Communication and the emergence of a unidimensional world

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

[1]Department of Economics, Univesity of Cyprus

[2]Department of Economics, Lancaster University Management School

May 17, 2018

Abstract

We provide theoretical and experimental support on the emergence of a unidimensional world through communication. Our 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 relative to others.

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 often 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 beyond the world of politics, and ideological cleavages spillover to preferences over leisure activities, consumption and art, as well as personal morality (see Kosinski et al., 2013, 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 these 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 a unidimensional world 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 point.[4] But DeMarzo, Vayanos, and Zwiebel (2003) observed that, if

[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. For the representation of preferences in multidimensional models see Eguia (2011, 2013).

[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.

[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,

opinions are communicated over a *fixed* communication structure, prior to opinions eventually converging there is disagreement at any point in time, which can be summarized by a unidimensional spectrum. Our theory shows that such 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 don't require individuals to communicate in every period with the same individuals, nor assign the same subjective weight to one's opinions in all periods.

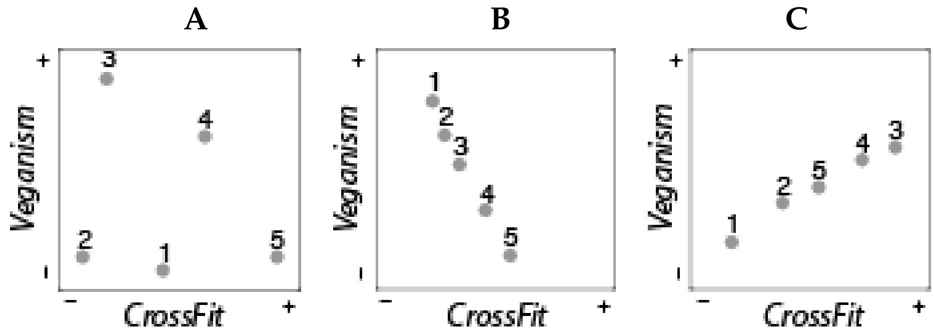


Figure 1: Possible distribution of opinions on two issues. Each point represents an individual's opinions on the two issues: CrossFit and veganism. Opinions can vary from extremely negative to extremely positive. Panel A shows an example of uncorrelated opinions. Panels B and C show examples of unidimensional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

While our approach is more general and potentially more realistic compared to DeMarzo et al. (2003), permitting a dynamic communication structure comes at a cost: our theory remains partly agnostic about the exact characteristics of the unidimensional world that may emerge. The following example elaborates on such difference. 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 correlation between issues. Panels *B* and *C* represent two different unidimensional worlds, where opinions on the

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 to illustrate the concept of unidimensionality. Notice that in our setup opinions can be defined in a very broad sense including beliefs, judgements, attitudes and all such fundamental drivers of behaviour that are amenable to social influence or the advent of new information.

two issues are strongly correlated. However, while we observe *unidimensionality* in both panels *B* and *C*, there are differences between the two. A first difference 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 remains an extremist in panel *C*, 5 has a moderate position. A second difference is the *direction of disagreement*, captured by the slope of the line indicating the importance of each issue 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 relatively flat in panel *C*.

In terms of our theoretical results, our model does predict that communication generically makes opinions move from a multidimensional world (panel *A*) towards a unidimensional one (panels *B* and *C*) even in the absence of a fixed communication structure. Nevertheless, stricter assumptions about the latter such as the presence of a fixed communication network 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 to predict its characteristics.

We also move 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 were linked through a fixed network. Networks differed minimally between these treatments. In a third treatment subjects listened 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.

Our experimental data permit also a closer analysis of individual updating patterns. Using clustering analysis, we identify different types of individuals that may coexist in a group. Once such heterogeneous behaviors are incorporated in a theoretical model, our simulations closely replicate our experimental results. Our results hence highlight that despite the "standard" theoretical assumption of agents updating their opinions in a homogeneous manner, heterogeneity in updating patterns can be critical.

Our experimental design and results complement previous experiments where communication is over a single issue and the analysis focuses on the individual updating process (see Corazzini et al. 2012; Chandrasekhar et al. 2012; Grimm and Mengel 2014; Brandts et al. 2015; Battiston and Stanca 2015). In terms of design, our work is the first to permit communication of opinions over multiple issues and therefore a focus on the emergence of a unidimensional world. In terms of results, we also show that some kind of opinion averaging is a prevalent updating process. However, some patterns in the data cannot be explained by the homogeneous model of DeMarzo et al. (2003). Nevertheless, this becomes feasible once we allow for heterogeneity in the updating behavior.

Our work further links to several 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). In a similar spirit, Melguizo (2017) allows the endogenous formation of the communication structure depending on individuals' characteristics and homophily.

2 Theory

2.1 The opinion updating process

Our theoretical framework pertains to the family of average–based updating processes introduced by DeGroot (1974), and builds on DeMarzo et al. (2003). The key difference with DeMarzo et al. (2003) is that we don't assume the presence of fixed communication networks. Our theory hence highlights the exact role of fixed networks in the shape of disagreement and provides testable predictions in the presence or absence of networks to be explored experimentally.

Consider a population D consisting of N agents, forming opinions on K different issues.[6] Agents 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

[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.

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 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 $\mathbf{T}(t) = \left(T_{i,j}(t)\right)$, which will be called the *listening matrix*. Moreover, all positive weights should be bounded away from zero, i.e. there is $b > 0$ such that $T_{i,j}(t) \geq b$ whenever $T_{i,j}(t) > 0$. This assumption (often mentioned as *uniform positivity*) rules out cases in which some weights converge to zero in the limit of the sequence of listening matrices.[8]

When at every period $t = 1, 2, \dots$ (i) the network of communication is strongly connected, (ii) all agents put strictly positive weights to their own current opinion and (iii) the listening matrices satisfy uniform positivity, then covergence of opinions to consensus is guaranteed in the long–run (see Nedic and Ozdaglar, 2009; Nedic and Liu, 2014).

It is useful to define a sequence of such listening matrices as $\mathcal{T}_t = \{\mathbf{T}(\tau)\}_{\tau=1}^{t}$

[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]In fact, this is the more general version of the assumption $\sum_{t=1}^{\infty} \lambda(t) = +\infty$ that appears in DeMarzo et al. (2003).

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 follows:

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

where (1) can be also written as follows:[9]

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

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

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

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, $\mathbf{PC_1}(t)$.[10] The major advantage of 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).

[10]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 component analysis see Jolliffe (2014).

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* $\mathbf{C}^{\kappa,\lambda}$ whose element $C^{\kappa,\lambda}_{i,j}$ is equal to 1 whenever i's opinion on the line relative to j's is concordant to κ's opinion relative to λ's, and equal to 0 otherwise.[11] As we use the principal component of opinions to define the comparison matrix, it is well-defined even in the absence of unidimensionality.If *unidimensionality* holds, the choice of the reference pair (κ, λ) is inconsequential, resulting in the same comparison matrix up to transposition.[12]

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 less 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 *strong position determinacy* which is the following property.

Property 2. (Strong Position Determinacy) An opinion formation process that

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

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

The opinion comparison matrix $\mathbf{C}^{\kappa,\lambda}(t)$ has elements $\mathbf{C}^{\kappa,\lambda}_{i,j}(t) = c(s^P_i(t), s^P_j(t); s^P_\kappa(t), s^P_\lambda(t))$, where the 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: $\mathbf{C}^{\kappa,\lambda}(t) = C(\mathcal{T}_t, \mathbf{s}(\mathbf{0}))$.

[12] 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.

can be described by (1) has *strong position determinacy* if:

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

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 *strong position determinacy* holds, relative positions in the long-run converge and this convergence occurs in a way that is captured by the long–run *opinion comparison matrix* and does not depend on the initial opinions. Note that, *strong position determinacy* does not imply *unidimensionality*, nor the other way around.

Alternatively, one could consider a weaker property of position determinacy, according to which relative positions eventually become independent of initial opinions, even if they do not converge. More specifically, consider a sequence of listening matrices $\mathcal{T}_t$ and two different initial opinion matrices $\mathbf{s_A}(t), \mathbf{s_B}(t)$ and let the matrices $\mathbf{C}_A^{\kappa,\lambda}(t)$ and $\mathbf{C}_B^{\kappa,\lambda}(t)$ denote the opinion comparison matrices for the updating process starting at the respective initial opinions, for some common reference pair (κ, λ). Then, the property *weak position determinacy* is defined as follows:

Property 3. (Weak Position Determinacy) An opinion formation process that can be described by (1) has *weak position determinacy* if:

$$\lim_{t\to\infty} \mathbf{C}_A^{\kappa,\lambda}(t) - \mathbf{C}_B^{\kappa,\lambda}(t) = \mathbf{0} \ , \quad \text{for any } \mathbf{s_A}(0), \mathbf{s_B}(0) \in \mathbb{R}^{N\times K}$$

It is important to notice that $\mathbf{C}_A^{\kappa,\lambda}(t)$ and $\mathbf{C}_B^{\kappa,\lambda}(t)$ need not converge for *weak position determinacy* to hold.

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, $\mathbf{PC}_1(t)$.

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 4. (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 *(strong/weak) 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.[13] However, despite the fact that opinions converge, at each point in time there is disagreement. *Unidimensionality* implies

[13]In our case, as it has been mentioned already, 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, and positive weights being bounded away from zero are sufficient to obtain the result (see Nedic and Ozdaglar, 2009; Nedic and Liu, 2014). Another interesting case is when the listening matrices are randomly drawn, for which asymptotic consensus is proven by Tahbaz-Salehi and Jadbabaie (2008).

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.[14] We denote by α_2 the second largest eigenvalue of this matrix. 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.[15]

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:

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

where $\lambda(t) \in (0, 1]$, $\mathbf{I}$ is the identity matrix 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"[16] 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, Strong Position Determinacy and Direction Determinacy.*

DeMarzo et al. (2003) 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.

[14] 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 $\|a + ib\| = \sqrt{a^2 + b^2}$.

[15] 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.

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

However, it turns out that *unidimensionality* is a much more general property that emerges even when the structure of the listening matrix is not rigid across rounds and is allowed to change. In fact, we show this to be true when the sequence of listening matrices satisfies certain conditions, most of which are satisfied generically. Moreover, under the same conditions, we show that the process also satisfies *weak position determinacy*, which suggests that relative positions, despite not stabilizing, eventually become independent of the initial opinions. For the economy of exposition we state the result informally and relegate the formal conditions and the proof to the Appendix. Namely:[17]

Theorem 2. *For a general class of sequences of listening matrices $\mathcal{T}_\infty$ and matrices of initial opinions* $\mathbf{s}(0)$*, the opinion process described by (1) satisfies Unidimensionality and Weak Position Determinacy.*

Notice that this result does not specify the way in which the elements of the sequence are selected. This means that the matrices could be determined either randomly or deterministically, as long as they satisfy the necessary conditions. 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.

The conditions we impose are mostly regulatory. In particular, we retain the conditions of individual listening matrices being simple –i.e. to have distinct eigenvalues–, thus also diagonalizable, and having second larger eigenvalues that are real and we consider these conditions to hold also in the limit of their sequence. To these we add two conditions on the whole sequence of listening matrices that essentially guarantee that the long–run opinion disagreements are governed by the second largest eigenvalues and the respective eigenvectors of the listening matrices. This suggests that the generalization of *unidimensionality*

[17]Unidimensionality 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 relation between correlation and the percentage of variance explained by PCA, despite the two measure being intuitively similar. Weak position determinacy is proven by showing that the relative opinions on two distinct issues, compared to a common reference pair, get close as time grows irrespectively of the initial opinions.

is again related to the size of the second eigenvalues of the sequence of listening matrices, yet (as becomes apparent in the formal analysis) in a much more complicated way than for processes where the listening matrix is constant. In fact, the result on *weak position determinacy* suggests that at a given period the values of the column eigenvector associated with the second largest eigenvalue provide a predictor of the individuals' relative positions at that period that becomes more and more accurate over time.

3 Experimental Design

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 in the different treatments.

3.1 The experiment

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.

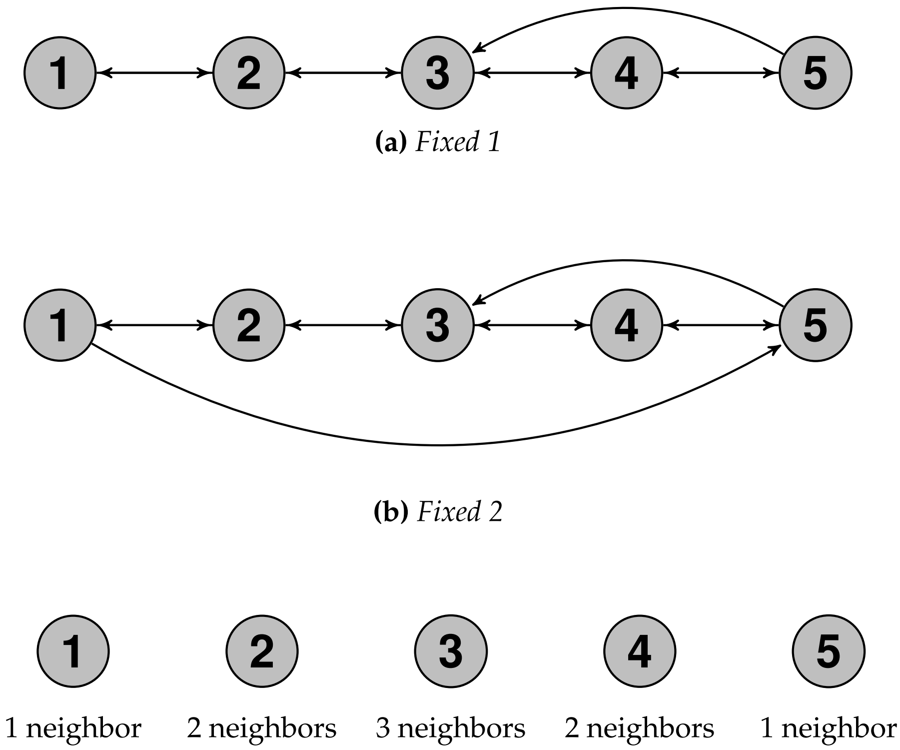
The number of red balls represents an unknown state and the guess represents the subject's opinion about what this state is. The high number of balls is chosen to avoid the necessity of decimal numbers for subjects to give finer guesses. For the main experiment we ask subjects to make guesses about two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

Subjects play the guessing game for three phases, consisting of 10 rounds each.[18] 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 purple 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. Subjects' experimental earnings depend on how close their guess is to the actual state.

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 subject's neighbors (group members whose guesses she observes) remains 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.

[18]The main experiment was preceded by two parts that aimed at familiarizing subjects with the information and communication environments. In part 1 subjects made guesses for a single tank without communication. In part 2 subjects could communicate but again made guesses for only one tank. Part 3 was the main part of the experiment described above. 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.

Subjects are not informed about the structure of their group's network.[19] 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.

[19]This as another difference in terms of design compared to related work on one dimension by Corazzini et al. (2012); Brandts et al. (2015); Battiston and Stanca (2015). We opted for an unknown network to guarantee comparability between the designs of fixed and random networks where in the latter keeping track of the continuously changing network would be practically infeasible. Notice also that the theory of DeMarzo et al. (2003) does not require knowledge of the network structure.

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 $\{0, 1, 2, ..., 100000\}$. The samples were i.i.d. draws from a binomial distribution with parameters $n = 100$ and $p = \frac{\theta_k}{100000}$. 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 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.

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.[20] In total we had 12 groups for each treatment. Average total payment was around £10.5 and the experiment lasted about 90 minutes.[21]

3.2 Hypotheses

Following Theorems 1 and 2, **Unidimensionality** should arise 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.*

Theorem 1 suggests that if in each fixed network groups use the same listening matrix and this differs between the two networks, then in general we

[20] 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.

[21] 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 the main part of the experiment in part 3. Subjects received an additional £3 participation fee.

should expect differences in subjects' **relative positions**. In other words, we expect relative positions to become the same within each fixed treatment and expect to see differences across fixed treatments.

Hypothesis 2 (a). *Relative positions as projected on to the long-run opinions' first principal component should generically differ across treatments.*

To obtain more crisp predictions about subjects' relative positions after some rounds of communication we focus on the theoretical predictions assuming that subjects assign an equal weight to their own and every other neighbor's opinion for each round of communication. While strong, this assumption seems the most natural as a theoretical benchamark, since subjects in the experiment are not aware of the network structure and therefore all neighbors are identical to them. Updating according to 3 then requires $\lambda(t) = 1$ for all t and listening matrices $\mathbf{T}_{F1}$ and $\mathbf{T}_{F2}$ defined as follows:

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

The difference between these two listening matrices is in the last row representing the weight that agent 5 assigns to her 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 in the fixed treatments are then determined by the ranking of agents' corresponding element in the second column eigenvector of $\mathbf{T}$.[22] 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 column 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 $(4, 3, 5, 2, 1)$).[23]

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

[23] 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 $\mathbf{T}_{F2}$.

In the *Random* treatment the updating matrix can not remain the same and Theorem 1 does not apply. However, assuming again equal weights, intuition and simulations suggest that 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. Therefore, while in *Random* it is not possible to predict on which 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. As it turns out 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 and equal weights we can calculate the expected opinion comparison matrix $E\left[\hat{\mathbf{C}}^{\kappa,\lambda}_{random}\right]$.

Hypothesis 2 (b). *Assuming that all subjects in a given treatment update their opinions by assigning equal weight to their own and their neigbors' opinions, relative positions as projected on to the long-run opinions' first principal component will converge to the following opinion comparison matrices:*

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

$$E\left[\hat{\mathbf{C}}^{1,3}_{random}\right] = \begin{pmatrix} . & . & . & . & . \\ 0.18 & . & . & . & . \\ . & 0.46 & . & . & . \\ 0.18 & 0.46 & 0.55 & . & . \\ 0.21 & 0.44 & 0.45 & 0.43 & . \end{pmatrix}$$

Notice that we present the opinion comparison matrices fixing $\kappa = 1$ and $\lambda = 3$. 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*. Such choice guarantees that subjects at these nodes in the two fixed networks remain at the same distance in both treatments and are the furthest apart from all such pairs of nodes hence making the exercise less sensitive to noise. Also notice that for illustrative purposes, we present only the nine elements of each matrix that are enough to accurately describe all subjects' relative opinions.[24]

Having presented our hypotheses, 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 Updating in the lab

Having described our design and hypotheses it must be clear that the aim of the experiment is to recreate the stylized communication setup of the model and test its predictions regarding the emergence of a unidimensional world and its characteristics. In other words, the experiment is not designed to understand how subjects update their opinions. Given recent experimental evidence (Corazzini et al., 2012; Chandrasekhar et al., 2012; Grimm and Mengel, 2014; Brandts et al., 2015; Battiston and Stanca, 2015) we rather take the averaging model as a reasonable approximation of individual behavior and test the robustness of our theoretical predictions to other behavioral elements.

Notice tough that our theoretical predictions depend on a) agents updating their opinions (at least for some rounds), b) agents updating their opinions in a

[24]That is, we exclude the diagonal and (1,3) elements that are zero by definition and all entries above the diagonal given that the matrix is symmetric.

similar manner for both dimensions, and c) updating occurring in the spirit of an averaging model (i.e.,).

For illustration of our data, Figure 3 shows the evolution of opinions across rounds in a phase for one of the groups in the experiment. There is heterogeneity in what we observe, but this example facilitates the understanding of the different measures we use to summarize the data.

First of all, guesses become closer over time and most of this convergence happens in the first rounds. In this example we also see the group's guesses quickly align, captured 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 $C^{1,3}$. 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. Such "jumps" or perturbations occur sometimes in the data. Interestingly, we observe that the process of convergence picks up again immediately, only now on a line 'tilted' by the jump.

4.1 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 4 depicts the mean percentage of variance explained by the first principal component β^P in our data, for each round t and tretment (by definition $\beta^P(t) \in [50, 100]$). The middle panel presents the benchmark theoretical prediction of our model if agents were to assign equal weights to their own and their neighbors' opinions (see $\mathbf{T}_{F1}$ and $\mathbf{T}_{F2}$). It is clear that in the theoretical benchmark $\beta^P(t)$ approaches one hundred percent as our theory dictates. In our data, despite not observing so high values for the average there

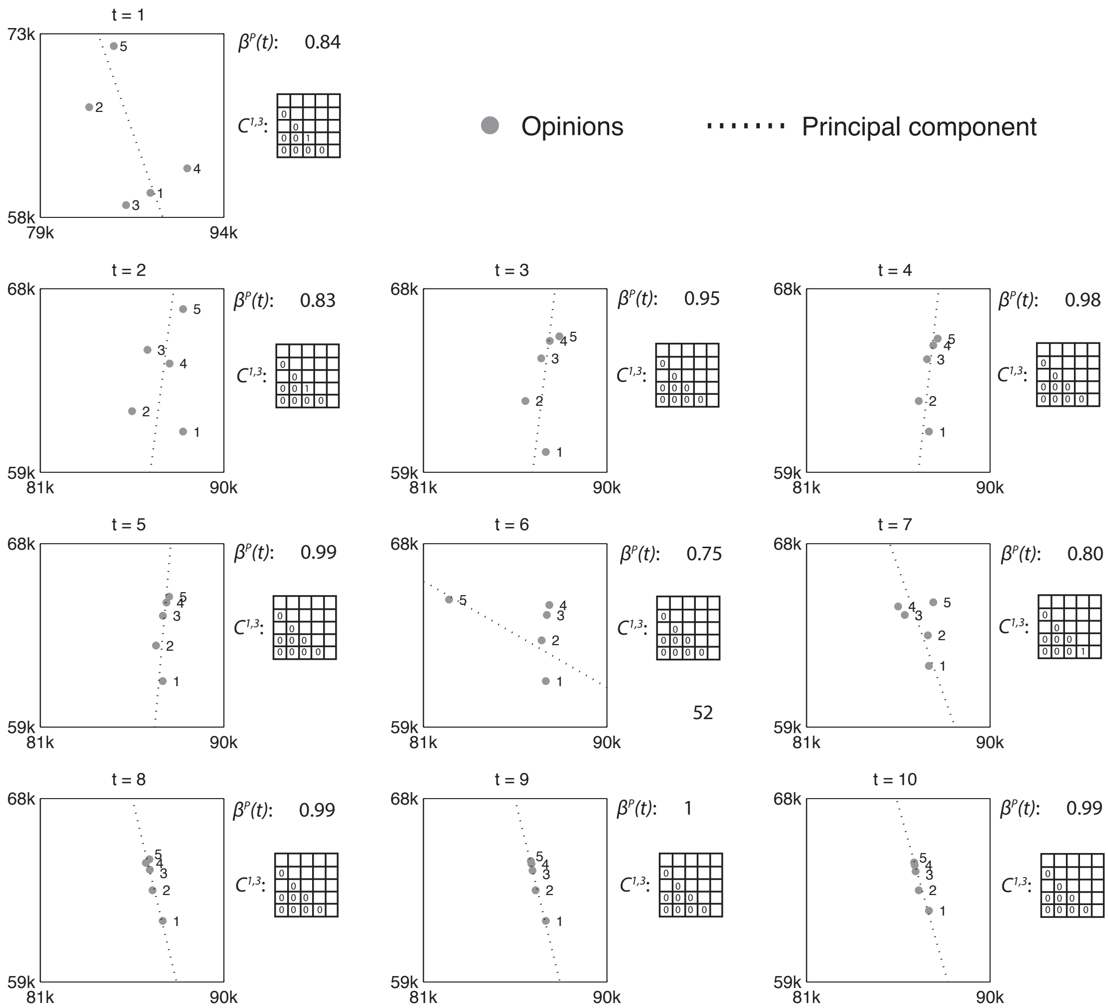


Figure 3: Guesses across rounds in a phase for a particular group in treatment *Fixed 1* in the experiment (*session 6, group 2, phase 2*). Each graph represents guesses in each round t, from 1 to 10. Each point represents a subject's guesses (in thousands) for each tank. Tank 1 in the horizontal axis 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. $\beta^P(t)$ is the variance explained by the first principal component in t. The opinion comparison matrix is $C^{1,3}$.

is an increase in $\beta^P(t)$ across all treatments.[25]

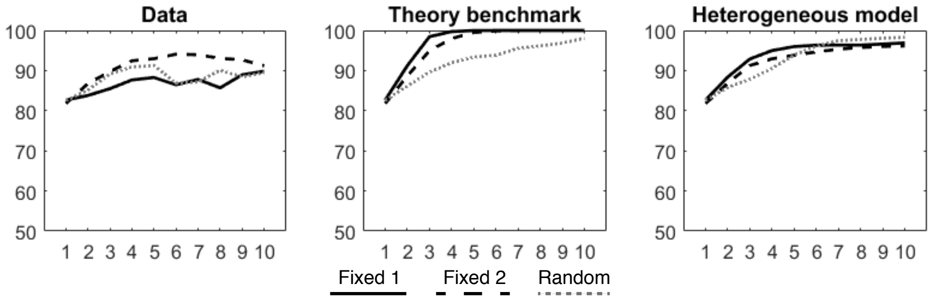


Figure 4: Mean percentage of variance across round and treatment in a) our data (left), b) theoretical predictions assuming all subjects behave the same (middle), c) theoretical predictions assuming an heterogenous behavior (right).

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.

Support: Figure 4 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* ($p = 0.002$).[26] Furthermore, the mean (median) value of $\beta^P(t)$ in rounds 6 to 10 is 87.4 (90.4) in *Fixed 1*, 93 (96.9) for *Fixed 2* and 88.4 (92.3) for *Random*.

There are a few things to note with respect to this result. First of all, while convergence to unidimensionality holds for all treatments, it is noticeably stronger in *Fixed 2*. As it turns out, in this treatment there is less noise due to perturbations, which could explain this difference.[27] Second, the graphs indicate 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 reject the null that there is no positive trend in all three treatments ($p < 0.001$ for all three tests). For rounds 6 to 10 we cannot reject the null of no trend in any of the treatments ($p = 0.484$

[25]We will refer to the *Heterogeneous model* on the right panel after presenting our main results.

[26]Note that the seasonal Mann Kendall test uses information from individual group observations, not just the aggregate data shown in Figure 4.

[27]This is discussed in more details in Appendix B

in *Fixed 1*, $p = 0.134$ in *Fixed 2*, $p = 0.388$ in *Random*). This "early action" in our data is in line with most updating taking place in the first rounds. As we show in Appendix B, convergence of opinions as measured by the normalized coefficient of variation also mainly occurs in the first 5 rounds. Third, the mean β^P in round 1 seems relatively high, which can lead to the concern that this may explain the high levels achieved in subsequent rounds. 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 $\beta^P(1)$ 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.

4.2 Relative Positions

To compare subjects' relative positions we rely on the *opinion comparison matrices*, $C^{1,3}_{i,j}$. Recall that by definition, any matrix $C^{\kappa,\lambda}_{i,j}$ is symmetric, the elements of the diagonal are 0 and the elements (κ, λ) and (λ, κ) are 1 and 0 respectively. 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).

The left panel of Figure 5 shows the average value for each of those elements across rounds for all observations in our data for each treatment. Taking into account subjects' initial guesses, the middle panel presents the theoretical predictions of Hypothesis 2(b), with all subjects updating their opinions in the same manner and putting equal weight to their own and their neighbors' opinions. Observing the graph, 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(b) -and the panel in the middle- precisely these two elements should contain a treatment effect across the two fixed network treatments. The exact prediction according to Hypothesis 2(b) would be for these two elements to converge to 1, which we do not perfectly observe here. Still, their average values do remain above 0.5. In *Random* we do not observe any tendency for elements to converge to extreme values. In fact, on average elements take values very close to the prediction obtained in Hypothesis 2(b).

Result 2. Hypothesis 2(a) can not be rejected, Hypothesis 2(b) is partially supported.

Support: Concerning Hypothesis 2(a) we compare the graph in the first row and column of Figure 5, with the one right below it. In particular, we compare the average values of elements in $C^{1,3}_{Fixed\ 1}$ and $C^{1,3}_{Fixed\ 2}$, averaged across rounds 6 to 10. A Fisher's exact test returns significant or slightly significant differences between elements (2,1) ($p = 0.019$), (4,3) ($p = 0.027$), (5,3) ($p = 0.089$) and (5,4) ($p < 0.001$). We therefore conclude that relative positions in these two treatments are different.

Concerning Hypothesis 2(b) we compare the three graphs in the first column of Figure 5 to the ones in the second column. In particular, we compare the average values in $C^{1,3}_{Fixed\ 1}$, $C^{1,3}_{Fixed\ 2}$ and $C^{1,3}_{Random}$, averaged across rounds 6 to 10, with the corresponding theoretical predictions from Hypothesis 2(b). For *Random* we only find significant differences for element (2,1) ($p = 0.014$): the value in the data is 0.30 when theory predicts 0.18. For both *Fixed 1* and *Fixed 2* all elements of the empirical comparison matrices differ significantly from the theoretical predictions. Nevertheless, as mentioned in the discussion above, some of the qualitative features of the theoretical predictions do manifest in the data.

4.3 Heterogeneity

As we show in the previous section, when comparing our experimental results to the theoretical benchmark we find evidence in support of the averaging model of DeGroot (1974) and the predictions about unidimensionality put forth by De Marzo et al. (2003) and ourselves. As in any empirical exercise, we observe some divergence from the theoretical bencharks. The question then is whether we can identify the reasons for such divergence when it is observed. In this

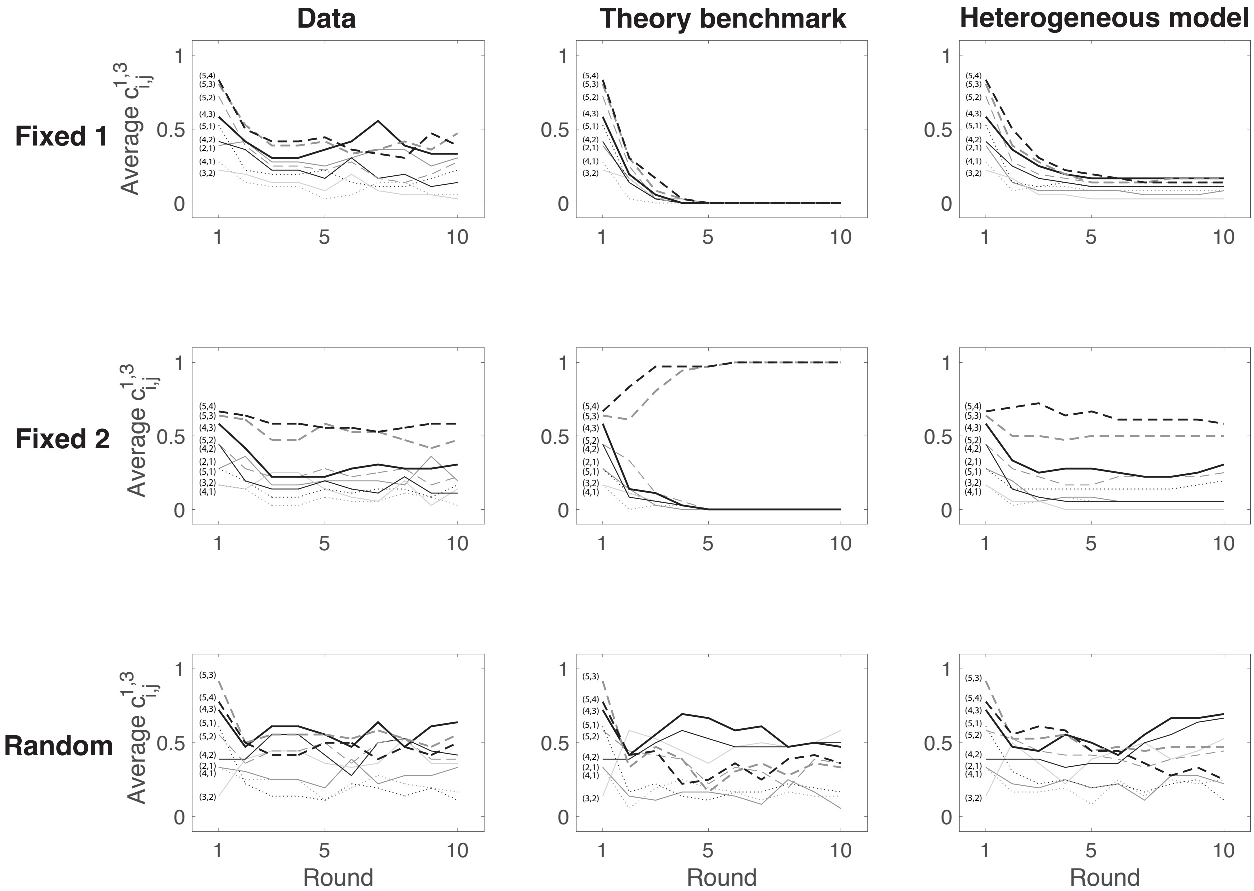


Figure 5: Each line corresponds to an element below the diagonal of the opinion comparison matrix, specified by the pairs in parenthesis on the left of each graph. Values indicate the average value for that element across all groups in each treatment in a given round. The graphs on the left show the values in our experimental data. The graphs on the middle simulate how the initial guesses in our data would evolve if all subjects behave the same and put equal weights to all their neighbors and their own previous guess as in Hypothesis 2(b). The graphs on the right simulate how the initial guesses in our data would evolve permitting an heterogeneous behavior among subjects.

section we show how a simple extension of the homogeneous averaging model introduced earlier, which allows for heterogeneity among individuals' updating behavior, can account for most of these cases and, when simulated, traces our experimental observations very well.

The theoretical benchmark we used for the design of the experiment and the formulation of hypothesis is given in equation (3). In that homogeneous model the updating behavior captured by $\lambda(t) \in (0, 1]$ must satisfy $\sum_{t=1}^{\infty} \lambda(t) = +\infty$. We use $\lambda(t) = 1, \ \forall \ t$ to obtain our theoretical benchmark. Using different updating types (λ functions) could offset some of the divergence of the data from the theory to a limited extend. Here we go a step further allowing different individuals $i \in D$ to have different updating behaviors, captured by a respective λ function. Formally equation (3) is replaced by:

$$T_{i,j}(t) = (1 - \lambda_i(t))\mathbf{1}_{i=j}(j) + \lambda_i(t)\frac{1}{|D_i(t) \cup i|} \tag{4}$$

All else remains as before. A value of $\lambda_i(t) = 1$ means that the individual places equal weight to all neighbors opinions including her own. A value of $\lambda_i(t) = 0$ means that an individual keeps the same opinion as in the previous round and does not perform any updating.

As it is, the model allows many degrees of freedom in choosing the number of different updating types as well as their exact specification. To determine these we turn to the data and use cluster analysis to identify a limited number of patterns of updating behavior.[28] The potential downside of using this approach is that of "over-fitting". As this exercise is mainly explorative we do not want to draw inferences about the empirical relevance of the identified updating types outside the domain of this experiment. The main objective is to obtain some feeling of the degree to which simply introducing heterogeneity in to the De Marzo et al framework can explain the divergence of the data from the theoretical benchmark. The interested reader may use our code to explore how some of the parameter choices involved in the cluster analysis exercise may result in identifying different updating types. Still, our main result, namely that

[28]The exact way in which this is done is described in the appendix. The codes used are available from the authors upon request.

the heterogeneous model explains much of the divergence between the data and the theoretical benchmark, is robust to any reasonable choice of parameters.

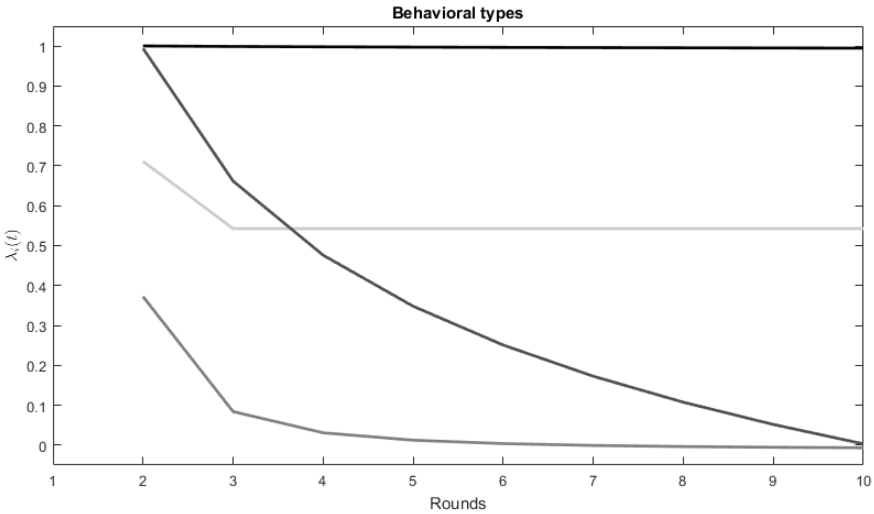


Figure 6: Updating types identified through cluster analysis and used in the Heterogeneous model.

Figure 4.3 shows the updating types identified through cluster analysis. The type represented by a horizontal line at 1 is essentially one that takes an exact average of his neighbors' and his own opinion in every round. This coincides with the behavior assumed for the theoretical benchmark. The type represented by the line starting near 0.4 and dropping to 0 fast represents an individual that only updates her initial opinion slightly in the first couple of rounds and essentially stops updating after that. Note that this behavior is not allowed in the De Marzo et al. model on which we built on. In particular it violates the condition $\sum_{t=1}^{\infty} \lambda(t) = +\infty$. The updating patterns of the other two types lie in between these two extremes. It is worth noting that the perfectly averaging and extremely stubborn types come up consistently in the cluster analysis exercise, independently of the choice of parameters used.

Next, we simulate the model using subjects' initial guesses as $s(1)$ in the same way we did for the Theory benchmark. The results of these simulations in terms of the percentage of variance explained by the principal component and the relative positions are shown in the respective panels of figures 4 and 5. In all these graphs one can see how allowing for heterogeneity gives predictions that are closer to the data than the homogeneous model. Perhaps more importantly, the heterogenenous model replicates the qualitative features of the data very well.

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.

Chandrasekhar, A. G., H. Larreguy, and J. P. Xandri (2012): "Testing models of social learning on networks: Evidence from a framed field experiment," *Work. Pap., Mass. Inst. Technol., Cambridge, MA*.

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.

Eguia, J. X. (2011): "Foundations of spatial preferences," *Journal of Mathematical Economics*, 47, 200–205.

——— (2013): "On the spatial representation of preference profiles," *Economic Theory*, 52, 103–128.

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.

Grimm, V. and F. Mengel (2014): "Experiments on belief formation in networks," *working paper*.

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.

Knight, K. (2006): "Transformations of the concept of ideology in the twentieth century," *American Political Science Review*, 100, 619–626.

Kosinski, M., D. Stillwell, and T. Graepel (2013): "Private traits and attributes are predictable from digital records of human behavior," *Proceedings of the National Academy of Sciences*, 110, 5802–5805.

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*.

Melguizo, I. (2017): "Homophily and the Persistence of Disagreement," *working paper*.

Nedic, A. and J. Liu (2014): "A Lyapunov approach to discrete-time linear consensus," in *Signal and Information Processing (GlobalSIP), 2014 IEEE Global Conference on*, IEEE, 842–846.

Nedic, A. and A. Ozdaglar (2009): "Distributed subgradient methods for multi-agent optimization," *IEEE Transactions on Automatic Control*, 54, 48–61.

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.

Taubinsky, D. (2011): "Network Architecture and the Left-Right Spectrum," *The B.E. Journal of Theoretical Economics*, 11.

Wilson, C. and J. Haidt (2014): "It's true: Liberals like cats more than conservatives do," *TIME, published on Feb. 18th 2014*.

A Formal Theoretical Results

In this section we provide the formal proof of Theorem 2. This is presented in the form of two separate results, one regarding *Unidimensionality* (Theorem 3) and one regarding *Weak Position Determinacy* (Proposition 1).

First, notice that for a matrix of opinions $\mathbf{s}(t)$ the following equivalence relation holds:

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

Hence, we can focus on showing under which conditions the square correlation between two issues tends to 1 as time grows.

In order to state the formal result we need some conditions regarding the listening matrices and their sequence. The conditions are jointly sufficient. We present the conditions gradually during our analysis, so as to connect them with the role they play in establishing the result.

Let us first introduce some necessary 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 matrix in period t is 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. For α_n^t being the nth largest eigenvalue of $\mathbf{T_t}$ the vectors $\mathbf{V_n^{c,t}}$, $\mathbf{V_n^{r,t}}$ will denote the column and row eigenvectors that correspond to this eigenvalue. Finally, the correlation coefficient in period t will be denoted by 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 that satisfy uniform positivity[29] 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 $\mathbf{X^0}$, $\mathbf{Y^0}$. By Perron-Frobenius Theorem, for each matrix $\mathbf{T}_\tau$ the spectral radius is equal to 1 and the largest eigenvalue is real, positive and simple –has algebraic multiplicity one–, i.e. $\alpha_1^t = 1$ *and* $\alpha_1^t > ||\alpha_2^t||$ –henceforth, condition **(0)**–.

[29]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.

The next two conditions are also assumed to be satisfied by all listening matrices:

(1) *For each* $\mathbf{T_t}$ *the second largest eigenvalue* α_2^t *is real*, and

(2) *Each* $\mathbf{T_t}$ *is simple, i.e. it has N distinct eigenvalues*.

Condition (2) ensures that there are no eigenvalues with equal moduli except of complex conjugates. This observation, together with Condition (1) ensures that for each listening matrix the second eigenvalue α_2^t will be strictly larger (in modulus) than all remaining eigenvalues. Despite this not being an additional condition, we will refer to it in what follows as Condition (2′), as it plays an important role. Namely,

(2′) $\alpha_2^t > \|\alpha_n^t\|$ *for all* $n > 2$.

Condition (2) also ensures that *each matrix* $\mathbf{T_t}$ *is diagonalizable*. This is important as it means that we can perform an eigendecomposition to it, i.e. to factorize it into a product of the matrices of its eigenvalues and its left and right eigenvectors.[30] Given that the listening matrices are in general different, we should also ensure that conditions (0) and (2′) should also hold at the limit of the sequence of listening matrices: Formally,

(3) $\lim\limits_{\tau\to\infty} \alpha_2^\tau \neq 1$, and

(4) $\lim\limits_{\tau\to\infty} \frac{\|\alpha_3^\tau\|}{\alpha_2^\tau} \neq 1$.

Conditions (3) and (4) imply that the said limits either do not exist, or in case they exist they are bounded away from one. Note that, Condition (3) is implied by uniform positivity and is stated explicitly for completeness, as it can be seen as the generalization of $\sum\limits_{t=1}^{\infty} \lambda(t) = +\infty$ in DeMarzo et al. (2003). Furthermore, Condition (4) is commonly true as it rules out only sequences that follow very specific patterns.

We can now present ρ_t^2 as a function of the initial opinions and the eigenvalues and eigenvectors of the listening matrices. Namely,

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

[30]Condition (2) could be slightly relaxed by considering explicitly that $\mathbf{T_t}$ is diagonalizable, since simple matrices are diagonalizable. Yet, in this case we would then need to add explicitly Condition (2′). In that sense, our condition is slightly less generic than the one of DeMarzo et al. (2003), yet as pointed out by Taubinsky (2011) the set of simple matrices is dense.

$$\rho_t^2 = \frac{\sum_{i=1}^{N}\sum_{j=1}^{N}(X_i^t - \overline{X^t})(Y_i^t - \overline{Y^t})(X_j^t - \overline{X^t})(Y_j^t - \overline{Y^t})}{\sum_{i=1}^{N}\sum_{j=1}^{N}(X_i^t - \overline{X^t})(X_i^t - \overline{X^t})(Y_j^t - \overline{Y^t})(Y_j^t - \overline{Y^t})} \tag{5}$$

We proceed by expressing the factors of the form $(X_i^t - \overline{X^t})$ through eigendecomposition of the listening matrices, which is possible given that –by Condition (2)– the listening matrices are diagonalizable. In particular, each diagonalizable listening matrix $\mathbf{T_t}$ can be factorized into the following 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 (right) and row (left) eigenvectors respectively, such that $\mathbf{V^{r,t}} = (\mathbf{V^{c,t}})^{-1}$.

Note that, for all irreducible, aperiodic, row–stochastic listening matrices $\mathbf{T_t}$ it holds that $\alpha_1^t = 1$, $\mathbf{V_1^{c,t}} = \mathbb{1}$ (column vector of ones) and $\mathbf{V_1^{r,t}} = \mathbf{w_t}$ (for some positive vector $\mathbf{w_t}$) –see for instance Karlin and Taylor (1981)–. Therefore, each listening matrix can be written as follows:

$$\mathbf{T_t} = \mathbf{V^{c,t}A^tV^{r,t}} = \sum_{n=1}^{N} \alpha_n^t \mathbf{V_n^{c,t}V_n^{r,t}} = \mathbb{1}\mathbf{w_t} + \sum_{n=2}^{N} \alpha_n^t \mathbf{V_n^{c,t}V_n^{r,t}} \tag{6}$$

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. Given this, a finite product of listening matrices can be written as follows:

$$\prod_{\tau=1}^{t} \mathbf{T}_\tau = \prod_{\tau=1}^{t} \left[\mathbb{1}\mathbf{w}_\tau + \sum_{n=2}^{N} \alpha_n^\tau \mathbf{V_n^{c,\tau}V_n^{r,\tau}} \right] \tag{7}$$

Expression (7) can 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 $\mathbf{B}$ 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 $\mathbf{B}$ imply that $\alpha_n = 1$ if and only if $n = 1$. □

Note that, Condition (0) ensures that $\alpha^t_n = 1 \Leftrightarrow n = 1$, which means that for all the other eigenvalues the elements of the respective eigenvector sum to zero. This, in turn, is important for the following lemma.

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 $\sum_{n=2}^{N} \alpha^\tau_n \mathbf{V^{c,\tau}_n V^{r,\tau}_n} \mathbb{1} \mathbf{w}_{\tau-1}$ for some τ will be equal to zero. Condition (4) guarantees that this holds also in the limit (as t grows). Therefore:

$$\prod_{\tau=1}^{t} \mathbf{T}_\tau = \mathbb{1} \mathbf{w_t} \left(\prod_{\tau'=1}^{t-1} \mathbf{T}_{\tau'} \right) + \prod_{\tau=1}^{t} \left(\sum_{n=2}^{N} \alpha^\tau_n \mathbf{V^{c,\tau}_n V^{r,\tau}_n} \right) \tag{8}$$

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, except of the ith element that is equal to 1.

Moreover, $\hat{\mathbf{w}} = (1/N)\mathbb{1}^T$ is a row vector with all of its elements being equal to $1/N$.[31] The following lemma will help us simplify the final expression.

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 $(X_i^t - \overline{X^t})$

$$X_i^t = \mathbf{e_i}\mathbf{X^t} = \mathbf{e_i}\prod_{\tau=1}^{t}\mathbf{T}_\tau\mathbf{X^0} = \mathbf{e_i}\left[\mathbb{1}\mathbf{w_t}\left(\prod_{\tau'=1}^{t-1}\mathbf{T}_{\tau'}\right)\mathbf{X^0}\right] + \mathbf{e_i}\left[\prod_{\tau=1}^{t}\left(\sum_{n=2}^{N}\alpha_n^\tau\mathbf{V_n^{c,\tau}}\mathbf{V_n^{r,\tau}}\right)\right]\mathbf{X^0} \quad (9)$$

$$\overline{X^t} = \hat{\mathbf{w}}\mathbf{X^t} = \hat{\mathbf{w}}\prod_{\tau=1}^{t}\mathbf{T}_\tau\mathbf{X^0} = \hat{\mathbf{w}}\left[\mathbb{1}\mathbf{w_t}\left(\prod_{\tau'=1}^{t-1}\mathbf{T}_{\tau'}\right)\mathbf{X^0}\right] + \hat{\mathbf{w}}\left[\prod_{\tau=1}^{t}\left(\sum_{n=2}^{N}\alpha_n^\tau\mathbf{V_n^{c,\tau}}\mathbf{V_n^{r,\tau}}\right)\right]\mathbf{X^0} \quad (10)$$

[31]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_t^2. They would also need to consider as a normalizing factor the standard deviation in each issue, instead of the issue–independent normalization that is enough for characterizing 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.

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

$$X_i^t - \overline{X^t} = \mathbf{e_i}\left[\prod_{\tau=1}^{t}\left(\sum_{n=2}^{N}\alpha_n^\tau \mathbf{V_n^{c,\tau}}\mathbf{V_n^{r,\tau}}\right)\right]\mathbf{X^0} - \mathbf{\hat{w}}\left[\prod_{\tau=1}^{t}\left(\sum_{n=2}^{N}\alpha_n^\tau \mathbf{V_n^{c,\tau}}\mathbf{V_n^{r,\tau}}\right)\right]\mathbf{X^0} \tag{11}$$

$$= \left[\sum_{n=2}^{N}\alpha_n^t\left(V_{i,n}^{c,t} - \mathbf{\hat{w}}\mathbf{V_n^{c,t}}\right)\mathbf{V_n^{r,t}}\right]\prod_{\tau=1}^{t-1}\left(\sum_{m=2}^{N}\alpha_m^\tau \mathbf{V_m^{c,\tau}}\mathbf{V_m^{r,\tau}}\right)\mathbf{X^0} \tag{12}$$

$$= \sum_{n_1,\ldots,n_t=2}^{N}\left[\alpha_{n_t}^t\left(V_{i,n_t}^{c,t} - \mathbf{\hat{w}}\mathbf{V_{n_t}^{c,t}}\right)\mathbf{V_{n_t}^{r,t}}\left(\prod_{\tau=1}^{t-1}\alpha_{n_\tau}^\tau \mathbf{V_{n_\tau}^{c,\tau}}\mathbf{V_{n_\tau}^{r,\tau}}\right)\mathbf{X^0}\right] \tag{13}$$

$$= \sum_{n_1,\ldots,n_t=2}^{N}\left[\left(\prod_{\tau=1}^{t}\alpha_{n_\tau}^\tau\right)\left(V_{i,n_t}^{c,t} - \mathbf{\hat{w}}\mathbf{V_{n_t}^{c,t}}\right)\left(\prod_{\tau=1}^{t-1}\mathbf{V_{n_{\tau+1}}^{r,\tau+1}}\mathbf{V_{n_\tau}^{c,\tau}}\right)\left(V_{n_1}^{r,1}\mathbf{X^0}\right)\right] \tag{14}$$

where $V_{i,n_t}^{c,t}$ denotes the ith element of the column eigenvector corresponding to the nth largest eigenvalue of $\mathbf{T_t}$. Notice that, $\left(V_{i,n_t}^{c,t} - \mathbf{\hat{w}}\mathbf{V_{n_t}^{c,t}}\right)$ is the only factor that depends on i, it is a scalar and does not depend on $\mathbf{X^0}$. Moreover, all the other factors that appear inside parentheses in Equation (14) are also scalars and do not depend on i.

At this point we impose an additional condition on the eigenvectors associated with the second largest eigenvalues of consecutive listening natrices. Namely,

(5) $\mathbf{V_2^{r,t}}\mathbf{V_2^{c,t-1}} \neq 0$ for all t.

Condition (5) will commonly hold for most pairs of matrices as for each $\mathbf{V_2^{c,t-1}}$ the vector $\mathbf{V_2^{r,t}}$ needs to be exactly perpendicular to it to violate the condition.

A similar sum of products can be defined for $X_j^t - \overline{X^t}$, $Y_i^t - \overline{Y^t}$ and $Y_j^t - \overline{Y^t}$. Given that each of these products appear both in the numerator and the denominator of ρ_t^2, we can divide each of them with $\prod_{\tau=1}^{t}\alpha_2^\tau$ and $\prod_{\tau=1}^{t-1}\mathbf{V_2^{r,\tau+1}}\mathbf{V_2^{c,\tau}}$ and ρ_t^2 will remain unchanged. The product of eigenvalues is real –by Condition (1)– and different from zero –by Condition (0)– and the product of eigenvectors is also real –as a consequence of Condition (1)– and different from zero –by Condition (5)–.This will be helpful when calculating the limit of ρ_t^2 as t grows. Yet, note that, as time grows, the number of elements in each sum as the one in Equation (14) grows, which means that one should be careful when intending to define the limit values of ρ_t^2 in relation to those sums, which is what we do in the next part.

More specifically, we define the set $\mathcal{S} = \mathcal{L}^{\mathbb{N}}$, where $\mathcal{L} = \{2, \ldots, N\}^4$, with typical element $(\mathbf{L^1}, \mathbf{L^2}, \ldots) \in \mathcal{S}$. Intuitively, $\mathbf{L^t} = (l_1^t, l_2^t, l_3^t, l_4^t)$, where l_k^t determines the eigenvalue n_t for each of the four parts of each product –corresponding to $X_i^t - \overline{X^t}$, $Y_i^t - \overline{Y^t}$, $X_j^t - \overline{X^t}$ and $Y_j^t - \overline{Y^t}$ for the numerator and analogously for the denominator–. As usual, $\mathcal{S}$ is endowed with the σ-algebra $\mathcal{F}$ generated by the cylinder sets, that is by all the events of the form $\{\mathbf{L^1}\} \times \cdots \times \{\mathbf{L^t}\} \times \mathcal{L} \times \mathcal{L} \times \cdots$.

Define the measure μ over the measurable space $(\mathcal{S}, \mathcal{F})$ as follows: Fix an arbitrary $t \in \mathbb{N}$ and take the algebra $\mathcal{F}_t$ which is generated by the events in the collection $\{\{\mathbf{L^1}\} \times \cdots \times \{\mathbf{L^t}\} \times \mathcal{L} \times \mathcal{L} \times \cdots | (\mathbf{L^1}, \ldots, \mathbf{L^t}) \in \mathcal{L}^t\}$. Then let μ_t be a measure that assigns equal weight $1/(N-1)^{4t}$ to every subset on the partition that generates $\mathcal{F}_t$. Essentially, this is a partition of the set $\mathcal{S}$ to subsets whose elements share the same elements in the first t rounds. By the Kolmogorov extension theorem there exists a unique measure over $(\mathcal{S}, \mathcal{F})$ that agrees with every μ_t in events in $\mathcal{F}_t$, which is the measure we call μ.

Now, we partition $\mathcal{S}$ into two subsets $\mathcal{S}_1$ and $\mathcal{S}_2$ as follows: The set $\mathcal{S}_1$ contains all elements of $\mathcal{S}$ for which all but finitely many vectors $\mathbf{L^t}$ are equal to $(2, 2, 2, 2)$. Formally:

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

Given this definition, we define $\mathcal{S}_2 := \mathcal{S} \setminus \mathcal{S}_1$, which contains the remaining elements of $\mathcal{S}$. It is apparent that both sets $\mathcal{S}_1$ and $\mathcal{S}_2$ are measurable and that they have a strictly positive measure on $\mathcal{S}$.

For each $s = (\mathbf{L^1}, \mathbf{L^2}, \ldots) \in \mathcal{S}$ we can define two sequences of complex numbers $\{f_{i,j}^t(s)\}_{t=1}^\infty$ and $\{g_{i,j}^t(s)\}_{t=1}^\infty$ as follows:

$$f_{i,j}^t(s) = f_{i,j}(\mathbf{L^1}, \mathbf{L^2}, \ldots, \mathbf{L^t}) = (X_i^t - \overline{X^t})(Y_i^t - \overline{Y^t})(X_j^t - \overline{X^t})(Y_j^t - \overline{Y^t}) \cdot \left(\prod_{\tau=1}^{t} \frac{1}{\alpha_2^\tau} \right)^4 \cdot \left(\prod_{\tau=1}^{t-1} \frac{1}{\mathbf{V_2^{r,\tau+1} V_2^{c,\tau}}} \right)^4 =$$

$$= \left[\prod_{k=1}^{2} \left(V_{i,l_k^t}^{c,t} - \hat{\mathbf{w}} \mathbf{V_{l_k^t}^{c,t}} \right) \prod_{k=3}^{4} \left(V_{j,l_k^t}^{c,t} - \hat{\mathbf{w}} \mathbf{V_{l_k^t}^{c,t}} \right) \right] \cdot \left[\prod_{k=1}^{4} \left(\prod_{\tau=1}^{t} \frac{\alpha_{l_k^\tau}^\tau}{\alpha_2^\tau} \right) \right] \cdot \left[\prod_{k=1}^{4} \left(\prod_{\tau=1}^{t-1} \frac{\mathbf{V_{l_k^{\tau+1}}^{r,\tau+1} V_{l_k^\tau}^{c,\tau}}}{\mathbf{V_2^{r,\tau+1} V_2^{c,\tau}}} \right) \right] \cdot \left[\left(V_{l_1^1}^{r,1} \mathbf{X^0} \right) \left(V_{l_2^1}^{r,1} \mathbf{Y^0} \right) \left(V_{l_3^1}^{r,1} \mathbf{X^0} \right) \left(V_{l_4^1}^{r,1} \mathbf{Y^0} \right) \right]$$

$$g_{i,j}^t(s) = g_{i,j}(\mathbf{L^1}, \mathbf{L^2}, \ldots, \mathbf{L^t}) = (X_i^t - \overline{X^t})(X_i^t - \overline{X^t})(Y_j^t - \overline{Y^t})(Y_j^t - \overline{Y^t}) \cdot \left(\prod_{\tau=1}^{t} \frac{1}{\alpha_2^\tau} \right)^4 \cdot \left(\prod_{\tau=1}^{t-1} \frac{1}{\mathbf{V_2^{r,\tau+1} V_2^{c,\tau}}} \right)^4 =$$

$$= \left[\prod_{k=1}^{2} \left(V_{i,l_k^t}^{c,t} - \hat{\mathbf{w}} \mathbf{V_{l_k^t}^{c,t}} \right) \prod_{k=3}^{4} \left(V_{j,l_k^t}^{c,t} - \hat{\mathbf{w}} \mathbf{V_{l_k^t}^{c,t}} \right) \right] \cdot \left[\prod_{k=1}^{4} \left(\prod_{\tau=1}^{t} \frac{\alpha_{l_k^\tau}^\tau}{\alpha_2^\tau} \right) \right] \cdot \left[\prod_{k=1}^{4} \left(\prod_{\tau=1}^{t-1} \frac{\mathbf{V_{l_k^{\tau+1}}^{r,\tau+1} V_{l_k^\tau}^{c,\tau}}}{\mathbf{V_2^{r,\tau+1} V_2^{c,\tau}}} \right) \right] \cdot \left[\left(V_{l_1^1}^{r,1} \mathbf{X^0} \right) \left(V_{l_2^1}^{r,1} \mathbf{X^0} \right) \left(V_{l_3^1}^{r,1} \mathbf{Y^0} \right) \left(V_{l_4^1}^{r,1} \mathbf{Y^0} \right) \right]$$

Given the previous definitions, the correlation in each period t is equal to:

$$\rho_t^2 = \frac{\sum_{i=1}^{N}\sum_{j=1}^{N}\left[\int_{\mathcal{S}_1} f_{i,j}^t d\mu + \int_{\mathcal{S}_2} f_{i,j}^t d\mu\right]}{\sum_{i=1}^{N}\sum_{j=1}^{N}\left[\int_{\mathcal{S}_1} g_{i,j}^t d\mu + \int_{\mathcal{S}_2} g_{i,j}^t d\mu\right]}$$

which can also be written as follows:

$$\rho_t^2 = 1 + \frac{\sum_{i=1}^{N}\sum_{j=1}^{N}\left[\int_{\mathcal{S}_1} f_{i,j}^t d\mu - \int_{\mathcal{S}_1} g_{i,j}^t d\mu + \int_{\mathcal{S}_2} f_{i,j}^t d\mu - \int_{\mathcal{S}_2} g_{i,j}^t d\mu\right]}{\sum_{i=1}^{N}\sum_{j=1}^{N}\left[\int_{\mathcal{S}_1} g_{i,j}^t d\mu + \int_{\mathcal{S}_2} g_{i,j}^t d\mu\right]} \tag{15}$$

Notice that both functions are integrable over the partition $\{\mathcal{S}_1, \mathcal{S}_2\}$ of $\mathcal{S}$, hence all integrals are well–defined.

We can now state the last required condition:

(6) There is $M \in \mathbb{R}$ such that $\left\| \prod_{\tau=1}^{t-1} \frac{\mathbf{V}_{\mathbf{l}_\mathbf{k}^{\tau+1}}^{\mathbf{r},\tau+1} \mathbf{V}_{\mathbf{l}_\mathbf{k}^{\tau}}^{\mathbf{c},\tau}}{\mathbf{V}_\mathbf{2}^{\mathbf{r},\tau+1} \mathbf{V}_\mathbf{2}^{\mathbf{c},\tau}} \right\| \leq M$ for all t and for all infinite sequences $\{l_k^1, l_k^2, \ldots\}$, with $l_k^t \in \{2, \ldots, N\}$.

Condition (6) implies that for all $s \in \mathcal{S}$ the fractions of products of eigenvectors of sequential matrices are uniformly bounded. This implies that for all $i, j \in N$ and for all t, we can find a constant C such that $\|f_{i,j}^t(s)\| \leq C$ and $\|g_{i,j}^t(s)\| \leq C$. This happens because all other factors of each $f_{i,j}^t$ and $g_{i,j}^t$ are bounded for any normalization of the eigenvectors. This observation allows us to use dominated convergence theorem.[32] It should be noted, that we could relax Condition (6) by considering a weaker condition that would include also the respective eigenvalues. That is, even if the product in Condition (6) diverges for some element of $\mathcal{S}$, the result would still be obtained as long as the respective product of fractions of eigenvalues $\left\| \prod_{\tau=1}^{t} \frac{\alpha_{l_k^\tau}^\tau}{\alpha_2^\tau} \right\|$ converges to zero faster.

Finally, consider the sequences $\{f_{i,j}^t(s)\}$ and $\{g_{i,j}^t(s)\}$ for $s \in \mathcal{S}_1$ and notice that

[32]One could even reform the problem in one containing only real numbers, by observing that for any two elements $s, s' \in \mathcal{S}$ such that for all complex eigenvalues that appear in s the element s' is chosen to contain their complex conjugates, the following properties hold: $f_{i,j}^t(s) = \overline{f_{i,j}^t(s')}$, $g_{i,j}^t(s) = \overline{g_{i,j}^t(s')}$ and $s \in \mathcal{S}_1 \Leftrightarrow s' \in \mathcal{S}_1$. Hence, one can consider these two elements together and observe that $f_{i,j}^t(s) + f_{i,j}^t(s') \in \mathbb{R}$ and similarly for $g_{i,j}^t$.

neither $\left[\prod_{k=1}^{4}\left(\prod_{\tau=1}^{t}\frac{\alpha^{\tau}_{l^{\tau}_{k}}}{\alpha^{\tau}_{2}}\right)\right]$ nor $\left[\prod_{k=1}^{4}\left(\prod_{\tau=1}^{t-1}\frac{\mathbf{V}^{\mathbf{r},\tau+1}_{\mathbf{l}^{\tau+1}_{\mathbf{k}}}\mathbf{V}^{\mathbf{c},\tau}_{\mathbf{l}^{\tau}_{\mathbf{k}}}}{\mathbf{V}^{\mathbf{r},\tau+1}_{\mathbf{2}}\mathbf{V}^{\mathbf{c},\tau}_{\mathbf{2}}}\right)\right]$ converge to zero pointwise as t grows. This means that in general $\{f^t_{i,j}(s)\}$ and $\{g^t_{i,j}(s)\}$ will not converge to zero, which in turn means that $\sum_{i=1}^{N}\sum_{j=1}^{N}\int_{\mathcal{S}_1} f^t_{i,j}d\mu$ and $\sum_{i=1}^{N}\sum_{j=1}^{N}\int_{\mathcal{S}_1} g^t_{i,j}d\mu$ will mostly consist of factors different than zero. Given that, these factors also depend on the vectors of initial opinions we can safely argue that for generic vectors of initial opinions $\mathbf{X^0}, \mathbf{Y^0}$ it will hold that $\lim_{t\to\infty}\sum_{i=1}^{N}\sum_{j=1}^{N}\int_{\mathcal{S}_1} f^t_{i,j}d\mu \neq 0$ and $\lim_{t\to\infty}\sum_{i=1}^{N}\sum_{j=1}^{N}\int_{\mathcal{S}_1} g^t_{i,j}d\mu \neq 0$, which means that the two limits either do not exist, or if they do exist they are different than zero. However, we will show that the same cannot be said for the quantity $\sum_{i=1}^{N}\sum_{j=1}^{N}\int_{\mathcal{S}_1}(f^t_{i,j} - g^t_{i,j})d\mu$.

Theorem 3. *Let $\rho(\mathbf{x},\mathbf{y})$ be the correlation coefficient of two vectors $\mathbf{x}$ and $\mathbf{y}$. Consider a sequence of listening matrices $\{\mathbf{T}(t)\}_{t=1}^{\infty}$ and two vectors $\mathbf{X}(t)$ and $\mathbf{Y}(t)$ that are updated according to (1), starting from some generic initial vectors $\mathbf{X}(0)$ and $\mathbf{Y}(0)$ with positive variance. If $\{\mathbf{T}(t)\}_{t=1}^{\infty}$ satisfies Conditions (0)–(6), then $\rho\left(\mathbf{X}(t),\mathbf{Y}(t)\right)^2 \to 1$ as $t\to\infty$.*

Proof. First, recall Expression (15) of the squared correlation coefficient:

$$\rho_t^2 = 1 + \frac{\sum_{i=1}^{N}\sum_{j=1}^{N}\left[\int_{\mathcal{S}_1}\left(f^t_{i,j} - g^t_{i,j}\right)d\mu + \int_{\mathcal{S}_2} f^t_{i,j}d\mu - \int_{\mathcal{S}_2} g^t_{i,j}d\mu\right]}{\sum_{i=1}^{N}\sum_{j=1}^{N}\left[\int_{\mathcal{S}_1} g^t_{i,j}d\mu + \int_{\mathcal{S}_2} g^t_{i,j}d\mu\right]}$$

Observe that, given that for generic initial vectors $\mathbf{X(0)}, \mathbf{Y(0)}$ the quantity $\sum_{i=1}^{N}\sum_{j=1}^{N}\int_{\mathcal{S}^t_1} g^t_{i,j}d\mu$ does not converge to zero, it suffices to show that the numerator of the fraction converges to zero, which we do by showing that each of the following three quantities: $\sum_{i=1}^{N}\sum_{j=1}^{N}\int_{\mathcal{S}^t_2} f^t_{i,j}d\mu$, $\sum_{i=1}^{N}\sum_{j=1}^{N}\int_{\mathcal{S}^t_2} g^t_{i,j}d\mu$ and $\sum_{i=1}^{N}\sum_{j=1}^{N}\int_{\mathcal{S}_1}\left(f^t_{i,j} - g^t_{i,j}\right)d\mu$ converges to zero. An obvious sufficient condition for these three quantities to converge to zero would be that the sequences $\{f^t_{i,j}(s)\}_{t=1}^{\infty}$ and $\{g^t_{i,j}(s)\}_{t=1}^{\infty}$ converge to zero for all $s\in\mathcal{S}_2$ and that the sequence $\{f^t_{i,j}(s) - g^t_{i,j}(s)\}_{t=1}^{\infty}$ converges to zero for all $s\in\mathcal{S}_1$, which is what we prove.

Consider any $s = (\mathbf{L^1}, \mathbf{L^2}, \dots)\in\mathcal{S}_1$. By definition of $\mathcal{S}_1$, there exists some $\hat{t}$ such that for all $\tau > \hat{t}$ it holds that $\mathbf{L}^{\tau} = (l^{\tau}_1, l^{\tau}_2, l^{\tau}_3, l^{\tau}_4) = (2,2,2,2)$. Now, consider $\underline{s} = (\underline{\mathbf{L^1}}, \underline{\mathbf{L^2}}, \dots)\in\mathcal{S}$ that is defined as follows: if $\mathbf{L^t} = (l^t_1, l^t_2, l^t_3, l^t_4)$ then $\underline{\mathbf{L^t}} = (l^t_1, l^t_3, l^t_2, l^t_4)$. Note that, the two middle elements of the vector are reversed. We can make two immediate observations regarding $\underline{s}$: First, $\underline{s} \in \mathcal{S}_1$ and second that for all $\tau > \hat{t}$ it holds that $\underline{\mathbf{L}}^\tau = (l_1^\tau, l_3^\tau, l_2^\tau, l_4^\tau) = (2,2,2,2)$ for the same $\hat{t}$ as for s. This last observation is important because it guarantees that for all $\tau > \hat{t}$ holds the equality $f_{i,j}^t(s) = g_{i,j}^t(\underline{s})$.

Given that, for each $s \in \mathcal{S}_1$ define a new sequence $\{h_{i,j}^t\}_{t=1}^\infty$ such that $h_{i,j}^t(s) = g_{i,j}^t(\underline{s})$ –observe that in some cases $s = \underline{s}$–. This new sequence is well–defined for all $s \in \mathcal{S}_1$ and by construction it satisfies that $\int_{\mathcal{S}_1} g_{i,j}^t d\mu = \int_{\mathcal{S}_1} h_{i,j}^t d\mu$ because μ is uniform. Therefore, $\int_{\mathcal{S}_1} \left(f_{i,j}^t - g_{i,j}^t\right) d\mu = \int_{\mathcal{S}_1} \left(f_{i,j}^t - h_{i,j}^t\right) d\mu$. Now, consider a new sequence $\{f_{i,j}^t(s) - h_{i,j}^t(s)\}_{t=1}^\infty$ for each $s \in \mathcal{S}_1$. By construction, for all $\tau > \hat{t}$ it holds that $f_{i,j}^\tau(s) = g_{i,j}^\tau(\underline{s}) = h_{i,j}^\tau(s)$, hence $f_{i,j}^\tau(s) - h_{i,j}^\tau(s) = 0$. Therefore, for all $s \in \mathcal{S}_1$ the sequence $\{f_{i,j}^t(s) - h_{i,j}^t(s)\}_{t=1}^\infty$ converges pointwise to 0 and given that by Condition (6) the values of $f_{i,j}^t(s) - h_{i,j}^t(s)$ are uniformly bounded for all t and s, we can use the dominated convergence theorem to conclude that $\lim\limits_{t\to\infty} \int_{\mathcal{S}_1} \left(f_{i,j}^t - h_{i,j}^t\right) d\mu = 0$, which in turn implies that $\lim\limits_{t\to\infty} \int_{\mathcal{S}_1} \left(f_{i,j}^t - g_{i,j}^t\right) d\mu = 0$.

Now, it remains to show that $\{f_{i,j}^t(s)\}_{t=1}^\infty$ and $\{g_{i,j}^t(s)\}_{t=1}^\infty$ converge pointwise to zero for all $s \in \mathcal{S}_2$. If we show this to be true, then we can use once again the dominated convergence theorem –given that Condition (6) guarantees that both $f_{i,j}^t(s)$ and $g_{i,j}^t(s)$ are uniformly bounded for all t and s– to conclude that $\lim\limits_{t\to\infty} \int_{\mathcal{S}_1} f_{i,j}^t d\mu = \lim\limits_{t\to\infty} \int_{\mathcal{S}_1} f_{i,j}^t d\mu = 0$, which completes our proof.

Hence, consider any $s \in \mathcal{S}_2$. By definition of $\mathcal{S}_2$ as the complement of $\mathcal{S}_1$, s should be such that for all τ exists $t > \tau$ such that $(l_1^t, l_2^t, l_3^t, l_4^t) \neq (2,2,2,2)$. Condition (6) ensures that the product $\left\| \prod\limits_{k=1}^{4} \left(\prod\limits_{\tau=1}^{t-1} \frac{\mathbf{V}^{\mathbf{r},\tau+1}_{\mathbf{l}_\mathbf{k}^{\tau+1}} \mathbf{V}^{\mathbf{c},\tau}_{\mathbf{l}_\mathbf{k}^{\tau}}}{\mathbf{V}^{\mathbf{r},\tau+1}_{\mathbf{2}} \mathbf{V}^{\mathbf{c},\tau}_{\mathbf{2}}} \right) \right\|$ is bounded for all t, and therefore for the result to hold it is sufficient to show that $\prod\limits_{\tau=1}^{\infty} \prod\limits_{k=1}^{4} \frac{\alpha^\tau_{l_k^\tau}}{\alpha_2^\tau} = 0$. For this to hold, $\sum\limits_{\tau=1}^{\infty} \sum\limits_{k=1}^{4} \log\left(\left\| \frac{\alpha^\tau_{l_k^\tau}}{\alpha_2^\tau} \right\| \right)$ must diverge to $-\infty$. By Condition (2'), for all τ exists $t > \tau$ such that $\left\| \frac{\alpha^t_{l_k^t}}{\alpha_2^t} \right\| < 1$. In addition to this, by Condition (4), if $\lim\limits_{t\to\infty} \left\| \frac{\alpha_3^t}{\alpha_2^t} \right\| = L < 1$ (the limit exists and is strictly less than 1), then there is τ such that for all $> \tau$ holds that $\left\| \frac{\alpha_3^t}{\alpha_2^t} \right\| \leq 1 - \epsilon$ for some $\epsilon > 0$. If the said limit does not exist, then there is $\epsilon > 0$ such that for all τ exists $t > \tau$ such that $\left| \frac{\|\alpha_3^t\|}{\|\alpha_2^t\|} - 1 \right| > \epsilon$ (otherwise 1 would be its limit), hence for infinitely many t we get that $\left\| \frac{\alpha_3^t}{\alpha_2^t} \right\| \leq 1 - \epsilon$ for some $\epsilon > 0$. Given that $\|\alpha_3^t\| \geq \|\alpha_k^t\|$ for $k > 3$, we conclude that there is $\epsilon > 0$ such that $\left\|\frac{\alpha_{l_k^t}^t}{\alpha_2^t}\right\| \leq (1-\epsilon)$ for infinitely many elements of s, whereas for the remaining elements $\left\|\frac{\alpha_{l_k^t}^t}{\alpha_2^t}\right\| = 1$. Let $\delta = \log(1-\epsilon) < 0$, then $\log\left(\left\|\frac{\alpha_{l_k^t}^t}{\alpha_2^t}\right\|\right) \leq \delta < 0$ for infinitely many elements of s and $\log\left(\left\|\frac{\alpha_{l_k^t}^t}{\alpha_2^t}\right\|\right) = 0$ for the remaining elements. $\sum_{\tau=1}^{\infty}\sum_{k=1}^{4}\log\left(\left\|\frac{\alpha_{l_k^t}^t}{\alpha_2^t}\right\|\right)$ diverges to $-\infty$, which completes the argument.

□

We now turn our attenton to Weak Position Determinacy. In what follows, let $\mathbf{X^0}$ and $\mathbf{Y^0}$ denote two distinct vectors of initial conditions regarding the same issue, rather than two different issues. In order to prove that Weak Position Determinacy holds it is sufficient to prove that for some reference pair (κ, λ) holds the following:[33]

$$\left(\frac{X_i^t - X_j^t}{X_\kappa^t - X_\lambda^t}\right) - \left(\frac{Y_i^t - Y_j^t}{Y_\kappa^t - Y_\lambda^t}\right) \to 0 \text{ as } t \to \infty$$

It is helpful to rewrite the quantity of interest as follows:

$$\left(\frac{X_i^t - X_j^t}{X_\kappa^t - X_\lambda^t}\right) - \left(\frac{Y_i^t - Y_j^t}{Y_\kappa^t - Y_\lambda^t}\right) = \frac{(X_i^t - X_j^t)(Y_\kappa^t - Y_\lambda^t) - (X_\kappa^t - X_\lambda^t)(Y_i^t - Y_j^t)}{(X_\kappa^t - X_\lambda^t)(Y_\kappa^t - Y_\lambda^t)}$$

Recall the Equation (14) from the proof of Theorem 3 and observe that for each t, we can write the differences that appear in the fraction as follows:

$$X_i^t - X_j^t = \sum_{n_1,\ldots,n_t=2}^{N}\left[\left(\prod_{\tau=1}^{t}\alpha_{n_\tau}^{\tau}\right)\left(V_{i,n_t}^{c,t} - V_{j,n_t}^{c,t}\right)\left(\prod_{\tau=1}^{t-1}\mathbf{V}_{\mathbf{n}_{\tau+1}}^{\mathbf{r},\tau+\mathbf{1}}\mathbf{V}_{\mathbf{n}_\tau}^{\mathbf{c},\tau}\right)\left(V_{n_1}^{r,1}\mathbf{X^0}\right)\right]$$

for a vector of initial opinions $\mathbf{X^0}$ and analogously the $Y_i^t - Y_j^t$ for a vector of initial opinions $\mathbf{Y^0}$.

In analogy to our approach in Theorem 3, let us define the set $\widetilde{\mathcal{S}} = \widetilde{\mathcal{L}}^{\mathbb{N}}$, where $\widetilde{\mathcal{L}} = \{2, \ldots, N\}^2$, with typical element $(\widetilde{\mathbf{L}}^{\mathbf{1}}, \widetilde{\mathbf{L}}^{\mathbf{2}}, \ldots) \in \widetilde{\mathcal{S}}$. As usual, $\widetilde{\mathcal{S}}$ is endowed with the σ-algebra $\widetilde{\mathcal{F}}$ generated by the cylinder sets, that is by all the events of

[33]Taking a fixed reference pair require the assumption that the opinions of individuals κ and λ are never exactly equal. To be precise, one could consider insted a sequence of reference pairs $\{(\kappa_t, \lambda_t)\}_{t=1}^{\infty}$, for all $i, j \in N$ that satisfy this condition.

the form $\{\widetilde{\mathbf{L}}^{\mathbf{1}}\} \times \cdots \times \{\widetilde{\mathbf{L}}^{\mathbf{t}}\} \times \widetilde{\mathcal{L}} \times \widetilde{\mathcal{L}} \times \cdots$.

Define the measure $\widetilde{\mu}$ over the measurable space $(\widetilde{\mathcal{S}}, \widetilde{\mathcal{F}})$ as follows: Fix an arbitrary $t \in \mathbb{N}$ and take the algebra $\widetilde{\mathcal{F}}_t$ which is generated by the events in the collection $\left\{\{\widetilde{\mathbf{L}}^{\mathbf{1}}\} \times \cdots \times \{\widetilde{\mathbf{L}}^{\mathbf{t}}\} \times \widetilde{\mathcal{L}} \times \widetilde{\mathcal{L}} \times \cdots | (\widetilde{\mathbf{L}}^{\mathbf{1}}, \dots, \widetilde{\mathbf{L}}^{\mathbf{t}}) \in \widetilde{\mathcal{L}}^t\right\}$. Then let $\widetilde{\mu}_t$ be a measure that assigns equal weight $1/(N-1)^{2t}$ to every subset on the partition that generates $\widetilde{\mathcal{F}}_t$. Essentially, this is a partition of the set $\widetilde{\mathcal{S}}$ to subsets whose elements share the same elements in the first t rounds. By the Kolmogorov extension theorem there exists a unique measure over $(\widetilde{\mathcal{S}}, \widetilde{\mathcal{F}})$ that agrees with every $\widetilde{\mu}_t$ in events in $\widetilde{\mathcal{F}}_t$, which is the measure we call $\widetilde{\mu}$.

Now, we partition $\widetilde{\mathcal{S}}$ into two subsets $\widetilde{\mathcal{S}}_1$ and $\widetilde{\mathcal{S}}_2$ as follows:

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

and $\widetilde{\mathcal{S}}_2 := \widetilde{\mathcal{S}} \setminus \widetilde{\mathcal{S}}_1$. Like before, both sets $\widetilde{\mathcal{S}}_1$ and $\widetilde{\mathcal{S}}_2$ are measurable and have a strictly positive measure on $\widetilde{\mathcal{S}}$.

For each $s = (\widetilde{\mathbf{L}}^{\mathbf{1}}, \widetilde{\mathbf{L}}^{\mathbf{2}}, \dots) \in \widetilde{\mathcal{S}}$ we define the sequences of complex numbers $\{\widetilde{f}^t_{i,j,\kappa,\lambda}(s)\}_{t=1}^\infty$ as follows:

$$\widetilde{f}^t_{i,j,\kappa,\lambda}(s) = \widetilde{f}_{i,j,\kappa,\lambda}(\widetilde{\mathbf{L}}^{\mathbf{1}}, \widetilde{\mathbf{L}}^{\mathbf{2}}, \dots, \widetilde{\mathbf{L}}^{\mathbf{t}}) = (X_i^t - X_j^t)(Y_\kappa^t - Y_\lambda^t)\left(\prod_{\tau=1}^t \frac{1}{\alpha_2^\tau}\right)^2 \left(\prod_{\tau=1}^{t-1} \frac{1}{\mathbf{V}_{\mathbf{2}}^{\mathbf{r},\tau+\mathbf{1}} \mathbf{V}_{\mathbf{2}}^{\mathbf{c},\tau}}\right)^2 =$$

$$= \left(V^{c,t}_{i,l_1^t} - V^{c,t}_{j,l_1^t}\right)\left(V^{c,t}_{\kappa,l_2^t} - V^{c,t}_{\lambda,l_2^t}\right)\left(\prod_{\tau=1}^t \frac{\alpha^\tau_{l_1^\tau}}{\alpha_2^\tau}\right)\left(\prod_{\tau=1}^t \frac{\alpha^\tau_{l_2^\tau}}{\alpha_2^\tau}\right)\left(\prod_{\tau=1}^{t-1} \frac{\mathbf{V}^{\mathbf{r},\tau+\mathbf{1}}_{\mathbf{l}_1^{\tau+1}} \mathbf{V}^{\mathbf{c},\tau}_{\mathbf{l}_1^\tau}}{\mathbf{V}_{\mathbf{2}}^{\mathbf{r},\tau+\mathbf{1}} \mathbf{V}_{\mathbf{2}}^{\mathbf{c},\tau}}\right)\left(\prod_{\tau=1}^{t-1} \frac{\mathbf{V}^{\mathbf{r},\tau+\mathbf{1}}_{\mathbf{l}_2^{\tau+1}} \mathbf{V}^{\mathbf{c},\tau}_{\mathbf{l}_2^\tau}}{\mathbf{V}_{\mathbf{2}}^{\mathbf{r},\tau+\mathbf{1}} \mathbf{V}_{\mathbf{2}}^{\mathbf{c},\tau}}\right) V^{r,1}_{l_1^1}\mathbf{X}^{\mathbf{0}} V^{r,1}_{l_2^1}\mathbf{Y}^{\mathbf{0}}$$

We are now ready to proceed to the proof. There are several steps of the proof that follow the same reasoning as the respective parts of the proof of Theorem 3. For economy of space, we refer the reader to these parts rather than repeating the same analysis.

Proposition 1. *Consider a sequence of listening matrices $\{\mathbf{T}(t)\}_{t=1}^\infty$ and two vectors $\mathbf{X}(t)$ and $\mathbf{Y}(t)$ that are updated according to (1), starting from some generic initial vectors $\mathbf{X}(0)$ and $\mathbf{Y}(0)$ with positive variance. If $\{\mathbf{T}(t)\}_{t=1}^\infty$ satisfies Conditions (0)–(6) then for some reference pair (κ, λ), for all $i, j \in N$ holds that:*

$$\left(\frac{X_i^t - X_j^t}{X_\kappa^t - X_\lambda^t}\right) - \left(\frac{Y_i^t - Y_j^t}{Y_\kappa^t - Y_\lambda^t}\right) \to 0 \text{ as } t \to \infty$$

Proof. Given the definition of $\widetilde{f}$, we can rewrite the quantity of interest as follows:

$$\left(\frac{X_i^t - X_j^t}{X_\kappa^t - X_\lambda^t}\right) - \left(\frac{Y_i^t - Y_j^t}{Y_\kappa^t - Y_\lambda^t}\right) = \frac{\int_{\widetilde{\mathcal{S}}} \widetilde{f}^t_{i,j,\kappa,\lambda}(s) d\widetilde{\mu} - \int_{\widetilde{\mathcal{S}}} \widetilde{f}^t_{\kappa,\lambda,,i,j}(s) d\widetilde{\mu}}{\int_{\widetilde{\mathcal{S}}} \widetilde{f}^t_{\kappa,\lambda,\kappa,\lambda}(s) d\widetilde{\mu}} = \tag{16}$$

$$= \frac{\int_{\widetilde{\mathcal{S}}_1} \left[\widetilde{f}^t_{i,j,\kappa,\lambda}(s) - \widetilde{f}^t_{\kappa,\lambda,i,j}(s)\right] d\widetilde{\mu} + \int_{\widetilde{\mathcal{S}}_2} \widetilde{f}^t_{i,j,\kappa,\lambda}(s) d\widetilde{\mu} - \int_{\widetilde{\mathcal{S}}_2} \widetilde{f}^t_{\kappa,\lambda,i,j}(s) d\widetilde{\mu}}{\int_{\widetilde{\mathcal{S}}_1} \widetilde{f}^t_{\kappa,\lambda,\kappa,\lambda}(s) d\widetilde{\mu} + \int_{\widetilde{\mathcal{S}}_2} \widetilde{f}^t_{\kappa,\lambda,\kappa,\lambda}(s) d\widetilde{\mu}} \tag{17}$$

Note that, both the numerator and the denominator of the fraction are divided by the same quantity, hence the fraction remains unaffected. Moreover, $\widetilde{f}$ is integrable over the partition $\{\widetilde{\mathcal{S}}_1, \widetilde{\mathcal{S}}_2\}$ of $\widetilde{\mathcal{S}}$, therefore all integrals are well–defined. Moreover, using the same reasoning as in Theorem 3, for generic initial vectors $\mathbf{X}(\mathbf{0}), \mathbf{Y}(\mathbf{0})$ the inegral $\int_{\widetilde{\mathcal{S}}_1} \widetilde{f}^t_{k,m,k,m}(s) d\widetilde{\mu}$ does not converge to zero.

Once again, following identical reasoning to the Theorem the integrals $\int_{\widetilde{\mathcal{S}}_2} \widetilde{f}^t_{i,j,\kappa,\lambda}(s) d\widetilde{\mu}$, $\int_{\widetilde{\mathcal{S}}_2} \widetilde{f}^t_{\kappa,\lambda,i,j}(s) d\widetilde{\mu}$ and $\int_{\widetilde{\mathcal{S}}_2} \widetilde{f}^t_{\kappa,\lambda,\kappa,\lambda}(s) d\widetilde{\mu}$ converge to zero as t grows. Namely, for all $i, j, \kappa, \lambda \in N$, $\widetilde{f}^t_{i,j,\kappa,\lambda}$, $\widetilde{f}^t_{\kappa,\lambda,i,j}$ and $\widetilde{f}^t_{\kappa,\lambda,\kappa,\lambda}$ converge pointwise to zero for all $s \in \widetilde{\mathcal{S}}_2$ and given that they are unformly bounded –as a result of Condition (6)– then by dominated convergence theorem the respective integrals converge to zero as well.

It remains to show that $\int_{\widetilde{\mathcal{S}}_1} \left[\widetilde{f}^t_{i,j,\kappa,\lambda}(s) - \widetilde{f}^t_{\kappa,\lambda,i,j}(s)\right] d\widetilde{\mu}$ converges to zero as t grows. It is important to notice that we take each part of the integral separately, then neither of these two integrals converges to zero (as they most probably do not converge at all). Consider any $s = (\mathbf{L^1}, \mathbf{L^2}, \dots) \in \widetilde{\mathcal{S}}_1$ and let $\underline{s} = (\underline{\mathbf{L^1}}, \underline{\mathbf{L^2}}, \dots) \in \mathcal{S}$ be defined as follows: if $\mathbf{L^t} = (l_1^t, l_2^t)$ then $\underline{\mathbf{L^t}} = (l_2^t, l_1^t)$. By definition of $\widetilde{\mathcal{S}}_1$, there exists some $\hat{t}$ such that for all $\tau > \hat{t}$ it holds that $\mathbf{L}^\tau = (l_1^\tau, l_2^\tau) = (2,2)$. This means that $\underline{s} \in \mathcal{S}_1$ as well and also that for all $\tau > \hat{t}$ it holds that $\underline{\mathbf{L}^\tau} = (l_2^\tau, l_1^\tau) = (2,2)$ for the same $\hat{t}$ as for s. This last observation is important because it guarantees that for all $\tau > \hat{t}$ holds the equality $\widetilde{f}^t_{i,j,\kappa,\lambda}(s) = \widetilde{f}^t_{\kappa,\lambda,i,j}(\underline{s})$.

Given that, for each $s \in \widetilde{\mathcal{S}}_1$ define a new sequence $\{h^t_{\kappa,\lambda,i,j}(s)\}_{t=1}^\infty$ such that $h^t_{\kappa,\lambda,i,j}(s) = \widetilde{f}^t_{\kappa,\lambda,i,j}(\underline{s})$. This new sequence is well–defined for all $s \in \widetilde{\mathcal{S}}_1$ and by construction it satisfies that $\int_{\widetilde{\mathcal{S}}_1} \widetilde{h}^t_{\kappa,\lambda,i,j} d\widetilde{\mu} = \int_{\widetilde{\mathcal{S}}_1} \widetilde{f}^t_{\kappa,\lambda,i,j} d\widetilde{\mu}$ because $\widetilde{\mu}$ is uniform. Therefore, it also holds that $\int_{\widetilde{\mathcal{S}}_1} \left[\widetilde{f}^t_{i,j,\kappa,\lambda} - \widetilde{f}^t_{\kappa,\lambda,i,j}\right] d\widetilde{\mu} = \int_{\widetilde{\mathcal{S}}_1} \left[\widetilde{f}^t_{i,j,\kappa,\lambda} - h^t_{\kappa,\lambda,i,j}\right] d\widetilde{\mu}$. Now, consider a new sequence $\{\widetilde{f}^t_{i,j,\kappa,\lambda}(s) - h^t_{\kappa,\lambda,i,j}(s)\}_{t=1}^\infty$ for each $s \in \widetilde{\mathcal{S}}_1$. By construction, for all $\tau > \hat{t}$ it holds that $\widetilde{f}^\tau_{i,j,\kappa,\lambda}(s) = \widetilde{f}^\tau_{\kappa,\lambda,i,j}(\underline{s}) = h^\tau_{\kappa,\lambda,i,j}(s)$, hence $\widetilde{f}^\tau_{i,j,\kappa,\lambda}(s) - h^\tau_{\kappa,\lambda,i,j}(s) = 0$. Therefore, for all $s \in \widetilde{\mathcal{S}}_1$ the sequence $\{\widetilde{f}^t_{i,j,\kappa,\lambda}(s) - h^t_{\kappa,\lambda,i,j}(s)\}_{t=1}^\infty$ converges pointwise to 0 and given that by Condition (6) the values of $\widetilde{f}^t_{i,j,\kappa,\lambda}(s) - h^t_{\kappa,\lambda,i,j}(s)$ are uniformly bounded for all t and s, we can use the dominated convergence theorem to conclude that $\lim_{t\to\infty} \int_{\widetilde{S}_1} \left[\widetilde{f}^t_{i,j,\kappa,\lambda} - h^t_{\kappa,\lambda,i,j}\right] d\widetilde{\mu} = 0$, which in turn implies that $\lim_{t\to\infty} \int_{\widetilde{S}_1} \left[\widetilde{f}^t_{i,j,\kappa,\lambda} - \widetilde{f}^t_{\kappa,\lambda,i,j}\right] d\widetilde{\mu} = 0$.

□

B Updating, convergence and perturbations

Notice that two main assumptions in our theoretical model are that a) agents update their opinions (in a way that resembles averaging over one's neighbors' opinions), and b) they do this in the same way for both dimensions. If the former holds then opinions need to come closer over time, something that would be reflected on a diminishing normalized coefficient of variation (NCV).[34] If the latter holds, then the decrease of the NCV should be the same across the two dimensions.

The three upper panels of Figure 7 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 mis-typing one's guess. Others may be deliberate, although there does not seem to be some systematic pattern of behavior to explain them.[35] We do observe that such perturbations are less common in *Fixed 2*. This can be seen in Figure 7 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",

[34]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 $NCV(t) = \frac{CV(t)}{CV(1)}$.

[35]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).

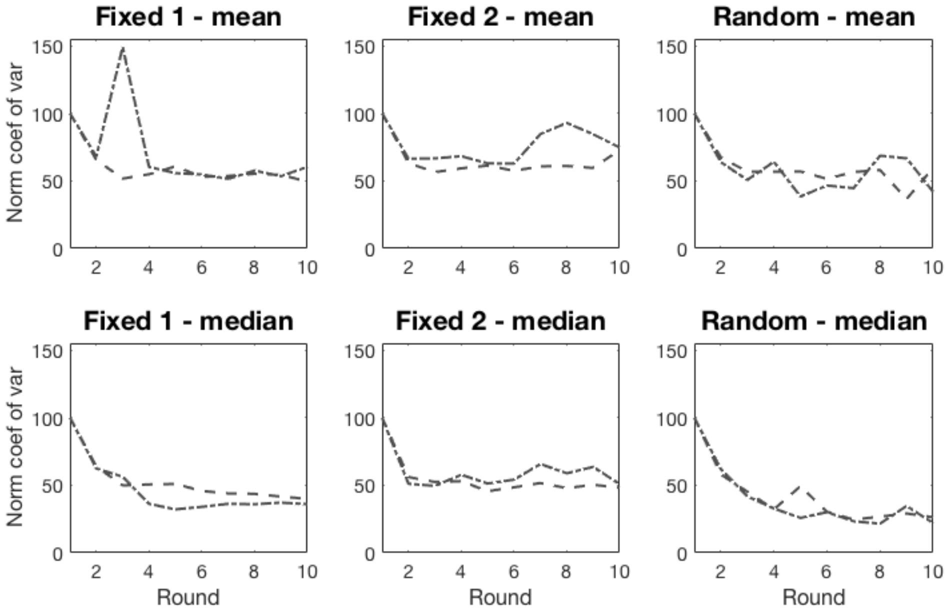


Figure 7: The graphs show the mean and median NCV values per round in each treatment. *Dashed lines* correspond to the NCV of guesses for *tank 1* and *dot-dashed lines* correspond to the same for *tank 2*.

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

The lower panels of Figure 7 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.[36] 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.[37]

[36]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 × 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.

[37]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*.

C Clustering analysis

Combining equations (1) and (4) gives us an individual's opinion on issue k at time $t+1$ as:

$$s_i^k(t+1) = (1-\lambda_i(t))s_i^k(t) + \lambda_i(t)\frac{1}{|D_i(t)|}\sum_{j\in D_i}[s_j^k(t)]$$

.

Each subject played 3 phases x 10 guesses for tanks 1 and 2. For each individual and each of these guesses we solve this equation for $\lambda_i(t)$ and obtain a $\hat{\lambda}_i^k(t_p)$, where $p \in \{1,2,3\}$ refers to the phase. As subjects guesses are not restricted, it happens that some of these values lie outside of the theoretical bounds: $0 \leq \lambda_i(t) \leq 1$. While this does happen, most values obtained are within or very close to the theoretical bounds. To avoid extreme values biasing the excercise, we set a censoring parameter $c = .3$ and apply the following set of rules: if $\hat{\lambda}_i^k(t_p) < -c$ or $\hat{\lambda}_i^k(t_p) > 1+c$ for one of the tanks but not the other, we only use the moderate value. If both values are extreme we consider them as missing. If none of the values are extreme, we take the average of the two: $\lambda_i(t_p) = (\hat{\lambda}_i^1(t_p) + \hat{\lambda}_i^2(t_p))/2$. We then average across phases (ignoring missing values) to obtain $\lambda_i(t) = \frac{1}{3}\sum_p \lambda_i(t_p)$. Any remaining missing values are imputed using the nearest-neighbor method.

We now have for each individual a vector λ_i with nine elements, one for each period from 2 to 10. We use the k-means algorithm to classify these in to separate clusters using the 'city-block' distance measure. This metric is preferred compared to the 'squared euclidean' metric, as the data is already in the same scale, but we do expect to have different variability across periods. Using a non-linear distance metric would then put more weight on specific periods. We repeat the clustering 1000 times with different random initializations and select the centroids that give the lowest within-cluster sums of data to centroid distances. To determine the number of clusters, we perform this exercise for different numbers of clusters and record the performance in terms of sum of distances. A higher number of clusters will always improve performance, but we observe that these gains are not substantial for more than 4 clusters (see figure 8).

To obtain the smooth decreasing functions showed in Figure 4.3, we fit a two-term power series model to each one of the four centroids.

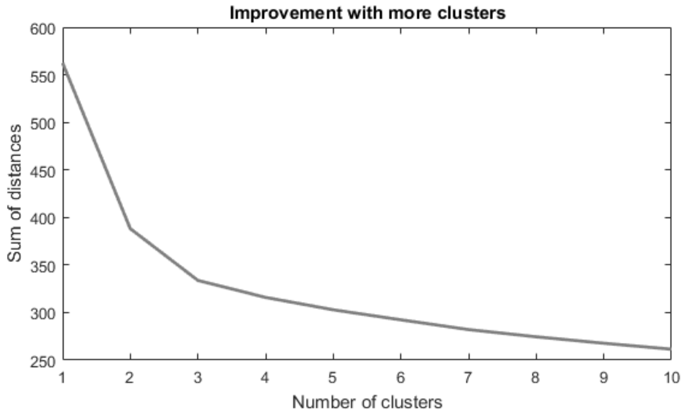


Figure 8: The graphs show the performance of the k-means clustering for different numbers of clusters. Lower values represent better performance.

D 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.[38]

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.[39]

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:

[38] **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.

[39] **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.**[40]
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.**

[40] **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.[41]

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.[42]

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.[43]

[41]**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.**

[42]**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.

[43]**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.[44]**
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.**

[44]**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.**