho(\mathbf{x}, \mathbf{y})$ be the correlation coefficient of two vectors $\mathbf{x}$ and $\mathbf{y}$. Consider a sequence of generic listening matrices , $\{\mathbf{T}(t)\}_{t=1}^{\infty}$, with $\alpha_2^t \in \mathbb{R}$ for all t, and two vectors $\mathbf{X}(t)$ and $\mathbf{Y}(t)$ that are updated according to (1), for some initial vectors $\mathbf{X}(0)$ and $\mathbf{Y}(0)$ with positive variance. Then $\rho\left(\mathbf{X}(t), \mathbf{Y}(t)\right)^2 \to 1$ as $t \to \infty$.*

Before stating the proof, we introduce some simplifying notation. Recall that bold letters denote matrices and normal letters denote scalars. The time parameter in an opinion vector is denoted as a superscript, i.e. $\mathbf{X^t} := \mathbf{X}(t)$ and $\mathbf{Y^t} := \mathbf{Y}(\mathbf{t})$, the scalar X_i^t will denote the ith element of the vector $\mathbf{X^t}$ and $\overline{X^t}$ will denote the average opinion on issue X in period t. The listening matrices in period t are denoted by $\mathbf{T_t}$ with eigenvalues $\alpha_1^t, \ldots, \alpha_N^t$, with typical element α_n^t and ranked in decreasing order according to their modulus, $\|\alpha_n^t\|$, as they may be complex numbers. The correlation coefficient in period t will be denoted as ρ_t, i.e. $\rho_t := \rho(\mathbf{X^t}, \mathbf{Y^t})$. Any other necessary quantity will be defined in the relevant part.

Recall that all elements of the sequence $\{\mathbf{T_t}\}_{t=1}^{\infty}$ are irreducible, aperiodic, row–stochastic listening matrices[25] and the opinion formation process is described by the dynamics $\mathbf{X^t} = \prod_{\tau=1}^{t} \mathbf{T}_\tau \cdot \mathbf{X^0}$ and $\mathbf{Y^t} = \prod_{\tau=1}^{t} \mathbf{T}_\tau \cdot \mathbf{Y^0}$, for some initial vectors $\mathbf{X^0}$, $\mathbf{Y^0}$. We also repeat the three assumptions regarding the listening matrices: 1) For each $\mathbf{T_t}$ the second largest eigenvalue α_2^t is real, 2) $\mathbf{T_t}$ is diagonalizable and 3) $\|\alpha_2^t\| > \|\alpha_n^t\|$ for all $n > 2$.

In the first part, we construct ρ_t^2 as a function of the initial opinions and the

[25]Strong connectivity of the networks together with strictly positive weights imply irreducibility of $\mathbf{T_t}$ and positive diagonal implies aperiodicity. The matrices are also row–stochastic by definition, given that their rows sum to 1.

eigenvalues and eigenvectors of the listening matrices.

$$\rho_t = \frac{\sum_{i=1}^{N}(X_i^t - \overline{X^t})(Y_i^t - \overline{Y^t})}{\sqrt{\sum_{i=1}^{N}(X_i^t - \overline{X^t})^2}\sqrt{\sum_{i=1}^{N}(Y_i^t - \overline{Y^t})^2}} \Rightarrow$$

$$\rho_t^2 = \frac{\sum_{i=1}^{N}\sum_{j=1}^{N}(X_i^t - \overline{X^t})(Y_i^t - \overline{Y^t})(X_j^t - \overline{X^t})(Y_j^t - \overline{Y^t})}{\sum_{i=1}^{N}\sum_{j=1}^{N}(X_i^t - \overline{X^t})(X_i^t - \overline{X^t})(Y_j^t - \overline{Y^t})(Y_j^t - \overline{Y^t})} \tag{4}$$

At this point, we need to construct the factors of the form $(X_i^t - \overline{X^t})$. By assumption 2 each listening matrix $\mathbf{T_t}$ can be diagonalized in the form $\mathbf{T_t} = \mathbf{V^{c,t}A^tV^{r,t}}$, where $\mathbf{A^t}$ is a diagonal matrix consisting of the N eigenvalues of $\mathbf{T_t}$ and $\mathbf{V^{c,t}}$ and $\mathbf{V^{r,t}}$ are the $N \times N$ matrices of column and row eigenvectors respectively. The expression can be restated as follows:

$$\mathbf{T_t} = \mathbf{V^{c,t}A^tV^{r,t}} = \sum_{n=1}^{N} \alpha_n^t \mathbf{V_n^{c,t}V_n^{r,t}} = \mathbb{1}\mathbf{w_t} + \sum_{n=2}^{N} \alpha_n^t \mathbf{V_n^{c,t}V_n^{r,t}}$$

where α_n^t is the nth largest eigenvalue of $\mathbf{T_t}$ and $\mathbf{V_n^{c,t}}$, $\mathbf{V_n^{r,t}}$ are the column and row eigenvectors that correspond to this eigenvalue. The last equality holds because for all irreducible, aperiodic, row–stochastic $\mathbf{T_t}$ it holds that $\alpha_{1,t} = 1$, $\mathbf{V_1^{c,t}} = \mathbb{1}$ (column vector of ones) and $\mathbf{V_1^{r,t}} = \mathbf{w_t}$ (see for instance Karlin and Taylor, 1981). Therefore,

$$\prod_{\tau=1}^{t} \mathbf{T}_\tau = \prod_{\tau=1}^{t} \left[\mathbb{1}\mathbf{w}_\tau + \sum_{n=2}^{N} \alpha_n^\tau \mathbf{V_n^{c,\tau}V_n^{r,\tau}} \right] \tag{5}$$

This expression can also be simplified further through the following two lemmas:

Lemma 1. *Let $\mathbf{V_n^r}$ be the row eigenvector corresponding to the nth largest eigenvalue, α_n, of a listening matrix $\mathbf{B}$. Then, either $\alpha_n = 1$ (i.e. $n = 1$) or $\sum_i v_{i,n}^r = 0$, where $v_{i,n}^r$ is the ith element of the eigenvector.*

Proof. The row eigenvector $\mathbf{V_n^r}$ and its associated eigenvalue α_n satisfy the matrix equation $\mathbf{V_n^rB} = \alpha_n\mathbf{V_n^r}$. Let $b_{i,j}$ be a typical element of $\mathbf{B}$ and rewrite the equation as follows:

$$
\begin{aligned}
v^r_{1,n}b_{1,1} + v^r_{2,n}b_{2,1} + \cdots + v^r_{n,n}b_{n,1} - \alpha_n v^r_{1,n} &= 0 \\
v^r_{1,n}b_{1,2} + v^r_{2,n}b_{2,2} + \cdots + v^r_{n,n}b_{n,2} - \alpha_n v^r_{2,n} &= 0 \\
&\vdots \\
v^r_{1,n}b_{1,n} + v^r_{2,n}b_{2,n} + \cdots + v^r_{n,n}b_{n,n} - \alpha_n v^r_{n,n} &= 0
\end{aligned}
$$

Summing all rows we obtain:

$$
v^r_{1,n}\sum_j b_{1,j} + v^r_{2,n}\sum_j b_{2,j} + \cdots + v^r_{n,n}\sum_j b_{n,j} - \alpha_n \sum_i v^r_{i,n} = 0
$$

and row stochasticity of **B** implies that $\sum_j b_{i,j} = 1$ for all i, hence

$$
\sum_i v^r_{i,n} - \alpha_n \sum_i v^r_{i,n} = 0 \Rightarrow \alpha_n = 1 \text{ or } \sum_i v^r_{i,n} = 0
$$

Irreducibility and aperiodicity of **B** imply that $\alpha_n = 1$ if and only if $n = 1$. □

Lemma 2. $\mathbf{V^c_n V^r_n} \mathbb{1} = \mathbf{0}$ *for all* $n \geq 2$*, where* $\mathbf{V^c_n}$ *and* $\mathbf{V^r_n}$ *are the column and row eigenvectors respectively associated to the nth largest eigenvalue and* $\mathbb{1}$*,* $\mathbf{0}$ *are column vectors consisting of ones and zeros respectively.*

Proof. The (i, j)-th element of $\mathbf{V^c_n V^r_n}$ is equal to $v^c_{n,i} v^r_{j,n}$. Hence, the ith element of the vector $\mathbf{V^c_n V^r_n}\mathbb{1}$ is equal to $v^c_{n,i}\left(\sum_j v^r_{j,n}\right) = 0$, where the last equality follows from the Lemma 1. □

Lemmas 1 and 2 imply that any product that contains a factor of the form $\sum_{n=2}^{N} \alpha_n^\tau \mathbf{V^{c,\tau}_n V^{r,\tau}_n} \mathbb{1}\mathbf{w}_{\tau-1}$ for some τ will be equal to zero. Therefore:

$$
\prod_{\tau=1}^{t} \mathbf{T}_\tau = \mathbb{1}\mathbf{w_t}\left(\prod_{\tau'=1}^{t-1} \mathbf{T}_{\tau'}\right) + \prod_{\tau=1}^{t}\left(\sum_{n=2}^{N} \alpha_n^\tau \mathbf{V^{c,\tau}_n V^{r,\tau}_n}\right) \tag{6}
$$

We are almost ready to construct the quantity $(X^t_i - \overline{X^t})$. Denote by $\mathbf{e_i}$ a row vector with all elements equal to 0 and ith element equal to 1. Moreover, $\hat{\mathbf{w}} = (1/N)\mathbb{1}^T$ is a row vector that will give us the average in each period.[26] The following lemma will help us simplify the final expression.

[26]DeMarzo et al. (2003) consider a general vector with non–negative elements that sum to one when calculating relative positions, but they would need this more exact definition to obtain the expression of ρ^2_t. They would also need to consider as normalizing factor the standard deviation in each issue, instead of the issue–independent normalization that is enough for characterizing

Lemma 3. *Let $\mathbb{1}$ be a $N \times 1$ column vector of ones, $\mathbf{w}$ be a $1 \times N$ row vector. then* $\mathbf{e_i}\mathbb{1}\mathbf{w} = \hat{\mathbf{w}}\mathbb{1}\mathbf{w} = \mathbf{w}$

Proof. Let $\mathbf{w} = (w_1, w_2 \ldots, w_N)$, then:

$$\mathbf{e_i}\mathbb{1}\mathbf{w} = (0, \ldots, 0, 1, 0, \ldots, 0) \begin{pmatrix} w_1 & w_2 & \ldots & w_n \\ w_1 & w_2 & \ldots & w_n \\ \vdots & \vdots & \ddots & \vdots \\ w_1 & w_2 & \ldots & w_n \end{pmatrix} = (w_1, w_2 \ldots, w_N) = \mathbf{w}$$

$$\hat{\mathbf{w}}\mathbb{1}\mathbf{w} = \frac{1}{N}(1, \ldots, 1) \begin{pmatrix} w_1 & w_2 & \ldots & w_n \\ w_1 & w_2 & \ldots & w_n \\ \vdots & \vdots & \ddots & \vdots \\ w_1 & w_2 & \ldots & w_n \end{pmatrix} = \frac{1}{N}(Nw_1, Nw_2 \ldots, Nw_N) = \mathbf{w}$$

□

Let us now construct $(X_i^t - \overline{X^t})$

$$X_i^t = \mathbf{e_i}\mathbf{X^t} = \mathbf{e_i}\prod_{\tau=1}^{t}\mathbf{T}_\tau\mathbf{X^0} = \mathbf{e_i}\left[\mathbb{1}\mathbf{w_t}\left(\prod_{\tau'=1}^{t-1}\mathbf{T}_{\tau'}\right)\mathbf{X^0}\right] + \mathbf{e_i}\left[\prod_{\tau=1}^{t}\left(\sum_{n=2}^{N}\alpha_n^\tau\mathbf{V_n^{c,\tau}}\mathbf{V_n^{r,\tau}}\right)\right]\mathbf{X^0} \tag{7}$$

$$\overline{X^t} = \hat{\mathbf{w}}\mathbf{X^t} = \hat{\mathbf{w}}\prod_{\tau=1}^{t}\mathbf{T}_\tau\mathbf{X^0} = \hat{\mathbf{w}}\left[\mathbb{1}\mathbf{w_t}\left(\prod_{\tau'=1}^{t-1}\mathbf{T}_{\tau'}\right)\mathbf{X^0}\right] + \hat{\mathbf{w}}\left[\prod_{\tau=1}^{t}\left(\sum_{n=2}^{N}\alpha_n^\tau\mathbf{V_n^{c,\tau}}\mathbf{V_n^{r,\tau}}\right)\right]\mathbf{X^0} \tag{8}$$

Lemma 3 shows that the first two factors are equal, therefore:

$$X_i^t - \overline{X^t} = \mathbf{e_i}\left[\prod_{\tau=1}^{t}\left(\sum_{n=2}^{N}\alpha_n^\tau\mathbf{V_n^{c,\tau}}\mathbf{V_n^{r,\tau}}\right)\right]\mathbf{X^0} - \hat{\mathbf{w}}\left[\prod_{\tau=1}^{t}\left(\sum_{n=2}^{N}\alpha_n^\tau\mathbf{V_n^{c,\tau}}\mathbf{V_n^{r,\tau}}\right)\right]\mathbf{X^0} \tag{9}$$

which can be further simplified as follows:

$$X_i^t - \overline{X^t} = \left[\sum_{n=2}^{N}\alpha_n^t\left(V_{i,n}^{c,t} - \hat{\mathbf{w}}\mathbf{V_n^{c,t}}\right)\mathbf{V_n^{r,t}}\right]\prod_{\tau=1}^{t-1}\left(\sum_{m=2}^{N}\alpha_m^\tau\mathbf{V_m^{c,\tau}}\mathbf{V_m^{r,\tau}}\right)\mathbf{X^0} \tag{10}$$

---

the relative positions. Having mentioned their proof, it is important to notice that in our case stabilization of relative positions is not to be expected, because even if unidimensionality arises, the agents' positions may change across periods based on the realized listening matrices.

where $V^{c,t}_{i,n}$ denotes the ith element of the column eigenvector corresponding to the nth largest eigenvalue of T_t. Notice that, $\left(V^{c,t}_{i,n} - \hat{\mathbf{w}}\mathbf{V^{c,t}_n}\right)$ is the only factor that depends on i, it is a scalar and does not depend on the issue. Moreover, for each n, $\mathbf{V^{r,t}_n} \prod_{\tau=1}^{t-1} \left(\sum_{m=2}^{N} \alpha^{\tau}_{m} \mathbf{V^{c,\tau}_m} \mathbf{V^{r,\tau}_m} \right) \mathbf{X^0}$ is also a scalar and does not depend on i. This expression can also be written as a sum of products, as follows:

$$X^t_i - \overline{X^t} = \sum_{n_1,\ldots,n_t=2}^{N} \left[\alpha^t_{n_t} \left(V^{c,t}_{i,n_t} - \hat{\mathbf{w}}\mathbf{V^{c,t}_{n_t}} \right) \mathbf{V^{r,t}_{n_t}} \left(\prod_{\tau=1}^{t-1} \alpha^{\tau}_{n_\tau} \mathbf{V^{c,\tau}_{n_\tau}} \mathbf{V^{r,\tau}_{n_\tau}} \right) \mathbf{X^0} \right] \tag{11}$$

A similar sum of products can be defined for $X^t_j - \overline{X^t}$, $Y^t_i - \overline{Y^t}$ and $Y^t_j - \overline{Y^t}$. Given that each of these products appears both in the numerator and the denominator of ρ^2_t, we can divide each of them with $\prod_{\tau=1}^{t} \alpha^{\tau}_2$ and ρ^2_t will remain unchanged. This will be helpful when calculating the limit of ρ^2_t as t grows.

For each t define the sequence $\{(l^1_1, l^1_2, l^1_3, l^1_4), (l^2_1, l^2_2, l^2_3, l^2_4), \ldots, (l^t_1, l^t_2, l^t_3, l^t_4)\}$, where its element $l^{\tau}_k \in \{2, \ldots, N\}$ determines n_τ for each of the four parts of each product (corresponding to $X^t_i - \overline{X^t}$, $Y^t_i - \overline{Y^t}$, $X^t_j - \overline{X^t}$ and $Y^t_j - \overline{Y^t}$ for the numerator and analogously for the denominator). The set that contains all these sequences is denoted by $\mathcal{S}^t$ with typical element s^t. According to this notation, we can denote the value of each of the above products (after being divided by $\prod_{\tau=1}^{t} \alpha^{\tau}_2$) in the numerator and denominator as $f_{i,j}(s^t)$ and $g_{i,j}(s^t)$ respectively. Notice that in general $f_{i,j}(s^t) \neq g_{i,j}(s^t)$, as the order of variables is different. Hence, we can rewrite ρ^2_t as follows:

$$\rho^2_t = \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} \left[\sum_{s^t \in \mathcal{S}^t} f_{i,j}(s^t) \right]}{\sum_{i=1}^{N} \sum_{j=1}^{N} \left[\sum_{s^t \in \mathcal{S}^t} g_{i,j}(s^t) \right]} \tag{12}$$

Moreover, define as $\mathcal{S}$ the set of all infinite sequences $\{(l^1_1, l^1_2, l^1_3, l^1_4), (l^2_1, l^2_2, l^2_3, l^2_4), \ldots\}$, with typical element s and observe that

$$\rho^2_t \to \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} \left[\sum_{s \in \mathcal{S}} f_{i,j}(s) \right]}{\sum_{i=1}^{N} \sum_{j=1}^{N} \left[\sum_{s \in \mathcal{S}} g_{i,j}(s) \right]} \text{ as } t \to \infty \tag{13}$$

The set $\mathcal{S}$ can be partitioned in two subsets $\mathcal{S}_1$ and $\mathcal{S}_2 = \mathcal{S} \setminus \mathcal{S}_1$, where the set $\mathcal{S}_1$ is defined as follows:

$$\mathcal{S}_1 := \left\{ s \in \mathcal{S} : \text{ exists } \hat{t} \text{ such that for all } t > \hat{t} \text{ it holds that } (l_1^t, l_2^t, l_3^t, l_4^t) = (2, 2, 2, 2) \right\}$$

In words, the set $\mathcal{S}_1$ contains all the sequences in which all but finitely many elements are equal to $(2, 2, 2, 2)$. The set $\mathcal{S}_2$ is just the complement of $\mathcal{S}_1$. Hence, the previous expression can be rewritten as follows:

$$\rho_t^2 \to \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} \left[\sum_{s_1 \in \mathcal{S}_1} f_{i,j}(s_1) + \sum_{s_2 \in \mathcal{S}_2} f_{i,j}(s_2) \right]}{\sum_{i=1}^{N} \sum_{j=1}^{N} \left[\sum_{s_1 \in \mathcal{S}_1} g_{i,j}(s_1) + \sum_{s_2 \in \mathcal{S}_2} g_{i,j}(s_2) \right]} \text{ as } t \to \infty \tag{14}$$

Observe that, for each i, j, the quantity $f_{i,j}(s)$ has the form:

$$f_{i,j}(s) = \left(\prod_{\tau=1}^{\infty} \prod_{k=1}^{4} \frac{\alpha_{l_k^\tau}^\tau}{\alpha_2^\tau} \right) C_{i,j}$$

where $C_{i,j}$ is a finite number, generically different than zero, that is determined by the rest of the parameters. The expression for $g_{i,j}(s)$ is analogous.

Now consider any $s_2 \in \mathcal{S}_2$ and notice that by its definition as complement of $\mathcal{S}_1$ it means that for all t exists $\tau > t$ such that $(l_1^\tau, l_2^\tau, l_3^\tau, l_4^\tau) \neq (2, 2, 2, 2)$. Therefore, by assumption (3), for all t exists $\tau > t$ such that $\left\| \frac{\alpha_{l_k^\tau}^\tau}{\alpha_2^\tau} \right\| < 1$. Hence, there is $\epsilon > 0$ such that $\left\| \frac{\alpha_{l_k^\tau}^\tau}{\alpha_2^\tau} \right\| \leq (1 - \epsilon)$ for infinitely many elements of s_2, whereas for the remaining elements it holds that $\left\| \frac{\alpha_{l_k^\tau}^\tau}{\alpha_2^\tau} \right\| = 1$. This observation combined with the finiteness of $C_{i,j}$ means that $f_{i,j}(s_2) = 0$ and analogously $g_{i,j}(s_2) = 0$ as well.

Finally, consider $s_1 \in \mathcal{S}_1$ and the sequence s_1^C that is defined in relation to s_1 as follows: for all t, if the t-th quadruple element of s_1 is $(l_1^t, l_2^t, l_3^t, l_4^t)$ then the t-th quadruple element of s_1^C is $(l_1^t, l_3^t, l_2^t, l_4^t)$ –Observe the interchanged positions of l_2^t and l_3^t–. Notice that, $s_1^C \in \mathcal{S}_1$ and in fact if t_1 is the period such that for all $t > t_1$ it holds that $(l_1^t, l_2^t, l_3^t, l_4^t) = (2, 2, 2, 2)$, then the same t_1 is such that for all $t > t_1$ it holds that $(l_1^t, l_3^t, l_2^t, l_4^t) = (2, 2, 2, 2)$. Moreover, for all $s_1 \in \mathcal{S}_1$ the sequence s_1^C exists and in some cases even coincides with s_1. Therefore we have defined a one–to–one relation from $\mathcal{S}_1$ to itself. It follows immediately from expressions (4) and (11) that $f_{i,j}(s_1) = g_{i,j}(s_1^C)$, for all i, j and for all $s_1 \in \mathcal{S}_1$. In fact, this is true not only in the limit, but also for all finite subsequences of s_1 and s_1^C with $t > t_1$.

Therefore, $\sum_{i=1}^{N}\sum_{j=1}^{N}\left[\sum_{s_1\in\mathcal{S}_1} f_{i,j}(s_1)\right] = \sum_{i=1}^{N}\sum_{j=1}^{N}\left[\sum_{s_1\in\mathcal{S}_1} g_{i,j}(s_1)\right]$. These sums are generically different than zero, given that $\mathcal{S}_1$ contains the element $\hat{s}$ for which $l_k^t = 2$ for all k, t, hence satisfies $\frac{\alpha^t_{l^t_k}}{\alpha^t_2} = 1$ for all k, t. Hence, we can conclude that:

$$\rho_t^2 \to 1 \text{ as } t \to \infty$$

□

6.2 Experimental Instructions

We present the instructions for the two fixed treatments F1 and F2. The relevant changes for the random treatment appear in footnotes.

INSTRUCTIONS

Thank you for participating in this session. Please remain quiet! You will be using the computer terminal for the entire experiment, and your decisions will be made via your computer terminals. Please DO NOT talk or make any other audible noises during the experiment. The use of mobile phones or other devices is prohibited. You are free to use the calculator provided. If you have any questions, raise your hand and your question will be answered so that everyone can hear.

General Instructions:

The experiment will take place in three parts. The remaining instructions refer to Part 1 of the experiment. Once part 1 is over you will be given instructions for Parts 2 and 3.

The experiment will involve a series of guesses. Each of you may earn different amounts. You also receive a 3 participation fee. Upon completion of the experiment, you will be paid individually and privately in room B33 upon presentation of the computer number you were assigned.

IMPORTANT:

The amount each participant earns, in today's experiment, depends only on his/her decisions and not on the decisions of other participants.

There is no specific time limit for making each guess. In order to finish the experiment on time we ask you to enter your guess in a reasonable amount of time. If a notification asking you to enter a guess appears on your screen please do so as soon as possible.

Part 1

The Task: In a tank there are 100,000 balls. These balls are either RED or BLUE. The number of balls of each colour is random and any combination is equally likely. You are asked to guess the number of RED balls in the tank. This number could be anywhere between 0 and 100,000.

Before making your guess, you observe a sample of 100 balls picked randomly from the tank. That is, the computer will inform you how many of the 100 balls in your sample were RED. You are then asked to enter a guess concerning the total number of red balls (between 0 and 100,000).

You will repeat this task three times. In other words you will make three guesses for three different tanks (filled with a different number of red and blue balls) and using a different sample each time. After each guess you will be given the correct number of RED balls and the points earned calculated as follows:

Points: For each guess you earn a maximum of 100 points if your guess is correct, while you earn fewer points for guessing wrong. The higher the difference between your guess and the correct number, the less points you earn. The exact number of points you earn in each round is given by the following formula:

$$\text{Points} = 100 - \text{Error Factor} \times \left(\frac{\text{Correct} - \text{Guess}}{1,000} \right)^2$$

If the result from the formula is negative, you earn zero points. You will be shown on your screen the exact value of the *Error Factor* for each guess you make.

Example
Suppose the number of red balls in the tank is 57,345. The following table illustrates examples of different guesses and the resulting number of points you would earn for different error factors.

Guess	**69,345**	**52,345**	**52,345**	**59,545**	**55,145**	**55,545**
Difference from correct	12,000	5,000	5,000	2,200	2,200	1,800
Error factor	1	1	10	10	25	25
Formula result	-44	75	-150	51.6	-21	19
Points	0	75	0	51.6	0	19

Given this formula, you maximize the expected number of points you earn in each round by making a guess that is as close as possible to your true estimate of the correct number of Red balls in the tank.

Out of the three guesses you will make in part 1, one will be selected randomly by the computer at the end of the experiment. The points you earned in the randomly selected guess will be transformed into monetary earnings. The exchange rate used is **1 for every 85 points**. Notice that since all guesses can be chosen with the same probability, you cannot know for which of the guesses you will be paid. Therefore you should treat all guesses the same and make a guess as if you are going to be paid for it.

Part 2:

You will be assigned to a group that has 5 members. These groups are formed randomly and anonymously. You will interact exclusively within each group without knowing the identity of the other group members.

The Task: The task in this part has 5 rounds. Like in part 1, in a tank there are 100,000 balls. These balls are either RED or BLUE. The number of balls of each colour is random and any combination is equally likely. Each group member will be asked to guess the number of RED balls in the tank in each of the 5 rounds. This number could be anywhere between 0 and 100,000 and remains the same for all 5 rounds.

The task proceeds as follows: Before making a first guess, **each member observes a different sample of 100 balls picked randomly from the tank.**

Round 1: On your screen you will see the amount of RED balls in your sample of 100.

You are asked to make a guess about the number of red balls in the tank, as in Part 1.

Round 2: On your screen you will see the guess you made in round 1, as well as the guess(es) made in round 1 by some of the other group members. **You may observe a subset of one, two, or three other members. Each group member observes the guess(es) of a different subset of group members.** Furthermore, the fact that you observe a group member X does not necessarily mean that X observes you.[27]

You are asked to make a new guess about the number of RED balls in the tank.

Rounds 3-5: These rounds are the same as round 2. You see the guess(es) made previously by the group members you observe, and are asked to make a new guess.[28]

Payoffs: Again, you can earn a maximum of 100 points in each round if your guess is correct, while you earn fewer points for guessing wrong. The higher the difference between your guess and the correct number, the less points you earn. The exact number of points you earn in each round is given by the same formula as before:

[27] **Round 2:** On your screen you will see the guess you made in round 1, as well as the guess(es) made in round 1 by some other randomly chosen group members. **You may observe one, two, or three other members.** The fact that you observe a group member X does not necessarily mean that X observes you.

[28] **Rounds 3-5:** These rounds are the same as round 2. Some group members are chosen randomly, you see their guess(es) from the previous round, and are asked to make a new guess.

$$\text{Points} = 100 - \text{Error Factor}_{Round} \times \left(\frac{\text{Correct} - \text{Guess}}{1,000}\right)^2$$

Now the *Error Factor* is different in each round:

Round	1	2	3	4	5
Error Factor	1	5	10	20	25

If in a round the result from the formula is negative, you earn zero points in that round. After the 5 rounds are over, you will be shown how many points you earned in each round. At the end of the experiment the computer will randomly choose 1 out of the 5 rounds. The points you earned in this randomly chosen round will be transformed in to monetary earnings. The exchange rate used is **1 for every 85 points**. Notice that since all 5 rounds can be chosen with the same probability, you cannot know for which of the rounds you will be paid. Therefore you should treat all rounds the same and make a guess as if you are going to be paid for it.

Remember:

1. **You will play 5 rounds. Your group and the members whose guesses you observe remain fixed during the whole time.**[29]
2. **There is a single tank and everybody in the group is guessing the number of RED balls in this tank. Each member observes a different sample of 100 balls and a different sample of group members.**
3. **Given the formula, you maximize the expected number of points you earn in each round by making a guess that is as close as possible to your true estimate of the correct number of Red balls in the tank.**
4. **The points you earn depend only on your guess and not on the guesses of other members.**

[29]**You will play 5 rounds. Your group and the number of members whose guesses you observe remain fixed during the whole time. The group members you observe are chosen randomly in each round.**

Part 3:

The composition of each group remains unchanged throughout all the experiment (same as in part 2). Remember that each group member observes the guess(es) of a different subset of group members, some will observe one, some two, and some three other group members.[30]

In this part there are 3 Phases of 10 rounds each.

The Task: Now there are two tanks filled with 100,000 balls each. Tank 1 contains RED and BLUE balls, while Tank 2 contains GREEN and PURPLE balls. The number of balls of each colour in each tank is random and any combination is equally likely. You are asked to guess the correct number of RED balls in Tank 1 and of GREEN balls in Tank 2. **The number of RED balls in Tank 1 is not related to the number of GREEN balls in Tank 2.** These two numbers could be anywhere between 0 and 100,000.

As in part 2, before making a first guess, each participant observes 2 samples of 100 balls picked randomly: one sample for Tank 1, and one sample for Tank 2. **Remember that each participant observes different random samples.** Each phase then proceeds as follows:

Round 1: On your screen you will see the amount of RED balls in your sample from Tank 1 and the number of GREEN balls in your sample from Tank 2. You are asked to make a guess about the correct number of balls of the corresponding colour in each tank.

Round 2: As in part 2, on your screen you will see your guesses for each tank from round 1. You will also see the guesses made for each tank by the group members you observe from your group. After seeing their guesses you are asked to make new guesses about the number of RED balls in Tank 1 and GREEN balls in Tank 2.[31]

Rounds 3-10: As before, these rounds are the same as round 2. You see the guesses made in the previous rounds by the group members you observe, and make new guesses.[32]

---

[30]**The composition of each group remains unchanged throughout all the experiment (same as in part 2). Remember that each group member observes the guess(es) of one, two, or three other group members in each round.**

[31]**Round 2:** As in part 2, on your screen you will see your guesses for each tank from round 1. You will also see the guesses made for each tank in the previous round by some other group members that are chosen randomly. After seeing their guesses you are asked to make new guesses about the number of RED balls in Tank 1 and GREEN balls in Tank 2.

[32]**Rounds 3-10:** As before, these rounds are the same as round 2. Some group members are chosen randomly, you see their guesses from the previous round, and are asked to make new guesses.

History: Starting from the 2^{nd} phase in this part, you will have the opportunity to see the history of all guesses you and the group members you observe have made in previous phases. To access the history you just have to press the button on the top of your screen.

Payoffs: As in Parts 1 and 2, your payoff in each round is determined using the same formula. The guess you make for each tank enters the formula along with the correct number of balls of the corresponding colour. The Error Factor in each round is shown in the table below.

Round	1	2	3	4	5	6	7	8	9	10
Error Factor	1	5	10	15	15	15	20	20	20	25

The points you earn from each tank (maximum 100) are added together to give the total number of points for the round. At the end of each phase you will be shown how many points you earned in each round and from each tank. At the end of the experiment the computer will randomly choose 1 out of the 10 rounds for each of the three phases. The points you earned in this randomly chosen round will be transformed into monetary earnings. The exchange rate used is **1 for every 85 points**. Notice that since all 10 rounds of each phase can be chosen with the same probability, you cannot know for which of the rounds you will be paid. Therefore you should treat all rounds the same and make a guess as if you are going to be paid for it. Your monetary earning from each of these 3 rounds (one for each phase) will be added to your earnings from parts 1 and 2 and the show-up fee of 3. A screen will inform you about your total monetary earnings at the end of the experiment.

Remember:

1. **You will play 3 phases of 10 rounds. The groups and the members whose guesses you observe remain the same. What changes in each phase is the amount of balls in each tank and the samples observed by each member before making the first guess.**[33]
2. **In each phase there is a different amount of balls in each tank. The samples observed by each member before making the first guess are also different for each phase.**
3. **You maximize the expected number of points you earn in each round by making guesses that are as close as possible to your true estimates of the correct number of RED balls in Tank 1 and GREEN balls in Tank 2.**
4. **The points you earn depend only on your guess and not on the guesses of other members.**

[33]**You will play 3 phases of 10 rounds. The groups and their members remain the same. In each round some group members are chosen randomly and you observe their guesses in the previous rounds.**