# Homophily and the Persistence of Disagreement

Melguizo, Isabel

Universidad Iberoamericana

January 2017

# Homophily and the Persistence of Disagreement*

Isabel Melguizo†

March 8, 2017

**Abstract**

We study a dynamic model of attitude formation in which individuals average others' attitudes to develop their own. We assume that individuals exhibit homophily in sociodemographic exogenous attributes, that is, the attention they pay to each other is based on whether they possess similar attributes. We also assume that individuals exhibit homophily in attitudes, at the group level. Specifically, attributes that are salient, that is, that exhibit a substantial difference in attitudes between the groups of individuals possessing and lacking them, deserve high attention. Since we allow attention to evolve over time we prove that when there is, initially, a unique most salient attribute, it deserves growing attention overtime in detriment of the remaining ones. As a result, individuals eventually interact only with others similar to them across this attribute and disagreement persists. It materializes in two groups of thinking defined according to this attribute.

Keywords: disagreement, homophily, salience, average-based updating
JEL classification: D83, D85, Z13

*I am grateful to my advisor, Miguel Ángel Ballester. I thank Jorge Alcalde Unzu, Antonio Cabrales, Francesco Cerigioni, Giacomo De Giorgi, Ben Golub, Matthew O. Jackson, Markus Kinateder, Konrad Mierendorff, Juan D. Moreno-Ternero, Pedro Rey-Biel, Tomás Rodríguez Barraquer, Javier Alejandro Rodríguez Camacho, Benjamín Tello, Jan Zápal and the participants at IDEA Micro-Theory Lab, UCL Student Seminar and Barcelona IDGP Workshop for their helpful comments and suggestions. Financial support from the Universitat Autònoma de Barcelona through PIF Scholarship 912-01-09/2010, from the Spanish Ministry of Economy and Competitiveness through FPI Scholarship BES-2013-06492 and from the Spanish Ministry of Science and Innovation through grants "Consolidated Group-C" ECO2008-04756 and FEDER is gratefully acknowledged.

†Universidad Iberoamericana. Economics Department. Prolongación Paseo de la Reforma 880. Lomas de Santa Fe, 01219, Ciudad de México. Email: isabelmelguizolopez@gmail.com. I certify that I have the right to deposit the contribution with MPRA.

# 1 Introduction

Disagreement is an everyday life phenomenon. When in 1987 the American public was confronted with the question of whether the government should guarantee every citizen enough to eat and a place to sleep, 80% of black people agreed whereas only 55% of white people. For around 25 years these percentages have remained almost constant. Disagreements tend to persist, most of the times, over non-factual issues. In fact, differences in attitudes regarding a wide range of topics of ethical and ideological content have persisted among the American public during the aforementioned period.[1]

Despite this evidence, existing models of communication and learning, regardless of whether individuals behave as Bayesian or use rules of thumb, typically lead to consensus results. This is the case in DeMarzo et al. (2003), Acemoglu et al. (2010), Golub and Jackson (2010), Golub and Jackson (2012), Smith and Sørensen (2000), Gale and Kariv (2003) and Banerjee and Fudenberg (2004).[2] They are, thus, not suitable for explaining the persistence of disagreement.

The purpose of this paper is to investigate intuitive processes allowing for persistent disagreements. To do so we study the dynamics of attitude formation following DeGroot (1974), a parsimonious and widely used framework in which individuals use the rule of thumb of averaging others' attitudes to develop their own over time. As pointed out by Ellison and Fudenberg (1993), Acemoglu and Ozdaglar (2011) and Golub and Jackson (2012), the computational requirements imposed on agents that behave as Bayesian, updating their priors regarding the true state of the nature according to all relevant information, have placed rules of thumb as a useful and powerful alternative for the understanding of learning and communication processes. This reason seems to be borne out by recent evidence supporting, in particular, averaging models as a consistent description of individuals' updating behavior. For instance, experimental results in Chandrasekhar et al. (2012) and Grimm and Mengel (2014) favor a DeGroot procedure over a Bayesian one.[3] Furthermore, we capture the natural idea that, in general, individuals form and update their attitudes regarding a given issue through own experiences, by observing others' actions and by communicating with others about their attitudes and behavior. That is, learning is social and takes place within the individuals' social network.

But in the canonical DeGroot procedure, which considers a time independent averaging rule, consensus is, almost always, the eventual outcome. Disagreement only persists in the extreme situation in which there are groups of individuals completely

[1] Detailed information is available at http://www.people-press.org/2012/06/04/section-2-demographics-and-american-values.

[2] We discuss notable exceptions at the end of this section.

[3] See also Corazzini et al. (2012) and Brandts et al. (2014).

ignoring each other.[4] In fact, communication models based on DeGroot procedure, as Golub and Jackson (2010) and Golub and Jackson (2012), work with strongly connected (and time independent) network structures in which individuals incorporate everyone else's attitudes, thus always deriving consensus results. In particular, Golub and Jackson (2012) discuss the effects of homophily, the robust tendency of individuals to associate disproportionately with similar others, on the speed of convergence to consensus, an eventual outcome that is never precluded no matter the level of homophily.[5] In contrast with these papers, we propose a version of DeGroot procedure in which we incorporate the natural idea that the intensity of individual interactions varies over time.[6]

We thus explore a particular mechanism allowing for the co-evolution of homophily in attributes and attitudes. In our approach the type of an individual is defined as a subset of attributes that are assumed to be exogenous. We also assume that individuals are homophilous with respect to them, that is, common attributes between any pair of individuals, for instance being of the same gender, guarantee that they relate to each other. Furthermore, the intensity of this relation is governed by the salience of attributes. The salience of an attribute is given by the difference in attitudes between individuals possessing and lacking it. The more salient an attribute the higher the attention that, on the basis of it, individuals pay among themselves, that is, the more homophilous towards this attribute individuals are. As a consequence, the lower the attention that these individuals pay to others not sharing this attribute with them. We allow differences in attitudes to feedback the homophilous behavior overtime, promoting, as stated, the co-evolution between homophily and attitudes.

There is a large literature in the context of consumer choice supporting the idea that individuals focus in aspects in which their alternatives differ more, that is, in aspects that are salient. For instance, in Bordalo et al. (2013) consumers' purchasing decisions are driven by either the price or the quality of products, depending on which aspect is furthest from prices and qualities of an average bundle.[7] There is also evidence suggesting a negative relationship between differences in attitudes and interactions among individuals. Specifically, Suanet and Van de Vijver (2009) study the relationship between perceived cultural distance, that is, individual reports of discrepancies in attitudes and values between the home and the host culture, and the acculturation of foreign students in Russia. They find a positive (respectively negative) relationship between perceived cultural distance and interactions of for-

[4] See Jackson (2008), chapter 8, for two characterizations of consensus.

[5] See McGuire et al. (1978) for a survey on homophily. In Golub and Jackson (2012) homophily is technically defined in as the second largest eigenvalue of the matrix of linking densities among types.

[6] See Kossinets and Watts (2006).

[7] See also Kőszegi and Szeidl (2013) and the references therein.

eign students with co-nationals (respectively host nationals). Support to the idea that repulsion to others with dissimilar attitudes is the main mechanism shaping homophily can be found in Rosenbaum (1986) and Singh and Ho (2000).

With this model at hand we answer the following questions:

Q1: Under which conditions does attributes' salience preclude consensus, and therefore, promote the persistence of disagreement?

Q2: How does eventual disagreement look like? In particular, which ones are the types exhibiting different attitudes?

Q3: How does salience relate to the speed of convergence to the eventual situation in which disagreement persist?

Our results are as follows. We find that disagreement persists if and only if there is, initially, a unique attribute for which the difference in average initial attitudes is the highest, that is, a unique most salient attribute. When this is the case, this attribute becomes increasingly salient, receiving growing attention over time in detriment of the remaining attributes. In other words, the ties among individuals sharing it, will progressively gain strength in detriment of the ties based on the remaining shared attributes. As a result, the society appears eventually divided in two groups of thinking, according to whether individuals possess or lack the initially most salient attribute. Thus disagreement across this attribute, that is, the difference in average eventual attitudes between the groups of individuals possessing and lacking it, persists while the differences in average eventual attitudes associated to the possession and lack of the remaining attributes vanish. This result however corresponds to a situation in which the value of homophily, i.e: the attention that individuals pay to others on the basis of shared attributes, is linked to differences in opinions by a particular functional form. In the extensions we offer a more general representation of homophily. Thus, as a more general statement, we conclude that disagreement persists when the dynamic of segregation in attention is fast enough. By fast enough we mean that the force that drives individuals to develop strong ties with specific individuals dominates the one that pushes them to pay attention to everyone else. Thus, the complete mixing of attitudes is precluded.

The process of segregation in attention can be understood as one by which individuals act (as if) they construct their identity. That is, initially confronted with several attributes at which they may stick, they progressively focus in only one of them, developing their relations and attitudes according to it. Our results can also

be understood as a theory of polarization, since two groups of individuals are in persistent disagreement.

With respect to the properties of disagreement, we find how the difference in average eventual attitudes between the groups of individuals possessing and lacking the initially most salient attribute is a proportion of the difference in average initial attitudes between these two groups.

With respect to the speed of convergence we find that, everything else equal, the higher the difference in average initial attitudes related to the initially most salient attribute or the lower the difference in average initial attitudes related to any other attribute, the higher the magnitude of disagreement and the quicker the convergence to a situation in which it persists.

Our work is related to previous papers discussing disagreement. Specifically, Krause (2000) and Hegselmann and Krause (2002) study disagreement in a model of bounded confidence in which individuals only consider others' attitudes when they are sufficiently close to their own. There are, at least, two differences with their approach. The first one is that while our primary source of attention are individual types as well as their attitudes, they directly focus on similarity in attitudes and do not explicitly model homophily in exogenous attributes. The second one is that they assume that the attitudes of the peers finally considered by any individual, matter at the same extent. This is not generally true in our case because homophily depends precisely on types. In Acemoglu et al. (2013) disagreement persists because of the presence of stubborn agents, interpreted as leaders or media sources, that never change their attitudes. We do not model the presence of such agents.

The rest of the paper is organized as follows. Section 2 formally presents the model. Section 3 derives the condition for disagreement to persist and provides its properties. Section 4 deals with the speed of convergence. Section 5 concludes. Section 6 discusses extensions. Section 7 contains the technical proofs.

## 2 Preliminaries

Let $I = \{1, 2, \ldots, n\}$ be a finite set of attributes. The type $A$ of an individual is defined by the attributes possessed by this individual, that is, $A \subseteq I$. Thus, there are $2^n$ types. Given two types $A$ and $B$, we say that they are $i$-similar whenever attribute $i$ is either present or absent in these two types. Otherwise, we say that they are $i$-dissimilar. Let us denote by $A^c$ the complementary set of $A$. We finally define $I(AB)$ as all the attributes for which $A$ and $B$ are similar, i.e., $I(AB) = (A \cap B) \cup (A^c \cap B^c)$. Notice that attributes are dichotomous, that is, either a type possesses an attribute

or lacks it.[8]

The (column) vector of attitudes at time $t \in \mathbb{Z}_+$ is denoted by $a_t \in [-1,1]^{2^n}$, where the component relative to type $A$ is $a_t^A$. The average attitude across all types is denoted $\overline{a}_t$ and the average attitude across all types possessing (respectively lacking) attribute $i$ is denoted $\overline{a}_t[i]$ (respectively $\overline{a}_t[-i]$).[9] Without loss of generality let us normalize the average initial attitude to zero, that is, $\overline{a}_0 = 0$.

Attitudes evolve according to an average-based process similar to DeGroot (1974). Namely, each of the components of $a_{t+1}$, that is, attitudes at $t+1$ is a weighted average of each of the components of $a_t$, that is, attitudes at $t$. Let $W_t$ be the $2^n \times 2^n$ matrix of weights describing the updating of attitudes from time $t$ to time $t+1$. We have that:

$$a_{t+1} = W_t a_t.^{10} \tag{1}$$

Notice that every entry of $W_t$ is the weight that type $A$ assigns to type $B$. Let $w_t^{A,B}$ denote this weight. As in Golub and Jackson (2012), individuals are homophilous, a behavior that can be captured as follows: every attribute $i$ has a non-negative value $\alpha_t^i$ and the weight that type $A$ assigns to type $B$, is the sum of values of the attributes they share, that is, $w_t^{A,B} = \sum_{i \in I(AB)} \alpha_t^i$. For normalization purposes we set $\sum_i \alpha_t^i = (2^{n-1})^{-1}$. That is the right normalization because a type $A$ is $i$-similar to exactly $2^{n-1}$ types. Then, $\sum_B w_t^{A,B} = 2^{n-1} \sum_i \alpha_t^i = 1$. In this paper, we study the case in which the value of homophily, namely, the magnitude of $\alpha_t^i$, co-evolves with attitudes. In particular, it depends, at every $t$, on the difference in average attitudes between individuals possessing attribute $i$ and individuals lacking it, that is, $\Delta_t[i] = \overline{a}_t[i] - \overline{a}_t[-i]$ (the salience of attribute $i$ at time $t$). We assume, without loss of generality, that the differences in average initial attitudes are non-negative and such that $\Delta_0[1] \geq \Delta_0[2] \geq \cdots \geq \Delta_0[n] \geq 0$.[11]

We link homophily and salience by the well-known Luce form, that is:

$$\alpha_t^i = \frac{1}{2^{n-1}} \frac{\Delta_t[i]}{\sum_j \Delta_t[j]}. \tag{2}$$

We endow this functional form with the following interpretation: the attention that individuals pay to other when they share a given attribute, depends on how big the differences in attitudes associated to this attribute are in relation to the differences associated to the remaining attributes.

[8] See Schelling (1969) for a discussion of this assumption. Also, as McGuire et al. (1978) point out, the distinction in terms of social distance appears to be of the type same versus different, and not on any more elaborated forms of stratification.

[9] Formally, $\overline{a}_t = (2^n)^{-1} \sum_A a_t^A$, $\overline{a}_t[i] = (2^{n-1})^{-1} \sum_{A: i \in A} a_t^A$ and $\overline{a}_t[-i] = (2^{n-1})^{-1} \sum_{A: i \notin A} a_t^A$.

[10] With this updating rule, $\overline{a}_0 = 0$ implies that at every time $t$, $\overline{a}_t = 0$. See section 6.

[11] In fact, they preserve this order and remain non-negative over time.

The following example illustrates the notation above:

**Example 1.** Consider the case in which types come from the combination of two attributes. Thus, there are four types, namely, $\{1,2\}$ $\{1\}$, $\{2\}$ and $\{\emptyset\}$. The structure of the $4 \times 4$ matrix of weights at an arbitrary time $t$ is:

$$\mathrm{W}_t = \begin{array}{c} \begin{matrix} \{1,2\} & \{1\} & \{2\} & \{\emptyset\} \end{matrix} \\ \begin{bmatrix} \alpha_t^1 + \alpha_t^2 & \alpha_t^1 & \alpha_t^2 & 0 \\ \alpha_t^1 & \alpha_t^1 + \alpha_t^2 & 0 & \alpha_t^2 \\ \alpha_t^2 & 0 & \alpha_t^1 + \alpha_t^2 & \alpha_t^1 \\ 0 & \alpha_t^2 & \alpha_t^1 & \alpha_t^1 + \alpha_t^2 \end{bmatrix} \end{array} \begin{matrix} \{1,2\} \\ \{1\} \\ \{2\} \\ \{\emptyset\} \end{matrix}$$

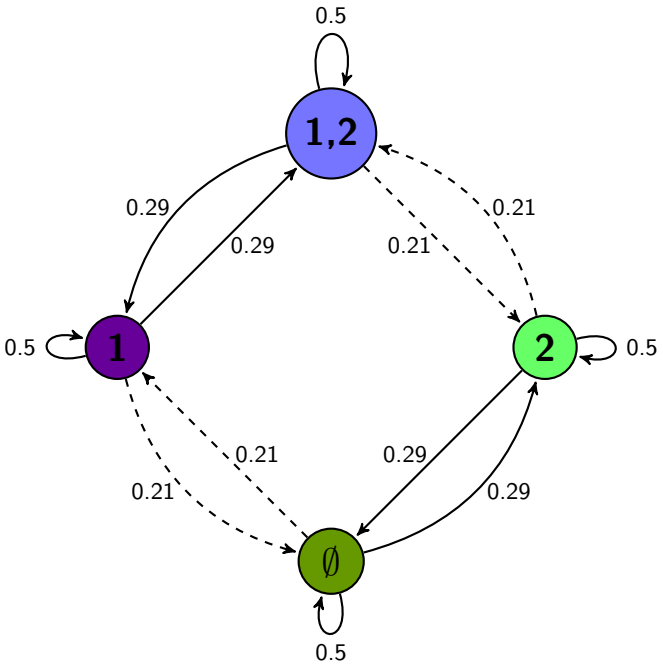
To make clear how homophily in exogenous attributes determines the structure of attention, let us consider type $\{2\}$. It is 1-similar to type $\{\emptyset\}$ and 2-similar to type $\{1,2\}$. Thus, it pays a non-negative amount of attention to both types. Also, it pays more attention to these types than to type $\{1\}$, with whom it does not share any attribute. Consider also type $\{1,2\}$. It is 1-similar to type $\{1\}$ and 2-similar to type $\{2\}$. Thus, it pays a non-negative amount of attention attention to them. It also pays zero attention to type $\{\emptyset\}$, with whom it does not share any attribute.

Suppose that initial attitudes are: $a_0^{\{1,2\}} = 0.8$, $a_0^{\{1\}} = 0.2$, $a_0^{\{2\}} = -0.05$ and $a_0^{\{\emptyset\}} = -0.95$. Thus, the differences in average initial attitudes associated to attribute 1 and 2 are $\Delta_0[1] = 0.5(0.8 + 0.2) - 0.5(-0.05 - 0.95) = 1$ and $\Delta_0[2] = 0.5(0.8 - 0.05) - 0.5(0.2 - 0.95) = 0.75$, respectively. The initial value of homophily, according to expression (2), is $\alpha_0^1 = 0.29$ and $\alpha_0^2 = 0.21$, for attributes 1 and 2, respectively. The interaction matrix above thus becomes:

$$\mathrm{W}_0 = \begin{array}{c} \begin{matrix} \{1,2\} & \{1\} & \{2\} & \{\emptyset\} \end{matrix} \\ \begin{bmatrix} 0.5 & 0.29 & 0.21 & 0 \\ 0.29 & 0.5 & 0 & 0.21 \\ 0.21 & 0 & 0.5 & 0.29 \\ 0 & 0.21 & 0.29 & 0.5 \end{bmatrix} \end{array} \begin{matrix} \{1,2\} \\ \{1\} \\ \{2\} \\ \{\emptyset\} \end{matrix}$$

In the following figure we depict this interaction structure. For this purpose, let us color types as follows: types possessing attribute 1 are blue and those lacking it are green. Types possessing attribute 2 are white while types lacking it are red. Thus, $\{1,2\}$ is a mixture of blue and white, $\{2\}$ is a mixture of green and white, $\{\emptyset\}$ is a mixture of green and red and $\{1\}$ is a mixture of blue and red. Interpret every row of the matrix above as individuals having one unit of time to devote to others. Thick lines then represent more intense relations than dashed lines:

Figure 1. Depicting initial interactions




We use this structure as a running example in subsequent sections.

## 3 The persistence and properties of disagreement

To analyze under which conditions disagreement persists, notice that expression (1) can be solved recursively to get $a_{t+1} = W^T a_0$ where $W^T = \prod_{t=0}^{T} W_{T-t}$. Thus one can express attitudes at an arbitrary point in time $t$ as a function of initial ones. Notice that if the matrix describing point-wise interactions, as the one in example 1, was constant over time, consensus will eventually emerge. The reason is that individuals would then be able to incorporate, directly or indirectly, everyone else's attitudes at every point in time. Formally, the (constant) matrix of interactions is strongly connected and aperiodic and thus consensus is guaranteed.[12] In our framework, this is equivalent to establish that all eventual attitudes will be equal to zero, that is, $a_\infty = \lim_{t \longrightarrow \infty} a_{t+1} = 0$.

Allowing for the intensity of the attention that individuals pay to each other to vary over time, opens the possibility of persistent disagreement. Clearly, the existence and properties of eventual attitudes can be understood by investigating the existence and properties of the limiting product of time dependent interactions matrices. We denote this limit by $W^\infty$, where, $W^\infty = \lim_{T \longrightarrow \infty} W^T$.

In order to state our main result, let us discuss the concept of *Dobrushin* coefficient of ergodicity. Ergodicity coefficients provide information about the extent to

[12] See Jackson (2008), chapter 8.

which all the rows of a matrix are equal.[13] The Dobrushin coefficient of ergodicity of a matrix $M$ is defined as:

$$\tau(M) = \frac{1}{2} \max_{ij} \sum_{k} |m_{ik} - m_{jk}|. \tag{3}$$

It lies between zero and one and is different from zero if and only if the rows of $M$ are not the same. Now, we present our main result, that describes the form and extent of disagreement in eventual attitudes. It is as follows:

**Theorem 1.** *For every configuration of initial attitudes, eventual ones always exist. They exhibit disagreement if and only if attribute* 1 *is, initially, the unique most salient (that is, if and only if* $\Delta_0[1] > \Delta_0[2]$*). In this case, eventual attitudes are such that, for every type A:*

$$|a^A_\infty| = \frac{1}{2}\tau(W^\infty)\Delta_0[1] \tag{4}$$

*where* $\tau(W^\infty) \in (0, 1]$*. Furthermore,* $a^A_\infty > 0$ *if and only if* $1 \in A$*.*

Several aspects merit further attention. First, disagreement is almost the unique outcome of this process.[14] When there is, initially, a unique most salient attribute, it gains increasing attention in detriment of the attention paid to the remaining attributes. Thus, eventual homophily is based upon one, and only one, dimension. Consensus would emerge if and only if there were, at least, two initially most salient attributes. In the extreme case in which all differences in average initial attitudes were equal, all attributes will deserve the same initial homophily, which will be also constant over time. Specifically, every attribute $i$ will be receiving always same amount of attention, $\alpha^i_t = (2^{n-1}n)^{-1}$.[15]

Second, disagreement persists between two groups. Specifically, types possessing attribute 1 have the same eventual attitudes and the same happens for types lacking attribute 1. The eventual attitudes between these two groups are different.

Third, eventual disagreement, measured as the difference in average eventual attitudes between the groups of types possessing and lacking attribute 1, is a proportion of the difference in average initial attitudes between the groups of types possessing and lacking attribute 1. This proportion is exactly given by the ergodicity coefficient of the infinite product of the point-wise matrices of weights.[16] The ergodicity coef-

[13] See Stachurski (2009) for a reference on the Dobrushin coefficient in the study of economic models with a Markovian structure. See also Ipsen and Selee (2011) and Chatterjee and Seneta (1977) for the study of convergence properties of inhomogeneous Markov chains by means of ergodicity coefficients.

[14] Since differences in average attitudes are real numbers, they are generally different. Also the results remain the same if we consider that initial attitudes are defined over the entire real line instead of belonging to $[-1, 1]$.

[15] When all differences in average initial attitudes are equal to zero, expression (2) is not defined. We set $\alpha^i_t = (2^{n-1}n)^{-1}$ in this case.

[16] Let $\overline{a}_\infty[i] = \lim_{t\longrightarrow\infty} \overline{a}_t[i]$ and $\overline{a}_\infty[-i] = \lim_{t\longrightarrow\infty} \overline{a}_t[-i]$. Since there are $2^{n-1}$ types possessing (respectively lacking) attribute 1, $\overline{a}_\infty[1] - \overline{a}_\infty[-1] = 2^{-1}(\tau(W^\infty)\Delta_0[1] + \tau(W^\infty)\Delta_0[1]) = \tau(W^\infty)\Delta_0[1]$.

ficient then characterizes the *distance to consensus in the long-run*. It is important to highlight that this coefficient is fully determined by the initial configurations of attitudes.[17] Also, the difference in average eventual attitudes of types possessing and lacking any attribute different from 1, is zero.[18]

The case with two attributes is pretty informative. In it, the deviation from consensus in the long-run is given by the ratio of the differences in average initial attitudes between attributes 2 and 1. Specifically, $\tau(W^\infty) = 1 - \Delta_0[2]/\Delta_0[1]$. The following example illustrates the results:

**Example 2.** Consider that $\Delta_0[1] = 1$ and $\Delta_0[2] = 0.75$, as in example 1. The entries in the interaction matrices evolve as follows:

$$W_0 = \begin{bmatrix} 0.5 & 0.29 & 0.21 & 0 \\ 0.29 & 0.5 & 0 & 0.21 \\ 0.21 & 0 & 0.5 & 0.29 \\ 0 & 0.21 & 0.29 & 0.5 \end{bmatrix}, \; W_1 = \begin{bmatrix} 0.5 & 0.32 & 0.18 & 0 \\ 0.32 & 0.5 & 0 & 0.18 \\ 0.18 & 0 & 0.5 & 0.32 \\ 0 & 0.18 & 0.32 & 0.5 \end{bmatrix}, ..., \lim_{t\to\infty} W_t = \begin{bmatrix} 0.5 & 0.5 & 0 & 0 \\ 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 \\ 0 & 0 & 0.5 & 0.5 \end{bmatrix}.$$

$$\text{Also, } W^\infty = \begin{bmatrix} 0.313 & 0.313 & 0.187 & 0.187 \\ 0.313 & 0.313 & 0.187 & 0.187 \\ 0.187 & 0.187 & 0.312 & 0.312 \\ 0.187 & 0.187 & 0.312 & 0.312 \end{bmatrix} \text{ and } W^\infty \text{ times } a_0 = \begin{bmatrix} 0.8 \\ 0.2 \\ -0.05 \\ -0.95 \end{bmatrix} \text{ is } a_\infty = \begin{bmatrix} 0.126 \\ 0.126 \\ -0.126 \\ -0.126 \end{bmatrix}.$$

Notice how on one hand, types $\{1, 2\}$ and $\{1\}$ and on the other hand, types $\{2\}$ and $\{\emptyset\}$ hold the same eventual attitudes, which are different between these two groups. In this case $\tau(W^\infty) = 0.25$.

Fourth, it follows that Theorem 1 goes through for the the general Luce form, $\gamma_t^i = \dfrac{\Delta_t[i]^\delta}{\sum_j \Delta_t[j]^\delta}$ where $\delta \in (0, \infty)$.[19] The literature, for instance Chen et al. (1997), interprets $\delta$ as a rationality parameter. In our case $\delta$ reflects the extent to which the difference in attitudes across attribute 1 is exacerbated. For $\delta \in (0, 1)$, this difference becomes less important than before, when $\delta = 1$, but disagreement still persists, being its magnitude also smaller. When exactly $\delta = 0$, the difference in attitudes associated to attribute 1 is as important as the difference associated to any other attribute, regardless of their magnitude. Notice that in this case $\gamma_t^i = n^{-1}$ for any attribute $i$ and at every time $t$, thus individuals always pay attention $\alpha_t^i = (2^{n-1}n)^{-1}$ to every attribute and consensus eventually emerges. Finally, when $\delta \in (1, \infty)$ the difference in attitudes associated to attribute 1 is exacerbated with respect to the

[17] See step 8 in the proof of Theorem 1.

[18] That is so because within the $2^{n-1}$ types possessing (respectively lacking) attribute 1, there are $2^{n-2}$ possessing (respectively lacking) any other attribute $i > 1$, thus the average eventual attitudes of i-similar types are the same and the difference between them cancels out.

[19] In this case $\alpha_t^i = (2^{n-1})^{-1}\gamma_t^i$.

case in which $\delta = 1$, thus disagreement is the eventual outcome and its magnitude increases.

### 3.1 Segregation in interactions and disagreement

As stated, 1-similar types eventually interact exclusively among themselves. They reach this situation by weakening their interactions with 1-dissimilar types.

To summarize this interaction information, we derive here the Spectral Segregation Index proposed by Echenique and Fryer (2007), for attribute $i$ at time $t$, henceforth $SSI_t^i$.[20] Being based on the nature of individual interactions, it is particularly suitable in our framework. Other indexes measuring segregation, as the Dissimilarity or the Isolation Index, are based on partitions (census) of a physical unit (a city). In our case individuals are not partitioned into physical units, thus, we do not interpret our interaction process in their terms.[21]

Before stating the result let us stress the fact that interactions within the groups of types possessing and lacking any attribute $i$, follow the same pattern at every time $t$, that is, these groups of individuals divide their time in the same way. This can be seen using the symmetric interaction matrix in example 1. Interactions among types possessing attribute 1, collapsed in the submatrix composed by $\{1,2\}$ and $\{1\}$, take the same form as those of types lacking it, collapsed in the submatrix composed by $\{2\}$ and $\{\emptyset\}$. The same is true for attribute 2. Thus, the $SSI_t^i$ describes interactions within both groups. Before the result let $\lambda_t^i = \dfrac{\Delta_t[i]}{\sum_j \Delta_t[j]}$. We now present the properties of the index:

**Proposition 1.** *At every time $t$ and for every attribute $i$, $SSI_t^i = \dfrac{1+\lambda_t^i}{2}$. Furthermore:*

*(i)* $SSI_0^1 > \dfrac{n+1}{2n}$, $SSI_{t+1}^1 > SSI_t^1$ *and* $\lim_{t\longrightarrow\infty} SSI_t^1 = 1$.

*(ii)* $SSI_0^i \leq \dfrac{n+1}{2n}$, $SSI_{t+1}^1 < SSI_t^i$ *and* $\lim_{t\longrightarrow\infty} SSI_t^i = 0.5$ *for every* $i > 1$.[22]

Due to our assumptions, the groups of 1-similar types have more intense overall relations than the groups of $i$-similar types for attributes $i > 1$. That comes from the fact that attribute 1 is always the most salient. Also, interactions among 1-similar types are gradually intensified on the basis of this attribute. We eventually observe the extreme situation in which individuals only interact with others if they are

[20] The Spectral Segregation Index has a static nature, we just repeat its computation at every $t$.

[21] See Echenique and Fryer (2007) for a discussion.

[22] The Spectral Index of Segregation at time $t$ is computed by looking only at interactions among $i$-similar types at that time. It is the largest eigenvalue of the matrix describing these interactions. Also, this result refers to the case in which $\lambda_t^i > 0$ for every attribute $i$. The results are the same when $\lambda_t^i = 0$ for some/all attributes $i > 1$. We address this case in the proof of this proposition.

1-similar. Thus, two disconnected groups, the one composed by types possessing attribute 1 and the one composed by types lacking it, emerge. In this case the segregation of 1-similar types ends up being maximal. In others words, the limiting value of the Spectral Segregation Index is equal to 1.

Attributes $i > 1$ become gradually irrelevant in shaping interactions and thus segregation according to them decreases over time. Thus, eventually individuals evenly split their time between others that are similar to them in these attributes and those that are different to them, that is why the limiting value of the index for attributes $i > 1$ is exactly ones half.

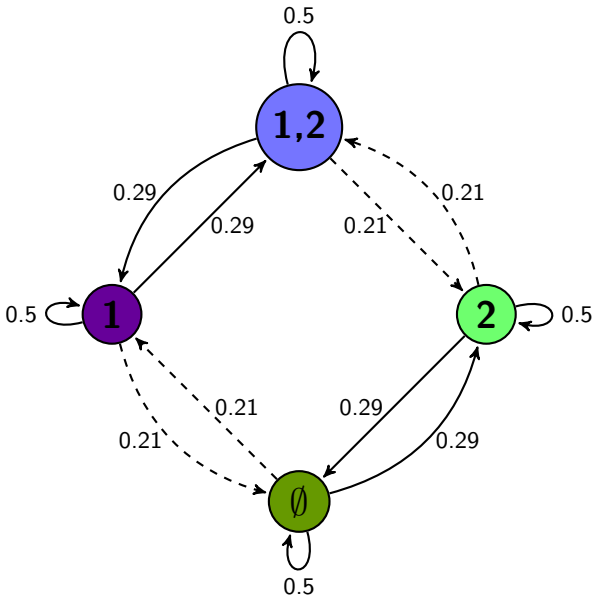
Finally, it is worth mentioning the relationship between the segregation of a group of $i$-similar types and the segregation of its members. By definition, the Spectral Segregation Index is the average of individual segregation indexes. Individual segregation indexes are computed by distributing the overall Spectral Segregation Index among the members of the group. Following Echenique and Fryer (2007), this distribution is done according to the entries of the eigenvector associated to the largest eigenvalue of the matrix describing interactions of $i$-similar types. In our case this eigenvector is composed by ones. It is then the case that, at every point in time $t$ and for every attribute $i$, the level of segregation of every type is the same, and equal to the overall Spectral Segregation Index. Intuitively, every type pays the same total amount of attention to $i$-similar types. As a consequence, it also pays the same total amount of attention to $i$-dissimilar types. In a nutshell, every type segregates its interactions at the same extent and thus equally contributes to the segregation of its group.

Another measure for the intensity of interactions is the so called Network Cohesion, proposed by Cavalcanti et al. (2012). Given a network, represented by a matrix of interactions, Network Cohesion measures how uneven relations are. In other words, how uniform or fragmented a network is. At every time $t$, Network Cohesion, henceforth $C_t$, can be computed as one minus the largest eigenvalue of the matrix of interactions $W_t$. It lies between zero and one, where zero and one represent the lowest and the largest cohesion, respectively. In our framework, $\lambda_t^1$, is indeed the largest eigenvalue of the matrix of interactions $W_t$, thus we have that $C_t = 1 - \lambda_t^1$. Network Cohesion decreases overtime and becomes eventually zero, reflecting the eventual emergence of two disconnected groups of individuals. For instance, in example 2 above we have that $\lambda_0^1 = 0.58$, $\lambda_1^1 = 0.64$ and $\lim_{t\to\infty} \lambda_t^1 = 1$, thus $C_0 = 0.42$, $C_1 = 0.36$ and $\lim_{t\to\infty} C_t = 0$.

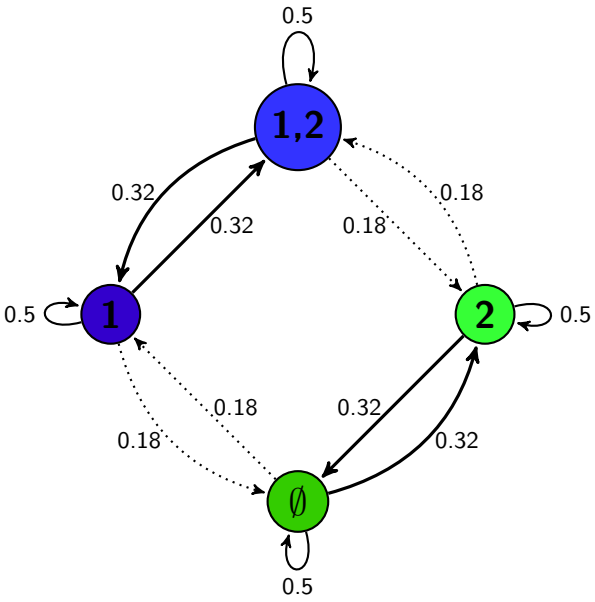
In the following figure we illustrate the evolution of interactions as time goes by and compute the Spectral Segregation Index. Observe how 1-similar types eventually interact exclusively among themselves. Observe also how types possessing (respec-

tively lacking) attribute 2, equally split their unit of attention between themselves and others lacking (respectively possessing) attribute 2.

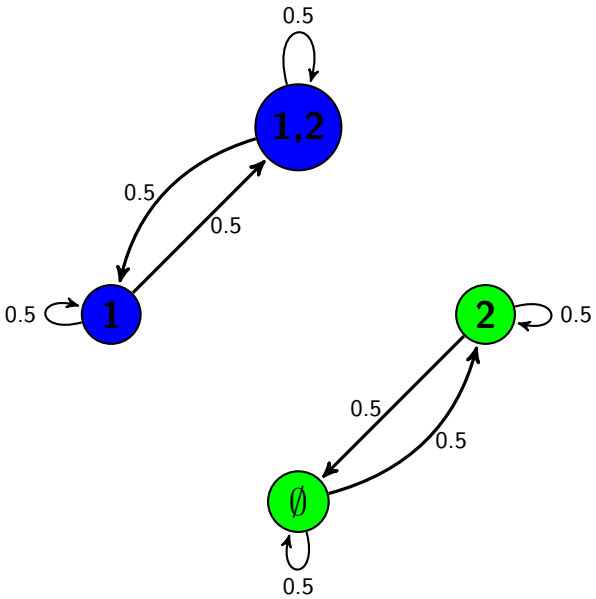
Figure 2. Segregation in interactions




(a) $SSI_0^1 = 0.79$, $SSI_0^2 = 0.71$




(b) $SSI_1^1 = 0.82$, $SSI_1^2 = 0.68$




(c) $\lim_{t\to\infty} SSI_t^1 = 1$, $\lim_{t\to\infty} SSI_t^2 = 0.5$

## 4 Speed of convergence

We focus here on the role of salience in determining the speed of convergence to the eventual disagreement. One reason as to why is relevant to study the speed of convergence is because disagreements might indeed have pernicious consequences. In the presence of a policy intervention aiming to recover consensus, it might be then important to know the timing for its implementation.

As Alesina and Tabellini (1990) point out, discrepancies between policymakers in ideological views about social welfare, specifically regarding the desired composition of government spending in public goods, might cause the accumulation of inefficient levels of public debt. Also, Voss et al. (2006) show how the organizational success of non-profit professional theatres was affected by the divergent views of their leaders regarding the values that should drive the organizations' behavior and Andreoni and Mylovanov (2012) discuss how, among other consequences, disagreement might create inefficient delays in bargaining. In a broad sense, Friedkin and Johnsen (1999) state that there might be difficulties in arriving at agreed decisions when individuals have fixed discrepant preferences.

The speed of convergence to the eventual disagreement is determined by the relation between the difference in average initial attitudes associated to attribute 1 and the ones associated to the remaining attributes. In other words, the initial relative salience of attribute 1 determines how long it takes for individuals to become sufficiently homophilous with respect to it. Recall that the expression that links homophily based on attribute 1 and the salience of this attribute is given by $\lambda_t^1 = \frac{\Delta_t[1]}{\sum_i \Delta_t[i]}$. As previously discussed, eventually 1-similar individuals interact exclusively among themselves which formally means that $\lim_{t \longrightarrow \infty} \lambda_t^1 = 1$. Thus, when we are sufficiently close to this interaction pattern, we can state that we are sufficiently close to the equilibrium in which disagreement persists. It turns out that every time $t$, $\lambda_t^1$ is the second largest eigenvalue of the point-wise matrix of interactions $W_t$. As deeply discussed in Golub and Jackson (2010) and Golub and Jackson (2012), the second largest eigenvalue of a stochastic matrix plays an important role in the analysis of the speed of convergence.

Our aim in this section is precisely to characterize the time it takes for individuals to become homophilous exclusively with respect to attribute 1, that is, the minimum time it takes for $\lambda_t^1$ to be above an $\epsilon > 0$ distance of its limit. For this purpose we formally define this minimum time as:

$$T_\epsilon = min\{t : \lambda_t^1 \geq 1 - \epsilon\}. \tag{5}$$

In what follows we describe the properties of $T_\epsilon$, specifically we define its bounds and analyze how it behaves in response to changes in the initial relative salience of attribute 1, that is, to changes in the relation between the difference in average initial attitudes associated to attribute 1 and the ones associated to the remaining attributes. For this purpose, we focus on the case in which all differences in average initial attitudes are strictly positive. We also consider the case in which the relative salience of attribute 1 is modified by altering the differences in average initial attitudes, for just one attribute at a time.[23]

Before stating the result let $r_0^i = \Delta_0[i]/\Delta_0[1]$ for every attribute $i > 1$. This ratio captures the initial relative salience of attribute 1 with respect to any other attribute $i > 1$. The smaller this ratio the more salient attribute 1 is with respect to any other attribute $i > 1$. Let us specifically set $\underline{r}_0 = \Delta_0[n]/\Delta_0[1]$ and $\overline{r}_0 = \Delta_0[2]/\Delta_0[1]$. These two ratios represent extreme cases. Specifically, $\underline{r}_0$ considers the difference in average initial attitudes associated to attribute $n$, which is the smallest one. In contrast, $\overline{r}_0$ considers the difference in average initial attitudes associated to attribute 2, which is the second highest one. Let us set $\overline{\lambda}_t^1 = [1+(n-1)(\underline{r}_0)^{2^t}]^{-1}$ and define $T_\epsilon^{min} = min\{t : \overline{\lambda}_t^1 \geq 1-\epsilon\}$ accordingly. Similarly, let $\underline{\lambda}_t^1 = [1+(n-1)(\overline{r}_0)^{2^t}]^{-1}$ and $T_\epsilon^{max} = min\{t : \underline{\lambda}_t^1 \geq 1-\epsilon\}$. Notice that both, $\overline{\lambda}_t^1$ and $\underline{\lambda}_t^1$, are constructed from the expression $\lambda_t^1 = [1 + \sum_{i>1}(r_0^i)^{2^t}]^{-1}$, by substituting all differences in average initial attitudes, by the smallest and second highest difference, respectively.[24] We now present the result:

**Proposition 2.** *For every configuration of initial attitudes such that disagreement persists, $T_\epsilon$ is non-increasing in the initial relative salience of attribute 1. Furthermore, $T_\epsilon \in [T_\epsilon^{min}, T_\epsilon^{max}]$.*

It directly follows that, everything else equal, the higher the difference in average initial attitudes associated to attribute 1, the higher the overall attention within the groups of 1-similar types. It is also the case that the lower the difference in average initial attitudes associated to an attribute $i > 1$, the higher the overall attention within the groups of 1-similar types. In particular, attribute 1 becomes relatively more salient than this other attribute $i > 1$, which is now a weaker competitor for attention. In general when attribute 1 is fairly salient, individuals exhibit high homophily with respect to 1-similar others and form completely inward-looking groups relatively fast.

[23] That is, for one attribute $i$, we alter $\Delta_0[i]$ such that $\Delta_0[1] > \Delta_0[2] \geq \cdots \geq \Delta_0[n] \geq 0$ is preserved in order, and in magnitude for differences associated to the remaining attributes $j \neq i$. In fact, when we can decrease or increase any $\Delta_0[i]$ by decreasing or increasing, in the same magnitude, initial attitudes of both, the type that possesses all attributes and the type that only possesses the considered attribute $i$, differences associated to attributes $j \neq i$, keep unaltered. The decrease or increase has to be such that the order above is preserved.

[24] Given the order of initial differences, when $\Delta_0[i] = 0$ for some attribute $i \leq n$ then, $\overline{\lambda}_t^1 = 1$. In this case $T_\epsilon^{min} = 0$. Similarly, when $\Delta_0[2] = 0$ then $\underline{\lambda}_t^1 = 1$. In this case $T_\epsilon^{max} = 0$. In this last case $\lambda_0^1 = 1$ and the equilibrium is reached at $t = 1$.

Not only the speed of convergence but the magnitude of disagreement is also sensitive to the aforementioned changes in differences in attitudes. To see this consider the eventual attitudes in expression (4) and notice that we can rewrite the ergodicity coefficient as $\tau(W^{\infty}) = \lim_{T\to\infty} \prod_{t=0}^{T}[1+\overline{r}_0^{2^t}+...+\underline{r}_0^{2^t}]^{-1}$. It is immediate that the proposed changes in the differences in attitudes decrease the ratios in the denominator of the expression above, making the elements of this product (and hence its limit) higher than before.

It is also worth mentioning how it is enough to focus on the evolution of the homophily value associated to attribute 1 to describe the minimum time of convergence for the system as a whole. The reason is that this homophily value is always further away from 1, its limiting value, than any of the homophily values associated to the remaining attributes is from 0, its limiting value. Then, the time it takes for it to be sufficiently close to one, is at least the same as the time it takes for the remaining homophily values to be sufficiently close to zero.[25]

We finally discuss how the configuration of initial attitudes matters in determining the speed of convergence. Consider the extreme case in which the difference in average initial attitudes associated to attribute 1 is fairly similar to the differences associated to the remaining attributes, for instance, $\Delta_0[1] \simeq \Delta_0[2] = \cdots = \Delta_0[n]$. In this case the initial relative salience of attribute 1 is fairly small and it would take a while for individuals to gradually redirect their homophilous behavior towards attribute 1. The time to reach the equilibrium would be considerably high in this case. The other extreme situation is such that the difference in average initial attitudes associated to attribute 1 is, by far, the highest one, for instance, $\Delta_0[1] >> \Delta_0[2] = \cdots = \Delta_0[n] \simeq 0$. Being the relative salience of attribute 1 fairly high, individuals would quickly conclude that the possession or lack of this attribute clearly defines two groups in society, or in other words, that this attribute is explanatory for social differences. Thus, it would not take much time for them to become homophilous exclusively with respect to it. The equilibrium will be reached much more faster than before. When the differences in attitudes associated to all attributes $i > 1$ are zero, the equilibrium is reached at $t = 1$.

## 5 Conclusions

On the basis of the observation that disagreement in attitudes is a common phenomenon, we propose a model of attitude evolution able to capture its persistence. In our approach individuals exhibit homophily and the attention they pay to similar others varies over time. Specifically, homophily co-evolves with attitudes governed by how determinant attributes are in shaping social differences.

[25] See the proof of Proposition 2.

We find that disagreement is the long-run outcome of this process if and only if there is a unique attribute that becomes increasingly salient as time goes by. This attribute is precisely the initially most salient one. Thus, eventual homophily is such that individuals only pay attention to others if they are similar to them in that particular attribute. As a product of this behavior, two groups of thinking emerge in the long-run. The time to convergence to this scenario is non-increasing in the initial relative salience of this attribute.

We consider our findings to be related to the phenomenon of unidimensionality in attitudes, a widely discussed topic in political economy. As DeMarzo et al. (2003) point out, there is a strong debate on whether voting records of Congress and Senate members can be explained by a unidimensional liberal-conservative model. There is, in fact, evidence strongly supporting this model. For instance, Poole and Daniels (1985) find that the voting behavior in the U.S. Congress can be mainly explained by a single liberal-conservative dimension. We also consider that our model has a direct application related to the persistence of the gender pay gap. It is sometimes argued that the reason as to why females consistently self-report to be happier at work than males, relies on the fact that they have traditionally held lower labor reward aspirations than males. This phenomenon is known as *The Paradox of Female Happiness*. Two references discussing this paradox and related aspects are Bertrand (2011) and Clark (1997). Divergent aspirations between males and females might be able to explain that part of the gender gap that remains unexplained even after controlling for relevant aspects such as skill levels. Our intuition is that a model of wage setting in which individuals are of both sexes and are endowed with gender biased aspirations, will deliver as a result a gender pay gap, provided that the updating of aspirations takes place with our mechanism. Specifically, females might end up self-selected into low payment jobs, even without discriminatory behavior from the part of employers.

Finally we would like to mention two aspects of the model that we left for future research: first, our model follows a *representative agent approach* in which there is one individual by type. We do not deal with the case in which individuals appear in society in different frequencies. Second, we have assumed that, in determining the intensity of relations individuals sum up the homophily values associated to shared attributes. It will be interesting to investigate the case in which when any pair of individuals share two (or more) attributes $i$ and $j$, the attention they pay to each other at time $t$ is given by a more general function (than the sum) of $\alpha_t^i$ and $\alpha_t^j$.

# 6 Appendix. Extensions

In this section we explore how previous findings react to natural modifications in the assumptions regarding individual behavior. To start with, it might be the case that individuals do not exhibit certain attitudes with respect to a given issue, but that these attitudes are subject to shocks, or are random. In contexts in which individuals aim 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 subjective probability distributions about the true state, as in DeGroot (1974). In situations in which individuals deal with ideological issues we might interpret attitudes' randomness as flexibility or lack of stubbornness. Regarding this point we consider, for the case in which types are defined by two attributes, that initial attitudes of every type are randomly drawn from symmetric continuous distributions. In the context of two attributes, we find that the persistence of disagreement is robust to randomness. In particular, disagreement may now persist across either attribute, being more likely to persist across the one for which the mean of the distribution of the initial difference in attitudes is the highest.

We also explore the more general question of what are the conditions that the evolution of homophily has to satisfy for disagreement to persist. Previously we used Luce as a particular rule for the evolution of homophily and discussed how this evolution gave raise to persistent disagreement. That was the case because homophily with respect to the initially most salient attribute increased over time in such a way that the convergence of attitudes to a common value was precluded. In contrast, the constant homophily feature in Golub and Jackson (2012), a somewhat different model, only affects the speed of convergence to consensus, an outcome that always emerges. We thus find relevant to go beyond in reconciling these two views and in the understanding of what are the homophily patterns that give raise to either persistent disagreement or consensus. We find that under a more general representation of homophily, disagreement persists if and only if the process by which individuals intensify their relations with others with whom they share the initially most salient attribute, is fast enough. More specifically, there are two forces playing a role: on the one hand individuals pay increasing attention to others on the basis of this attribute but on the other hand, they also always pay a positive amount of (possibly indirect) attention to everyone else. For disagreement to persist it has to be that the first force dominates the second.

### 6.1 Random Attitudes

In this section we consider random instead of certain attitudes in a context in which individuals are composed by two attributes, namely, 1 and 2. Specifically, let $\tilde{a}_0^A$ be the initial attitude of a type $A$. Let it follow a symmetric continuous distribution, with mean $a_0^A$ and variance $\sigma_A^2$. Initial attitudes of all types are assumed to be independent although not necessarily identically distributed. Let $\tilde{\Delta}_0[1] = 2^{-1}(\tilde{a}_0^{\{1,2\}} - \tilde{a}_0^{\{\emptyset\}} + (\tilde{a}_0^{\{1\}} - \tilde{a}_0^{\{2\}}))$ and $\tilde{\Delta}_0[2] = 2^{-1}(\tilde{a}_0^{\{1,2\}} - \tilde{a}_0^{\{\emptyset\}} + (\tilde{a}_0^{\{2\}} - \tilde{a}_0^{\{1\}}))$ be the distributions of the initial differences in attitudes associated to attribute 1 and 2, respectively. They have means $\Delta_0[1] = 2^{-1}(a_0^{\{1,2\}} - a_0^{\{\emptyset\}} + (a_0^{\{1\}} - a_0^{\{2\}}))$ and $\Delta_0[2] = 2^{-1}(a_0^{\{1,2\}} - a_0^{\{\emptyset\}} + (a_0^{\{2\}} - a_0^{\{1\}}))$, respectively, and the same variance, $\sum_A \sigma_A^2/4$. Let us assume without loss of generality that $\Delta_0[1] \geq \Delta_0[2] \geq 0$. In linking homophily and salience we discuss the Luce form, as in the main body. Thus:

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

Notice that the homophily values could, in principle, be positive or negative, depending on the realization of the random variables $\tilde{\Delta}_t[1]$ and $\tilde{\Delta}_t[2]$. As we assume that the only aspect that matters is the magnitude of the differences in attitudes and not their sign, we work with its absolute value.

As a preview of the results, we find that the persistence of disagreement is robust to randomness. In contrast with the deterministic case, in general disagreement persists across either attribute with positive probability. Disagreement will persist across attribute 1 with probability equal to one when the minimum among all possible realizations of $|\tilde{\Delta}_0[1]|$ is higher than the maximum among all possible realizations of $|\tilde{\Delta}_0[2]|$. In what follows we discuss how the likelihood that disagreement persists across either attribute depends on the features of the distributions of initial differences in attitudes associated to these attributes. The results are as follows:

**Proposition 3.** *In general disagreement persists across either attribute* 1 *or* 2 *with positive probability. Also, disagreement across attribute* 1 *is at least as likely as disagreement across attribute* 2*. The (non-negative) expression accounting for the difference in probabilities is:*

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

*Thus, both events are equally likely if and only if the initial expected differences in attitudes are equal (that is, if and only if* $\Delta_0[1] = \Delta_0[2]$*, or equivalently,* $a_0^{\{1\}} - a_0^{\{2\}} = 0$*) whereas disagreement across attribute* 1 *is the most likely event if and only if the initial expected difference is the highest across this attribute (that is, if and only if,* $\Delta_0[1] > \Delta_0[2]$*, or equivalently,* $a_0^{\{1\}} - a_0^{\{2\}} > 0$*).*

Expression (6) is non-negative because 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.[26] Notice also that fixing the variance of the aforementioned distributions, the probability that they take non-negative values, increases with their mean, so does the likelihood of disagreement across attribute 1.

It is also worth mentioning that since we focus on perturbations in initial attitudes, once they are realized, eventual attitudes acquire the same form as the ones in Theorem 1 in the main body. The following remark formally states the point. For this purpose let $\tilde{a}_\infty^A$ denote the eventual attitude of a type $A$:

**Remark.** *Suppose that disagreement persists across attribute $i = \{1, 2\}$. Then:*

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

*and*

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

.

The following examples illustrate previous findings and related aspects. In example 1 we pin down the probability that disagreement persists across either attribute when initial attitudes are uniformly distributed. In example 2, initial attitudes are normally distributed. We illustrate how the probability that disagreement persists across either attribute is not only sensitive to the mean, as previously stated, but we offer insights on how the variance of these distributions may play a role:

**Example 3.** First, let initial attitudes be such that $a_0^{\{1,2\}} \sim U[0,1]$, $\tilde{a}_0^{\{1\}} \sim U[-1,1]$, $\tilde{a}_0^{\{2\}} \sim U[-1,1]$ and $\tilde{a}_0^{\{\emptyset\}} \sim U[-1,1]$. Thus, $\tilde{\Delta}_0[1]$ and $\tilde{\Delta}_0[2]$ have means $\Delta_0[1] = \Delta_0[2] = 0.25$. Thus, from Proposition 3, disagreement across either attribute is equally likely. To see this 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 mean zero, this difference in probabilities is zero. Second, let initial attitudes be such that $a_0^{\{1,2\}} \sim U[0,1]$, $\tilde{a}_0^{\{1\}} \sim U[0,1]$, $\tilde{a}_0^{\{2\}} \sim U[-1,1]$ 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 3, disagreement across attribute 1 is the most likely event. As above we focus 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:

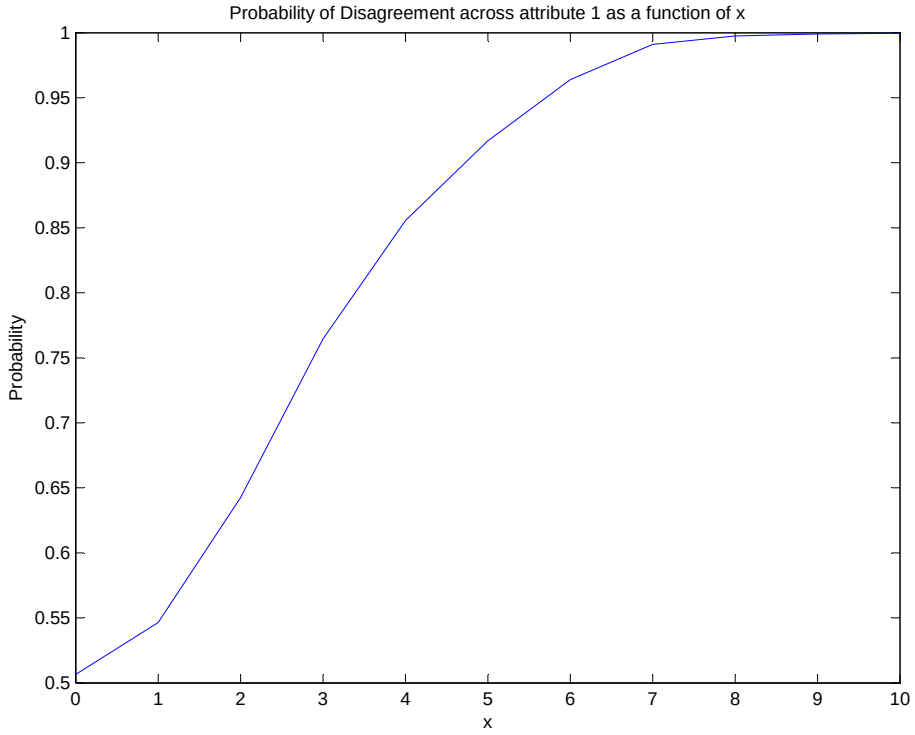
[26] See the Proof of Proposition 3.

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

Thus, $P(y \geq 0) = 1 - \int_{y=-1}^{y=0} \frac{1+y}{2} = 0.75$. Also, let $z \equiv \tilde{a}_0^{\{1\}} - \tilde{a}_0^{\{2\}}$. Notice that it follows the same distribution as $y$. Thus, $P(z \geq 0) = 0.75$ as well. In this case expression (6) equals 0.25. Thus disagreement persists across attribute 1 and 2 with probabilities 0.625 and 0.375, respectively.

**Example 4.** Let initial attitudes be normally distributed with means such that $\Delta_0[1] \geq \Delta_0[2] > 0$ and variances equal to one.[27] In the first figure we depict the probability that disagreement persist across attribute 1, as a function of the mean of the distribution of difference in attitudes associated it. In particular we keep $\Delta_0[2]$ constant and increase $\Delta_0[1]$. On the x-axis we depict the difference $x = \Delta_0[1] - \Delta_0[2]$ and on the y-axis, the probability of disagreement across attribute 1. We observe a positive relation.

Figure 3. Probability that disagreement persists across attribute 1




[27] As stated in the main body the results go through when initial attitudes are defined over the entire real line.

[28] Specifically, we simulate the model 2000 times for every configuration of the initial differences in attitudes, starting from the case in which $\Delta_0[1] = 0$ and $\Delta_0[2] = 0$ and thus $x = 0$ and ending in the case in which $\Delta_0[1] = 10$ and $\Delta_0[2] = 0$ thus, $x = 10$. The Matlab code is available upon request.

Within this example, it is also worth illustrating how the variances of the distributions of initial attitudes may play a role in determining the likelihood of disagreement. Let us consider that initial attitudes are normally distributed with the same means as above, 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 $\tilde{a}_0^{\{1\}'} - \tilde{a}_0^{\{2\}'}$ and/or $\tilde{a}_0^{\{1,2\}'} - \tilde{a}_0^{\{\emptyset\}'}$ have higher variances than (mean preserving spreads of) $\tilde{a}_0^{\{1\}} - \tilde{a}_0^{\{2\}}$ and/or $\tilde{a}_0^{\{1,2\}} - \tilde{a}_0^{\{\emptyset\}}$, respectively. Then in this case, $0.5 < P(\tilde{a}_0^{\{1\}'} - \tilde{a}_0^{\{2\}'} \geq 0) < P(\tilde{a}_0^{\{1\}} - \tilde{a}_0^{\{2\}} \geq 0)$ and/or $0.5 < P(\tilde{a}_0^{\{1,2\}'} - \tilde{a}_0^{\{\emptyset\}'} \geq 0) < P(\tilde{a}_0^{\{1,2\}} - \tilde{a}_0^{\{\emptyset\}} \geq 0)$.[29] Notice that expression (6) in Proposition 3 has now lower value than before, meaning that disagreements across either attribute are closer to being equally likely.

To conclude, it is also the case that disagreement manifests as two groups holding different eventual attitudes. Specifically, the difference in average eventual attitudes associated to attribute 1 (respectively 2) persists whereas the one associated to attribute 2 (respectively 1) is zero.[30]

### 6.2 A general representation of homophily

In this section we relate the persistence of disagreement to the properties and evolution of homophily. For this purpose we define homophily values in broader terms. Specifically, let $\gamma_t^i$ be the homophily value associated to attribute $i$ at time $t$. Let this value depend on the differences in average attitudes associated to (possibly) all attributes. As in the main body we assume that at every time $t$, $\gamma_t^i$ is non-negative and we normalize to one the total amount of attention that every individual devotes to others. It then has to be the case that at every time $t$ the sum of these homophily values is one, that is, $\sum_i \gamma_t^i = 1$.[31] We finally assume that the homophily values satisfy two properties dealing with the monotonicity aspects of attention with respect to the differences in attitudes. The first one states that the attention that every attribute enjoys is positive if and only if the difference in attitudes across it, is positive. The second one states that if attribute $i$ exhibits a higher difference in attitudes than attribute $j$ then the former enjoys higher attention than the latter:

[29] 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 3.

[30] When disagreement persists across attribute 1, the difference in average eventual attitudes associated to attribute 1 is $\tilde{\Delta}_\infty[1] = (2^{n-1})^{-1}(\sum_{A:i\in A} \tilde{a}_\infty^A - \sum_{A:1\notin A} \tilde{a}_\infty^A) = (2^{n-1})^{-1}2^{n-1}(\tilde{a}_\infty^A : 1 \in A - \tilde{a}_\infty^A : 1 \notin A) = |\tilde{\Delta}_0[1]| - |\tilde{\Delta}_0[2]|$ (respectively $|\tilde{\Delta}_0[2]| - |\tilde{\Delta}_0[1]|$) when $\tilde{\Delta}_0[1] \geq 0$ (respectively $\tilde{\Delta}_0[1] < 0$). Since disagreement across attribute 1 persists when $|\tilde{\Delta}_0[1]| > |\tilde{\Delta}_0[2]|$, $\tilde{\Delta}_\infty[1]$ has either positive or negative support. Furthermore, the distribution of the difference in average eventual attitudes associated to attribute 2 is degenerated at zero. To see this notice that within the $2^{n-1}$ types possessing attribute 1 there are $2^{n-2}$ types possessing and lacking attribute 2, respectively. The same happens within the $2^{n-1}$ types lacking attribute 1, hence, $\tilde{\Delta}_\infty[2] = (2^{n-1})^{-1}(\sum_{A:2\in A} \tilde{a}_\infty^A - \sum_{A:2\notin A} \tilde{a}_\infty^A) = 2^{-1}(2^{n-2})^{-1}2^{n-2}(\tilde{a}_\infty^A : 1 \in A, 2 \in A + \tilde{a}_\infty^A : 1 \notin A, 2 \in A) - \tilde{a}_\infty^A : 1 \in A, 2 \notin A + \tilde{a}_\infty^A : 1 \notin A, 2 \notin A) = 2^{-1}2(\overline{a}_0 - \overline{a}_0) = 0$. The analysis is the same when disagreement persists across attribute 2.

[31] See Section 2.

**Within differences monotonicity (WDM).** $\Delta_t[i] = 0$ implies that $\gamma_t^i = 0$ and $\Delta_t[i] > 0$ implies that $\gamma_t^i > 0$.[32]

**Across differences monotonicity (ADM).** $\Delta_t[1] \geq \Delta_t[2] \geq \cdots \geq \Delta_t[n] \geq 0$ implies that $\gamma_t^1 \geq \gamma_t^2 \geq \cdots \geq \gamma_t^n \geq 0$.

We also set the technical condition that $\lim_{t \longrightarrow \infty} \gamma_t^i$ exists for every attribute $i$ and that $\lim_{t \longrightarrow \infty} \sum_i \gamma_t^i = \sum_i \lim_{t \longrightarrow \infty} \gamma_t^i = 1$.

We now state the condition for the persistence of disagreement and provide its form:

**Theorem 2.** *For every configuration of initial attitudes, eventual ones always exist. They exhibit disagreement if and only if homophily based on attribute* 1*, approaches value* 1 *sufficiently fast (that is, if and only if* $\sum_{t=0}^{\infty} log\gamma_t^1$ *exists). In this case, eventual attitudes are such that, for every type A:*

$$|a_\infty^A| = \frac{1}{2}\tau(W^\infty)\Delta_0[1]$$

*where* $\tau(W^\infty) \in (0, 1]$*. Furthermore,* $a_\infty^A > 0$ *if and only if* $1 \in A$*.*

Disagreement persists whenever the process by which individuals progressively intensify their relations with others similar to them in attribute 1 is fast enough. Intuitively there are two forces playing a role: on the one hand individuals pay increasing attention to others on the basis of attribute 1 but on the other hand, they also pay a positive amount of (possibly indirect) attention to everyone else. For disagreement to persist, it has to be that the first force dominates the second. Needless to say that when the value of homophily is linked to differences in attitudes by the Luce form, as in the main body, the updating process satisfies these requirements. As an illustration, for the case with two attributes $\sum_{t=0}^{\infty} log\gamma_t^1 = log(1 - \Delta_0[2]/\Delta_0[1])$ and $\tau(W^\infty) = 1 - \Delta_0[2]/\Delta_0[1]$, with $\Delta_0[1] > \Delta_0[2] \geq 0$.

Disagreement materializes in two groups of thinking, defined according to whether individuals possess or lack attribute 1. We cannot specify the closed form expression for the ergodicity coefficient $\tau(W^\infty)$ in this case, since it depends on the particular functional form for the homophily values. We just say that $\tau(W^\infty) = \lim_{T \to \infty} \prod_{t=0}^{T} \gamma_t^1 \in (0, 1]$.

The following examples illustrate the requirement in the Theorem above. For this purpose, we consider updating rules that are mainly based on modifications of the Luce form in expression (2), with the exception of example 8. Example 5 deals with a scenario in which consensus is achieved whereas in examples 6 to 8 disagreement persists.

**Example 5. Eventual consensus.** Consider the following updating rule:

[32] When $\Delta_t[i] = 0$ for every attribute $i$, we set $\gamma_t^i = 1/n$.

$$\gamma_t^1 = \begin{cases} \dfrac{\Delta_t[1]}{\Delta_t[1]+\Delta_t[2]} & if \quad \gamma_{t-1}^1 < H \in [0,1) \\ \gamma_{t-1}^1 & if \quad \gamma_{t-1}^1 \geq H \in [0,1) \end{cases}.$$

Let $\gamma_t^2 = 1-\gamma_t^1$ at every time $t$. Under this rule individuals use Luce to determine the attention they pay to others, but whenever a level H of homophily has been reached, they are no longer sensitive to changes in differences in attitudes. In this case interactions become static from some point in time on, and thus, individuals do not become homophilous exclusively with respect to attribute 1. The requirements in the proposition above are therefore not satisfied and consensus will eventually emerge.

**Example 6. The persistence of disagreement.** Let initial attitudes be $a_0' = [0.8 \;\; 0.2 \;\; -0.05 \;\; -0.95]$. Thus, the difference in average initial attitudes associated to attribute 1 is $\Delta_0[1] = 0.5(0.8+0.2) - 0.5(-0.05-0.95) = 1$ and the one associated to attribute 2 is $\Delta_0[2] = 0.5(0.8-0.05) - 0.5(0.2-0.95) = 0.75$. Consider the generalized Luce form, $\gamma_t^1 = \dfrac{\Delta_t[1]^\delta}{\Delta_t[1]^\delta + \Delta_t[2]^\delta}$ and $\gamma_t^2 = \dfrac{\Delta_t[2]^\delta}{\Delta_t[1]^\delta + \Delta_t[2]^\delta}$. When $\delta = 1.2$ we have that $\gamma_0^1 = 0.58$ and $\gamma_0^2 = 0.42$. The entries in the interaction matrices evolve as follows:

$$W_0 = \begin{bmatrix} 0.5 & 0.3 & 0.2 & 0 \\ 0.3 & 0.5 & 0 & 0.2 \\ 0.3 & 0 & 0.5 & 0.3 \\ 0 & 0.2 & 0.3 & 0.5 \end{bmatrix}, \; W_1 = \begin{bmatrix} 0.5 & 0.34 & 0.16 & 0 \\ 0.34 & 0.5 & 0 & 0.16 \\ 0.16 & 0 & 0.5 & 0.34 \\ 0 & 0.16 & 0.34 & 0.5 \end{bmatrix}, ..., \lim_{t\to\infty} W_t = \begin{bmatrix} 0.5 & 0.5 & 0 & 0 \\ 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 \\ 0 & 0 & 0.5 & 0.5 \end{bmatrix}.$$

$$\text{Also, } W^\infty = \begin{bmatrix} 0.33 & 0.33 & 0.17 & 0.17 \\ 0.33 & 0.33 & 0.17 & 0.17 \\ 0.17 & 0.17 & 0.33 & 0.33 \\ 0.17 & 0.17 & 0.33 & 0.33 \end{bmatrix} \text{ and } W^\infty \text{ times } a_0 = \begin{bmatrix} 0.8 \\ 0.2 \\ -0.05 \\ -0.95 \end{bmatrix} \text{ is } a_\infty = \begin{bmatrix} 0.16 \\ 0.16 \\ -0.16 \\ -0.16 \end{bmatrix}.$$

In this case $\tau(W^\infty) = 0.32$.[33] Notice that in the relation between attributes 1 and 2, summarized in $(\Delta_0[2]/\Delta_0[1])^{1.2}$, the difference in attitudes associated to attribute 1 exacerbates with respect to the case in which $\delta = 1$.

**Example 7. The persistence of disagreement.** Consider the case in which $\gamma_t^1 = \dfrac{\beta\Delta_t[1]}{\beta\Delta_t[1] + \delta\Delta_t[2]}$ and $\gamma_t^2 = \dfrac{\delta\Delta_t[2]}{\beta\Delta_t[1] + \delta\Delta_t[2]}$, with $\beta > \delta > 0$. Notice that the relation between attributes 1 and 2, that is, $\delta\Delta_0[2]/\beta\Delta_0[1]$, exacerbates with respect to the case in which $\beta = \delta$. This process leads to disagreement for any configuration

[33] The entries of $W^\infty$ are a function of $\tau(W^\infty)$, thus we can recover its value from the expression of $W^\infty$. See the proof of Theorem 1.

of initial attitudes, that is, even in the case in which $\Delta_0[1] = \Delta_0[2]$. The reason is that $\gamma_0^1 > \gamma_0^2$ and $\Delta_1[1] = \gamma_0^1 \Delta_0[1] > \Delta_1[2] = \gamma_0^2 \Delta_0[2]$, thus form $t = 1$ Theorem 1 in the main body applies.

**Example 8. The persistence of disagreement.** Let initial attitudes be $a_0' = [0.8 \;\; 0.2 \;\; -0.05 \;\; -0.95]$. Thus, the difference in average initial attitudes associated to attribute 1 is $\Delta_0[1] = 1$ and the one associated to attribute 2 is $\Delta_0[2] = 0.75$, as in example 6. Consider the following updating rule:

$$\gamma_t^1 = \begin{cases} 0.5 & if \quad \Delta_t[1] = \Delta_t[2] \geq 0 \\ 1 & if \quad \Delta_t[1] > \Delta_t[2] = 0 \\ 0 & if \quad \Delta_t[2] > \Delta_t[1] = 0 \\ \left(\dfrac{\Delta_t[1]}{\Delta_t[2]}\right)^{\alpha} & if \quad \Delta_t[2] > \Delta_t[1] > 0 \\ 1 - \left(\dfrac{\Delta_t[2]}{\Delta_t[1]}\right)^{\beta} & if \quad \Delta_t[1] > \Delta_t[2] > 0 \end{cases}.$$

We set $\alpha$ and $\beta$ such that $(\Delta_t[1]/\Delta_t[2])^{\alpha}$ and $(\Delta_t[2]/\Delta_t[1])^{\beta}$ are smaller than one half. Let $\gamma_t^2 = 1 - \gamma_t^1$, at every time $t$. Let us set, for instance, $\beta = 2.5$.[34] The entries in the interaction matrix evolve as follows:

$$W_0 = \begin{bmatrix} 0.5 & 0.26 & 0.24 & 0 \\ 0.26 & 0.5 & 0 & 0.24 \\ 0.24 & 0 & 0.5 & 0.26 \\ 0 & 0.24 & 0.26 & 0.5 \end{bmatrix}, \; W_1 = \begin{bmatrix} 0.5 & 0.28 & 0.22 & 0 \\ 0.28 & 0.5 & 0 & 0.22 \\ 0.22 & 0 & 0.5 & 0.28 \\ 0 & 0.22 & 0.28 & 0.5 \end{bmatrix}, ..., \lim_{t\to\infty} W_t = \begin{bmatrix} 0.5 & 0.5 & 0 & 0 \\ 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 \\ 0 & 0 & 0.5 & 0.5 \end{bmatrix}.$$

$$\text{Also, } W^{\infty} = \begin{bmatrix} 0.31 & 0.31 & 0.19 & 0.19 \\ 0.31 & 0.31 & 0.19 & 0.19 \\ 0.19 & 0.19 & 0.31 & 0.31 \\ 0.19 & 0.19 & 0.31 & 0.31 \end{bmatrix} \text{ and } W^{\infty} \text{ times } a_0 = \begin{bmatrix} 0.8 \\ 0.2 \\ -0.05 \\ -0.95 \end{bmatrix} \text{ is } a_{\infty} = \begin{bmatrix} 0.12 \\ 0.12 \\ -0.12 \\ -0.12 \end{bmatrix}.$$

In this case $\tau(W^{\infty}) = 0.24$.

# 7 Appendix. Proofs

First of all let $\lambda_t^i = \Delta_t[i] / \sum_j \Delta_t[j]$ for every attribute $i$ and at every time $t$.

*Proof of Theorem 1.* The proof is composed by several steps. In step 1 we show how, at every time $t$, $\lambda_0^i > 0$ and $\lambda_0^i = 0$ imply that $\lambda_t^i > 0$ and $\lambda_t^i = 0$, respectively. Steps 2-8 analyze disagreement when $\lambda_0^i > 0$ for every attribute $i$. In particular, Steps 2-5

[34] Since at every time $t$ it is the case that $\Delta_t[1] > \Delta_t[2] > 0$, we do not specify any value for $\alpha$. Also, since $\Delta_t[1]/\Delta_t[2]$ (respectively $\Delta_t[2]/\Delta_t[1]$) is decreasing over time whenever $\Delta_t[2] > \Delta_t[1]$ (respectively $\Delta_t[2] < \Delta_t[1]$) we set $\alpha$ and $\beta$ to be constant. See steps 1 and 7 in the proof of Theorem 1.

identify the eigenvalues and eigenvectors of $W_t$ and diagonalize it. Step 6 deals with the existence of the limiting product of point-wise stochastic matrices, that is, $W^\infty$. Step 7 provides its form. Step 8 establishes the necessary and sufficient condition for disagreement to persist, qualifying it. Finally, step 9 elaborates on the case in which $\lambda_0^i = 0$ for some/all attributes $i > 1$.

Step 1. We prove that $\lambda_0^i > 0$ implies that $\lambda_t^i > 0$ and $\lambda_0^i = 0$ implies that $\lambda_t^i = 0$, at every $t$. We proceed by decomposing $\Delta_t[i] = (2^{n-1})^{-1}\left[\sum_{A:i\in A} a_t^A - \sum_{A:i\notin A} a_t^A\right]$. Consider a type $A$ such that $i \in A$. By (1), $a_t^A = \sum_B w_{t-1}^{A,B} a_{t-1}^B$. Since $w_{t-1}^{A,B} = (2^{n-1})^{-1}\sum_{i\in I(AB)} \lambda_{t-1}^i$, then:

$$a_t^A = \sum_B w_{t-1}^{A,B} a_{t-1}^B = \frac{1}{2^{n-1}} \sum_{B:i\in B} \lambda_{t-1}^i a_{t-1}^B + \frac{1}{2^{n-1}} \sum_{j\neq i} \lambda_{t-1}^j \sum_{B:j\in I(AB)} a_{t-1}^B.$$

Since there are $2^{n-1}$ types $A$ possessing attribute $i$, $\sum_{A:i\in A} a_t^A = \sum_{B:i\in B} \lambda_{t-1}^i a_{t-1}^B + \sum_{j\neq i} \lambda_{t-1}^j \sum_{B:j\in I(AB)} a_{t-1}^B$. By a similar reasoning, for types $A$ such that $i \notin A$, $\sum_{A:i\notin A} a_t^A = \sum_{B:i\notin B} \lambda_{t-1}^i a_{t-1}^B + \sum_{j\neq i} \lambda_{t-1}^j \sum_{B:j\in I(AB)} a_{t-1}^B$. Therefore:

$$\frac{1}{2^{n-1}} \sum_{A:i\in A} a_t^A - \frac{1}{2^{n-1}} \sum_{A:i\notin A} a_t^A = \frac{1}{2^{n-1}} \sum_{B:i\in B} \lambda_{t-1}^i a_{t-1}^B - \frac{1}{2^{n-1}} \sum_{B:i\notin B} \lambda_{t-1}^i a_{t-1}^B$$

or equivalently, $\Delta_t[i] = \lambda_{t-1}^i \Delta_{t-1}[i]$.[35] From the definition of $\lambda_t^i$, it follows that at every $t$, $\Delta_t[i] \geq 0$ implies that $\lambda_t^i \geq 0$. Also, $\Delta_t[i] \geq 0$ if and only if $\lambda_{t-1}^i \geq 0$ and $\Delta_{t-1}[i] \geq 0$. With these two observations we conclude that $\Delta_0[i] > 0$ and $\lambda_0^i > 0$ imply that at every time $t$, $\Delta_t[i] > 0$ and $\lambda_t^i > 0$, respectively. Also $\Delta_0[i] = 0$ and $\lambda_0^i = 0$ imply that at every time $t$, $\Delta_t[i] = 0$ and $\lambda_0^i = 0$, respectively.[36]

Step 2. At every time $t$, 1 is an eigenvalue of $W_t$, with right-eigenvector $u$ of size $2^n \times 1$, where $u$ has all components equal to 1. This directly follows from the stochasticity of $W_t$. Notice that $u$ is time independent. We thus omit the time subscript.

Step 3. At every time $t$ and for every attribute $i$, $\lambda_t^i$ is an eigenvalue of $W_t$, with right-eigenvector $u^i$ of size $2^n \times 1$, where $u^i$ has the following form: the component of $u^i$ associated to type $A$ is equal to 1 if $i \in A$ and equal to $-1$ otherwise. We prove that by showing that the pair $(\lambda_t^i, u^i)$ satisfies the eigenvalue equation, $W_t u^i = \lambda_t^i u^i$. Consider an attribute $i$ and an arbitrary type $A$. Suppose first that $i \in A$. Notice that there are exactly $2^{n-1}$ types $B$ possessing attribute $i$. Also, notice that for every $j \neq i$, there are exactly $2^{n-2}$ types $B$ possessing attribute $i$ that are $j$-similar to $A$ and $2^{n-2}$ types $B$ lacking attribute $i$ that are $j$-similar to $A$. Therefore, the

[35] As this expression holds at every $t$, we recursively write $\Delta_t[i] = \prod_{s=0}^{t-1} \lambda_s^i \Delta_0[i]$.

[36] Also from the definition of $\lambda_t^i$ it follows that at every $t$, $\Delta_t[1] \geq \Delta_t[2] \geq \cdots \geq \Delta_t[n] \geq 0$ implies that $\lambda_t^1 \geq \lambda_t^2 \geq \cdots \geq \lambda_t^n \geq 0$. Additionally, $\Delta_t[1] \geq \Delta_t[2] \geq \cdots \geq \Delta_t[n] \geq 0$ if and only if $\lambda_{t-1}^1 \Delta_{t-1}[1] \geq \lambda_{t-1}^2 \Delta_{t-1}[2] \geq \cdots \geq \lambda_{t-1}^n \Delta_{t-1}[n] \geq 0$. Since by assumption $\Delta_0[1] \geq \Delta_0[2] \geq \cdots \geq \Delta_0[n] \geq 0$ then, at every $t$, $\Delta_t[1] \geq \Delta_t[2] \geq \cdots \geq \Delta_t[n] \geq 0$ and $\lambda_t^1 \geq \lambda_t^2 \geq \cdots \geq \lambda_t^n \geq 0$ hold. This also implies that $\prod_{t=0}^T \lambda_t^1 > \prod_{t=0}^T \lambda_t^2 \geq \cdots \geq \prod_{t=0}^T \lambda_t^n \geq 0$.

row in $W_t$ corresponding to type $A$, multiplied by $u^i$, is equal to:

$$\sum_{B:i\in B} w_t^{A,B} - \sum_{B:i\notin B} w_t^{A,B} = \frac{2^{n-1}\lambda_t^i + 2^{n-2}\sum_{j\neq i}^{n}\lambda_t^j - 2^{n-2}\sum_{j\neq i}^{n}\lambda_t^j}{2^{n-1}} = \lambda_t^i.$$

Since every type $A$ is such that $i \in A$, the RHS of the eigenvalue equation also equals $\lambda_t^i$. Thus, we conclude that $(\lambda_t^i, u^i)$ is a pair of eigenvalue and right-eigenvector of $W_t$. The proof for the case in which $A$ is such that $i \notin A$ is analogous and hence omitted. As in step 2, the eigenvectors $u^i$ corresponding to every $\lambda_t^i$ are also time independent.

Step 4. At every time $t$, the remaining eigenvalues of $W_t$ are zero. Consider any type $B$ such that $|B| \geq 2$. We start by proving that for every type $A$, $w_t^{A,B} = \sum_{i\in B} w_t^{A,\{i\}} - [|B|-1]\, w_t^{A,\emptyset}$. By doing so, we are proving that the column vector of weights associated to type $B$ is a linear combination of the column vectors of weights associated to types containing at most one attribute and hence, there are at most $n+1$ independent columns in $W_t$. Notice that getting rid of the normalization $1/2^{n-1}$, we are left with:

$$\sum_{i\in B} w_t^{A,\{i\}} = \sum_{i\in B\cap A}(\lambda_t^i + \sum_{j\in A^c}\lambda_t^j) + \sum_{i\in B\cap A^c}\sum_{j\in A^c, j\neq i}\lambda_t^j$$

and that this is equivalent to:

$$\sum_{i\in B\cap A}(\lambda_t^i + \sum_{j\in A^c}\lambda_t^j) + \sum_{i\in B\cap A^c}(\sum_{j\in A^c}\lambda_t^j - \lambda_t^i) = \sum_{i\in B\cap A}\lambda_t^i + \sum_{i\in B}\sum_{j\in A^c}\lambda_t^j - \sum_{i\in B\cap A^c}\lambda_t^i. \quad (7)$$

Second, notice that:

$$(|B|-1)w_t^{A,\emptyset} = (|B|-1)\sum_{j\in A^c}\lambda_t^j. \quad (8)$$

Thus, (7) minus (8) is equal to $\sum_{i\in B\cap A}\lambda_t^i + \sum_{j\in A^c}\lambda_t^j - \sum_{i\in B\cap A^c}\lambda_t^i$. This expression can be rewritten as $\sum_{i\in B\cap A}\lambda_t^i + \sum_{j\in I\cap A^c}\lambda_t^j - \sum_{i\in (I\setminus B^c)\cap A^c}\lambda_t^i = \sum_{i\in B\cap A}\lambda_t^i + \sum_{i\in B^c\cap A^c}\lambda_t^i$. This is equivalent to $w_t^{A,B} = \sum_{i\in I(AB)}\lambda_t^i$ where $I(AB) = (B\cap A)\cup(B^c\cap A^c)$ as defined in section 2. Thus, $rank(W_t) \leq n+1$.

Recall that the rank of a matrix is equal to the number of non-zero eigenvalues. Since steps 2 and 3 already identified $n+1$ of them, indeed $rank(W_t) = n+1$. Thus, the rest of the $2^n - (n+1)$ eigenvalues are zero.

Step 5. We prove here that $W_t$ is always diagonalizable and provide its form. From symmetry of $W_t$ there is an orthogonal diagonalization $W_t = U\Lambda_t U'$, where $U$

is an orthonormal basis. Orthonormal eigenvectors have unitary euclidean norm and are orthogonal to each other. Therefore, for the zero eigenvalues, there exist eigenvectors $u^0$ with $\|u^0\| = 1$, orthogonal to each other and to both, $u/\|u\|$ and every $u^i/\|u^i\|$, where $\|u\| = 2^{n/2}$ and $\|u^i\| = 2^{n/2}$, for every $i$. Since by steps 2 and 3 $u$ and every $u^i$ are time independent, every $u^0$ is also time independent. Now, fix the following order of eigenvalues: first eigenvalue 1, afterwards eigenvalues $\lambda_t^i$, by type, and finally the zero eigenvalues in a fixed order. Then $U = \left[\frac{u}{\|u\|} \frac{u^i}{\|u^1\|} ... \frac{u^n}{\|u^n\|} u^0 \dots u^0\right]$, and the diagonal matrix of eigenvalues at time $t$ is:

$$\Lambda_t = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \lambda_t^1 & \cdots & \cdots & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & \cdots & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & \cdots & \lambda_t^n & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Since at every time $t$ the matrix $W_t$ is diagonalizable over the same eigenspace, hence $W^T = U\Lambda^T U'$ where $\Lambda^T = \prod_{t=0}^{T} \Lambda_t$ with diagonal entries: 1, $\prod_{t=0}^{T} \lambda_t^i$ for every attribute $i$ and zeros.

Step 6. Here we deal with the existence of $W^\infty$ and $a_\infty$. By step 5, $W^\infty = U \lim_{T \longrightarrow \infty} \Lambda^T U'$, provided that the RHS of this expression exists. We confirm here that this is, in fact, the case. In computing $\lim_{T \longrightarrow \infty} \Lambda^T$ we focus on the non-zero diagonal entries of $\Lambda^T$. Eigenvalue 1 is constant over time, thus its limiting product is 1. Since at every time $t$, $\lambda_t^i \in (0,1)$ for every $i$ then $\prod_{t=0}^{\infty} \lambda_t^i$ exists in $[0,1)$. Thus, $U \lim_{T \longrightarrow \infty} \Lambda^T U'$ exists and defines both, $W^\infty$ and $a_\infty = W^\infty a_0$, for every $a_0$.

Step 7. We provide here the specific form of $W^\infty$. Suppose that $\Delta_0[1] > \Delta_0[2]$. Consider attribute 1 first. Let $r_t^i = \Delta_t[i]/\Delta_t[1]$ for every attribute $i$ and at every time $t$. We then rewrite $\lambda_t^1 = \Delta_t[1](\Delta_t[1] + \sum_{i>1} \Delta_t[i])^{-1} = [1 + \sum_{i>1} r_t^i]^{-1}$. By step 1, $r_t^i = \lambda_{t-1}^i \Delta_{t-1}[i] / \lambda_{t-1}^1 \Delta_{t-1}[1]$. From the expression of $\lambda_t^i$ it follows that $\lambda_{t-1}^i/\lambda_{t-1}^1 = \Delta_{t-1}[i]/\Delta_{t-1}[1] = r_{t-1}^i$. Thus, $r_t^i = (r_{t-1}^i)^2$ and recursively we get that $r_t^i = (r_0^i)^{2^t}$. Thus, $\lambda_t^1 = [1 + \sum_{i>1} (r_0^i)^{2^t}]^{-1}$. It is important to notice that, $0 < r_0^i < 1$ for attributes $i > 1$. It then follows that $\lim_{t \longrightarrow \infty} \lambda_t^1 = 1$. This opens the possibility for $\prod_{t=0}^{\infty} \lambda_t^1 \neq 0$. We prove that this is indeed the case by equivalently stating that $\sum_{t=0}^{\infty} log(\lambda_t^1)$ exists.

In order to do it, we consider $r_0^2$, the highest ratio smaller than one, and construct a new homophily value as follows: we replace $r_0^i$, for attributes $i > 1$, with $r_0^2$ in $\lambda_t^1$. Specifically, we have that $\underline{\lambda}_t^1 = [1 + (n-1)(r_0^2)^{2^t}]^{-1}$. Since $r_0^2 \geq r_0^i$ for every $i > 1$, then $\lambda_t^1 \geq \underline{\lambda}_t^1$ at every time $t$. We prove that $\sum_{t=0}^{\infty} log(\underline{\lambda}_t^1)$ exists, so does $\sum_{t=0}^{\infty} log(\lambda_t^1)$, by comparison. We proceed by testing the absolute

convergence (and hence the convergence) of $\sum_{t=0}^{\infty} log(\underline{\lambda}_t^1)$, using the ratio test. It is well known that an adaptation of the L'Hopital rule can be used to find limits of sequences. We thus define $f(x)$ and $g(x)$ as functions of a real variable $x$ and $\{s_t\}$ such that at every $t$, $s_t = f(t)/g(t)$. Then, we evaluate $\lim_{x\to\infty} f(x)/g(x) = \lim_{x\to\infty} \dfrac{log(1+(n-1)(r_0^2)^{2^{x+1}})}{log(1+(n-1)(r_0^2)^{2^x})}$. Since $0 < r_0^2 < 1$, this limit is indeterminate. By L'Hopital $\lim_{x\to\infty} f(x)/g(x) = \lim_{x\to\infty} f'(x)/g'(x) = \dfrac{2(r_0^2)^{2^x}(1+(n-1)(r_0^2)^{2^x})}{(1+(n-1)(r_0^2)^{2^{x+1}})} = 0$. Thus, $\lim_{t\longrightarrow\infty} s_t = \lim_{x\to\infty} f(x)/g(x) = 0$. This implies that $\sum_{t=0}^{\infty} |log(\underline{\lambda}_t^1)|$ exists. Since at every $t$, $\lambda_t^1 \geq \underline{\lambda}_t^1$, then $|log(\lambda_t^1)| \leq |log(\underline{\lambda}_t^1)|$. Thus, by comparison $\sum_{t=0}^{\infty} |log(\lambda_t^1)|$ exists, so does $\sum_{t=0}^{\infty} log(\lambda_t^1)$.

Consider now attributes $i > 1$. For a given $i > 1$, let $j$ denote attributes other than it and let $r_t^j = \Delta_t[j]/\Delta_t[i]$. Then $\lambda_t^i = [1 + \sum_{j\neq i}(r_0^j)^{2^t}]^{-1}$. Notice that $r_0^1 > 1$. Then $\lim_{t\longrightarrow\infty} \lambda_t^i = 0$ and $\prod_{t=0}^{\infty} \lambda_t^i = 0$ for attributes $i > 1$. Summing up we have that $\prod_{t=0}^{\infty} \lambda_t^1 = \mu^1 \in (0,1)$ and $\prod_{t=0}^{\infty} \lambda_t^i = 0$ for $i > 1$. Under this scenario:

$$\lim_{T\longrightarrow\infty} \Lambda^T = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \mu^1 & \vdots & \vdots & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 0 & \vdots & \vdots & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \cdots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad U \lim_{T\longrightarrow\infty} \Lambda^T = \frac{1}{2^{n/2}} \begin{bmatrix} 1 & \mu^1 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \cdots & 0 \\ 1 & \mu^1 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 1 & -\mu^1 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \cdots & 0 \\ 1 & -\mu^1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

and thus,

$$W^{\infty} = U \lim_{T\longrightarrow\infty} \Lambda^T U' = \frac{1}{2^n} \begin{bmatrix} 1+\mu^1 & \cdots & 1+\mu^1 & 1-\mu^1 & \cdots & 1-\mu^1 \\ 1+\mu^1 & \cdots & 1+\mu^1 & 1-\mu^1 & \cdots & 1-\mu^1 \\ \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ 1+\mu^1 & \cdots & 1+\mu^1 & 1-\mu^1 & \cdots & 1-\mu^1 \\ 1+\mu^1 & \cdots & 1+\mu^1 & 1-\mu^1 & \cdots & 1-\mu^1 \\ 1-\mu^1 & \cdots & 1-\mu^1 & 1+\mu^1 & \cdots & 1+\mu^1 \\ 1-\mu^1 & \cdots & 1-\mu^1 & 1+\mu^1 & \cdots & 1+\mu^1 \\ \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ 1-\mu^1 & \cdots & 1-\mu^1 & 1+\mu^1 & \cdots & 1+\mu^1 \\ 1-\mu^1 & \cdots & 1-\mu^1 & 1+\mu^1 & \cdots & 1+\mu^1 \end{bmatrix}$$

.

Since the eigenvectors $u^0$, associated to the zero eigenvalues, occupy the last columns (respectively rows) of $U$ (respectively $U'$), they do not play any role in the products above.

Suppose now that $\Delta_0[1] = \Delta_0[i]$ for some attributes $i > 1$. Let $e$ be the number of attributes $i > 1$ such that $\Delta_0[1] = \Delta_0[i]$. Then $\lim_{t\longrightarrow\infty} \lambda_t^1 = [e+1]^{-1} < 1$,

implying that $\prod_{t=0}^{\infty}\lambda_t^1 = 0$. By the same reasoning this is also the case for attributes $i > 1$ such that $\Delta_0[1] = \Delta_0[i]$. For attributes $i > 1$ such that $\Delta_0[1] > \Delta_0[i]$, then $\lim_{t\longrightarrow\infty}\lambda_t^i = 0$ by similar arguments as above, thus $\prod_{t=0}^{\infty}\lambda_t^i = 0$. Under this scenario, every entry in $W^{\infty}$ is $(2^n)^{-1}$.

Before concluding let us consider the general case in which $\gamma_t^i = \Delta_t[i]^{\delta}/\sum_j \Delta_t[j]^{\delta}$ with $\delta \in (0,\infty)$. We can rewrite $\gamma_t^1 = [1 + \sum_{i>1}(r_0^i)^{\delta(\delta+1)^t}]^{-1}$ in this case. Notice that $\lim_{t\to\infty}\gamma_t^1 = 1$ and thus, $\lim_{t\to\infty}\gamma_t^i = 0$ for $i > 1$. We study the convergence of $\sum_{t=0}^{\infty} log(\gamma_t^1)$ using the same reasoning as before, where now $\underline{\gamma}_t^1 = [1 + (n-1)(r_0^2)^{\delta(\delta+1)^t}]^{-1}$. Using similar algebra and reasoning as above we conclude that $\lim_{x\to\infty} f'(x)/g'(x) = \dfrac{2(r_0^2)^{\delta^2(\delta+1)^x}(1+(n-1)(r_0^2)^{\delta^2(\delta+1)^x})}{(1+(n-1)(r_0^2)^{\delta^2(\delta+1)^{x+1}})} = 0$ for $\delta \neq 0$. It is then the case that $\sum_{t=0}^{\infty} log(\gamma_t^1)$ converges, meaning that disagreement persists. When $\Delta_0[1] = \Delta_0[i]$ for some attribute(s) $i > 1$ consensus emerges as above.

Step 8. We establish here the necessary and sufficient condition for disagreement to persist. We also qualify disagreement. Recall that $\overline{a}_0 = 0$. The eventual attitude of a type $A$ is the result of multiplying its corresponding row in $W^{\infty}$ times the column vector of initial attitudes. Consider first that $\Delta_0[1] > \Delta_0[2]$. Then $W^{\infty}$ is the one derived in step 7. For the first $2^{n-1}$ rows of $W^{\infty}$, corresponding to types $A$ such that $1 \in A$, we thus have that $a_{\infty}^A = 2^{-1}\mu^1\left[\frac{1}{2^{n-1}}\sum_{A:i\in A} a_0^A - \frac{1}{2^{n-1}}\sum_{A:i\notin A} a_0^A\right] = 2^{-1}\mu^1\Delta_0[1]$. For the subsequent $2^{n-1}$ rows corresponding to types $A$ such that $1 \notin A$, $a_{\infty}^A = -\frac{1}{2}\mu^1\Delta_0[1]$. Thus, in general, for every type $A$, $|a_{\infty}^A| = 2^{-1}\mu^1\Delta_0[1]$ and eventual attitudes are positive if and only if $A$ is such that $1 \in A$. That is, disagreement persists.

We are left to prove that $\tau(W^{\infty}) = \mu^1$. Consider expression (3). Fixing any column in $W^{\infty}$, the maximum distance between any two rows is $\mu^1/2^{n-1}$ , which summing across the $2^n$ columns and dividing by 2 yields $\mu^1$. Finally, since $\mu^1 = \prod_{t=0}^{\infty}\lambda_t^i = \lim_{T\to\infty}\prod_{t=0}^{T}\left[1 + \sum_{i>1}\left(\Delta_0[i](\Delta_0[1])^{-1}\right)^{2^t}\right]^{-1}$, we have that, $|a_{\infty}^A| = 2^{-1}\tau(W^{\infty})\Delta_0[1]$.

Consider now that $\Delta_0[1] = \Delta_0[i]$ for some attributes $i > 1$. By step 7, every entry of $W^{\infty}$ is $(2^n)^{-1}$. In this case $a_{\infty}^A = 0$ for every type $A$. That is, consensus eventually emerges.

We then conclude that disagreement persists if and only if attribute 1 is, initially, the unique most salient attribute.

Step 9. We consider the case in which $\lambda_0^i = 0$ for some/all attributes $i > 1$. Step 1 relies on the linearity of the updating process. Thus, it still holds. Since at every $t$, $W_t$ remains stochastic, step 2 holds. For the attributes $i$ such that $\lambda_t^i > 0$, the statement in step 3 holds as well. Step 4 holds with the difference that now there are $2^n - (n+1-N)$ zero eigenvalues, where $N$ is the number of attributes $i$ such

that $\lambda_t^i = 0$. In the extreme case in which $\lambda_t^i = 0$ for every attribute $i > 1$, the column corresponding to the empty type and the $n-1$ columns corresponding to the singleton types with attributes different from 1, are the same. In such a case $N = n - 1$ and there are 2 independent columns. The eigenvalues different from zero at every $t$ are 1 because of stochasticity and $\lambda_t^1 = 1$. Since at every $t$, $W_t$ remains symmetric, step 5 holds. Step 6 deals with the existence of $W^\infty$, which is based on the existence of the limiting product of non-zero eigenvalues. It also goes through. Since the form of $W^\infty$ depends only on whether $\Delta_0[1] > \Delta_0[2]$, despite of $\lambda_t^i$ being 0 for some/all attributes $i > 1$, step 7 holds. Finally step 8, that establishes the necessary and sufficient condition for disagreement to persist, qualifying it, also holds. Notice that when $\lambda_0^i = 0$ for all attributes $i > 1$ then $W^\infty = W_0$. Also, $\mu^1 = 1$ and the equilibrium is reached at $t = 1$.

■

*Proof of Proposition 1.* We compute here the Spectral Index of Segregation at every time $t$. For this purpose we directly follow Echenique and Fryer (2007). Before proceeding recall that by step 1 in the proof of Theorem 1, positive (respectively zero) homophily values remain positive (respectively zero) all along the process. Recall also that $\sum_i \lambda_t^i = 1$ at every time $t$. Consider first the case in which for every attribute $i$, $\lambda_t^i > 0$.

Consider only types possessing attribute 1. Denote the matrix of their interactions by $\mathbf{1}_t$. Since all types have attribute 1 in common, they pay a positive amount of attention to each other, thus $\mathbf{1}_t$ has only one connected component composed by all individuals in $\mathbf{1}_t$. We now compute the largest eigenvalue of $\mathbf{1}_t$. Our claim is that $\lambda_t = \lambda_t^1 + 2^{-1}\sum_{j\neq i}^n \lambda_t^j$, with associated time independent right-eigenvector $u$ of size $2^{n-1} \times 1$, where $u$ is composed by ones, is the largest eigenvalue of $\mathbf{1}_t$. We first prove that $(\lambda_t, u)$ is a pair of eigenvalue and right-eigenvector of $\mathbf{1}_t$. Second, we argue that $\lambda_t$ is the largest eigenvalue of $\mathbf{1}_t$.

First, notice that every type $A$ shares attribute 1 with $2^{n-1}$ types. It also shares the rest of attributes with $2^{n-2}$ types. Thus, any row of $\mathbf{1}_t$ by $u$ reads $(2^{n-1}\lambda_t^1 + 2^{n-2}\sum_{j\neq i}^n \lambda_t^j)(2^{n-1})^{-1}$. This is equivalent to $\lambda_t \times 1$. Therefore, the eigenvalue equation is satisfied and $(\lambda_t, u)$ is a pair of eigenvalue and (column) eigenvector of $\mathbf{1}_t$. Second, by Perron-Froebenius Theorem, being $\mathbf{1}_t$ a positive matrix, it has a unique largest eigenvalue, which is strictly positive (that is, the spectral radius of $\mathbf{1}_t$). It is bounded above by the maximum sum of the entries of a row in $\mathbf{1}_t$ (see Meyer (2000), chapter 8). Notice that every row of $\mathbf{1}_t$ sums up to the same value, which is precisely $\lambda_t$. Suppose that there is other positive real eigenvalue, different than $\lambda_t$, which is the largest. Then it has to be also larger than the maximum sum of the entries of a row in $\mathbf{1}_t$, contradicting the Perron-Froebenius Theorem. Then, $\lambda_t$ has

to be the largest eigenvalue. We rewrite it as $\lambda_t = \lambda_t^1 + 2^{-1}(1-\lambda_t^1) = 2^{-1}(1+\lambda_t^1)$. Let us denote $SSI_t^1 = \lambda_t$. Finally, it directly follows that $\lambda_t$ increases with $\lambda_t^1$. Since $\lim_{t\longrightarrow\infty} \lambda_t^1 = 1$ then $\lim_{t\longrightarrow\infty} SSI_t^1 = 1$ as well. Also, if every attribute $i$ was initially equally salient, then $\lambda_0^i = 1/n$ for each of them. Since attribute 1 is the initially most salient, it has to be that $\lambda_0^i > 1/n$. Thus, $SSI_0^1 > (n+1)(2n)^{-1}$. Notice that the analysis is exactly the same when we consider interactions of types lacking attribute 1. In fact, the matrix of interactions is exactly $\mathbf{1}_t$. Also, in computing the $SSI_t^i$ for attributes $i > 1$, we follow similar arguments. Thus, we omit the proofs.

Consider now the case in which for attribute 1, $\lambda_t^1 > 0$ and for some/all attributes $i > 1$, $\lambda_t^i = 0$.[37] We prove here that when for an attribute $i$, $\lambda_t^i = 0$ then the $SSI_t^i$ is, at every $t$, equal to one half.[38] Given the evolution of the homophily values, as described in the proof of Theorem 1, this is also its limiting value. Recall that, by step 1 in the proof of Theorem 1, when for an attribute $i$ such that $2 \leq i \leq n$, $\lambda_t^i = 0$, this implies that $\lambda_t^j = 0$ for all attributes $j > i$. Let us focus on types possessing attribute $i$. The analysis is exactly the same when we consider interactions of types lacking attribute $i$. Two cases arise:

**C.1.** Suppose that for every attribute $j$ such that $1 < j < i$, then $\lambda_t^j = 0$, then interactions among types possessing attribute $i$ are defined by two connected components, based on the lack or possession of attribute 1. The matrices defining these two connected components are the same and have all their entries positive. One of the matrices has $2^{n-2}$ types possessing attribute 1 and the other has $2^{n-2}$ types lacking it. The analysis within each matrix is exactly the same as before. In each of them, the sum of every row is $2^{n-2}(2^{n-1})^{-1}\lambda_t^1 = 0.5$. Thus, within each component, $SSI_t^i$ equals to one half at every time $t$. Thus, the average of the $SSI_t^i$ of each component is also equal to one half.

**C.2.** Suppose that for some/all attributes $j$ such that $1 < j < i$ then $\lambda_t^j > 0$. In this case there is only one connected component. The reason is that types possessing (respectively lacking) attribute 1 are always connected among themselves and these two groups are connected between them since both contain types that are similar in attributes $j < i$, with $\lambda_t^j > 0$. The sum of the entries of every row of the matrix of interactions is $2^{n-2}(\lambda_t^1 + \sum_{j\neq i}\lambda_t^j)(2^{n-1})^{-1} = 0.5$. Thus, the index is equal to one half at every time $t$.[39] ■

*Proof of Proposition 2.* Consider the case in which all differences in average initial

[37] Recall that when all differences in average initial attitudes are equal, either positive or zero, then $\lambda_t^i = 1/n$ for every $i$ and at every $t$. Then, $SSI_t^i = (n+1)(2n)^{-1}$ for every attribute $i$ and at every $t$. See the proof of Theorem 1.

[38] When computing the $SSI_t^i$ for an attribute $i$ such that $\lambda_t^i > 0$ in the presence of attributes $j \neq i$ such that $\lambda_t^j = 0$, the interaction matrix for $i$-similar types has all its entries positive. Thus, the analysis is the same as before.

[39] In this case the matrix of interactions is just non-negative. Since it is irreducible, the Perron-Froebenius eigenvalue is equal to the sum of entries of any row in the interaction matrix, which is here always the same. The associated time independent eigenvector is $u$ of size $2^{n-1} \times 1$ with unitary entries.

attitudes are positive. In section 4 we comment on the case in which some/all differences associated to attributes $i > 1$ are zero.

To start with, we set the bounds for $T_\epsilon$ in expression (5). For this purpose recall that $T_\epsilon^{min} = min\{t : \overline{\lambda}_t^1 \geq 1-\epsilon\}$ and $T_\epsilon^{max} = min\{t : \underline{\lambda}_t^1 \geq 1-\epsilon\}$. First, let $\underline{r}_0 = \Delta_0[n]/\Delta_0[1]$. Now, consider $\lambda_t^1 = [1 + \sum_{i>1}(r_0^i)^{2^t}]^{-1}$ and replace every $r_t^i = \Delta_t[i]/\Delta_t[1]$ for attributes $i > 1$, with $\underline{r}_0$ to obtain $\overline{\lambda}_t^1 = [1 + (n-1)(\underline{r}_0)^{2^t}]^{-1}$. Notice that $\overline{\lambda}_t^1 \geq \lambda_t^1$ at every $t$. Solving $\overline{\lambda}_t^1 \geq 1-\epsilon$ for $t$, we get the expression for $T_\epsilon^{min}$, that is, $t = log\left(log\left(\frac{\epsilon}{(1-\epsilon)(n-1)}\right) log(\underline{r}_0)^{-1}\right)\frac{1}{log(2)}$. At every $t' < t$ it follows that $\overline{\lambda}_t^1 \leq 1-\epsilon$, implying that $\lambda_t^1 \leq 1-\epsilon$. Therefore, $T_\epsilon^{min}$ is a lower bound for $T_\epsilon$. Second, let $\overline{r}_0 = \Delta_0[2]/\Delta_0[1]$. Replace every $r_0^i$, for attributes $i > 1$, with $\overline{r}_0$ in $\lambda_t^1$. We get $\underline{\lambda}_t^1 = [1 + (n-1)(\overline{r}_0)^{2^t}]^{-1}$. Notice that $\underline{\lambda}_t^1 \leq \lambda_t^1$ at every $t$. Solving $\underline{\lambda}_t^1 \geq 1-\epsilon$ for $t$, we get the expression for $T_\epsilon^{max}$, that is, $t = log\left(log\left(\frac{\epsilon}{(1-\epsilon)(n-1)}\right) log(\overline{r}_0)^{-1}\right)\frac{1}{log(2)}$. At every $t' > t$ it follows that $\lambda_t^1 \geq \underline{\lambda}_t^1 \geq 1-\epsilon$. Thus, $T_\epsilon^{max}$ is an upper bound for $T_\epsilon$. Notice that making $T_\epsilon^{min}$ and $T_\epsilon^{max}$ positive is always possible, for small enough $\epsilon > 0$.

We now focus on how $T_\epsilon$ behaves with respect to changes in the the initial relative salience of attribute 1. Specifically, we do so by proving that $\lambda_t^1$ is decreasing in $r_0^i$. Recall that we consider that the variation in $r_0^i$ comes from varying $\Delta_0[i]$, for one attribute $i$ at a time. This is done in such a way that $\Delta_0[1] > \Delta_0[2] \geq ... \geq \Delta_0[n] \geq 0$ is preserved in order, as well as in magnitude for differences associated to attributes $j \neq i$. Consider the expression of $\lambda_t^1$ above. We have that $\partial\lambda_t^1/\partial r_0^i = -2^t(r_0^i)^{2^t-1}[1+\sum_{i>1}(r_0^i)^{2^t}]^{-2} < 0$. Thus, when $r_0^i$ decreases, at every time $t$ it turns out that $\lambda_t^1$ is higher than before. Therefore, the time it takes for it to be sufficiently close to its limit has to be smaller than before the change. Being $T_\epsilon$ an integer, we thus state that the time it takes for $\lambda_t^1$ to be sufficiently close to its limit cannot be higher than before. Finally, notice that $\lambda_t^1$ determines the minimum time of convergence for the system as a whole. The reason is that at every time $t$, $\lambda_t^1$ is further away from 1, its limiting value, than any of the remaining homophily values is from 0, its limiting value. To see this notice that at every time $t$, $\sum_i \lambda_t^i = 1$, then $\lambda_t^1 \geq 1-\epsilon$ implies that $\sum_{i>1}\lambda_t^i \leq \epsilon$. When only $\lambda_t^1$ and $\lambda_t^2$ are different from zero, then $\lambda_t^1 \geq 1-\epsilon$ implies that $\lambda_t^2 \leq \epsilon$. When $\lambda_t^i$ is also different from zero for some attributes $i > 2$, then $\lambda_t^1 \geq 1-\epsilon$ implies that $\epsilon/(n-1) \leq \lambda_t^2 < \epsilon$, with $\lambda_t^2 \geq \lambda_t^i$ for any attribute $i > 2$.

■

*Proof of footnote 10.* We show that $\overline{a}_0 = 0$ implies that $\overline{a}_t = 0$, at every $t$. By step 5 in the proof of Theorem 1, at every $t$, $W_t$ is diagonalizable over the same eigenspace.

Let $G$ be the projection onto the eigenspace of $W_t$ corresponding to eigenvalue 1. Let $G^i$ be the projection onto the eigenspace of $W_t$ corresponding to eigenvalue $\lambda_t^i$. By the Spectral Theorem, $W^T a_0 = Ga_0 + \sum_{i=1}^{n}(\prod_{t=0}^{T}\lambda_t^i)G^i a_0$ (see Meyer (2000), pages 517-520). We proceed by describing how row $j$ of $G^i$ looks like. Denote by $G^i_{jk}$ the $jk$ entry of $G^i$. It is constructed using eigenvectors in $U$, in step 5 in the proof of Theorem 1, as follows: $G^i_{jk} = U_{j(i+1)}U'_{(i+1)k}$. In constructing row $j$ of $G^i$, we fix column $i+1$ in $U$, i.e., the eigenvector corresponding to $\lambda_t^i$, and consider its $j$ entry. Entry $j$ takes value $1/2^{n/2}$ if $i \in A$ and $-1/2^{n/2}$ otherwise. Entry $j$ is multiplied, by the $k$ entries corresponding to row $i+1$ in $U'$, one in a turn. Notice that row $i+1$ of $U'$ is the (transposed) eigenvector associated to $\lambda_t^i$. Thus, row $j$ of $G^i$ is just the eigenvector associated to $\lambda_t^i$, divided by $1/2^{n/2}$, whenever $i \in A$ and its negative otherwise. Matrix $G$ is constructed in the same way and is composed by ones. Thus, $a_s^A = \overline{a}_0 + 2^{-1}\sum_{i=1}^{n}(-1)^{1+\mathbf{1}_i}\Delta_0[i]\prod_{t=0}^{s}\lambda_t^i$, where $\mathbf{1}_i$ is the indicator of type $A$ possessing attribute $i$. Since there are $2^{n-1}$ types possessing and lacking every attribute $i$, respectively, when summing $a_s^A$ for all types, the second term in the previous expression cancels out. Specifically, $\sum_A a_s^A = \sum_A \overline{a}_0 = 2^n\overline{a}_s$. Since $\overline{a}_0 = 0$ then at every time $s$, $\overline{a}_s = 0$. ■

*Proof of Proposition 3.* We compute here the probability that disagreement persists across attribute 1 and across attribute 2. For this purpose, let us first focus on the case in which $\Delta_0[1] > \Delta_0[2] \geq 0$, 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 $|\tilde{\Delta}_0[1]| = |\tilde{\Delta}_0[2]|$ and disagreement persists across attribute 1 (respectively attribute 2) whenever $|\tilde{\Delta}_0[1]| > |\tilde{\Delta}_0[2]|$ (respectively $|\tilde{\Delta}_0[1]| < |\tilde{\Delta}_0[2]|$). To see this notice that once initial attitudes are realized, the process exactly mimics the one presented in the main body. In what follows we describe the probability that either consensus emerges or disagreement persists. The probability that $|\tilde{\Delta}_0[1]| = |\tilde{a}_0^{\{1,2\}} - \tilde{a}_0^{\{\emptyset\}} + \tilde{a}_0^{\{1\}} - \tilde{a}_0^{\{2\}}| = |\tilde{a}_0^{\{1,2\}} - \tilde{a}_0^{\{\emptyset\}} + \tilde{a}_0^{\{2\}} - \tilde{a}_0^{\{1\}}| = |\tilde{\Delta}_0[2]|$ 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, this event has zero probability.[40] Disagreement persists across attribute 1 whenever $|\tilde{\Delta}_0[1]| = |\tilde{a}_0^{\{1,2\}} - \tilde{a}_0^{\{\emptyset\}} + \tilde{a}_0^{\{1\}} - \tilde{a}_0^{\{2\}}| > |\tilde{a}_0^{\{1,2\}} - \tilde{a}_0^{\{\emptyset\}} + \tilde{a}_0^{\{2\}} - \tilde{a}_0^{\{1\}}| = |\tilde{\Delta}_0[2]|$. 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, $P(|\tilde{\Delta}_t[1]| > |\tilde{\Delta}_t[2]|) = P(\tilde{a}_0^{\{1,2\}} - \tilde{a}_0^{\{\emptyset\}} \geq 0 \cap \tilde{a}_0^{\{1\}} - \tilde{a}_0^{\{2\}} \geq 0) + P(\tilde{a}_0^{\{1,2\}} - \tilde{a}_0^{\{\emptyset\}} < 0 \cap \tilde{a}_0^{\{1\}} - \tilde{a}_0^{\{2\}} < 0)$. Since $\tilde{a}_0^A$ are independent to each other, this is equivalent to:

[40] Notice that as $\tilde{a}_0^{\{2\}}$ is continuous, so is $-\tilde{a}_0^{\{2\}}$ as well as $\tilde{a}_0^{\{1\}} + (-\tilde{a}_0^{\{2\}})$. See Sheldon et al. (2002).

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

On the contrary, disagreement persists across attribute 2 whenever $|\tilde{\Delta}_0[1]| = |\tilde{a}_0^{\{1,2\}} - \tilde{a}_0^{\{\emptyset\}} + \tilde{a}_0^{\{1\}} - \tilde{a}_0^{\{2\}}| < |\tilde{a}_0^{\{1,2\}} - \tilde{a}_0^{\{\emptyset\}} + \tilde{a}_0^{\{2\}} - \tilde{a}_0^{\{1\}}| = |\tilde{\Delta}_0[2]|$. This expression is satisfied when $\tilde{a}_0^{\{1,2\}} - \tilde{a}_0^{\{\emptyset\}} < 0$ and $\tilde{a}_0^{\{1\}} - \tilde{a}_0^{\{2\}} \geq 0$ , or $\tilde{a}_0^{\{1,2\}} - \tilde{a}_0^{\{\emptyset\}} \geq 0$ and $\tilde{a}_0^{\{1\}} - \tilde{a}_0^{\{2\}} < 0$ hold. Then $P(|\tilde{\Delta}_0[1]| < |\tilde{\Delta}_0[2]|)$ is:

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

We can thus rewrite, $P(|\tilde{\Delta}_0[1]| > |\tilde{\Delta}_0[2]|) - P(|\tilde{\Delta}_0[1]| < |\tilde{\Delta}_0[2]|) = P(\tilde{a}_0^{\{1,2\}} - \tilde{a}_0^{\{\emptyset\}} \geq 0)(P(\tilde{a}_0^{\{1\}} - \tilde{a}_0^{\{2\}} \geq 0) - P(\tilde{a}_0^{\{1\}} - \tilde{a}_0^{\{2\}} < 0)) + P(\tilde{a}_0^{\{1,2\}} - \tilde{a}_0^{\{\emptyset\}} < 0)(P(\tilde{a}_0^{\{1\}} - \tilde{a}_0^{\{2\}} < 0) - P(\tilde{a}_0^{\{1\}} - \tilde{a}_0^{\{2\}} \geq 0))$. This expression is equivalent to $(2\mathrm{P}(_0^{\{1,2\}} - \tilde{a}_0^{\{\emptyset\}} \geq 0) - 1)(2P(\tilde{a}_0^{\{1\}} - \tilde{a}_0^{\{2} \geq 0) - 1)$. Since $\tilde{a}_0^{\{1\}} - \tilde{a}_0^{\{2\}}$ follows a symmetric distribution with positive mean (recall that $a_0^{\{1\}} - a_0^{\{2\}} > 0$), then $P(\tilde{a}_0^{\{1\}} - \tilde{a}_0^{\{2\}} \geq 0) > 0.5$.[41] A parallel argument applies in stating that $P(\tilde{a}_0^{\{1,2\}} - \tilde{a}_0^{\{\emptyset\}} \geq 0) > 0.5$. This implies that the expression above is positive. Since $P(|\tilde{\Delta}_0[1]| > |\tilde{\Delta}_0[2]|) = 1 - P(|\tilde{\Delta}_0[1]| < |\tilde{\Delta}_0[2]|)$, disagreement across attribute 1 is the most likely. In the extreme case in which $P(\tilde{a}_0^{\{1,2\}} - \tilde{a}_0^{\{\emptyset\}} \geq 0) = P(\tilde{a}_0^{\{1\}} - \tilde{a}_0^{\{2\}} \geq 0) = 1$ the, probability that disagreement takes place across attribute 1 is exactly one.

Let us consider now the case in which $\Delta_0[1] = \Delta_0[2] \geq 0$. We have that $\Delta_0[1] - \Delta_0[2] = a_0^{\{1\}} - a_0^{\{2\}} = 0$. Also, since differences are non-negative, $a_0^{\{1,2\}} - a_0^{\{\emptyset\}} \geq 0$ has to hold. This implies, again by symmetry, that $P(|\tilde{\Delta}_0[1]| > |\tilde{\Delta}_0[2]|) - P(|\tilde{\Delta}_0[1]| < |\tilde{\Delta}_0[2]|) = 0$. In this case, disagreement across either attribute is equally likely.

■

*Proof of the Remark.* Suppose that disagreement persists across attribute 1. The expression for eventual attitudes mimics the one of deterministic ones for every realization of initial attitudes. That is, $a_\infty^A = 2^{-1}\tau(W^\infty)\Delta_0[1]$ if $1 \in A$ and $a_\infty^A = -2^{-1}\tau(W^\infty)\Delta_0[1]$ if $1 \notin A$, with $\tau(W^\infty) = 1 - \Delta_0[2]/\Delta_0[1]$. That is so because both, the homophily values and the differences in attitudes, preserve their properties when these differences enter in absolute value in the Luce form. Specifically, $\Delta_t[i] = \lambda_{t-1}^i \Delta_{t-1}[i]$ holds by linearity of the process. Thus, $\Delta_t[i] \neq 0$ if and only if $\lambda_{t-1}^i \neq 0$ and $\Delta_{t-1}[i] \neq 0$. Also, given the Luce form, $\lambda_t^i \neq 0$ if and only if $\Delta_t[i] \neq 0$. Furthermore, $\Delta_0[i] > (<)\ 0$ implies that at every time $t$, $\Delta_t[i] > (<)\ 0$ and $\lambda_0^i > 0$ implies that at every time $t$, $\lambda_t^i > 0$. Also, $\Delta_0[i] = 0$ and $\lambda_0^i = 0$, imply that at every time $t$, $\Delta_t[i] = 0$ and $\lambda_t^i = 0$, respectively.[42] For every realization of

[41] The difference of independent symmetric random variables is symmetric. See Stroock (2010).

[42] For more details, see steps 1 and 7 in the proof of Theorem 1.

initial attitudes the ergodicity coefficient, $\tau(W^{\infty})$, now becomes $1-|\Delta_0[2]|/|\Delta_0[1]|$. We thus have that $\tilde{a}^A_{\infty} = \tilde{\bar{a}}_0 + 2^{-1}(1-|\tilde{\Delta}_0[2]|/|\tilde{\Delta}_0[1]|)\tilde{\Delta}_0[1]$ if $1 \in A$ and $\tilde{a}^A_{\infty} = \tilde{\bar{a}}_0 - 2^{-1}(1-|\tilde{\Delta}_0[2]|/|\tilde{\Delta}_0[1]|)\tilde{\Delta}_0[1]$ if $1 \notin A$.[43]

*Proof of Theorem 2.* It follows the same steps as the one of Theorem 1. We proceed to explain, one in a row, which of these steps still hold here. Step 1 describes a property that relies on both, the linearity of the updating process and on the Luce form. Specifically, Luce guarantees that at every $t$, $\Delta_t[i] \geq 0$ implies that $\lambda^i_t \geq 0$. By **(WDM)**, this step holds. Steps 2-5, dealing with the diagonalization of $W_t$, do not depend on the Luce form, we thus, they apply here. Step 6 only relies on non-negativity of homophily values and on the fact that $\sum_i \gamma^i_t = 1$, not on their specific form, thus it also holds. Step 7 absolutely relies on the Luce form and requires a slightly different elaboration. It is as follows:

Step 7. We first consider the case in which the limiting product of the homophily values is different from zero for one attribute $i$. Notice that this limiting product cannot be different from zero for more than one attribute. The reason is that in this case $\sum_i \lim_{t\longrightarrow\infty} \gamma^i_t > 1$ contradicting the properties of $\gamma^i_t$. We prove that $\prod_{t=0}^{\infty} \gamma^i_t = \mu^i$ with $\mu^i \in (0,1)$ for attribute $i$ implies that $\prod_{t=0}^{\infty} \gamma^j_t = 0$ for attributes $j \neq i$. We also show that if there is such an attribute, it has to be attribute 1. Second, we consider the case in which the limiting product of homophily values is zero for all attributes. Before proceeding, recall that $\lim_{t\longrightarrow\infty} \gamma^i_t$ exists for every attribute $i$ and $\lim_{t\longrightarrow\infty} \sum_i \gamma^i_t = \sum_i \lim_{t\longrightarrow\infty} \gamma^i_t = 1$.

First, suppose that $\prod_{t=0}^{\infty} \gamma^i_t = \mu^i$ with $\mu^i \in (0,1)$, or equivalently, that $\sum_{t=0}^{\infty} log(\gamma^i_t)$ exists. This implies that $\lim_{t\longrightarrow\infty} \gamma^i_t = 1$. Thus, $\lim_{t\longrightarrow\infty} \gamma^j_t = 0$ for every attribute $j$, implying that for non of them, $\prod_{t=0}^{\infty} \gamma^j_t = \mu^j$ with $\mu^j \in (0,1)$, but $\prod_{t=0}^{\infty} \gamma^j_t = 0$. Now, recall that by **(ADM)** and step 1 in the proof of Theorem 1, at every $t$, $\Delta_t[1] \geq \Delta_t[2] \geq \cdots \geq \Delta_t[n] \geq 0$ and $\gamma^1_t \geq \gamma^2_t \geq \cdots \geq \gamma^n_t \geq 0$ hold. Suppose, that there is an attribute $i > 1$ for which $\prod_{t=0}^{\infty} \gamma^i_t = \mu^i$ with $\mu^i \in (0,1)$ holds. This implies that $\lim_{t\longrightarrow\infty} \gamma^i_t = 1$ and $\lim_{t\longrightarrow\infty} \gamma^1_t = 0$. Thus, for high enough $t$, $\gamma^i_t$ would be arbitrarily close to 1 while $\gamma^1_t$ would be arbitrarily close to 0. Given that, at every $t$, $\gamma^1_t \geq \gamma^2_t \geq \cdots \geq \gamma^n_t \geq 0$ holds, the former statement cannot be true. We therefore conclude that $\prod_{t=0}^{\infty} \gamma^1_t = \mu^1$ with $\mu^1 \in (0,1)$ and $\prod_{t=0}^{\infty} \gamma^i_t = 0$ for attributes $i > 1$. Under this scenario we have that:

[43] We do not impose here that for every realization of initial attitudes their average is equal to zero.

$$W^{\infty} = \frac{1}{2^n} \begin{bmatrix} 1+\mu^1 & \cdots & 1+\mu^1 & 1-\mu^1 & \cdots & 1-\mu^1 \\ 1+\mu^1 & \cdots & 1+\mu^1 & 1-\mu^1 & \cdots & 1-\mu^1 \\ \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ 1+\mu^1 & \cdots & 1+\mu^1 & 1-\mu^1 & \cdots & 1-\mu^1 \\ 1+\mu^1 & \cdots & 1+\mu^1 & 1-\mu^1 & \cdots & 1-\mu^1 \\ 1-\mu^1 & \cdots & 1-\mu^1 & 1+\mu^1 & \cdots & 1+\mu^1 \\ 1-\mu^1 & \cdots & 1-\mu^1 & 1+\mu^1 & \cdots & 1+\mu^1 \\ \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ 1-\mu^1 & \cdots & 1-\mu^1 & 1+\mu^1 & \cdots & 1+\mu^1 \\ 1-\mu^1 & \cdots & 1-\mu^1 & 1+\mu^1 & \cdots & 1+\mu^1 \end{bmatrix},$$

where $\mu^1 = \prod_{t=0}^{\infty} \gamma_t^1$.

Second, suppose that $\prod_{t=0}^{\infty} \gamma_t^1 = 0$ and $\lim_{t \longrightarrow \infty} \gamma_t^1 = 1$. It implies that $\lim_{t \longrightarrow \infty} \gamma_t^j = 0$ for every attribute $i > 1$. Thus, no attribute $i > 1$ is such that $\prod_{t=0}^{\infty} \gamma_t^i = \mu^i$ with $\mu^i \in (0,1)$. Suppose now that $\prod_{t=0}^{\infty} \gamma_t^1 = 0$ and $\lim_{t \longrightarrow \infty} \gamma_t^1 = \alpha$ with $\alpha \in (0,1)$. Then, $\sum_{i>1} \lim_{t \longrightarrow \infty} \gamma_t^i = 1 - \alpha$ with $1 - \alpha \in (0,1)$. Thus, no attribute $i$ is such that $\prod_{t=0}^{\infty} \gamma_t^i = \mu^i$ with $\mu^i \in (0,1)$. Finally, suppose that $\prod_{t=0}^{\infty} \gamma_t^1 = 0$ and $\lim_{t \longrightarrow \infty} \gamma_t^1 = 0$. Then, either for exactly one attribute $i > 1$, $\lim_{t \longrightarrow \infty} \gamma_t^i = 1$, or for some (possibly all) attributes $i > 1$, $\sum_{i>1} \lim_{t \longrightarrow \infty} \gamma_t^i = 1$. None of these cases can hold. The reason is that for high enough $t$, some $\gamma_t^i$ would be arbitrarily close to a positive number (which is 1 when for exactly one attribute $i > 1$, $\lim_{t \longrightarrow \infty} \gamma_t^i = 1$) while $\gamma_t^1$ would be arbitrarily close to 0. Since, at every $t$, $\gamma_t^1 \geq \gamma_t^2 \geq \cdots \geq \gamma_t^n \geq 0$ holds, this cannot be true. We therefore conclude that in all these cases $\prod_{t=0}^{\infty} \gamma_t^i = 0$ for all attributes. Under this scenario all entries of $W^{\infty}$ are $(2^n)^{-1}$.

Step 8. By step 7, the necessary and sufficient condition for disagreement to persist is that $\sum_{t=0}^{\infty} log \gamma_t^1$ exists. When this is the case, eventual attitudes take the same form as in Theorem 1, where now $\tau(W^{\infty}) = \lim_{T \to \infty} \prod_{t=0}^{T} \gamma_t^1 \in (0,1)$.

Step 9. It depends partially on Luce since it guarantees that, at every time $t$, $\Delta_t[i] \geq 0$ implies that $\lambda_t[i] \geq 0$. We use **(WDM)** to cover this aspect, thus this step holds. ■

## References


Daron Acemoglu and Asuman Ozdaglar. Opinion dynamics and learning in social networks. *Dynamic Games and Applications*, 1(1):3–49, 2011.

Daron Acemoglu, Asuman Ozdaglar, and Ali ParandehGheibi. Spread of (mis) information in social networks. *Games and Economic Behavior*, 70(2):194–227, 2010.

Daron Acemoglu, Giacomo Como, Fabio Fagnani, and Asuman Ozdaglar. Opinion fluctuations and disagreement in social networks. *Mathematics of Operations Research*, 38(1):1–27, 2013.

Alberto Alesina and Guido Tabellini. A positive theory of fiscal deficits and government debt. *The Review of Economic Studies*, 57(3):403–414, 1990.

James Andreoni and Tymofiy Mylovanov. Diverging opinions. *American Economic Journal: Microeconomics*, pages 209–232, 2012.

Abhijit Banerjee and Drew Fudenberg. Word-of-mouth learning. *Games and Economic Behavior*, 46(1):1–22, 2004.

Marianne Bertrand. New perspectives on gender. *Handbook of labor economics*, 4: 1543–1590, 2011.

Pedro Bordalo, Nicola Gennaioli, and Andrei Shleifer. Salience and consumer choice. *Journal of Political Economy*, 121(5):803–843, 2013.

Jordi Brandts, Ayça Ebru Giritligil, and Roberto A. Weber. An experimental study of persuasion bias and social influence in networks. Technical report, BELIS, Istanbul Bilgi University, 2014.

Tiago VV Cavalcanti, Chryssi Giannitsarou, and Charles R Johnson. Network cohesion. *Economic Theory*, pages 1–21, 2012.

Arun Chandrasekhar, Horacio Larreguy, and Juan Pablo Xandri. Testing models of social learning on networks: Evidence from a framed field experiment. *Work. Pap., Mass. Inst. Technol., Cambridge, MA*, 2012.

Samprit Chatterjee and Eugene Seneta. Towards consensus: some convergence theorems on repeated averaging. *Journal of Applied Probability*, pages 89–97, 1977.

Hsiao-Chi Chen, James W. Friedman, and Jacques-Francois Thisse. Boundedly rational nash equilibrium: a probabilistic choice approach. *Games and Economic Behavior*, 18(1):32–54, 1997.

Andrew E. Clark. Job satisfaction and gender: why are women so happy at work? *Labour economics*, 4(4):341–372, 1997.

Luca Corazzini, Filippo Pavesi, Beatrice Petrovich, and Luca Stanca. Influential listeners: An experiment on persuasion bias in social networks. *European Economic Review*, 56(6):1276–1288, 2012.

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

Peter M. DeMarzo, Jeffrey Zwiebel, and Dimitri Vayanos. Persuasion bias, social influence, and unidimensional opinions. *The Quarterly Journal of Economics*, 118 (3):909–968, 2003.

Federico Echenique and Roland G. Fryer. A measure of segregation based on social interactions. *The Quarterly Journal of Economics*, 122(2):441–485, 2007.

Glenn Ellison and Drew Fudenberg. Rules of thumb for social learning. *Journal of political Economy*, 101(4):612–643, 1993.

Noah E. Friedkin and Eugene C. Johnsen. Social influence networks and opinion change. *Advances in group processes*, 16(1):1–29, 1999.

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

Benjamin Golub and Matthew O. Jackson. Naive learning in social networks and the wisdom of crowds. *American Economic Journal: Microeconomics*, 2:112–149, 2010.

Benjamin Golub and Matthew O. Jackson. How homophily affects the speed of learning and best-response dynamics. *The Quarterly Journal of Economics*, 127 (3):1287–1338, 2012.

Veronika Grimm and Friederike Mengel. An experiment on belief formation in networks. *Available at SSRN 2361007*, 2014.

Rainer Hegselmann and Ulrich Krause. Opinion dynamics and bounded confidence models, analysis, and simulation. *Journal of Artificial Societies and Social Simulation*, 5(3), 2002.

Ilse C. F. Ipsen and Teresa M. Selee. Ergodicity coefficients defined by vector norms. *SIAM Journal on Matrix Analysis and Applications*, 32(1):153–200, 2011.

Matthew O. Jackson. *Social and economic networks*. Princeton University Press, 2008.

Gueorgi Kossinets and Duncan J. Watts. Empirical analysis of an evolving social network. *Science*, 311(5757):88–90, 2006.

Botond Kőszegi and Adam Szeidl. A model of focusing in economic choice. *The Quarterly Journal of Economics*, 128(1):53–104, 2013.

Ulrich Krause. A discrete nonlinear and non-autonomous model of consensus formation. *Communications in difference equations*, pages 227–236, 2000.

William J McGuire, Claire V McGuire, Pamela Child, and Terry Fujioka. Salience of ethnicity in the spontaneous self-concept as a function of one's ethnic distinctiveness in the social environment. *Journal of personality and social psychology*, 36(5):511, 1978.

Carl D. Meyer. *Matrix analysis and applied linear algebra*. SIAM, 2000.

Keith T. Poole and R. Steven Daniels. Ideology, party, and voting in the us congress, 1959–1980. *American Political Science Review*, 79(2):373–399, 1985.

Milton E. Rosenbaum. The repulsion hypothesis: On the nondevelopment of relationships. *Journal of Personality and Social Psychology*, 51(6):1156, 1986.

Thomas C. Schelling. Models of segregation. *The American Economic Review*, 59 (2):488–493, 1969.

Ross Sheldon et al. *A first course in probability*. Pearson Education India, 2002.

Ramadhar Singh and Soo Yan Ho. Attitudes and attraction: A new test of the attraction, repulsion and similarity-dissimilarity asymmetry hypotheses. *British Journal of Social Psychology*, 39(2):197–211, 2000.

Lones Smith and Peter Sørensen. Pathological outcomes of observational learning. *Econometrica*, 68(2):371–398, 2000.

John Stachurski. *Economic dynamics: theory and computation*. MIT Press, 2009.

Daniel W. Stroock. *Probability theory: an analytic view*. Cambridge university press, 2010.

Irina Suanet and Fons J.R. Van de Vijver. Perceived cultural distance and acculturation among exchange students in russia. *Journal of Community & Applied Social Psychology*, 19(3):182–197, 2009.

Zannie Giraud Voss, Daniel M. Cable, and Glenn B. Voss. Organizational identity and firm performance: What happens when leaders disagree about who we are?. *Organization Science*, 17(6):741–755, 2006.