Disagreement in a Naïve Model of Attitude Formation: Comparative Statics Results

Received: September 25th, 2019 – Accepted: February 27th, 2020

Doi: https://doi.org/10.12804/revistas.urosario.edu.co/economia/a.9112

Isabel Melguizo*

Abstract

This paper further studies the persistence of disagreement in a model similar to Melguizo (2019), by relaxing two essential assumptions. First, attitudes are random variables, and second, individuals may be homophilous at different extents. Regarding the first extension, the finding is that disagreement persists with the highest probability across the attribute that exhibits the highest mean of the distribution of initial differences in average attitudes. Regarding the second one, the finding is that the magnitude of disagreement and the speed of convergence to it, increase with respect to the case in which individuals are homophilous to the same extent.

Keywords: Disagreement, homophily, average-based updating.
JEL classification: D83, D85, Z13.

* Centro de Investigación y Docencia Económicas (CIDE), Ciudad de México. Carretera México-Toluca 3655, Lomas de Santa Fe, 01210. Email: isabel.melguizo@cide.edu, isabelmelguizolopez@gmail.com

To quote this article: Melguizo, I. (2020). Disagreement in a Naïve Model of Attitude Formation: Comparative Statics Results. *Revista de Economía del Rosario, 23*(2), 1-24. https://doi.org/10.12804/revistas.urosario.edu.co/economia/a.9112

ISSN 0123-5362 - ISSNE 2145-454x

El desacuerdo en modelos de formación de opiniones: resultados de estática comparativa

Resumen

Este artículo estudia la persistencia del desacuerdo en un modelo similar a Melguizo (2019), relajando dos supuestos importantes. Primero, las opiniones iniciales de los individuos son variables aleatorias, y segundo, los individuos pueden tener distintos grados de homofilia. Con respecto a la primera extensión, se encuentra que el desacuerdo persiste con más probabilidad en el atributo que exhibe la mayor media de la distribución de las diferencias en actitudes medias iniciales. Respecto a la segunda, la magnitud del desacuerdo y la velocidad de convergencia en ´el incrementan con respecto al modelo original.

Palabras clave: desacuerdo, homofilia, actualización de actitudes basada en medias.
Clasificación JEL: D83, D85, Z13.

O desacordo em modelos de formação de opiniões: resultados de estática comparativa

Resumo

Este artigo estuda a persistência do desacordo em um modelo similar a Melguizo (2019), flexibilizando dois supostos importantes. Primeiro, as opiniões iniciais dos indivíduos são variáveis aleatórias, e segundo, os indivíduos podem ter distintos graus de homofília. Em relação à primeira extensão, se encontra que o desacordo persiste com mais probabilidade no atributo que exibe a maior média da distribuição das diferenças em atitudes médias iniciais. Com relação à segunda, a magnitude do desacordo e a velocidade de convergência nele incrementam respeito ao modelo original.

Palavras-chave: desacordo, homofilia, atualização de atitudes baseada em médias.
Classificação JEL: D83, D85, Z13.

Introduction

Melguizo (2019) proposes a model of attitude formation where individuals update their attitudes by averaging those of their friends. This model captures the persistence of disagreement, that is, the persistence of a situation in which individuals hold different attitudes about an issue. In that approach, a collection of dichotomous attributes defines the individuals' types. Also, individuals are more prone to interact with others similar to them in those attributes; that is, individuals exhibit homophily.[1] Interactions co-evolve with attitudes, and this feature is central to the persistence of disagreement. The main finding is that disagreement persists if, and only if, individuals develop sufficiently intense relations over time with others, similar in one specific attribute. Thus, society polarizes according to this dimension.

In that framework, individuals update their attitudes as in DeGroot (1974), further: (i) individuals' attitudes are known with certainty, and (ii) homophily relations were symmetric; that is, all individuals were homophilous with the same intensity. This paper explores how previous findings react to natural modifications in these two assumptions. The context is one in which individual types come from the combination of two dichotomous attributes. Some insights for the case of $n \geq 2$ attributes are provided. The following is a preview of the results.

(i) *Random attitudes.* It might be that it is better to describe attitudes as random variables. In contexts in which the aim is to learn the true state of the world, randomness might be interpreted as lack of information (noise) regarding the issue at hand, as in Golub and Jackson (2010); as the degree of attitudes' precision, as in DeMarzo, et al. (2003), or as experts having probability distributions about the true state of the world, as in DeGroot (1974). In situations in which individuals deal with ideological issues, randomness might be interpreted as flexibility or lack of stubbornness. In line with these observations, this paper goes further in proposing that initial attitudes draw from symmetric continuous distributions. This is in line with DeMarzo, et al. (2003), which allow for randomness only in the first period. The persistence of disagreement is robust to this modification. In particular, disagreement may now persist across one of the two attributes. It is, however, more likely to persist across the one for which

1 For a survey on homophily as a pervasive phenomenon in real life interactions, see McPherson et al. (2001).

the mean of the distribution of the initial differences in attitudes is the highest.

(ii) *Non-symmetric homophily.* It might be also natural to think that the (intensity of) relations that individuals establish with others depend(s) on the specific nature of the shared attributes. In fact, McPherson, et al. (2001) document how gender homophily is lower when people are younger than older. Gender homophily is also lower for high educated than for low educated people and for Anglos than for African Americans. This might imply, in particular, that pairs of individuals no longer devote the same amount of attention to each other. As an example, suppose there are four types of individuals, that is, an individual can be either young or old and either a female or a male. Consider that young people establish less intense relations with same-gender others than seniors. This behavior could emerge in the model when individuals have different sensitivity to differences in attitudes between groups. Specifically, when confronted with information about differences in attitudes between males and females, seniors exacerbate the differences in attitudes by gender with respect to young people. Notice then that the intensity of gender relations depends on another attribute defining the individuals involved; that is, on youth. The finding is that, when disagreement persists across the attribute for which initial differences in attitudes are the highest, its magnitude is higher than in the case in which homophilous relations are symmetric. The process also converges faster to the eventual attitudes.

The remaining of the paper is as follows. First, it presents the model in Melguizo (2019) to clarify the baseline setup. Then, it discusses random attitudes and then the non-symmetric homophily. Next, it includes some comments on the co-existence of random attitudes and non-symmetric homophily. After that, it presents the conclusions, and finally, it shows the proofs.

1. A Model on Homophily and Disagreement

Before the extensions, let us first introduce the baseline model in Melguizo (2019).

Let $I = \{1, 2, \dots, n\}$ be a finite set of attributes. The type A of an individual is defined by the attributes possessed by this individual, that is, $A \subseteq I$. Two types, A and B, are i-similar whenever attribute i is either present or absent in these two types. Otherwise, they are i-dissimilar. Let A^c be the complementary set of A. Then $I(AB) \equiv (A \cap B) \cup (A^c \cap B^c)$ is the set of shared attributes between A and B.

The (column) vector of attitudes at time $t \in Z_+$ is denoted $a_t \in [-1, 1]^{2^n}$. Let a_t^A be a typical component of a_t, denoting the attitude of type A. Notice that there are 2^{n-1} types possessing (resp. lacking) any attribute. Thus, the average attitude across types possessing (resp. lacking) attribute i is $\bar{a}_t[i] \equiv \left(2^{n-1}\right)^{-1} \sum_{A:i\in A} a_t^A$ $\left(resp.\ \bar{a}_t[-i] \equiv \left(2^{n-1}\right)^{-1} \sum_{A:i\notin A} a_t^A\right)$. The difference between average attitudes across attribute i is denoted $\Delta_t[i] \equiv \bar{a}_t[i] - \bar{a}_t[-i]$.

The following example illustrates the notation for the two-attribute case.

Example 1. Let $I = \{1, 2\}$. Types are {1, 2} {1}, {2} and {∅}. Observe, as an illustration, that types {1, 2} and {1} are 1-similar and 2-disimilar. Let $a'_t = \left[a_t^{\{1,2\}} a_t^{\{1\}} a_t^{\{2\}} a_t^{\{\emptyset\}}\right]$ such that $a'_0 = [0.8\ 0.2 - 0.05 - 0.95]$. Notice that $\bar{a}_0[1] = 0.5$, $\bar{a}_0[-1] = -0.5$, and $\Delta_0[1] = 0.5 - (-0.5) = 1$. Analogously, $\bar{a}_0[2] = 0.375$, $\bar{a}_0[-2] = -0.375$, and $\Delta_0[2] = 0.375 - (-0.375) = 0.75$.

Attitudes evolve according to an average-based process similar to DeGroot (1974). That is, current attitudes are weighted averages of previous ones. Let W_t be the 2^n x 2^n weighting matrix describing the updating of attitudes from t to $t + 1$. Thus:

$$a_{t+1} = W_t a_t. \tag{1}$$

The interpretation of every row in W_t is that every type A has one unit of attention to devote to others (and to itself). Then every entry of W_t is the weight, i.e., the share of attention, that type A assigns to type B at time t. Let $w_t^{A,B}$ denote this weight. Individuals are homophilous, a behavior that can be captured as follows; every attribute i has a non-negative value α_t^i. The weight that type A assigns to type B is the sum of the values of the shared attributes between A and B, that is, $w_t^{A,B} \equiv \sum_{i\in I(AB)} \alpha_t^i$. For normalization purposes let $\sum_i \alpha_t^i = \left(2^{n-1}\right)^{-1}$. That is the right normalization because any type A is i-similar to exactly 2^{n-1} types. Then, $\sum_B w_t^{A,B} \equiv 2^{n-1} \sum_i \alpha_t^i \equiv 1$. Let $\lambda_t^i \equiv 2^{n-1} \alpha_t^i$, so that $\lambda_t^i \in [0,1]$ and $\sum_i \lambda_t^i \equiv 1$. Notice that sharing at least one attribute is necessary for any pair of types A and B to hold a relation.

The following example describes the attention structure.

Example 2. In the two-attribute case, the interaction matrix at time t is:

$$\begin{array}{cc} & \begin{array}{cccc} \{1,2\} & \{1\} & \{2\} & \{\emptyset\} \end{array} \\ W_t = 2^{-1} & \begin{bmatrix} \lambda_t^1+\lambda_t^2 & \lambda_t^1 & \lambda_t^2 & 0 \\ \lambda_t^1 & \lambda_t^1+\lambda_t^2 & 0 & \lambda_t^2 \\ \lambda_t^2 & 0 & \lambda_t^1+\lambda_t^2 & \lambda_t^1 \\ 0 & \lambda_t^2 & \lambda_t^1 & \lambda_t^1+\lambda_t^2 \end{bmatrix} \begin{array}{c} \{1,2\} \\ \{1\} \\ \{2\} \\ \{\emptyset\} \end{array}. \end{array}$$

To clarify this structure, notice that relations are symmetric and focus on type {2}. It is 1-similar and 2-similar to types {Ø} and {1, 2}, respectively. Thus, it pays attention to them, on the basis of attributes 1 and 2, respectively. Since type {2} and {1} does not share any attribute, they pay no (direct) attention to each other.

Let λ_t^i depend on the difference in average attitudes between the individuals possessing and lacking attribute i, that is, on $\Delta_t[i]$, and on (possibly) all the differences associated with the remaining attributes, that is, on $\Delta_t[j]$ for every attribute $j \neq i$. Let, w.l.o.g,

$$\Delta_0[1] \geq \Delta_0[2] \geq \cdots \Delta_0[n] \geq 0. \tag{2}$$

Let further λ_t^i satisfy three properties:

Within differences monotonicity (WDM): for every attribute i, $\Delta_t[i] = 0$ implies that $\lambda_t^i = 0$, and $\Delta_t[i] > 0$ implies that $\lambda_t^i > 0$.

Across differences monotonicity (ADM): $\Delta_t[1] \geq \Delta_t[2] \geq \cdots \geq \Delta_t[n] \geq 0$ implies that $\lambda_t^1 \geq \lambda_t^2 \geq \cdots \geq \lambda_t^n \geq 0$. When $\Delta_t[i] = 0$ for every attribute i at time t, set $\lambda_t^i = 1/n$.

Well defined limit: for every attribute i, $\lim_{t\to\infty}\lambda_t^i \in [0,1]$ and $\sum_i \lim_{t\to\infty}\lambda_t^i = 1$. A functional form for λ_t^i that satisfies the above properties is:

$$\lambda_t^i = \frac{\Delta_t[i]^\delta}{\sum_j \Delta_t[j]^\delta}, \tag{3}$$

with $\delta \in [0,\infty)$.

The co-evolution of interactions and attitudes is a key feature for the persistence of disagreement. The main finding is that disagreement persists if and only if individuals develop sufficiently intense relations over time with others, similar in one specific attribute. Thus, society polarizes according to this attribute, and two groups of thinking eventually emerge. Within each of these two groups, there is consensus, a situation in which everyone holds the same attitude. The attribute according to which society polarizes is the one for which initial differences in average attitudes are the highest, that is, attribute 1. Specifically, the eventual attitude of an individual that possesses attribute 1 is $a_{\infty}^{A} = \bar{a}_0 - 2^{-1}(1-r)\Delta_0[1]$ where $r = \Delta_0[2] \,/\, \Delta_0[1]$.[2]

In what follows, we analyze the robustness of the model to the proposed extensions, one in a row.

2. Random Attitudes

This section follows the environment described in the previous section, except for the fact that initial attitudes are random variables. In particular, for every individual A, $\tilde{a}_0^A$ follows a continuous symmetric distribution with mean a_0^A and variance σ_A^2. Let initial attitudes be independent of each other but not necessarily identically distributed. The vector of random attitudes at t is denoted by $\tilde{a}_t$.

The analysis is for the case of two attributes, denoted by 1 and 2. We provide insights for the case in which an arbitrary finite set of n attributes is considered. Individuals update attitudes over time, using averages of previous period attitudes. The updating process is, thus, summarized in an analogous way to (1). The only difference is that now, the attitudes, as well as the weights in the updating matrix, are random variables. That is:

$$\tilde{a}_t = \widetilde{W}_t \tilde{a}_{t-1}, \tag{4}$$

where

$$\widetilde{W}_t = \frac{1}{2} \begin{array}{c} \\ \left[\begin{array}{cccc} \{1,2\} & \{1\} & \{2\} & \{\emptyset\} \\ \tilde{\lambda}_t^1 + \tilde{\lambda}_t^2 & \tilde{\lambda}_t^1 & \tilde{\lambda}_t^2 & 0 \\ \tilde{\lambda}_t^1 & \tilde{\lambda}_t^1 + \tilde{\lambda}_t^2 & 0 & \tilde{\lambda}_t^2 \\ \tilde{\lambda}_t^2 & 0 & \tilde{\lambda}_t^1 + \tilde{\lambda}_t^2 & \tilde{\lambda}_t^1 \\ 0 & \tilde{\lambda}_t^2 & \tilde{\lambda}_t^1 & \tilde{\lambda}_t^1 + \tilde{\lambda}_t^2 \end{array} \right] \end{array} \begin{array}{c} \{1,2\} \\ \{1\} \\ \{2\} \\ \{\emptyset\} \end{array}.$$

2 See the main theorem in Melguizo (2019).

Also, $\tilde{\lambda}_t^i \in [0,1]$ is the weight, i.e., the intensity of attention, that an individual assigns to others that share attribute $i = 1, 2$ with her at time t. In other words, $\tilde{\lambda}_t^i$ accounts for the intensity of homophilous relations. Recall that conditional on paying attention, the intensity of this attention evolves over time, i.e., $\tilde{\lambda}_t^1$ and $\tilde{\lambda}_t^2$ are time-dependent. In particular, the average initial differences across attributes 1 and 2 determine the evolution of attention.

We, thus, describe these average initial differences. Let $\tilde{\Delta}_0[1] = 2^{-1}\left(\tilde{a}_0^{\{1,2\}} - \tilde{a}_0^{\{\emptyset\}} + \left(\tilde{a}_0^{\{1\}} - \tilde{a}_0^{\{2\}}\right)\right)$ and $\tilde{\Delta}_0[2] = 2^{-1}\left(\tilde{a}_0^{\{1,2\}} - \tilde{a}_0^{\{\emptyset\}} + \left(\tilde{a}_0^{\{2\}} - \tilde{a}_0^{\{1\}}\right)\right)$ be the distributions of the average initial differences associated to attributes 1 and 2, respectively. They have means $\Delta_0[1] = 2^{-1}\left(a_0^{\{1,2\}} - a_0^{\{\emptyset\}} + \left(a_0^{\{1\}} - a_0^{\{2\}}\right)\right)$ and $\Delta_0[2] = 2^{-1}\left(a_0^{\{1,2\}} - a_0^{\{\emptyset\}} + \left(a_0^{\{2\}} - a_0^{\{1\}}\right)\right)$, respectively, and the variance, $\sum_A \sigma_A^2 / 4$. Let, w.l.o.g, $\Delta_0[1] \geq \Delta_0[2] \geq 0$, an assumption which is the analogous counterpart of condition (2), in the deterministic case. The link between homophily and differences in attitudes is given by:

$$\tilde{\lambda}_t^1 = \frac{\left|\tilde{\Delta}_t[1]\right|}{\left|\tilde{\Delta}_t[1]\right| + \left|\tilde{\Delta}_t[2]\right|} \text{ and } \tilde{\lambda}_t^2 = \frac{\left|\tilde{\Delta}_t[2]\right|}{\left|\tilde{\Delta}_t[1]\right| + \left|\tilde{\Delta}_t[2]\right|}.$$ [3]

analogously to (3), with $\delta = 1$.

Regarding the persistence of disagreement, the main finding is as follows:

Proposition 1. *In general, disagreement persists across either attribute* 1 *or* 2 *with positive probability. Disagreement across attribute* 1 *is at least as likely as disagreement across attribute* 2. *Specifically:*

1. *Both events are equally likely if, and only if the differences in initial attitudes across both attributes have the same mean (that is, if, and only if,* $\Delta_0[1] = \Delta_0[2]$*).*
2. *Disagreement across attribute 1 is the most likely event if, and only if, the mean of its difference in initial attitudes is the highest (that is, if, and only if,* $\Delta_0[1] > \Delta_0[2]$*).*
3. *The difference between the probability of disagreement persisting across attributes 1 and 2 is non-negative. Its expression is:*

3 Notice that the realization of $\tilde{\Delta}_t[1]$ and $\tilde{\Delta}_t[2]$ can be positive or negative. The only aspect that matters is the magnitude of dierences, that is why the absolute value is considered.

$$\left(2P\left(\tilde{a}_0^{\{1,2\}} - \tilde{a}_0^{\{\emptyset\}} \geq 0\right) - 1\right)\left(P\left(\tilde{a}_0^{\{1\}} - \tilde{a}_0^{\{2\}} \geq 0\right) - P\left(\tilde{a}_0^{\{1\}} - \tilde{a}_0^{\{2\}} < 0\right)\right). \tag{5}$$

As in the deterministic case, differences in initial attitudes play a crucial role in disagreement. In contrast to the deterministic case, in which disagreement takes place only on attribute 1, in general, now disagreement persists across either attribute. Disagreement will persist across attribute 1 with probability equal to one when the minimum among all possible realizations of $\left|\tilde{\Delta}_0[1]\right|$ is higher than the maximum among all possible realizations of $\left|\tilde{\Delta}_0[2]\right|$. As the (symmetric) distributions of $\tilde{a}_0^{\{1,2\}} - \tilde{a}_0^{\{\emptyset\}}$ and $\tilde{a}_0^{\{1\}} - \tilde{a}_0^{\{2\}}$ have non-negative means, expression (5) is non-negative.[4]

It is worth noticing that since the focus is on initial randomness, once attitudes realize, eventual attitudes are as in the main Theorem in Melguizo (2019) main Theorem. Let $\tilde{a}_\infty^A$ be the eventual attitude of a type A. Also, let $i \in A$ (respectively $i \notin A$) express the fact that type A possesses (respectively lacks) attribute i. The observation is as follows:

Observation. *Let disagreement persist on attribute* $i = \{1, 2\}$. *Then,*

$$\tilde{a}_\infty^A = \bar{a}_0 + 2^{-1}\left(1 - \frac{\left|\tilde{\Delta}_0[j]\right|}{\left|\tilde{\Delta}_0[i]\right|}\right)\tilde{\Delta}_0[i] \;\; \textit{if } i \in A$$

and

$$\tilde{a}_\infty^A = \bar{a}_0 + 2^{-1}\left(1 - \frac{\left|\tilde{\Delta}_0[j]\right|}{\left|\tilde{\Delta}_0[i]\right|}\right)\tilde{\Delta}_0[i] \;\; \textit{if } i \notin A$$

Thus, disagreement manifests as two groups, holding different limit attitudes.[5] Examples 3 to 5 illustrate these findings and related aspects:

---

4 See the proof of Proposition 1.

5 For disagreement on attribute 1, the difference in average limit attitudes is $\tilde{\Delta}_\infty[1] = \left(2^{n-1}\right)^{-1}\left(\sum_{A:1\in A}\tilde{a}_\infty^A - \sum_{A:1\notin A}\tilde{a}_\infty^A\right) = \left(2^{n-1}\right)^{-1}2^{n-1}\left(\tilde{a}_\infty^A : 1 \in A - \tilde{a}_\infty^A : 1 \notin A\right) = \left|\tilde{\Delta}_0[1]\right| - \left|\tilde{\Delta}_0[2]\right|$ (respectively $\left|\tilde{\Delta}_0[2]\right| - \left|\tilde{\Delta}_0[1]\right|$) when $\tilde{\Delta}_0[1] \geq 0$ (respectively $\tilde{\Delta}_0[1] < 0$). Since disagreement across attribute 1 persists when $\left|\tilde{\Delta}_0[1]\right| > \left|\tilde{\Delta}_0[2]\right|$, $\tilde{\Delta}_\infty[1]$ has either positive or negative support. The distribution of the difference in average limit attitudes for attribute 2 degenerates at zero.

Example 3. First, let $\tilde{a}_0^{\{1,2\}} \sim U[0,1], \tilde{a}_0^{\{1\}} \sim U[-1,1], \tilde{a}_0^{\{2\}} \sim U[-1,1]$ and $\tilde{a}_0^{\{\emptyset\}} \sim U[-1,1]$. Thus, $\tilde{\Delta}_0[1]$ and $\tilde{\Delta}_0[2]$ have means $\Delta_0[1] = \Delta_0[2] =0.25$. From Proposition 1, disagreement across either attribute is equally likely. Recall that the probability that disagreement persists across attribute 1 minus the probability that it does across attribute 2 depends on $\tilde{a}_0^{\{1,2\}} - \tilde{a}_0^{\{\emptyset\}}$ and $\tilde{a}_0^{\{1\}} - \tilde{a}_0^{\{2\}}$. As $\tilde{a}_0^{\{1\}} - \tilde{a}_0^{\{2\}}$ follows a (symmetric) triangular distribution with zero mean, this difference is zero.

Second, let $\tilde{a}_0^{\{1,2\}} \sim U[0,1]$, $\tilde{a}_0^{\{1\}} \sim U[0,1]$, $\tilde{a}_0^{\{2\}} \sim U[-1,1]$ and $\tilde{a}_0^{\{\emptyset\}} \sim U[-1,1]$. Thus, $\tilde{\Delta}_0[1]$ and $\tilde{\Delta}_0[2]$ have means $\Delta_0[1] = 0.5$ and $\Delta_0[2] = 0$, respectively. From Proposition 1, disagreement across attribute 1 is the most likely event. As above the focus is on the distributions of $\tilde{a}_0^{\{1,2\}} - \tilde{a}_0^{\{\emptyset\}}$ and $\tilde{a}_0^{\{1\}} - \tilde{a}_0^{\{2\}}$. Let, $y \equiv \tilde{a}_0^{\{1,2\}} - \tilde{a}_0^{\{\emptyset\}}$. It follows a triangular distribution with density:

$$f(y)=\begin{cases} \dfrac{1+y}{2} & if \quad -1<y<0 \\ 0.5 & if \quad 0 \le y \le 1 \\ 1-\dfrac{y}{2} & if \quad 1<x<2 \end{cases}.$$

Thus, $P(y \ge 0) = 1 - \int_{y=-1}^{y=0} 2^{-1}(1+y) = 0.75$. Also, let $z \equiv \tilde{a}_0^{\{1\}} - \tilde{a}_0^{\{2\}} \ge 0$. Notice that z follows the same distribution as y; thus, $P(z \ge 0) = 0.75$. In this case, expression (5) equals 0.25. Disagreement persists across attribute 1 (respectively 2) with probability 0.625 (respectively 0.375).

Example 4. Let initial attitudes be normally distributed with means $\Delta_0[1] \ge \Delta_0[2] > 0$ and unitary variances. Figure 1 depicts the probability that disagreement persists across attribute 1, as a function of the means of the distribution of the initial differences in attitudes. In particular, $\Delta_0[2]$ is kept constant, while $\Delta_0[1]$ increases. The x-axis represents the ratio $x \equiv \Delta_0[1] / \Delta_0[2]$. The y-axis represents the probability of disagreement. The graph suggests a positive relation between the mean of the initial differences in attitudes associated

---

That is so since out of the 2^{n-1} types possessing attribute 1, there are 2^{n-2} types possessing attribute 2 and 2^{n-2} types lacking attribute 2. The same happens within the 2^{n-1} types lacking attribute 1, hence, $\tilde{\Delta}_\infty[2] = \left(2^{n-1}\right)^{-1}\left(\sum_{A:2\in A}\tilde{a}_\infty^A - \sum_{A:2\notin A}\tilde{a}_\infty^A\right) = 2^{-1}\left(2^{n-2}\right)^{-1}2^{n-2}\left(\tilde{a}_\infty^A : 1\in A, 2\in A + \tilde{a}_\infty^A : 1\notin A, 2\in A\right) - \tilde{a}_\infty^A : 1\in A, 2\notin A + \tilde{a}_\infty^A : 1\notin A, 2\notin A) = 2^{-1}2\left(\tilde{a}_0 - \tilde{a}_0\right) = 0$. The analysis is parallel when disagreement is on attribute 2.

with attribute 1, relative to the mean for attribute 2, and the probability of disagreement across attribute 1.

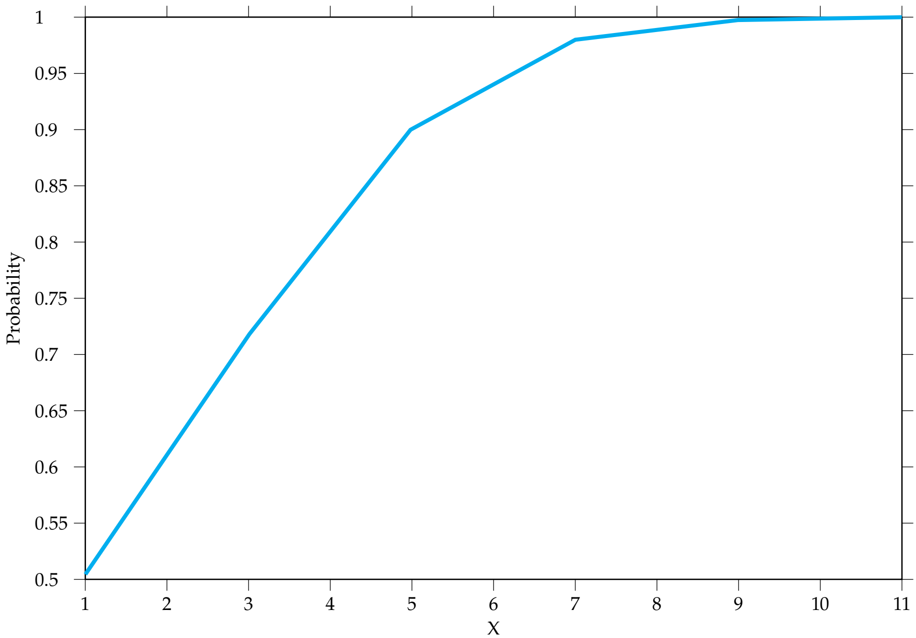

Figure 1. Probability that Disagreement Persists across Attribute 1

Example 5. The variances of the distributions of initial attitudes may affect the probability of disagreement. Consider that initial attitudes are normally distributed with the same means as in the previous example, for the cases in which $\Delta_0[1] > \Delta_0[2] \geq 0$. In contrast, let the variances of these random variables, instead of being all equal to one, be such that and $\tilde{a}_0^{\{1\}'} - \tilde{a}_0^{\{2\}'}$ and/or $\tilde{a}_0^{\{1,2\}'} - \tilde{a}_0^{\{\emptyset\}'}$ are mean preserving spreads of $\tilde{a}_0^{\{1\}} - \tilde{a}_0^{\{2\}}$ or $\tilde{a}_0^{\{1,2\}} - \tilde{a}_0^{\{\emptyset\}}$, respectively. Thus, $0.5 < P\left(\tilde{a}_0^{\{1\}'} - \tilde{a}_0^{\{2\}'} \geq 0\right) < P\left(\tilde{a}_0^{\{1\}'} - \tilde{a}_0^{\{2\}'} \geq 0\right)$ or $0.5 < P\left(\tilde{a}_0^{\{1,2\}'} - \tilde{a}_0^{\{\emptyset\}'} \geq 0\right) < P\left(\tilde{a}_0^{\{1,2\}} - \tilde{a}_0^{\{\emptyset\}} \geq 0\right)$.[6] Notice that expression (5) has now lower value than before, meaning that disagreements across either attribute are closer to being equally likely.

The case of an arbitrary set of $n > 2$ attributes is a natural extension of the case of two attributes. In this case, the probability of disagreement across a

6 Notice that $\Delta_0[1] = \Delta_0[2] \geq 0$ holds when $a_0^{\{1\}} - a_0^2 = 0$. In this case, regardless of the variances, disagreement across either attribute is equally likely. See the proof of Proposition 1.

particular attribute i is the probability that this attribute generates the highest average initial differences, that is, $P(|\Delta_0[i]| > |\Delta_0[k]|, \forall k \neq i)$. For any attribute i, the computation of this probability requires taking into account all the possible orders in which the differences of the $n-1$ remaining attributes may appear, when that particular attribute exhibits the highest initial difference. Thus, the computation of this probability entails some technical difficulties.

However, the intuition is that the workings of the model with an arbitrary set of attributes are parallel to the ones of the two-attribute case. More specifically, the conjectures are that: (*i*) disagreement is going to persist across any attribute with positive probability and (*ii*) the higher the mean of the distribution of initial differences in attitudes across a particular attribute, the higher the probability that disagreement persists across it. The following remark points toward this direction. Recall that we assume that the mean of the distributions of initial differences in attitudes is the analogous counterpart of expression (2), that is, $\Delta_0[1] \geq \Delta_0[2] \geq \cdots \geq \Delta_0[n] \geq 0$.

Remark 1. *For any pair of attributes* $i, i+1 \in \mathbb{N}$, *the probability* $|\tilde{\Delta}_0[1]| > |\tilde{\Delta}_0[i+1]|$ *is at least one half.*

This remark establishes that, for any pair of consecutive attributes, the probability of observing realizations of attitudes such that their initial differences preserve the order prescribed by their means is the most likely event. In particular, this probability is higher than one half when the means are strictly different and exactly one half when they are equal. The proof is the analogous counterpart of the one of Proposition 1. That result also suggests that the event in which we observe the same ranking of initial differences in attitudes than the one dictated by the means of the distributions of initial differences could be the most likely. Thus, disagreement across attribute 1 will be the most likely event.

3. Non-Symmetric Homophily

Again, consider the environment described in section 1. Let the four individuals be the result of the combination of attribute 1, which refers to the gender (possessing it means being female, whereas lacking it means being male), and attribute 2, which refers to youth (possessing it means being young, whereas lacking it means being old). Let old types (those lacking attribute 2) be more homophilous with respect to gender than young types (those possessing attribute 2). In this context, that means that old types are more sensitive to differences in attitudes associated with gender than young types. Given the structure of attention, as old types are more homophilous with respect to gender than young types, they are also less homophilous regarding youth than young types. Specifically, at each t, let β_t^1 and β_t^2 depend on $\Delta_t[1]$ and $\Delta_t[2]$ and be such that seniors exacerbate homophily towards gender. Specifically, let $\beta_t^1, \beta_t^2 \in [0, 1]$ be such that $\beta_t^1 + \beta_t^2 = 1$ and $\beta_t^2 > \lambda_t^1$, $\beta_t^1 < \lambda_t^2$. In particular, $\beta_t^1 = 1$ when $\lambda_t^1 = 1$.[7] The interaction matrix is:

$$W_t = \frac{1}{2}\begin{array}{c} \begin{array}{cccc} \{1,2\} & \{1\} & \{2\} & \{\emptyset\} \end{array} \\ \left[\begin{array}{cccc} \lambda_t^1+\lambda_t^2 & \lambda_t^1 & \lambda_t^2 & 0 \\ \beta_t^1 & \beta_t^1+\beta_t^2 & 0 & \beta_t^2 \\ \lambda_t^2 & 0 & \lambda_t^1+\lambda_t^2 & \lambda_t^1 \\ 0 & \beta_t^2 & \beta_t^1 & \beta_t^1+\beta_t^2 \end{array}\right] \end{array} \begin{array}{c} \\ \{1,2\} \\ \{1\} \\ \{2\} \\ \{\emptyset\} \end{array}.$$

Interactions, thus, become non-symmetric. Recall, that the law of motion of attitudes is given by (1). The consequences that this type of non-symmetric interactions has on the magnitude of disagreement, measured as the size of eventual average differences in attitudes across attribute 1, on segregation, and on the speed of convergence are summarized as follows.

Proposition 2. *For any vector of initial attitudes, the vector of limit attitudes is well defined and exhibits disagreement across attribute 1. Specifically, types hold the same limit attitude if and only if they are similar in attribute* 1. *Furthermore:*

1. *The magnitude of disagreement is larger than in the symmetric case where, at every t, $\beta_t^1 = \lambda_t^1$ and $\beta_t^2 = \lambda_t^2$.*
2. *Convergence is faster than in the symmetric case.*
3. *At every t, the groups of individuals similar in attribute* 1*(respectively attribute* 2*) are more (respectively less) segregated than in the symmetric case.*

As the relations on the basis of attribute 1 get more intense at every point in time, the process leads to a larger disagreement, measured as the magnitude of eventual average differences in attitudes across attribute 1, and converges faster to it. At each point in time, individuals that share attribute 1 are more segregated than in the symmetric case, according to the Spectral Segregation Index by Echenique and Fryer (2007). Also, this process will give

---

7 One can consider that β^1 and β^2 are transformations of λ^1 and λ^2, respectively. See example 6

rise to disagreement across attribute 1 even in the case in which the differences in average initial attitudes are equal, that is, even when $\Delta_0[1] = \Delta_0[2]$. That is so because attribute 1 deserves even more attention now that in the case in which relations are symmetric. An extreme version is such that at every time *t*, and regardless of the magnitude of differences in attitudes, $\beta_t^1 = 1$ and thus, $\beta_t^2 = 0$. Then, attribute 2 (youth) does not play any role for old types, namely, for types {1} and {∅}. They only pay attention to others based on the gender dimension - attribute 1-, that is, to types {1, 2} and {2}, respectively. The following example illustrates the findings:

Example 6. Let $a_0^{'} = [0.8\ 0.2\ {-0.05}\ {-0.95}]$ be the vector of initial attitudes. In this case, $\Delta_0[1] = 1$ and $\Delta_0[2] = 0.75$, and thus, as stated in Proposition 1 in Melguizo (2019), disagreement persists and manifests in two groups of thinking, defined according to the possession or lack of attribute 1. That is, the vector of limit attitudes is $a_\infty^{'} = [0.126\ 0.126 - 0.126 - 0.126]$. Let $\alpha > 1$ and set $\beta_t^2 = \left(\lambda_t^2\right)^\alpha < \lambda_t^2$, thus $\beta_t^1 = 1 - \left(\lambda_t^2\right)^\alpha > 1 - \lambda_t^2 = \lambda_t^1$. In particular, for $\alpha = 2$ we have that $a_\infty^{'} = [0.341\ 0.341 - 0.226 - 0.226]$. The magnitude of disagreement, measured as the difference in average eventual attitudes between the groups of individuals possessing and lacking attribute 1, is 0.25 in the symmetric case and 0.567 in the case described here. The later would also increase with the value of α.

For the case in which attribute 1 is, initially, at least as salient as attribute 2 —that is, $\Delta_0[1] \geq \Delta_0[2]$— but individuals pay more attention than before to others similar to them in attribute 2, that is, $\beta_t^2 > \lambda_t^2$ —and thus, $\beta_t^1 > \lambda_t^1$—, it might be that differences in attitudes associated with attribute 2 become the highest at some point in time. In this case disagreement would persist across that attribute. Consider two extreme cases: (i) let $\Delta_0[1] = \Delta_0[2] > 0$ but $\beta_t^1 > \lambda_t^1$. Then it directly follows that $\Delta_1[1] < \Delta_1[2]$. Thus, $\lambda_1^1 < 2^{-1} < \lambda_1^2$ and $\beta_1^1 < 2^{-1} < \beta_1^2$. The analysis from $t = 1$ on is the same than the one of Proposition 2, now applied to differences across attribute 2. Thus, disagreement persists across attribute 2. (ii) let $\Delta_0[1] > \Delta_0[2]$, but $\beta_t^1 = 0$ and thus $\beta_t^2 = 1$, at every *t*. In this case, disagreement also persists across attribute 2.[8]

8 From the proof of Proposition 2, we have that $\Delta_t[1] = \left(\lambda_{t-1}^1/2\right)\Delta_{t-1}[1]$ and $\Delta_t[2] = \left(\left(\lambda_{t-1}^2 + 1\right)/2\right)\Delta_{t-1}[2]$. Thus, $\lambda_1^1 = \dfrac{\Delta_1[1]}{\Delta_1[1] + \Delta_1[2]} = \dfrac{\left(\lambda_{t-1}^1/2\right)\Delta_0[1]}{\left(\lambda_{t-1}^1/2\right)\Delta_0[1] + \left(1 - \lambda_{t-1}^1/2\right)\Delta_0[2]}$. This is equivalent to $\lambda_1^1 = \left(1 + r_0^2\left(2 - \lambda_0^1\right)/\lambda_0^1\right)^{-1}$ where $r_0^2 = \Delta_0[2]/\Delta_0[1]$. Notice that $\lambda_0^1 = \left(1 + r_0^2\right)^{-1}$. Since $\lambda_1^1 < \lambda_0^1$ it follows that $\lambda_2^1 < \lambda_1^1$. In general, at every *t* time, $\lambda_{t+1}^1 < \lambda_t^1$, so that λ_t^1 tends to

Another possibility is to study the case in which some types exhibit constant homophily, that is, the attention they pay to similar others does not co-evolve with attitudes. This is the case if, for instance, $\beta_t^1 = \alpha$ in (0,1), at every time t. In this case, when the limiting matrix of interactions, namely, $\lim_{t\to\infty} W_t$, exists with $\lim_{t\to\infty} \lambda_t^1 \in (0,1)$, it would be such that no pair of rows are orthogonal. Following Leizarowitz (1992), consensus will eventually emerge in this case. In other words, when $\lim_{t\to\infty} \lambda_t^1 \in (0,1)$, eventual interactions are described by a strongly connected graph. That allows attitudes to flow from every individual to any other.

Notice that in the case of $n > 2$ attributes, the results would go through as long as individuals exacerbate the attention they pay to the attribute with the highest initial mean, that is, to attribute 1, which is gender, in this case. The conjecture is that this result will also hold even if individuals exacerbate attention at different extents. As for the case of two attributes, when individuals exacerbate the attention they pay to attributes different from attribute 1, it may be that one of these attributes becomes the one with the highest differences in attitudes at some point in time. Then, the conjecture is that disagreement will persist across it.

4. Comments on the Co-Existence of Random Attitudes and Non-Symmetric Homophily

The previous sections explored the implications of extending the model in two different directions, one at a time. We showed that for the case of two attributes, disagreement across attribute 1 was the most likely event. In the case of non-symmetric homophily, as long as individuals exacerbate their homophilous behavior towards 1-similar others, disagreement across attribute 1 still takes place.

An interesting question is what the implications are of facing both extensions simultaneously in action. In what follows, there are comparisons with the cases in which attitudes are random (and homophily symmetric), and the opposite case in which attitudes are non-random but homophily is non-symmetric.

---

0 and λ_t^2 tends to 1. Consequently, for sufficiently large t, it has to be that $\lambda_t^2 > \lambda_t^1$ and thus $\Delta_t[2] > \Delta_t[1]$. From this point on, apply the proof of Proposition 2 to attribute 2.

Remark 2. *Consider the case of two attributes in which attitudes are random and homophily is non-symmetric. Then, the probability that disagreement persists across attribute* 1 *is, at least, as high as in the case in which attitudes are random but homophily is symmetric.*

This remark tells that adding instances of non-symmetric homophily that exacerbate the attention paid to attribute 1, may increase the probability of disagreement on attribute 1, with respect to the case in which attitudes are random but homophily symmetric. The basic intuition is that with random attitudes, there may be instances in which initial differences are the highest on attribute 2. As the attention towards attribute 1 exacerbates, when this attention is sufficiently important, the process may end up showing disagreement on it.

Consider now the opposite case in which attitudes are non-random and homophily non-symmetric. In this case, there was a unique outcome, which was disagreement on attribute 1 because initial differences were the highest across it. Adding randomness over non-symmetric homophily implies that the initial differences may not be always the highest on attribute 1. That possibly reduces the probability that disagreement takes place on it. In other words, the probability that disagreement persists across attribute 2 may be positive. The reason is that with non-symmetric homophily, the evolution of average differences in attitudes basically relies on a linear combination between the original attention paid to attribute 1 in the symmetric case, λ_t^i, and the exacerbated attention paid to attribute 1, β_t^i. Intuitively, when differences across attribute 2 are the highest, it turns out that λ_t^i is small; thus, β_t^i should be high enough to compensate the former. If that is not the case, then disagreement may take place on attribute 2. That very much depends on the assumptions made on the relation between the aforementioned parameters.

Conclusions

This note explores natural modifications of the setting proposed by Melguizo (2019). Regarding random initial attitudes, the finding is that disagreement persists with the highest probability across the attribute that exhibits the highest mean of the distribution of initial differences in attitudes. Taking into account non-symmetric homophily, the magnitude of disagreement, and the speed of convergence to it, increases with respect to the case in which individuals' interactions are symmetric.

One interesting aspect to consider is the implications of having individuals seeking for information that supports their views, a phenomenon which is known as confirmation bias.[9] One model that may be interpreted in these terms is analyzed in Hegselmann and Krause (2002). In it, individuals pay attention to others whose attitudes are not so far from their own. To do so, they define what they call levels of confidence to others' opinions. The authors find instances of consensus (for high confidence levels) as well as of fragmentation (for low confidence levels), that is, the persistence of multiple groups of thinking.

In the context of the current setting, in which initial differences in attitudes determine disagreement, confirmation biases could be modeled as individuals not only considering initial differences but also how close *i*-similar individuals' attitudes are from their own. One way to incorporate this aspect is to consider that the attention, λ_t^1, each individual pays to 1-similar others at time t would be increasing in the initial differences as before and decreasing in the variance of the distribution of attitudes of the group of 1-similar types. Intuitively, small variances suggest that attitudes are sufficiently closed to the mean, and hence sufficiently closed among themselves. Thus, the confirmation bias effect may be captured in this setting.

To illustrate the idea, consider that $a_0^{'}$ = [0.75 0.5 0.25 0]. Notice that $\overline{a}_0[1] = 0.625$, $\overline{a}_0[-1] = 0.125$, and $\Delta_0[1] = 0.625 - 0.125 = 0.5$. Thus, in the original setting disagreement takes place across attribute 1. Consider the distribution of attitudes of 1-similar types. The variance of the distribution of these attitudes for individuals who have attribute 1 is $2^{-1}[(0.75 - 0.625)^2 + (0.5 - 0.625)^2] = 0.0078$. The variance of the distribution of these attitudes for individuals that lack attribute 1 is the same, that is, $2^{-1}[(0.25 - 0.125)^2 + (0 - 0.125)^2] = 0.0078$. Doing analogous computation, one sees that the variances of the distribution of attitudes for 2-similar individuals are the same and equal to 0.625. Thus, in this case, if attention is sensitive also to variances, that would exacerbate the attention individuals pay to 1-similar others. Intuitively the lower the variance, the most informative is the average attitude about the cohesion within a society of individuals that share an attribute. Of course, that is only an example. However, it suggests that the particular distribution of initial attitudes, and not only the average differences, would play a role in reinforcing disagreement across attribute 1, reverting this outcome to disagreement on attribute 2, or even reaching consensus.

Finally, it may also be that individuals possessing and lacking one specific attribute, say 1, observe different variances within the attitudes of their group.

9 See Nickerson (1998), Zollo, et al. (2015), and Quattrociocchi, et al. (2016).

Thus, they will be paying different attention to 1-similar others. That is a different setting that may be linked to the case of non-symmetric homophily.

Proofs

Proof of Proposition 1. Compute the probability that disagreement persists across attributes 1 and 2 when, first, $\Delta_0[1] > \Delta_0[2] \geq 0$ or, equivalently, $a_0^{\{1,2\}} - a_0^{\{\emptyset\}} + a_0^{\{1\}} - a_0^{\{2\}} > a_0^{\{1,2\}} - a_0^{\{\emptyset\}} + a_0^{\{2\}} - a_0^{\{1\}}$. Notice that $\Delta_0[1] - \Delta_0[2] = a_0^{\{1\}} - a_0^{\{2\}}$. Thus, $a_0^{\{1\}} - a_0^{\{2\}} > 0$ has to hold. Since $\Delta_0[2] \geq 0$ and $a_0^{\{2\}} - a_0^{\{1\}} < 0$, then $a_0^{\{1,2\}} - a_0^{\{\emptyset\}} \geq a_0^{\{1\}} - a_0^{\{2\}} > 0$ has to hold as well.

Now, consensus emerges whenever $|\widetilde{\Delta}_0[1]| = |\widetilde{\Delta}_0[2]|$, and disagreement persists across attribute 1 (respectively attribute 2) whenever $|\widetilde{\Delta}_0[1]| > |\widetilde{\Delta}_0[2]|$ (resp. $|\widetilde{\Delta}_0[1]| < |\widetilde{\Delta}_0[2]|$. To see this, notice that once initial attitudes are realized, the process exactly mimics the one in Melguizo (2019). In what follows, we describe the probability that either consensus emerges or disagreement persists. The probability that $|\widetilde{\Delta}_0[1]| = \left|\tilde{a}_0^{\{1,2\}} - \tilde{a}_0^{\{\emptyset\}} + \tilde{a}_0^{\{1\}} - \tilde{a}_0^{\{2\}}\right| = \left|\tilde{a}_0^{\{1,2\}} - \tilde{a}_0^{\{\emptyset\}} + \tilde{a}_0^{\{2\}} - \tilde{a}_0^{\{1\}}\right| = |\widetilde{\Delta}_0[2]|$ is zero. That is so because this expression holds when exactly $\tilde{a}_0^{\{1\}} - \tilde{a}_0^{\{2\}} = 0$ and/or $\tilde{a}_0^{\{1,2\}} - \tilde{a}_0^{\{\emptyset\}} = 0$. Since these differences follow continuous distributions, the probability that this happens is zero.[10] Disagreement persists across attribute 1 whenever $|\widetilde{\Delta}_0[1]| = \left|\tilde{a}_0^{\{1,2\}} - \tilde{a}_0^{\{\emptyset\}} + \tilde{a}_0^{\{1\}} - \tilde{a}_0^{\{2\}}\right| > \left|\tilde{a}_0^{\{1,2\}} - \tilde{a}_0^{\{\emptyset\}} + \tilde{a}_0^{\{2\}} - \tilde{a}_0^{\{1\}}\right| = |\widetilde{\Delta}_0[2]|$. This expression is satisfied when $\tilde{a}_0^{\{1,2\}} - \tilde{a}_0^{\{\emptyset\}} \geq 0$ and $\tilde{a}_0^{\{1\}} - \tilde{a}_0^{\{2\}} \geq 0$, or $\tilde{a}_0^{\{1,2\}} - \tilde{a}_0^{\{\emptyset\}} < 0$, and $\tilde{a}_0^{\{1\}} - \tilde{a}_0^{\{2\}} < 0$ hold. Thus, $P\left(|\widetilde{\Delta}_0[1]| > |\widetilde{\Delta}_0[2]|\right) = P\left(\tilde{a}_0^{\{1,2\}} - \tilde{a}_0^{\{\emptyset\}} \geq 0 \cap \tilde{a}_0^{\{1\}} - \tilde{a}_0^{\{2\}} \geq 0\right) + P\left(\tilde{a}_0^{\{1,2\}} - \tilde{a}_0^{\{\emptyset\}} < 0 \cap \tilde{a}_0^{\{1\}} - \tilde{a}_0^{\{2\}} < 0\right)$. Since $\tilde{a}_0^{A}$ are independent to each other, this is equivalent to:

$$P\left(\tilde{a}_0^{\{1,2\}} - \tilde{a}_0^{\{\emptyset\}} \geq 0\right) P\left(\tilde{a}_0^{\{1\}} - \tilde{a}_0^{\{2\}} \geq 0\right) + P\left(\tilde{a}_0^{\{1,2\}} - \tilde{a}_0^{\{\emptyset\}} < 0\right) P\left(\tilde{a}_0^{\{1\}} - \tilde{a}_0^{\{2\}} < 0\right) \tag{6}$$

On the contrary, disagreement persists across attribute 2 whenever $|\widetilde{\Delta}_0[1]| = \left|\tilde{a}_0^{\{1,2\}} - \tilde{a}_0^{\{\emptyset\}} + \tilde{a}_0^{\{1\}} - \tilde{a}_0^{\{2\}}\right| < \left|\tilde{a}_0^{\{1,2\}} - \tilde{a}_0^{\{\emptyset\}} + \tilde{a}_0^{\{2\}} - \tilde{a}_0^{\{1\}}\right| = |\widetilde{\Delta}_0[2]|$. This expression

10 As $\tilde{a}_0^{\{2\}}$ is continuous, so is $-\tilde{a}_0^{\{2\}}$. The sum $\tilde{a}_0^{\{1\}} + \left(-\tilde{a}_0^{\{2\}}\right)$ is, accordingly, continuous. See Sheldon et al. (2002).

is satisfied when $\tilde{a}_0^{\{1,2\}} - \tilde{a}_0^{\{\emptyset\}} < 0$ and, or $\tilde{a}_0^{\{1\}} - \tilde{a}_0^{\{2\}} \geq 0$, or $\tilde{a}_0^{\{1,2\}} - \tilde{a}_0^{\{\emptyset\}} \geq 0$ and $\tilde{a}_0^{\{1\}} - \tilde{a}_0^{\{2\}} < 0$ hold. Then $P\left(\left|\tilde{\Delta}_0[1]\right| < \left|\tilde{\Delta}_0[2]\right|\right)$ is

$$P\left(\tilde{a}_0^{\{1,2\}} - \tilde{a}_0^{\{\emptyset\}} \geq 0\right)P\left(\tilde{a}_0^{\{1\}} - \tilde{a}_0^{\{2\}} < 0\right) + P\left(\tilde{a}_0^{\{1,2\}} - \tilde{a}_0^{\{\emptyset\}} < 0\right)P\left(\tilde{a}_0^{\{1\}} - \tilde{a}_0^{\{2\}} \geq 0\right) \tag{7}$$

Thus (6) minus (7) results in . This expression is equivalent to:

$$\left(2P\left(\tilde{a}_0^{\{1,2\}} - \tilde{a}_0^{\{\emptyset\}} \geq 0\right) - 1\right)\left(P\left(\tilde{a}_0^{\{1\}} - \tilde{a}_0^{\{2\}} \geq 0\right) - P\left(\tilde{a}_0^{\{1\}} - \tilde{a}_0^{\{2\}} < 0\right)\right) \tag{8}$$

Since $\tilde{a}_0^{\{1\}} - \tilde{a}_0^{\{2\}}$ has positive mean (recall that $\tilde{a}_0^{\{1\}} - \tilde{a}_0^{\{2\}} > 0$) and the difference of independent symmetric random variables is symmetric, then $P\left(\tilde{a}_0^{\{1\}} - \tilde{a}_0^{\{2\}} \geq 0\right) > 0.5$.[11] The same argument holds for $\tilde{a}_0^{\{1,2\}} - \tilde{a}_0^{\{\emptyset\}}$ and, consequently, $P\left(\tilde{a}_0^{\{1,2\}} - \tilde{a}_0^{\{\emptyset\}} \geq 0\right) > 0.5$. This implies that (8) is positive. Since $P\left(\left|\tilde{\Delta}_0[1]\right| > \left|\tilde{\Delta}_0[2]\right|\right) = 1 - P\left(\left|\tilde{\Delta}_0[1]\right| < \left|\tilde{\Delta}_0[2]\right|\right)$, disagreement across attribute 1 is the most likely. In the extreme case in which $P\left(\tilde{a}_0^{\{1,2\}} - \tilde{a}_0^{\{\emptyset\}} \geq 0\right) = P\left(\tilde{a}_0^{\{1\}} - \tilde{a}_0^{\{2\}} \geq 0\right) = 1$ the probability that disagreement takes place across attribute 1 is exactly one. Let now $\Delta_0[1] = \Delta_0[2] \geq 0$ or, equivalently, $\Delta_0[1] - \Delta_0[2] = a_0^{\{1\}} - a_0^{\{2\}} = 0$. Since the mean of differences is non-negative, $a_0^{\{1,2\}} - a_0^{\{\emptyset\}} \geq 0$ holds. That implies, by symmetry, that $P\left(\left|\tilde{\Delta}_0[1]\right| > \left|\tilde{\Delta}_0[2]\right|\right) - P\left(\left|\tilde{\Delta}_0[1]\right| < \left|\tilde{\Delta}_0[2]\right|\right) = 0$. In this case, disagreement across either attribute is equally likely.

Proof of Remark 1. Consider n attributes with initial means $\Delta_0[1] \geq \Delta_0[2] \geq \cdots \geq \Delta_0[n] \geq 0$. For any pair of attributes i, $i + 1$, there are 2^{n-1} individuals that have i and 2^{n-1} that lack it. The same holds for $i + 1$.

As in the proof of Proposition 1, the ranking of initial means implies that for any pair $i, i+1: \Delta_0[i] - \Delta_0[i+1] = \sum_{A|i\in A, i+1\notin A} a_0^A - \sum_{A|i+1\in A, k\notin A} a_0^A > 0$. Since, differences are positive: $\sum_{A|i,i+1\in A} a_0^A - \sum_{A|i,i+1\notin A} a_0^A > 0$ must also hold. Analogously to expression (6) in the proof of Proposition 1:

11 See Stroock (2010).

$$
\begin{aligned}
&P\left(\left|\tilde{\Delta}_0[i]\right| > \left|\tilde{\Delta}_0[i+1]\right|\right) \\
&= P\left(\sum_{A|i,i+1\in A} \tilde{a}_0^A - \sum_{A|k,k+1\notin A} \tilde{a}_0^A \geq 0 \cap \sum_{A|i\in A,i+1\notin A} \tilde{a}_0^A - \sum_{A|i+1\in A,k\notin A} \tilde{a}_0^A \geq 0\right) \\
&+ P\left(\sum_{A|i,i+1\in A} \tilde{a}_0^A - \sum_{A|i,i+1\notin A} \tilde{a}_0^A < 0 \cap \sum_{A|i\in A,i+1\notin A} \tilde{a}_0^A - \sum_{A|i+1\in A,i\notin A} \tilde{a}_0^A < 0\right).
\end{aligned}
$$

We can also compute $P\left(\left|\tilde{\Delta}_0[i]\right| < \left|\tilde{\Delta}_0[i+1]\right|\right)$ as in (7). Notice that $P\left(\left|\tilde{\Delta}_0[i]\right| > \left|\tilde{\Delta}_0[i+1]\right|\right) - P\left(\left|\tilde{\Delta}_0[i]\right| < \left|\tilde{\Delta}_0[i+1]\right|\right)$ yields an expression analogous to (8). Specifically, this expression is, in short, $(2P-1)(2P'-1)$. Here P stands for the probability that the sum of attitudes of individuals that have both i and $i+1$, minus the sum of attitudes of individuals that lack both i and $i+1$ is non-negative. Analogously, P' stands for the probability that the sum of attitudes of individuals that have i and lack $i+1$, minus the sum of attitudes of individuals with $i+1$ and without i is non-negative.

Analogously to the proof of Proposition 1, the distribution of the sum of attitudes of individuals that have i and lack $i+1$ minus the sum of attitudes of individuals that with $i+1$ but without i have positive or zero means, depending on whether $\Delta_0[i] - \Delta_0[i+1] > 0$ or $\Delta_0[i] - \Delta_0[i+1] = 0$, respectively. The same holds for the distribution of the sum of attitudes of individuals that have both i and $i+1$ minus the sum of attitudes of individuals that lack both i and $i+1$.

That, together with symmetry implies that both, P and P', are higher than one half whenever $\Delta_0[i] - \Delta_0[i+1] > 0$ and that P' is one half —and thus $(2P-1)(2P'-1)$ is zero— whenever $\Delta_0[i] - \Delta_0[i+1] = 0$. Thus, $P\left(\left|\tilde{\Delta}_0[i]\right| - \left|\tilde{\Delta}_0[i+1]\right|\right)$ is higher than one half whenever $\Delta_0[i] - \Delta_0[i+1] > 0$, and one half otherwise.

Proof of Proposition 2. First, there is a statement regarding the evolution of homophily values. Let $\Delta_0[1] > \Delta_0[2]$. It then follows that $\lambda_0^1 > 2^{-1} > \lambda_0^2$. Let $\lambda_t^* = \left(\lambda_t^1 + \beta_t^1\right)2^{-1} \in (0,1]$. Notice that as $\Delta_t[1] = 2^{-1}\left(\left(a_t^{\{1,2\}} + a_t^{\{1\}}\right) - \left(a_t^{\{2\}} + a_t^{\{\emptyset\}}\right)\right)$, using W_t in the main body, we can rewrite $\Delta_t[1] = \lambda_{t-1}^* \Delta_{t-1}[1]$. Similarly, $\Delta_t[2] = \left(1 - \lambda_{t-1}^*\right)\Delta_{t-1}[2]$. As these expressions hold for every t, $\Delta_t[1] = \prod_{s=0}^{t-1} \lambda_{s-1}^* \Delta_t[1]$ and $\Delta_t[2] = \prod_{s=0}^{t-1}\left(1 - \lambda_s^*\right)\Delta_0[2]$ hold. So that disagree-ment persists whenever $\lim_{t\to\infty} \Delta_t[1] \neq 0$ or $\lim_{t\to\infty} \Delta_t[2] \neq 0$.

For the ease of exposition, let $\hat{\lambda}_t^i$ for $i = \{1,2\}$ be the homophily value of attribute i when W_t is symmetric. It is important to notice that $\hat{\lambda}_t^i$ with symmetric W_t is going to be different from λ_t^i in the non- symmetric W_t in the main body, as is made clear below. Since by assumption $\beta_t^1 > \lambda_t^1$ at every t, then, $\lambda_1^1 = \frac{\lambda_0^*\Delta_0[1]}{\lambda_0^*\Delta_0[1]+(1-\lambda_0^*)\Delta_0[2]} > \hat{\lambda}_1^1$. Thus, $\Delta_2[1]=\lambda_1^*\Delta_1[1]$ is higher than in the symmetric case, while $\Delta_2[2]=(1-\lambda_1^*)\Delta_1[2]$ is smaller than in the symmetric case. As a consequence, $\lambda_2^1 > \hat{\lambda}_2^1$. In general at every t, $\beta_t^1 > \lambda_t^1 > \hat{\lambda}_t^1$ and $\beta_t^2 < \hat{\lambda}_t^2 < \hat{\lambda}_2^1$, with $\lambda_0^1 = \hat{\lambda}_0^1$.

Consider a sequence of ones. Since at every t, $1 > \lambda_t^1 \geq \hat{\lambda}_t$ by step 7 in the proof of the main Therem in Melguizo (2019), as $\lim_{t\to\infty}\hat{\lambda}_t^1 = 1$ then $\lim_{t\to\infty}\lambda_t^1 = 1$. Also since $1 \geq \beta_t^1 > \lambda_t^1$, thus $\lim_{t\to\infty}\beta_t^1 = 1$. By the same step, $\sum_{t=0}^{\infty}\left|log\left(\hat{\lambda}_t^1\right)\right|$ was convergent. Since at every t, $\lambda_t^* = (\lambda_t^1+\beta_t^1)2^{-1} > \hat{\lambda}_t^1$ then $\left|log\left(\lambda_t^*\right)\right| \leq \left|log\left(\hat{\lambda}_t^*\right)\right|$. Thus, by comparison $\sum_{t=0}^{\infty}\left|log\left(\lambda_t^*\right)\right|$ converges and, hence, $\prod_{t=0}^{\infty}\lambda_t^* = \delta \in (0,1]$. Since at every t, $\lambda_t^1 > \hat{\lambda}_t^1$, then $\delta > \mu = \prod_{t=0}^{\infty}\hat{\lambda}_t^1 \in (0,1]$. Since $lim_{t\to\infty} 1-\lambda_t^* = 0$ then $\prod_{t=0}^{\infty}(1-\lambda_t^*) = 0$ holds. Recall that $a_{t+1} = W_t a_t$. Rewrite W_t as:

$$W_t = \frac{1}{4}\begin{array}{cc} \begin{array}{cccc} \{1,2\} & \{1\} & \{2\} & \{\emptyset\} \end{array} & \\ \begin{bmatrix} 1+\lambda_t^1+\lambda_t^2 & 1+\lambda_t^1-\lambda_t^2 & 1-\lambda_t^1+\lambda_t^2 & 1-\lambda_t^1+\lambda_t^2 \\ 1+\beta_t^1-\beta_t^2 & 1+\beta_t^1+\beta_t^2 & 1-\beta_t^1-\beta_t^2 & 1-\beta_t^1+\beta_t^2 \\ 1-\lambda_t^1+\lambda_t^2 & 1-\lambda_t^1-\lambda_t^2 & 1+\lambda_t^1+\lambda_t^2 & 1+\lambda_t^1-\lambda_t^2 \\ 1-\beta_t^1-\beta_t^2 & 1-\beta_t^1+\beta_t^2 & 1+\beta_t^1-\beta_t^2 & 1+\beta_t^1+\beta_t^2 \end{bmatrix} & \begin{array}{c} \{1,2\} \\ \{1\} \\ \{2\} \\ \{\emptyset\} \end{array} \end{array}.$$

The weights associated with attribute 1 (respectively attribute 2), i.e., λ_t^1 and β_t^1 (respectively λ_t^2 and β_t^2), enter with positive sign if, and only if, type A possesses attribute 1 (respectively attribute 2). It then follows that $a_{t+1}^A = \bar{a}_t + 2^{-1}\left((-1)^{1+1_i}\lambda_t^1\Delta_t[1]+\lambda_t^2\Delta_t[2]\right)$ if $2 \in A$ and $a_{t+1}^A = \bar{a}_t + 2^{-1}\left((-1)^{1+1_i}\beta_t^1\Delta_t[1]+\beta_t^2\Delta_t[2]\right)$ if $2 \notin A$, where 1_i is the indicator of type A possessing attribute 1. Using the recursive expressions for $\Delta_t[1]$ and $\Delta_t[2]$ above, let $a_{t+1}^A = \bar{a}_t + 2^{-1}\left((-1)^{1+1_i}\lambda_t^1\prod_{s=0}^{t-1}\lambda_s^*\Delta_0[1]+\lambda_t^2\prod_{s=0}^{t-1}\lambda_s^*\Delta_0[2]\right)$ if $2 \in A$ and $a_{t+1}^A = \bar{a}_t + 2^{-1}\left((-1)^{1+1_i}\beta_t^1\prod_{s=0}^{t-1}\lambda_s^*\Delta_0[1]+\beta_t^2\prod_{s=0}^{t-1}\lambda_s^*\Delta_0[2]\right)$ if $2 \notin A$. Since $\prod_{t=0}^{\infty}\lambda_t^* = \delta$, $\prod_{t=0}^{\infty}(1-\lambda_t^*) = 0$, $\lim_{t\to\infty}\beta_t^1 = 1$ and $\lim_{t\to\infty}\lambda_t^1 = 1$ it follows that $\lim_{t\to\infty}a_{t=0}^a = \lim_{t\to\infty}\bar{a}_t + 2^{-1}\delta\Delta_0[1]$ if

$1 \in A$ and $\lim_{t\to\infty} a^a_{t=1} = \lim_{t\to\infty} \bar{a}_t - 2^{-1}\delta\Delta_0[1]$ if $1 \notin A$, provided that $\lim_{t\to\infty} \bar{a}_t$ exists. Simple algebra yields $\bar{a}_t = \bar{a}_0 + 4^{-1}\sum_{s=0}^{t-1}\left(\lambda_s^2 - \beta_s^2\right)\Delta_s[2] = \bar{a}_0 + 4^{-1}\Delta_0[2]\sum_{s=0}^{t-1}\left(\lambda_s^2 - \beta_s^2\right)$ $\prod_{m=0}^{s-1}\left(1-\lambda_m^*\right)$. Thus, to prove the existence of the aforementioned limit, it is left to prove that $\sum_{s=0}^{\infty}\left(\lambda_s^2 - \beta_s^2\right)\lambda_s^2 - \beta_s^2$ converges. That is done by proving that $\sum_{s=0}^{\infty}\left(\lambda_s^2 - \beta_s^2\right)$ converges. Ergo, given that at every time s, $\lambda_s^2 - \beta_s^2 > \left(\lambda_s^2 - \beta_s^2\right)\prod_{m=0}^{s-1}\left(1-\lambda_m^*\right)$, by comparison, the conclusion is that $\sum_{s=0}^{t-1}\left(\lambda_s^2 - \beta_s^2\right)\prod_{m=0}^{s-1}\left(1-\lambda_m^*\right)$ converges. Notice that $\sum_{s=0}^{\infty}\left(\lambda_s^2 - \beta_s^2\right)$ converges if, and only if, $\prod_{t=0}^{\infty}\left(1-\left(\lambda_s^2 - \beta_s^2\right)\right)$ converges, i.e., this limit product is a number different from zero, see Apostol (1977) chapter 8, Theorem 8.55. Rewrite this last expression as $\sum_{s=0}^{\infty}\left(\lambda_s^1 - \beta_s^1\right)$. From step 7 in the proof of the main theorem in Melguizo (2019), $\prod_{s=0}^{\infty}\hat{\lambda}_s^1 \in (0,1]$. Recall that since $\beta_s^2 < \lambda_s^2$, then $0 \le \lambda_s^1 + \beta_s^2 < 1$. Also, since at each s $\lambda_s^1 \ge \hat{\lambda}_s^1$ then $\lambda_s^1 + \beta_s^2 \ge \hat{\lambda}_s^1$. Thus, $\prod_{s=0}^{\infty}\left(\lambda_s^1 + \beta_s^2\right) \in (0,1]$. Thereupon, if $\sum_{s=0}^{\infty}\left(\lambda_s^2 - \beta_s^2\right) \in (0,1]$ converges, so does $\sum_{s=0}^{\infty}\left(\lambda_s^2 - \beta_s^2\right)\prod_{m=0}^{s-1}\left(1-\lambda_m^*\right)$ by comparison. Now let $\alpha \in [0,\infty)$ be the value of this infinite sum, then $\lim_{t\to\infty} a^A_{t+1} = \bar{a}_0 4^1 \alpha\Delta_0[2] + 2^{-1}\delta\Delta_0[1]$ if $1 \in A$ and $\lim_{t\to\infty} a^A_{t+1} = \bar{a}_0 4^1 \alpha\Delta_0[2] - 2^{-1}\delta\Delta_0[1]$ if $1 \notin A$. As stated above, $\delta > \mu$. Thus, disagreement persists across attribute 1 and the magnitude of disagreement in the non-symmetric case, that is, $\lim_{t\to\infty}\Delta_t[1] = \delta\Delta_0[1]$ is higher than the magnitude of disagreement in the symmetric case: $\lim_{t\to\infty}\Delta_t[1] = \mu\Delta_0[1]$. Finally, since $\beta_t^1 > \lambda_t^1 > \hat{\lambda}_t^1$ and λ_t^1 are, at every time t, closer to 1 than $\hat{\lambda}_t^1$, the convergence is also faster.

Let now $\Delta_0[1] = \Delta_0[2]$. Then $\hat{\lambda}_t^1 = 2^{-1}$ at every time t by the proof of the main theorem in Melguizo (2019). By the same arguments as above, $\lambda_1^0 = \hat{\lambda}_t^1$ and $\lambda_t^1 > \hat{\lambda}_t^1$ at every time $t > 1$. Thus, $\Delta_t[1]$ is higher than the one in the symmetric case at every time $t > 1$. As a result, consider the process as starting at time $t = 1$ with $\lambda_t^1 > \hat{\lambda}_t^1$ and the same arguments as above follow. Thus, disagreement persists across attribute 1 and is higher than in the symmetric case, being the convergence faster.

The computation of the spectral segregation index directly follows the proof of Proposition 2 in Melguizo (2019). Its value is $2^{-1}\left(1+\beta_t^1\right)$ for the groups of individuals similar in attribute 1. Since $\beta_t^1 > \hat{\lambda}_t^1$ at every time, the segregation of this group is higher than in the symmetric case. The opposite holds for individuals similar in attribute 2.

Proof of Remark 2. Consider that attitudes are random and homophily is non-symmetric. Let the means of the initial distributions of attitudes be:

1. $\Delta_0[1] > \Delta_0[2]$. In this case, from Proposition 1, it holds that $P\left(\left|\tilde{\Delta}_0[1]\right| > \left|\tilde{\Delta}_0[2]\right|\right)$ is higher than one half. From Proposition 2, in all these events, disagreement takes place across attribute 1.
2. $\Delta_0[1] > \Delta_0[2]$. From Proposition 1, it holds that $P\left(\left|\tilde{\Delta}_0[1]\right| > \left|\tilde{\Delta}_0[2]\right|\right)$ is one half. From Proposition 2, in all these events, disagreement also takes place across attribute 1.

With respect to the case in which only random attitudes are in action, in the case in which, in addition, non-symmetric homophily plays a role, individuals exacerbate the attention they pay to attribute 1. That happens also in the events in which $P\left(\left|\tilde{\Delta}_0[1]\right| > \left|\tilde{\Delta}_0[2]\right|\right)$. Thus, it may be that the differences across attribute 1 become at least as high as the differences in attribute 2, at some point in time $\tilde{t}$. If such a exists, then the analysis in the proof of Proposition 2 applies, from that point in time $\tilde{t}$ on, to attribute 1. Thus, disagreement persists across that attribute. The conclusion is, therefore, that the probability that disagreement persists across attribute 1 is at least as high in the case with random variables and non-symmetric homophily as in the case in which only random variables are in action.

References

Apostol, T. M. (1977). *Análisis Matemático*. Barcelona: Revert´e.

DeGroot, M. H. (1974). Reaching a consensus. *Journal of the American Statistical Association, 69*(345), 118-121.

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

Echenique, F., & Fryer, R. G. (2007). A measure of segregation based on social interactions. *The Quarterly Journal of Economics, 122*(2), 441-485.

Golub, B., & Jackson, M. O. (2010). Naïve learning in social networks and the wisdom of crowds. *American Economic Journal: Microeconomics, 2*(1), 112-149.

Hegselmann, R., & Krause, U. (2002). Opinion dynamics and bounded confidence models, analysis, and simulation. *Journal of Artificial Societies and Social Simulation, 5*(3), 1-33.

Leizarowitz, A. (1992). On infinite products of stochastic matrices. *Linear algebra and its applications, 168,* 189-219.

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology, 27,* 415-444.

Melguizo, I. (2019). Homophily and the persistence of disagreement. *The Economic Journal, 129*(619), 1400-1424.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of general psychology, 2*(2), 175-220.

Quattrociocchi, W., Scala, A., & Sunstein, C. R. (2016). Echo chambers on Facebook. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2795110

Sheldon, R. et al. (2002). *A first course in probability*. Pearson Education India.

Stroock, D. W. (2010). *Probability theory: an analytic view*. Cambridge University Press.

Zollo, F., Kralj Novak, P., del Vicario, M., Bessi, A., Mozetiˇc, I., Scala, A., Caldarelli, G., & Quattrociocchi, W. (2015). Emotional dynamics in the age of misinformation. *PloS One*, *10*(9). http://dx.doi.org/10.1371/journal.pone.0138740