Anufriev, Mikhail; Borissov, Kirill; Pakhnin, Mikhail

Working Paper

Dissonance Minimization and Conversation in Social Networks

CESifo Working Paper, No. 9433

Provided in Cooperation with:
Ifo Institute – Leibniz Institute for Economic Research at the University of Munich

Suggested Citation: Anufriev, Mikhail; Borissov, Kirill; Pakhnin, Mikhail (2021) : Dissonance Minimization and Conversation in Social Networks, CESifo Working Paper, No. 9433, Center for Economic Studies and ifo Institute (CESifo), Munich

This Version is available at:
https://hdl.handle.net/10419/248978

CESifo WORKING PAPERS

9433
2021

November 2021

Dissonance Minimization and Conversation in Social Networks

Mikhail Anufriev, Kirill Borissov, Mikhail Pakhnin

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 9433*

Dissonance Minimization and Conservation in Social Networks

Abstract

We study a model of social learning in networks where the dynamics of beliefs are driven by conversations of dissonance-minimizing agents. Given their current beliefs, agents make statements, tune them to the statements of their associates, and then revise their beliefs. We characterize the long-run beliefs in a society, provide the necessary and sufficient conditions for a society to reach a consensus, and show that agents' social influences (weights on the consensus belief) are decreasing in their dissonance sensitivities. Comparing the outcomes of two models, with and without conversation, we show that conversation leads to a redistribution of social influences in favor of agents with higher self-confidence. Finally, we provide analytical insights for the model where agents minimize dissonance by revising both beliefs and network, and show that an endogenous change of network may prevent a society from reaching a consensus.

JEL-Codes: D830, D850, D910, Z130.

Keywords: social networks, DeGroot learning, social influence, dissonance minimization, conversation.

Mikhail Anufriev
Economics Discipline Group, Business School
University of Technology Sydney / Australia
mikhail.anufriev@uts.edu.au

Kirill Borissov
Department of Economics
European University St. Petersburg / Russia
kirill@eu.spb.ru

Mikhail Pakhnin
Department of Economics
European University St. Petersburg / Russia
mpakhnin@eu.spb.ru

This draft: November 2021
The authors are grateful to Jasmina Arifovic, Berno Buechel, Andrea Giovannetti and Paolo Pin for stimulating discussion and comments on earlier drafts of this paper. Mikhail Anufriev gratefully acknowledges the hospitality of the European University at St. Petersburg and the financial support of the Australian Research Council through Discovery Project DP170100429. Mikhail Pakhnin is grateful to PAO Severstal for its support.

1 Introduction

Opinions and beliefs affect many economic decisions, such as whether to buy a new product or to invest in a financial asset, or even which political candidate to vote for. A central role in the formation of opinions and beliefs is played by social networks; that is, by the sets of those individuals with whom people regularly interact and communicate. We learn about the quality of a new product, the prospects of financial assets, the political views of a candidate, and so on, while conversing with relatives, friends, co-workers, and acquaintances. However, people often misrepresent their opinions in conversation. Social psychologists suggest that one of the main reasons for this is the cognitive dissonance that people experience when faced with opposing opinions, as a result of which, in response, they tailor their own messages to their audience (Higgins, 1999; Echterhoff et al., 2005). In this paper, we incorporate conversation within the dissonance minimization framework, and study the role of both conversation and cognitive dissonance parameters in belief formation.

A large and growing body of economic literature has investigated the process of belief formation in social networks, focusing on such questions as whether and how quickly individual beliefs converge, whether society reaches consensus, and whether this consensus reflects the true state of the world. Surveys in Jackson (2008, chapter 8) and Golub and Sadler (2016) divide contributions in this field into two categories: those based on Bayesian learning models (e.g., Bala and Goyal, 1998) and those that adopt a repeated linear updating setting of DeGroot (1974). Recent examples of the former approach include Acemoglu et al. (2011) and Eyster and Rabin (2014). The Bayesian approach is attractive since it serves as a full rationality benchmark. Bayesian agents, in particular, would adjust the information they receive from their associates on a possible common source or on repetitions of information over subsequent communications. However, experimental and empirical evidence (Choi et al., 2008; Corazzini et al., 2012; Chandrasekhar et al., 2020) suggests that, in many situations, the less sophisticated DeGroot learning describes individual behavior better.

In the DeGroot (1974) model, belief dynamics are given by an average-based updating process. Agents are embedded in a network described by the interaction matrix, and agents' new beliefs are a weighted average of their current beliefs where the weights (trust parameters) are given by the interaction matrix. DeMarzo et al. (2003) modelled communication on a social network, using the DeGroot model, and argued that implied "persuasion bias" (agents' failure to discount repetitions of in-

formation) can explain such phenomena as political propaganda, censorship, and unidimensionality of political beliefs. Further important contributions in this stream of literature were made by Golub and Jackson (2010, 2012) and Buechel et al. (2015).[1] The DeGroot learning model is highly tractable due to its close connection to the well-established Markov chain theory, which greatly helps in analyzing the long-run properties of the model.

An important part of the story is that people often hear beliefs or views that contradict their own.[2] There is evidence that the amount of disagreement in society over many important issues has increased over time. For instance, according to the General Social Survey, in 1972 the percentage of the US population in favor of gun permits was 71% among Democrats and 70% among Republicans, while by 2018 those figures had altered to 84% and 59% respectively. Similarly, between the 1970s and 2018, the gap between the percentage of Democrats and Republicans in favor of abortion or the death penalty for murder grew substantially.[3] Moreover, it has been reported that the average gap between the views of Democrats and Republicans on 10 different political values increased from 15 p.p. to 36 p.p. between 1994 and 2017.[4]

Following Festinger (1957), the literature of social psychology suggests that people suffer from "cognitive dissonance" when confronted with opposing beliefs and that they react by minimizing the dissonance. Social psychologists have found that many decisions and attitudes are determined by people's desire to reduce cognitive dissonance (e.g., McGrath, 2017; Harmon-Jones, 2019). In particular, it has been argued that cognitive dissonance is essentially a *social phenomenon*, which stems from disagreement with others in a social group (Matz and Wood, 2005; McKimmie, 2015). Thus social networks are sources of both belief formation and dissonance arousal. In this paper, we link these two phenomena. We model people's conversations and network formations as motivated by their dissonance minimization reaction to the disagreement on the views expressed by associates in their network. We investigate how these psychological factors affect convergence of beliefs, possibilities of reaching consensus, and the social influences of different people.

A recent paper by Arifovic et al. (2015) developed a computational model of learn-

[1]See also closely related models on evolution of different social phenomena in Merlone and Radi (2014); Buechel et al. (2014); Panebianco and Verdier (2017); Olcina et al. (2017); Della Lena (2019); Ushchev and Zenou (2020).

[2]This may be caused by various preferences, sources of information, and embedded values. We abstract away from the source of disagreement but focus on the psychological consequences of it.

[3]The General Social Survey key trends, https://gssdataexplorer.norc.org/trends.

[4]Pew Research Center, October, 2017, "The Partisan Divide on Political Values Grows Even Wider".

ing in a social network, where agents minimize the dissonance arising from disagreement by not expressing their genuine opinions, instead tailoring their statements to the statements of their associates.[5] We build our paper in the framework proposed by Arifovic et al. (2015) and go beyond numerical simulations. Our goal is to formalize the role of conversation in belief evolution, to clarify assumptions, and to generalize the model in different directions. For these purposes, we provide an analytical solution to the dissonance minimization model with conversation, fully characterize the dynamics of beliefs, including the convergence of society to a consensus and the social influences of different members of society, and study the influence of dissonance sensitivities on these outcomes. Furthermore, we disentangle the effects of conversation and dissonance minimization by comparing the outcomes of the model with conversation with those of the model without conversation, the latter having been known as "folk wisdom" in the literature on the DeGroot model (see Groeber et al., 2014; Golub and Sadler, 2016).

We now summarize our main contributions. First, we show that repeated conversations, motivated by minimization of the dissonance, lead to the DeGroot learning model of opinion averaging. We relate the "trust" parameters in the DeGroot interaction matrix to behavioral parameters that reflect the sensitivity to cognitive dissonance. Intriguingly, we show that conversation leads to a much broader propagation of beliefs than the network of dissonance-arousing associates implies. Specifically, the belief of an agent is affected not only by the beliefs of their associates, but also by the beliefs of the associates of their associates, and by *their* associates' beliefs, and so on. We also show that the block structure of the DeGroot interaction matrix, which is studied in particular in DeMarzo et al. (2003) and Buechel et al. (2015), can be obtained by partitioning a society into communication classes and remaining agents on the basis of dissonances. In this sense, our model provides certain microfoundations for the DeGroot model in both connected and non-connected societies.

Second, we prove that, in the model with conversation, beliefs always converge to the long-run values (an outcome which is not guaranteed in the standard DeGroot model). We fully characterize the dynamics of beliefs, providing necessary and sufficient conditions for a society with diverse initial beliefs to reach a consensus, when the driving forces behind the evolution of beliefs are dissonance minimization and conversation. Within each communication class, agents have the same long-run

[5]This effect is known in the social psychology literature as "audience tuning". Higgins (1999) and Echterhoff et al. (2005) argue that the stated opinion also alters one's own memory (the so-called "saying is believing" effect).

belief, and their weights on this belief are agents' social influences. The long-run belief of each remaining agent is determined by the beliefs of communication classes, and the weights of communication classes on each such belief are the classes' external impacts. We relate both social influences and external impacts to the dissonance sensitivities of agents. Moreover, we highlight the role of conversation by showing that social influences in models with and without conversation are in general different. Interestingly, conversation leads to a redistribution of social influences in favor of agents with higher self-confidence. At the same time, external impacts are the same in both models.

Third, we provide analytical insights for the model where the dynamics of both beliefs and network are driven by dissonance minimization. We prove that when a network is endogenous, an initially polarized society may remain polarized even in the long run. We show that this outcome is non-monotonic in the dissonance sensitivity.

The paper is organized as follows. Section 2 studies the dissonance minimization model with conversation and discusses its relation to the DeGroot model. Section 3 characterizes long-run beliefs in terms of social influences and external impacts. Section 4 studies the effects of conversation and dissonance sensitivities on the outcome of the model. Section 5 provides an example of dynamics for the model with endogenous network. Section 6 concludes. Appendix A discusses the numerical results of Arifovic et al. (2015). Appendix B contains the proofs of the main results. Appendix C provides detailed description of and proofs for the model with endogenous network.

2 Dissonance minimization in conversation

A society consists of a set $\mathcal{N} = \{1, \ldots, N\}$ of agents updating their opinions about a certain issue in discrete time. In each period t, each agent i forms a belief $b_i(t)$ about the issue.[6] Let $\mathbf{b}(t)$ denote the (column) vector of agents' beliefs. The vector of initial beliefs $\mathbf{b}(0)$ is given.

Motivated by the literature of social psychology and the computational study of Arifovic et al. (2015) (AEW henceforth), we consider the setting where agents discuss the issue and experience dissonance from disagreement. Every agent i has a set of *associates*; that is, of agents who affect i's dissonance. Within each period $t \geq 1$,

[6]We use the term "belief" in a broad sense, spanning opinions, judgments, and estimations. For instance, one can think of $b_i(t)$ as a probability of some event. We assume that beliefs are real numbers, but the results can be extended to the case where they are elements of an arbitrary normed space.

the agents first participate in conversation and then form beliefs. Before presenting the full model, it will be useful to consider the setting without conversation, which is sometimes used to justify the DeGroot learning model (cf. Golub and Sadler, 2016).

Model without conversation. Suppose that in each period t, given $\mathbf{b}(t-1)$, agent i faces the following problem

$$\min_{b_i(t)} \left\{ (b_i(t) - b_i(t-1))^2 + \sum\nolimits_{j=1}^{N} d_{ij}(b_i(t) - b_j(t-1))^2 \right\}, \tag{1}$$

where coefficients $d_{ij} \geq 0$ are given. An interpretation of this problem is that agent i experiences dissonance when the new belief differs from the previous beliefs in a society. For tractability, the dissonance dis-utility in (1) is quadratic and additively separable across agents. The first term is the *private dissonance* arising from the agent's inconsistecy, and the second term is the *social dissonance* arising from the agent's disagreement with their associates. We normalize the sensitivity to private dissonance to 1 and set $d_{ii} = 0$ for every i. For two distinct agents i and j, parameter d_{ij} measures the sensitivity of agent i to disagreement with agent j, with $d_{ij} = 0$ indicating that agent j is not among the associates of i. Therefore, matrix $\mathbf{D} = \{d_{ij}\}_{i,j=1}^{N}$ characterizes both the sets of associates and the sensitivity to them. We call $\mathbf{D}$ the *dissonance matrix*.

The first-order condition of problem (1) implies that the new belief of agent i is a weighted average of i's previous belief and the previous beliefs of i's associates:

$$b_i(t) \;=\; t_{ii} b_i(t-1) + \sum\nolimits_{j=1}^{N} t_{ij} b_j(t-1)\,,$$

where the self-weight, t_{ii}, and the weights to the associates, t_{ij}, are given by

$$t_{ii} = \frac{1}{1 + \sum_k d_{ik}} \qquad \text{and} \qquad t_{ij} = \frac{d_{ij}}{1 + \sum_k d_{ik}}\,, \quad \text{for } i \neq j\,. \tag{2}$$

Thus, when agents minimize their dissonances caused by belief differences, the dynamics of beliefs follow the DeGroot learning process in the form $\mathbf{b}(t) = \mathbf{T}\mathbf{b}(t-1)$,

where the *row-stochastic* matrix $\mathbf{T}$ is defined as[7]

$$\mathbf{T} = \left(\mathbf{I}_N + \operatorname{diag}\left\{\mathbf{D}\mathbb{1}_N\right\}\right)^{-1}\left(\mathbf{I}_N + \mathbf{D}\right) . \tag{3}$$

The dissonance matrix $\mathbf{D}$, capturing all dissonance sensitivities, defines this so-called $\mathbf{T}$-model. In the corresponding $\mathbf{T}$-network, the links are directed from the agents to themselves and are weighted by coefficients given in (2). Note that in the $\mathbf{T}$-model the self-weights that measure agents' self-importance are all positive, $t_{ii} > 0$.

The dissonance minimization procedure leading to the $\mathbf{T}$-model assumes that, in each period, the beliefs of agents are known to their associates, as if there has been a conversation where the agents stated their beliefs truthfully. We now consider the model in which the distinction between beliefs and statements is made explicit.

2.1 Model with conversation

Suppose that in period t, each agent i first makes a statement, $s_i(t)$, and then forms a belief, $b_i(t)$, about the issue. The statement and the belief are affected by the previous belief of i and the contemporaneous statements of i's associates. We model the conversation stage as a game where all N agents simultaneously choose their statements. Agent i chooses the statement $s_i(t)$ by solving

$$\min_{s_i(t)}\left\{(s_i(t) - b_i(t-1))^2 + \sum\nolimits_{j=1}^{N} d_{ij}(s_i(t) - s_j(t))^2\right\} . \tag{4}$$

This game has a unique Nash equilibrium, the vector of statements $\mathbf{s}(t)$, as we show below. Dissonance dis-utility in (4) has both *private dissonance* arising from the agent being dishonest, and *social dissonance* arising from them disagreeing with their associates in conversation. The dissonance matrix $\mathbf{D}$ is a primitive of the model.[8] It contains sensitivity parameters and defines the set of associates of any agent i as those agents for whom $d_{ij} > 0$.

After the statements are made, agent i forms a new belief by solving

$$\min_{b_i(t)}\left\{(b_i(t) - b_i(t-1))^2 + \sum\nolimits_{j=1}^{N} d_{ij}(b_i(t) - s_j(t))^2\right\} . \tag{5}$$

[7]Throughout the paper, we use the following notation: $\mathbf{I}_N$ is the identity matrix of size N, $\mathbb{1}_N$ is the N-vector of ones, $\operatorname{diag}\{\mathbf{D}\mathbb{1}_N\}$ is the diagonal matrix whose entries are the row sums of $\mathbf{D}$. A *row-stochastic* matrix is the non-negative matrix with the sum of elements in each row equal to 1.
[8]That is, the model is defined by any non-negative square matrix $\mathbf{D}$ with zeros on the diagonal.

Note that this optimization problem coincides with problem (1) in the **T**-model only if the agents state their beliefs truthfully. This is generally not the case, however. To see this, consider the outcome of the conversation stage.

Due to strict convexity of the objective function in $s_i(t)$, problem (4) implies

$$s_i(t) \;=\; \frac{1}{1+\sum_k d_{ik}} b_i(t-1) + \sum\nolimits_{j=1}^{N} \frac{d_{ij}}{1+\sum_k d_{ik}} s_j(t)\,, \qquad \text{for any } i\,. \tag{6}$$

The linear system of these N equations can be written in the matrix form as

$$\mathbf{b}(t-1) = \left(\mathbf{I}_N + \operatorname{diag}\left\{\mathbf{D}\mathbb{1}_N\right\} - \mathbf{D}\right)\mathbf{s}(t). \tag{7}$$

Let us now define matrix $\mathbf{P}$, which plays the key role in the belief dynamics of the model with conversation,

$$\mathbf{P} = \left(\mathbf{I}_N + \operatorname{diag}\left\{\mathbf{D}\mathbb{1}_N\right\} - \mathbf{D}\right)^{-1}\,. \tag{8}$$

As we show further, this matrix $\mathbf{P}$ is well defined. Then it follows from (7) that there is a unique Nash equilibrium at the conversation stage, $\mathbf{s}(t) = \mathbf{P}\mathbf{b}(t-1)$.

Proposition 1 (**P-model**). *Let* $\mathbf{D}$ *be a non-negative square matrix of size* N *with zeros on the main diagonal. Beliefs in the model with conversation induced by* $\mathbf{D}$ *evolve as* $\mathbf{b}(t) = \mathbf{P}\mathbf{b}(t-1)$*, where matrix* $\mathbf{P}$ *defined in* (8) *is row-stochastic.*

Proof. In Appendix B.1, we show that the matrix $\mathbf{P}$ is well defined by (8) and row-stochastic. Since problems (5) for the belief and (4) for the statement are identical for any agent, we have $\mathbf{b}(t) = \mathbf{s}(t)$. The belief dynamics are then obtained from (7). ■

Proposition 1 establishes that dissonance minimization in conversation gives rise to linear updating of the beliefs in a society; that is, to the DeGroot learning process. We call this model the **P**-model and the associated network the **P**-network, as the beliefs are updated by matrix $\mathbf{P}$ defined in (8). Dissonance minimization, thus, microfounds DeGroot learning, connecting belief dynamics to the sensitivity parameters given exogenously. Our goal will be to express the properties of the belief dynamics in terms of the elements of the dissonance matrix $\mathbf{D}$.

It will also be informative to compare the **P**-model with the **T**-model. Indeed, both models lead to the DeGroot belief updating, but it is the **P**-model with conversation that highlights the role of dissonance that people experience in discussions.

Eq. (6) shows that when agents make statements, they attach positive weights to the statements of their associates. In other words, agents "tune to an audience", tailoring their statements to make them closer to the statements of their associates.[9] Due to audience tuning, $\mathbf{s}(t)$ and $\mathbf{b}(t-1)$ are in general different, and this distinguishes the model with conversation from the **T**-model.[10] As the following example demonstrates, conversation affects the structure of the interaction matrix **P** much more strongly than a simple normalization of the dissonance matrix in the **T**-model, given by (2).

Example 1. Consider the "chain" society of $N \geq 2$ agents, ordered from 1 to N in such a way that any agent experiences social dissonance to the previous agent only, except for the first agent, who has no social dissonance. All sensitivities are set to 1. In matrix **D**, thus, the non-zero elements are only $d_{i+1,i} = 1$ for any $i < N$.

Matrices **T** and **P** are computed using (3) and (8), respectively. In both models, the first agent experiences no social dissonance, attaching self-weight 1. In the **T**-model, the remaining agents weight themselves and the previous agent with equal weights 1/2. But in the **P**-model, agents weight *all* previous agents. The self-weight is 1/2, the weight of the previous agent is 1/4, the weight of the agent before the previous agent is 1/8, and so on, with agents 1 and 2 always being given the same weight, $1/2^{N-1}$. See illustrations in Figs. 1 and 2 for $N = 3$ and $N = 4$.

In matrices **T** and **P**, the first two rows are identical, but all other rows are very different. In the model with conversation, agent 3 places weight 1/4 to the belief of agent 1, who is not their associate and may not even be known to them. This becomes even more striking when N becomes large: note that for any size of society, the last agent N places positive weight to agent 1, even if they are separated by $N-2$ agents!

Role of audience tuning. Mutual audience tuning plays a crucial role in the **P**-model. In Example 1, the dissonance minimization problem of agent 3 includes the

[9]Thus dissonance minimization leads to conformity but not to counter-conformity. Buechel et al. (2015) study both. In contrast to that paper, agents in our model can have different sensitivities to disagreement with different associates.

[10]However, the belief update stage implies that $\mathbf{b}(t) = \mathbf{s}(t)$, the so-called "*saying is believing*" effect according to AEW. We follow their setting, where the statements of associates affect both the agent's statement and their newly formed belief via problems (4) and (5), respectively. Anufriev et al. (2021) consider a general case where associates causing an agent's dissonance in conversation differ from the people to whom the agent listens when forming their beliefs. In this case, the "saying is believing" effect disappears.

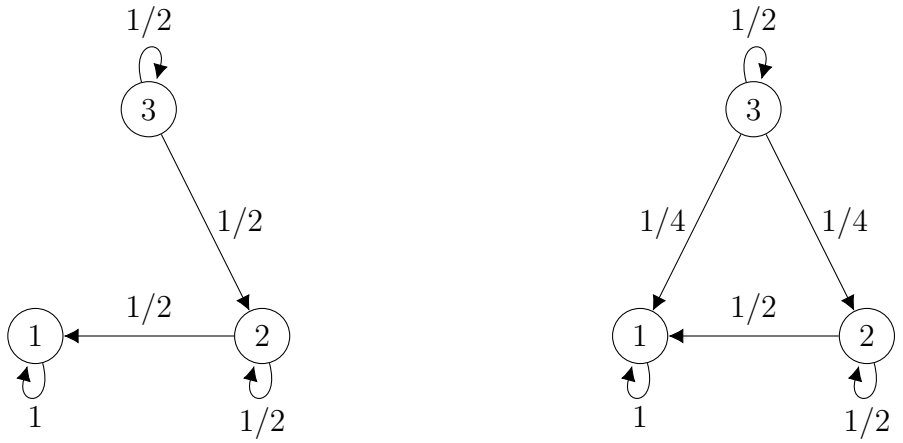

Figure 1: **T**-network (left) and **P**-network (right) in Example 1 with $N = 3$.

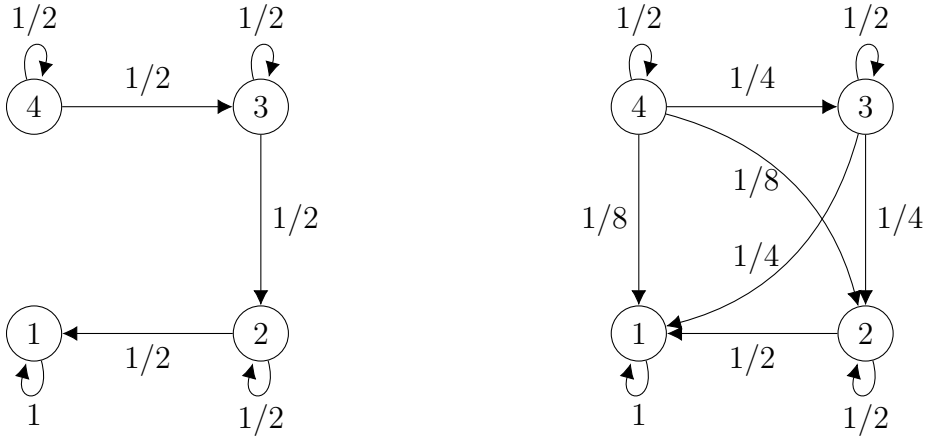

Figure 2: **T**-network (left) and **P**-network (right) in Example 1 with $N = 4$.

statement of agent 2, whose dissonance is affected by the statement of agent 1. This makes the belief of agent 3 dependent on the previous belief of agent 1.

Another assumption of the model is that all statements are determined simultaneously. This assumption is not crucial, in the sense that the statements can alternatively be obtained as a limit of a naïve adjustment process. Suppose that within period t there is a fast timescale $\tau \geq 0$. At each date τ, agent i chooses statement $s_i^\tau(t)$ to minimize the dissonance with their own belief $b_i(t-1)$ (which is fixed in this fast timescale) and the *previous* statements of their associates $s_j^{\tau-1}(t)$:

$$\min_{s_i^\tau(t)} \left\{ (s_i^\tau(t) - b_i(t-1))^2 + \sum\nolimits_{j=1}^{N} d_{ij}(s_i^\tau(t) - s_j^{\tau-1}(t))^2 \right\}.$$

The first-order conditions imply that during period t the statements evolve as

$$\mathbf{s}^{\tau}(t) = \left(\mathbf{I}_N + \operatorname{diag}\left\{\mathbf{D}\mathbb{1}_N\right\}\right)^{-1} \left(\mathbf{b}(t-1) + \mathbf{D}\mathbf{s}^{\tau-1}(t)\right) .$$

It turns out that these dynamics converge to $\mathbf{s}(t)$ defined in the model with conversation via (7). This justifies our use of the Nash equilibrium in the model.

Proposition 2 (**Nash equilibrium justification**). *For any initial* $\mathbf{s}^0(t)$*, we have* $\mathbf{s}^{\tau}(t) \xrightarrow[\tau\to\infty]{} \mathbf{s}(t)$*, where* $\mathbf{s}(t) = \mathbf{P}\mathbf{b}(t-1)$*.*

Proof. See Appendix B.2. ■

2.2 P-network topology

The dissonance matrix $\mathbf{D}$ is a primitive in our model and can, in principle, be observed. The results in this section characterize the dependence of $\mathbf{P}$ on the sensitivity parameters that are collected in matrix $\mathbf{D}$.

The following notion will be useful.[11] We say that "a path leads from i to j in $\mathbf{D}$" if agent i is linked with agent j via a chain of successive associates; that is, there is a sequence of agents $i_1 = i, \ldots, i_J = j$ such that $d_{i_n i_{n+1}} > 0$ for each $n = 1, \ldots, J-1$.

Proposition 3 (**P-network topology**). *Consider a society with the dissonance matrix* $\mathbf{D}$*. Let* $\mathbf{P}$ *be defined by* (8)*. Then* $p_{ij} > 0$ *iff there is a path from* i *to* j *in* $\mathbf{D}$*.*

Proof. See Appendix B.3. ■

This is a key result for characterizing the structure of the $\mathbf{P}$-network. The audience tuning phenomenon cuts the distance between agents unknown to each other, as we saw in Example 1. The distinguishing property of the model with conversation is that agents may directly impact the beliefs of non-associates. Indeed, agent i takes into account the belief of j (and so $p_{ij} > 0$) not only when i experiences dissonance from disagreement with j (that is, $d_{ij} > 0$), but also when there is a chain of intermediate agents leading from i to j, in which every agent experiences dissonance from disagreement with the next agent in the chain. Note that this property does not hold in the $\mathbf{T}$-network, where only the beliefs of associates matter: $t_{ij} > 0$ iff $d_{ij} > 0$.

[11] Matrix $\mathbf{D}$ induces a directed network. We do not use this network in the paper, as the belief dynamics are governed by the other matrices, either $\mathbf{T}$ or $\mathbf{P}$, depending on the model. However, the notion of a directed path in network $\mathbf{D}$ is useful and we introduce it here.

Corollary 1. *The **P**-model with conversation has the following properties:*

(i) If $d_{ij} = 0$ for all j, then $p_{ij} = 0$ for $j \neq i$, and $p_{ii} = 1$.

(ii) If $d_{hi} = 0$ for all h, then $p_{hi} = 0$ for $h \neq i$, and $p_{ii} = 1/(1 + \sum_k d_{ik})$.

(iii) $p_{ii} > p_{hi} \geq 0$ for all $h \neq i$.

(iv) For any i, $p_{ii} \geq t_{ii}$.

Proof. See Appendix B.3. ■

The first two statements immediately follow from Proposition 3. An agent with no associates (a zero row in **D**) has no outgoing paths of any length. Such agents do not experience social dissonance and therefore do not update their beliefs. These agents are called *stubborn* agents. An agent who is not among anyone's associates (a zero column in **D**) does not cause dissonance in others and hence does not impact the beliefs of other agents.

The third statement notes that self-weights p_{ii} (measuring self-importance) are positive in the **P**-model. Moreover, and in contrast to the **T**-model, agent i weights their own belief more than others weight i's belief. The intuitive reason for this is the following. The belief of i affects the beliefs of all agents who have paths leading to i. The belief of i propagates backward along any such path; however, audience tuning dampens the impact of i's belief. This belief, as a result, matters most for agent i, who is closest to i in any such path.

The fourth statement says that self-importance in the **P**-network is no lower than it is in the **T**-network for any agent. Cases *(i)* and *(ii)* are the cases in which the self-importance of i coincides in the two models.[12] In the general case, as soon as an agent both causes and experiences social dissonance, conversation tends to increase the weight that this agent attaches to their own belief.

How do the sensitivity parameters in **D** affect the weights of the interaction matrix? In the benchmark **T**-model without conversation, if the sensitivity of agent i to j increases, the weights of the learning dynamics adjust in a simple way. Agent i will place a greater weight to the belief of j, simultaneously decreasing the weights placed to their own belief and to the beliefs of all their other associates.[13] However, in the

[12] As $t_{ii} \leq p_{ii} \leq 1$ for any i, these cases provide the boundaries for self-importance.

[13] Formally, if d_{ij} increases, from (2) we have that t_{ij} increases and positive t_{ik} (for $k \neq j$) decreases. The other weights in **T** are unaffected.

model with conversation, according to Proposition 3, the beliefs of j, an associate of i, affect not only i's belief, but also the beliefs of all agents with a path leading to i in $\mathbf{D}$.

Denote by H_i the set consisting of i and all agents having a path to i in $\mathbf{D}$. We can describe all consequences of a change in one of the sensitivity parameters.

Proposition 4 (**P-model comparative statics**). *Let j be an associate of i, and suppose that d_{ij} increases and all other sensitivities remain the same. Then:*

(i) p_{hi} decreases and p_{hj} increases for any agent $h \in H_i$.

(ii) p_{hk} changes iff agent $h \in H_i$ and agent k is such that $p_{ik} \neq p_{jk}$. In this case, p_{hk} increases if $p_{ik} < p_{jk}$, and p_{hk} decreases if $p_{ik} > p_{jk}$.

Proof. See Appendix B.4. ■

The direct effect of a higher sensitivity to the dissonance that i experiences from disagreement with j is an increase of weight of i to j's belief. This is intuitive and, as in the model without conversation, can be interpreted as an increase in conformity of i to j. The self-importance of i also decreases. However, the similarity with the **T**-model ends here. Above all, in the **P**-model, there are spillovers to all other agents in the set H_i; that is, to all those who are affected by j's belief via i. All such agents also increase their weight to j and decrease their weight to i.

Furthermore, all affected agents (i and other agents from H_i) adjust the weights they place to *all* agents, and they all do so in the same direction for any agent k. The direction depends on the relative weights that agents i and j place to k. As j's belief is growing more important than the belief of i, agents from H_i increase their weights to those agents, whom j weighted stronger than i, and *vice versa*.

Proposition 4 will be important for Section 4, where the role of dissonance in the long-run beliefs is discussed. We illustrate Propositions 3 and 4 with two examples.

Example 2. Consider the chain society with $N = 3$ agents, and where $d_{21} = d > 0$ and $d_{32} = 1$. Direct computations show that

$$\mathbf{D} = \begin{pmatrix} 0 & 0 & 0 \\ d & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \Rightarrow \mathbf{T} = \begin{pmatrix} 1 & 0 & 0 \\ \frac{d}{1+d} & \frac{1}{1+d} & 0 \\ 0 & 1/2 & 1/2 \end{pmatrix}, \quad \mathbf{P} = \begin{pmatrix} 1 & 0 & 0 \\ \frac{d}{1+d} & \frac{1}{1+d} & 0 \\ \frac{d}{2+2d} & \frac{1}{2+2d} & 1/2 \end{pmatrix}.$$

The difference between the **T**- and **P**-models is that in the former the changes in d_{21} affect only the weights of agent 2, whereas in the latter they also affect the weights

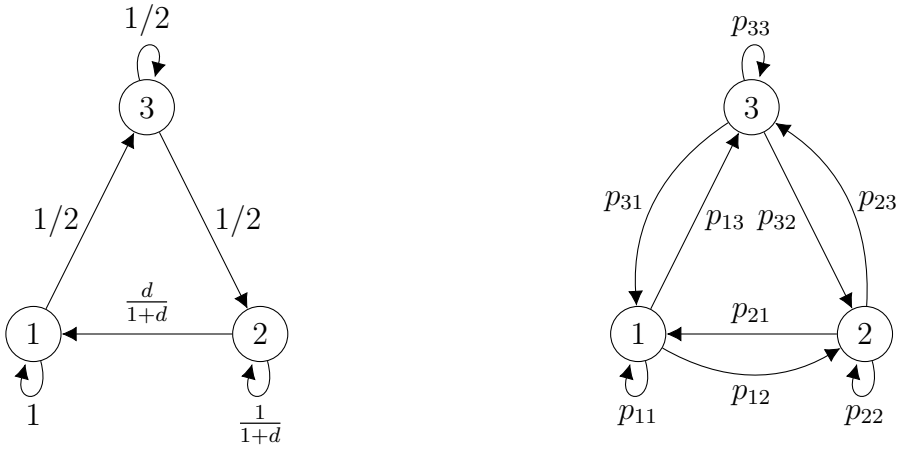

Figure 3: **T**-network (left) and the structure of the **P**-network (right) in Example 3.

that agent 3 (from H_2) attaches to all agents. The weights that agent 3 attaches to agents 1 and 2 move in the same direction as for agent 2. Moreover, since agents 1 and 2 attach equal (zero) weight to 3, the self-weight of 3 does not change.

Example 3. We modify Example 2 to make the "ring" society with $N = 3$ agents, where $d_{21} = d > 0$ and $d_{32} = d_{13} = 1$. Direct computations show that

$$\mathbf{D} = \begin{pmatrix} 0 & 0 & 1 \\ d & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \Rightarrow \mathbf{T} = \begin{pmatrix} 1/2 & 0 & 1/2 \\ \frac{d}{1+d} & \frac{1}{1+d} & 0 \\ 0 & 1/2 & 1/2 \end{pmatrix}, \quad \mathbf{P} = \frac{1}{4+3d} \begin{pmatrix} 2+2d & 1 & 1+d \\ 2d & 4 & d \\ d & 2 & 2+2d \end{pmatrix}.$$

Agent 1 is no longer stubborn and experiences dissonance to agent 3. With this modification, there is a path in **D** from any agent to any other agent. Hence, all the agents are linked in the **P**-model, see Fig. 3 and compare the structure of the **P**-network with that of the chain society in Fig. 1. Moreover, all the weights depend on $d_{21} = d$. When $d = 1$, all agents are homogeneous, their self-importance is $4/7$, the weight to the associate is $2/7$, and the weight to the remaining agent is $1/7$. When $d_{21} > 1$ and agent 2 further increases the dissonance sensitivity wrt agent 1, the weight of every agent to agent 2 decreases and to agent 1 increases. The effect of weights on agent 3 depends, according to Proposition 4, on the relative weights that agents 1 and 2 place on agent 3. Since $p_{23} > p_{13}$, the weight of every agent to agent 3 should increase, which is indeed the case.

3 Long-run beliefs

We are interested in the long-run properties of the model with conversation, defined by the dissonance matrix $\mathbf{D}$. Proposition 1 established that the belief dynamics are governed by the DeGroot averaging process. As is standard in the literature, the dynamics can then be studied with the theories of networks and Markov chains applied to the interaction matrix $\mathbf{P}$.[14]

There are two particularly useful properties of our model that follow from Proposition 3. First, *every agent has a positive self-importance, and therefore the belief dynamics converge to a steady state starting from any vector of initial beliefs.*[15] Second, a path from agent i to another agent j in $\mathbf{D}$ implies a link from i to j in the $\mathbf{P}$-network, and *vice versa*. The structure of the $\mathbf{P}$-network in terms of the so-called *communication classes*[16] plays a key role in characterizing the steady states, and this property allows us to relate this structure to the dissonance matrix.

Let us partition a society into different sets, based on matrix $\mathbf{D}$. Specifically, consider a set $\mathcal{S}$ of all stubborn agents, a number of communication classes $\mathcal{D}_m$ each with several agents and indexed by $m = 1, \ldots, M$, and a set $\mathcal{R}$ of all remaining agents. The numbers of agents in these sets are S, D_m, and R, respectively.[17] Then dissonance matrix $\mathbf{D}$ has the following block structure:

$$\mathbf{D} = \begin{pmatrix} \mathbf{0}_S & \mathbf{0} & \ldots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_1 & \ldots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \ldots & \mathbf{D}_M & \mathbf{0} \\ \mathbf{D}_{\mathcal{RS}} & \mathbf{D}_{\mathcal{R}1} & \ldots & \mathbf{D}_{\mathcal{R}M} & \mathbf{D}_{\mathcal{RR}} \end{pmatrix}, \tag{9}$$

where $\mathbf{0}_S$ is the square matrix of zeros of size S; $\mathbf{D}_1, \ldots, \mathbf{D}_M$ are the strongly con-

[14] The papers using similar approaches are DeMarzo et al. (2003), Golub and Jackson (2010), and Buechel et al. (2015). See Jackson (2008, chapter 8) for an overview.

[15] As $p_{ii} > 0$ for all i, matrix $\mathbf{P}$ has a cycle of length one and thus is aperiodic, when restricted to every group of agents. Convergence then follows from the standard results, as stated in Theorem 2 in Golub and Jackson (2010). Note that, for the same reason, dynamics converge in the $\mathbf{T}$-model.

[16] In a given network, a set of nodes is *strongly connected* if there is a directed path from any node to any other node in the set. If this holds for the whole network, then this network (and its corresponding matrix) is *strongly connected*. A set is *closed* if there is no link from a node in the set to a node outside the set. A *communication class* is the strongly connected and closed set. Any network can be partitioned into one or more disjoint communication classes and the set of all nodes that do not belong to any communication class.

[17] Thus $N = S + \sum_m D_m + R$. Any set can be empty, but the society cannot coincide with $\mathcal{R}$. Thus, if $M = 0$ (no communication classes in $\mathbf{D}$), then there is at least one stubborn agent.

nected matrices of sensitivity parameters of agents from $\mathcal{D}_m$ wrt themselves; $\mathbf{D}_{\mathcal{RS}}$, $\mathbf{D}_{\mathcal{R}m}$, and $\mathbf{D}_{\mathcal{RR}}$ are matrices of sensitivity parameters of agents from $\mathcal{R}$ wrt agents from $\mathcal{S}$, $\mathcal{D}_m$, and $\mathcal{R}$, respectively; and symbol $\mathbf{0}$ stands for matrices of various sizes with zero entries.

Using the relation between paths in $\mathbf{D}$ and links in $\mathbf{P}$, we obtain that there are two types of communication class in the $\mathbf{P}$-network: singleton sets formed by agents in $\mathcal{S}$, and sets $\mathcal{D}_m$, in each of which any two agents have links in both directions. All agents not belonging to any communication class are in the set $\mathcal{R}$. Thus, *interaction matrix* $\mathbf{P}$ *has essentially the same block structure as matrix* $\mathbf{D}$:

$$\mathbf{P} = \begin{pmatrix} \mathbf{I}_S & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{P}_1 & \dots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{P}_M & \mathbf{0} \\ \mathbf{P}_{\mathcal{RS}} & \mathbf{P}_{\mathcal{R}1} & \dots & \mathbf{P}_{\mathcal{R}M} & \mathbf{P}_{\mathcal{RR}} \end{pmatrix},$$

where $\mathbf{P}_m = (\mathbf{I}_{D_m} + \operatorname{diag}\{\mathbf{D}_m \mathbb{1}_{D_m}\} - \mathbf{D}_m)^{-1}$ is a positive matrix.[18]

Once matrix $\mathbf{P}$ is partitioned, the standard result (e.g., Theorem 3 in Golub and Jackson, 2010) can be applied to derive the vector of the long-run beliefs $\mathbf{b}^*$. We formulate this in Proposition 5 below. However, before this, let us consider a special but informative case of a society forming a single communication class.

Consensus for strongly connected dissonance matrix. A whole society forms a single communication class iff $\mathbf{D}$ is strongly connected. It is well known that a converging society then reaches *consensus*, with all agents having the same belief in the long run. This consensus belief, b^*, is the weighted average of agents' initial beliefs,

$$b^* \;=\; \pi_1 b_1(0) + \ldots + \pi_N b_N(0)\,. \tag{10}$$

The row-vector of weights, $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_N)$ with $\sum_i \pi_i = 1$, is a unique left eigenvector of $\mathbf{P}$ corresponding to eigenvalue 1 whose entries sum to 1. The components

[18] A matrix is positive if all of its elements are positive. The last result follows from Proposition 3 and from (8). Proposition 3 also leads to other properties of the blocks in $\mathbf{P}$. If agent $r \in \mathcal{R}$ does not have a path to an agent $s \in \mathcal{S}$, then in $\mathbf{P}_{\mathcal{RS}}$, $p_{rs} = 0$; if $r \in \mathcal{R}$ does not have a path to agents in $\mathcal{D}_m$, then row r in $\mathbf{P}_{\mathcal{R}m}$ is zero. If $r \in \mathcal{R}$ has at least one associate in $\mathcal{D}_m$, then row r in $\mathbf{P}_{\mathcal{R}m}$ is positive. If two agents in $\mathcal{R}$ have paths to each other, then their rows in $\mathbf{P}_{\mathcal{R}m}$ have zero and positive elements in the same positions.

of $\boldsymbol{\pi}$ are called the *social influences* of agents.[19]

Example 3 (continued). Consider the "ring" society with N agents, where the only non-zero elements of $\mathbf{D}$ are $d_{i+1,i}$ for $i = 1, \ldots, N-1$ and $d_{1,N}$. For any N, the dissonance matrix $\mathbf{D}$ is strongly connected. In the special case where $N = 3$ and $d_{21} = d > 0$, $d_{32} = d_{13} = 1$, the row eigenvector of $\mathbf{P}$ corresponding to eigenvalue 1 is $\boldsymbol{\pi} = \frac{1}{1+2d}(d, 1, d)$; its entries are the social influences. Hence, for given initial beliefs $\mathbf{b}(0) = (b_1(0), b_2(0), b_3(0))'$, agents converge to the common long-run belief:

$$b^* = \pi_1 b_1(0) + \pi_2 b_2(0) + \pi_3 b_3(0) = \frac{d}{1+2d} b_1(0) + \frac{1}{1+2d} b_2(0) + \frac{d}{1+2d} b_3(0)\,.$$

Note that when $d = 1$, all agents' initial beliefs are equally weighted in b^*.

General case. Applying the above result for any communication class, we have

Proposition 5 (**P-model long-run beliefs**). *Consider the society with the dissonance matrix* $\mathbf{D}$ *having block structure given by* (9)*. Let* $\mathbf{b}(0)$ *be an arbitrary vector of the initial beliefs with* $\mathbf{b}(0)|_{\mathcal{S}}$ *and* $\mathbf{b}(0)|_{\mathcal{D}_m}$ *denoting its restrictions to the sets* $\mathcal{S}$ *and* $\mathcal{D}_m$*, respectively. There exists* $\mathbf{b}^* = \lim_{t\to\infty} \mathbf{b}(t)$*; vector* $\mathbf{b}^*$ *of the long-run beliefs is given by*

$$\mathbf{b}^* = \begin{pmatrix} \mathbf{b}(0)|_{\mathcal{S}} \\ \mathbf{b}_1^* \\ \vdots \\ \mathbf{b}_M^* \\ \mathbf{b}_{\mathcal{R}}^* \end{pmatrix}, \tag{11}$$

where the block structure corresponds to the block structure of $\mathbf{D}$*. In Eq.* (11)*, vector* $\mathbf{b}_m^*$ *with* $m = 1, \ldots, M$ *has identical entries* b^{*m} *defined as*

$$b^{*m} = \boldsymbol{\pi}^m \mathbf{b}(0)|_{\mathcal{D}_m} = \pi_1^m b_1^m(0) + \ldots + \pi_{D_m}^m b_{D_m}^m(0)\,, \tag{12}$$

where $\boldsymbol{\pi}^m$ *is the left eigenvector of matrix* $\mathbf{P}_m$ *corresponding to eigenvalue* 1 *whose*

[19] See Chapter 8 of Berman and Plemmons (1979) or Proposition 1 in Golub and Jackson (2010). As beliefs converge, there is a limiting matrix, $\mathbf{P}^\infty = \lim_{t\to\infty} \mathbf{P}^t$. Row i of $\mathbf{P}^t$ is composed of impacts of agents' initial beliefs on agent i's belief at time t. Because of consensus, all the rows of $\mathbf{P}^\infty$ are identical and consist of the impacts of all agents on the long-run belief. But then the row $\boldsymbol{\pi}$ satisfies $\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P}$, and thus it is a left eigenvector of $\mathbf{P}$ corresponding to eigenvalue 1. There is a unique such positive eigenvector whose entries sum to 1, according to the Perron–Frobenius theorem.

entries sum to 1. *The components of the vector* $\mathbf{b}^*_{\mathcal{R}}$ *are given by*

$$b^*_r = \sum\nolimits_{s\in\mathcal{S}} \bar{\gamma}_{rs} b_s(0) + \sum\nolimits_{m=1}^{M} \gamma_{rm} b^{*m}\,, \quad r \in \mathcal{R}\,, \tag{13}$$

where coefficients form the $R \times (S+M)$ *matrix* $\mathbf{\Gamma} = \left(\{\bar{\gamma}_{rs}\}_{r,s=1}^{R,S} \;\; \{\gamma_{rm}\}_{r,m=1}^{R,M}\right)$ *given by*

$$\mathbf{\Gamma} = \left(\mathbf{I}_R - \mathbf{P}_{\mathcal{R}\mathcal{R}}\right)^{-1} \left(\mathbf{P}_{\mathcal{R}\mathcal{S}} \quad \mathbf{P}_{\mathcal{R}1}\mathbb{1}_{D_1} \quad \cdots \quad \mathbf{P}_{\mathcal{R}M}\mathbb{1}_{D_M}\right). \tag{14}$$

Proposition 5 states, first, that every stubborn agent sticks to their own initial belief. Second, within each communication class $\mathcal{D}_m$, there is a consensus.[20] This is a generalization of the case with strongly connected $\mathbf{D}$, where the whole society reaches a consensus. The common long-run belief in $\mathcal{D}_m$ is given in (12) as an average of the initial beliefs of agents from $\mathcal{D}_m$, weighted by the components of the vector that satisfies $\boldsymbol{\pi}^m = \boldsymbol{\pi}^m \mathbf{P}_m$ and is properly normalized. The entries of $\boldsymbol{\pi}^m$ are the agents' social influences within the class $\mathcal{D}_m$.

Third, the initial beliefs of agents in $\mathcal{R}$ do not matter. Instead, as stated in (13), the long-run beliefs of these agents are determined by the beliefs of the stubborn agents and the long-run beliefs of the communication classes.[21] The long-run belief of agent $r \in \mathcal{R}$ is the average of those others' beliefs with weights $\bar{\gamma}_{rs}$ and γ_{rm}. We call these weights the *external impacts* of the stubborn agent s and of the communication class $\mathcal{D}_m$, respectively, on agent r.

In Section 4, we use these results to discuss how agents' dissonance sensitivities affect the social influences and external impacts in our model. Let us now illustrate Proposition 5 with several examples. In Example 1, there is a chain of N agents, where agent 1 is stubborn. Since there are no communication classes in $\mathbf{D}$, all other agents are in the set $\mathcal{R}$. Hence, agent 1 keeps their own initial belief and all other agents converge to this belief; that is, $b^*_r = b_1(0)$ for any $r = 2, \ldots, N$. In the chain society, the structure of the $\mathbf{D}$-network is such that the exact values of the sensitivity parameters do not affect the long-run belief. For instance, in Example 2, we have $b^*_1 = b^*_2 = b^*_3 = b_1(0)$ regardless of parameter d.

In all previous examples, a consensus was reached. We now consider a society with no consensus.

[20] The long-run beliefs of different stubborn agents and in different communication classes $\mathcal{D}_m$ are different, for generic initial beliefs. Thus, there is no consensus in the society, as soon as $S + M > 1$.

[21] Cf. Theorem 10 in DeMarzo et al., 2003, or Proposition A.1 in Buechel et al., 2015. Matrix $\mathbf{I}_R - \mathbf{P}_{\mathcal{R}\mathcal{R}}$ in (14) is invertible because the remaining agents are linked to stubborn agents or communication classes, and hence the spectral radius of $\mathbf{P}_{\mathcal{R}\mathcal{R}}$ is less than one.

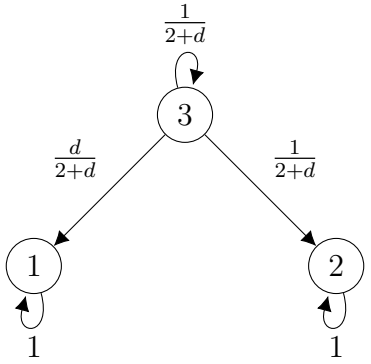

Figure 4: **T**- and **P**-networks in Example 4.

Example 4. Suppose that agents 1 and 2 have no associates, whereas agent 3 has sensitivities $d_{31} = d > 0$ wrt agent 1 and $d_{32} = 1$ wrt agent 2. In this case, the **T**- and **P**-networks (and matrices) are identical; see Fig. 4 for an illustration. Therefore, the long-run outcome is the same in both models. The stubborn agents 1 and 2 keep their initial beliefs, whereas agent 3 from the set $\mathcal{R}$ converges to their weighted average:

$$b_3^* = \frac{d}{1+d} b^{*1} + \frac{1}{1+d} b^{*2} = \frac{d}{1+d} b_1(0) + \frac{1}{1+d} b_2(0)\,.$$

The external impacts are $\gamma_{31} = d/(1+d)$ and $\gamma_{32} = 1/(1+d)$. Thus, the more agent 3 conforms to agent 1, the higher the external impact of agent 1 is on agent 3, and the lower the external impact of agent 2 is on agent 3.

Long-run beliefs in the model without conversation. The long-run properties of the **T**-model are now briefly discussed. For a given dissonance matrix **D**, the structure of communication classes in the **T**-model is the same as in the **P**-model:[22]

$$\mathbf{T} = \begin{pmatrix} \mathbf{I}_S & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{T}_1 & \dots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{T}_M & \mathbf{0} \\ \mathbf{T}_{\mathcal{RS}} & \mathbf{T}_{\mathcal{R}1} & \dots & \mathbf{T}_{\mathcal{R}M} & \mathbf{T}_{\mathcal{RR}} \end{pmatrix},$$

[22]It follows from (3) that, apart from the stubborn agents that are added as separate communication classes, the structure of **T** is identical to the structure of **D**. This leads to the same block structure for **T** as for **P**. The only difference is that whereas matrices $\mathbf{P}_m$ for communication classes $\mathcal{D}_m$ are positive, this is not the case for the corresponding matrices $\mathbf{T}_m$. However, since any agent has a positive self-importance in the **T**-model and any $\mathbf{T}_m$ is strongly connected, the convergence results within each communication class are the same.

where $\mathbf{T}_m = (\mathbf{I}_{D_m} + \operatorname{diag}\{\mathbf{D}_m \mathbb{1}_{D_m}\})^{-1} (\mathbf{I}_{D_m} + \mathbf{D}_m)$ is matrix $\mathbf{T}$ restricted to the class $\mathcal{D}_m$. We then can state the following (cf. Proposition 5). First, the stubborn agents are unaffected by conversations and keep their initial beliefs. Second, for each communication class $\mathcal{D}_m$ with $m = 1, \ldots, M$, there is a consensus belief

$$b^{*m} \;=\; \boldsymbol{\theta}^m \mathbf{b}(0)|_{\mathcal{D}_m} \;=\; \theta_1^m b_1^m(0) + \ldots + \theta_{D_m}^m b_{D_m}^m(0)\,,$$

where $\boldsymbol{\theta}^m = (\theta_1^m, \ldots, \theta_{D_m}^m)$ is the left eigenvector of $\mathbf{T}_m$ corresponding to eigenvalue 1 whose entries sum to 1. Its elements are the social influences of agents from $\mathcal{D}_m$ in the $\mathbf{T}$-network. Third, the long-run belief of any other agent $r \in \mathcal{R}$ is given by

$$b_r^* = \sum\nolimits_{s \in \mathcal{S}} \bar{\delta}_{rs} b_s(0) + \sum\nolimits_{m=1}^{M} \delta_{rm} \boldsymbol{\theta}^m \mathbf{b}(0)|_{\mathcal{D}_m}\,,$$

where $\bar{\delta}$ and δ's are the external impacts of the stubborn agents and communication classes, respectively, on agents in $\mathcal{R}$. Matrix $\boldsymbol{\Delta} = \left(\{\bar{\delta}_{rs}\}_{r,s=1}^{R,S} \quad \{\delta_{rm}\}_{r,m=1}^{R,M}\right)$ of the external impacts is defined as in (14) but based on the matrix $\mathbf{T}$, so that

$$\boldsymbol{\Delta} = (\mathbf{I}_R - \mathbf{T}_{\mathcal{R}\mathcal{R}})^{-1} \left(\mathbf{T}_{\mathcal{R}\mathcal{S}} \quad \mathbf{T}_{\mathcal{R}1}\mathbb{1}_{D_1} \quad \cdots \quad \mathbf{T}_{\mathcal{R}M}\mathbb{1}_{D_M}\right). \tag{15}$$

4 Main results

In the previous section, we characterized long-run beliefs in the $\mathbf{P}$-model; see Proposition 5. These beliefs depend on the exact positions the agents occupy in the network, their sensitivity parameters, and their initial beliefs. Within each communication class of $\mathbf{D}$, agents' social influences weight their initial beliefs to determine their long-run belief.[23] The long-run beliefs of the remaining agents are the weighted averages of the long-run beliefs of the stubborn agents and the communication classes, with weights given by the corresponding external impacts.

Now we discuss the roles that social influences and external impacts play in the model with conversation, and how they are affected by the sensitivity parameters of the dissonance matrix $\mathbf{D}$. As before, we also highlight the role of conversation by comparing the $\mathbf{P}$-model with the $\mathbf{T}$-model for the same matrix $\mathbf{D}$.

[23]Recall that stubborn agents are the singleton communication classes in $\mathbf{P}$ (but not in $\mathbf{D}$). They can be considered as the special case of this result, having social influence 1 within their class.

Consensus. The first question of interest is whether society reaches a consensus. For a given dissonance matrix, matrices $\mathbf{P}$ and $\mathbf{T}$ have the same structure in terms of their partitioning into the communication classes and the remaining agents. A consensus is reached for any initial beliefs iff the number of communication classes in the interaction matrix is 1. Reformulating this condition in terms of $\mathbf{D}$, we have

Proposition 6 (**Consensus**). *In a society with the dissonance matrix* $\mathbf{D}$*, a consensus in the* $\mathbf{P}$*-model is reached iff a consensus in the* $\mathbf{T}$*-model is reached. Consensus is reached iff* $\mathbf{D}$ *has either one stubborn agent or one communication class, but not both.*

This consensus equivalence does not mean that the consensus beliefs are identical in the two models, as they depend on two different social influence vectors, $\boldsymbol{\pi}$ and $\boldsymbol{\theta}$.

4.1 Social influences

According to Proposition 5, agents affect the consensus belief in strongly connected societies via their social influences; see Eq. (10). Similarly, in the general case their social influences affect the consensus belief within each communication class; see Eq. (12). For the purpose of finding the long-run beliefs, we can, therefore, treat each class as a separate strongly connected society. For this reason, in the discussion below we do not introduce the index of the communication class. The results can be applied *verbatim* to the strongly connected society of size N, but they also apply (up to the notation) to any communication class m.

Consider the vector $\boldsymbol{\pi}$ of social influences in the $\mathbf{P}$-model. As it is the left eigenvector of $\mathbf{P}$, we have $\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P}$, which is equivalent to $\boldsymbol{\pi}\mathbf{D} = \boldsymbol{\pi}\,\mathrm{diag}\,\{\mathbf{D}\mathbb{1}_N\}$, because of (8). We rewrite it as

$$\pi_i = \frac{1}{\sum_k d_{ik}} \sum\nolimits_{h=1}^{N} d_{hi}\ \pi_h\,. \tag{16}$$

Thus, social influences are interrelated in the way that makes an agent influential, if the agent is an associate of other influential agents who experience strong dissonance from disagreement with this agent.[24] Interestingly, even if matrix $\mathbf{P}$ is positive in the model with conversation, so that each agent is immediately affected by the previous beliefs of every other agent, these connections are only implicit in this characterization of the social influences in (16), where we return to matrix $\mathbf{D}$.

[24]It follows that the agents' social influences are their eigenvector centralities in a network determined by the sensitivity parameters and described by some properly defined adjacency matrix.

Consider the vector $\boldsymbol{\theta}$ of social influences in the **T**-model. As $\boldsymbol{\theta} = \boldsymbol{\theta}\mathbf{T}$, we have from (3) that

$$\theta_i = \frac{1 + \sum_k d_{ik}}{\sum_k d_{ik}} \sum\nolimits_{h=1}^{N} \frac{d_{hi}}{1 + \sum_k d_{hk}} \theta_h \,. \tag{17}$$

In the **T**-model, we have the same intuitive result as in the **P**-model, that the influence of agent i depends positively on how influential i's associates are. However, now the *relative* sensitivities of these associates to i matter. This means that the social influences in the models with and without conversation are generally different. Hence, even if consensus is reached in both models, the long-run beliefs are different.

By comparing (16) and (17), we come to the following

Theorem 1 (**Social influences in two models**). *Social influences in the* **P**- *and* **T**-*models are related as*

$$\frac{\pi_i/\theta_i}{\pi_j/\theta_j} \;=\; \frac{1 + \sum_k d_{jk}}{1 + \sum_k d_{ik}} \;=\; \frac{t_{ii}}{t_{jj}}\,, \quad \textit{for any pair of agents } i \textit{ and } j\,. \tag{18}$$

Theorem 1 shows, first, that the consensus beliefs are identical in the **P**- and **T**-models iff all agents have the same self-importance in the **T**-model (which is inversely proportional to the sum of dissonance sensitivities wrt other agents). In particular, this would be the case in a *homogeneous society*, defined as a society where all agents have the same number of associates and identical sensitivity parameters. Thus, agent heterogeneity is crucial for the effects of conversation to appear.

Second, Theorem 1 implies that conversation leads to a redistribution of social influences in favor of agents whose self-importance is high. It follows that conversation helps to propagate the belief of the most self-confident agent: the social influence of the agent with the highest self-importance in the **T**-model becomes even higher in the **P**-model. By contrast, the social influence of the agent with the lowest self-importance in the **T**-model becomes lower in the **P**-model.

The next question is how the social influences (of a society restricted to the communication class) are affected by changes in the sensitivity parameters.

Proposition 7 (**Comparative statics for social influences**). *Let j be an associate of i, and suppose that d_{ij} increases and all other sensitivities remain the same. Then:*

(i) In the **P**-*model, π_i strictly decreases and π_j strictly increases.*

(ii) In the **T**-*model, θ_i strictly decreases and θ_j strictly increases.*

In both models, the relative change in social influence of any other agent is not greater than that of agent j, and is not lesser than that of agent i.

Proof. See Appendix B.5. ■

Thus, a change in one sensitivity parameter in general affects all agents. The social influence of i is negatively associated with this change because an increase in sensitivity to dissonance wrt j leads to a higher conformity of i to j. At the same time, due to relation (16), agent j becomes more influential. The social influences of all other agents can change in any direction, but within the limits.

Moreover, in the **P**-model in the special case where each agent has the same sensitivities to their associates, when one agent increases their dissonance sensitivity to others, the social influences of all other agents become greater.

Proposition 8 (**Social influences for identical sensitivities**)**.** *If for all i, j either $d_{ij} = 0$ or $d_{ij} = d_i$, then π_i is decreasing in d_i and increasing in d_j for any $j \neq i$.*

Proof. See Appendix B.6. ■

Example 3 (continued). We have seen that in the "ring" society with $N = 3$ agents, with $d_{21} = d > 0$ and $d_{32} = d_{12} = 1$, the social influences in the **P**-model are $\pi_2 = 1/(1 + 2d)$ and $\pi_1 = \pi_3 = d/(1 + 2d)$. In the **T**-model, the social influences are $\theta_2 = (1 + d)/(1 + 5d)$ and $\theta_1 = \theta_3 = 2d/(1 + 5d)$, resulting in a different long-run consensus belief than in the **P**-model.

In both models, when agents have the same sensitivity parameters ($d = 1$), their social influences are the same and equal to $1/3$. Additionally, in both models, the higher d_{21} is, the lower the social influence of agent 2 is, and the higher the social influences of the two other agents are, illustrating Proposition 7. When $d > 1$ we have $t_{22} = 1/(1 + d) < 1/2 = t_{11} = t_{33}$, and hence the social influence of agent 2 in the **P**-model is lower than it is in the **T**-model ($\pi_2 < \theta_2$), while the opposite is true for agents 1 and 3. This illustrates the general result of Theorem 1.

4.2 On the speed of convergence in a homogeneous society

It follows from Theorem 1 that in a (strongly connected) homogeneous society the social influences are not affected by conversation, and the **P**- and **T**-models lead to the same long-run consensus belief. In this case, the important question is whether

conversation speeds up the convergence to consensus. We now address the question of the speed of convergence in such a society.

Recall from Jackson (2008, chapter 8) that the speed of convergence is inversely related to the absolute value of the second-largest eigenvalue λ_2 of the interaction matrix. When all agents have the same number of associates A and the same sensitivity parameters d, it can be seen (see Appendix B.7) that the second-largest eigenvalues of $\mathbf{P}$ and $\mathbf{T}$ are related to the second-largest eigenvalue of $\mathbf{D}$ as follows:

$$\lambda_2(\mathbf{P}) = \frac{1}{1 + Ad - \lambda_2(\mathbf{D})}, \qquad \text{and} \qquad \lambda_2(\mathbf{T}) = \frac{1 + \lambda_2(\mathbf{D})}{1 + Ad}. \tag{19}$$

It is natural to conjecture that conversation tends to accelerate the convergence to a consensus belief, and that the speed of convergence is higher in the $\mathbf{P}$-model; that is, we always have $|\lambda_2(\mathbf{P})| < |\lambda_2(\mathbf{T})|$. However, this is not the general case for any homogeneous society, as the following result shows.

Consider the case of a complete network where each agent is linked to every other agent in the society; that is, each agent has $N - 1$ associates. Then in matrix $\mathbf{D}$ for all $i \neq j$, $d_{ij} = d$, and it is easily seen that $\lambda_2(\mathbf{D}) = -d$. Using (19), we can compare the speed of convergence in the $\mathbf{P}$- and $\mathbf{T}$-models in the complete network.

Proposition 9 (**Speed of convergence in complete network**). *Suppose that each agent has $N - 1$ associates and all agents have identical sensitivity parameters d. Let $d^* = 1 - \frac{1}{N} + \sqrt{1 + \frac{1}{N^2}}$. Then:*

(i) For $d < d^*$, $|\lambda_2(\mathbf{T})| < |\lambda_2(\mathbf{P})|$, and $\mathbf{T}$-model converges faster than $\mathbf{P}$-model.

(ii) For $d > d^*$, $|\lambda_2(\mathbf{P})| < |\lambda_2(\mathbf{T})|$, and $\mathbf{P}$-model converges faster than $\mathbf{T}$-model.

Proof. See Appendix B.7. ■

In a society where everyone knows everyone, and where sensitivity to disagreement is sufficiently high, conversation helps the society to converge to consensus faster. However, the effect is opposite when the dissonance sensitivity is very low. The threshold value of the sensitivity d^* depends on the number of agents in the society, and monotonically increases with N from $d^* = 1.618$ for $N = 2$ to $d^* = 2$ for $N \to \infty$.

4.3 External impacts

We will now discuss the long-run beliefs of the agent from set $\mathcal{R}$: those that are neither stubborn nor belong to any communication class.

It follows from (8) that Eq. (14) can be rewritten as

$$\gamma_{im} \;=\; \frac{1}{\sum_k d_{ik}} \left(\sum_{r \in \mathcal{R}} d_{ir} \gamma_{rm} + \sum_{n \in \mathcal{D}_m} d_{in} \right), \qquad \bar{\gamma}_{is} \;=\; \frac{1}{\sum_k d_{ik}} \left(\sum_{r \in \mathcal{R}} d_{ir} \bar{\gamma}_{rs} + d_{is} \right).$$

Thus, an external impact of the class $\mathcal{D}_m$ on agent $i \in \mathcal{R}$ is a weighted average of the external impacts of $\mathcal{D}_m$ on all other agents from $\mathcal{R}$ (with weights proportional to the sensitivity parameters that agent i has wrt agents from $\mathcal{R}$), and a value 1 which can be interpreted as an external impact of $\mathcal{D}_m$ on $\mathcal{D}_m$ (with the weight proportional to the sum of sensitivity parameters that agent i has wrt all agents from $\mathcal{D}_m$). The same interpretation holds for an external impact of the stubborn agent s on agent $i \in \mathcal{R}$.

Moreover, it immediately follows from (15) that

$$\delta_{im} \;=\; \frac{1}{\sum_k d_{ik}} \left(\sum_{r \in \mathcal{R}} d_{ir} \delta_{rm} + \sum_{n \in \mathcal{D}_m} d_{in} \right), \qquad \bar{\delta}_{is} \;=\; \frac{1}{\sum_k d_{ik}} \left(\sum_{r \in \mathcal{R}} d_{ir} \bar{\delta}_{rs} + d_{is} \right),$$

which are the same equations as above. Therefore, we have

Theorem 2 (**Comparison of external impacts**). *The matrices of external impacts are the same in the* $\mathbf{P}$- *and* $\mathbf{T}$-*models:* $\mathbf{\Gamma} = \mathbf{\Delta}$.

Thus, while social influences are in general affected by conversation, as was shown in Theorem 1, external impacts do not depend on conversation and are always the same in the two models.

Again, we are interested in how external impacts are affected by changes in the sensitivity parameters. For $i \in \mathcal{R}$, consider the set H_i consisting of i and all agents from $\mathcal{R}$ having a path to i in $\mathbf{D}_{\mathcal{R}\mathcal{R}}$. The comparative statics results for external impacts depend on whether an agent from $\mathcal{R}$ wants to conform to a stubborn agent, to an agent from some communication class, or to an agent who is also from $\mathcal{R}$.

Proposition 10 (**Comparative statics for external impacts**). *Let j be an associate of $i \in \mathcal{R}$, and suppose that d_{ij} increases and all other sensitivities remain the same. Then:*

(i) If $j \in \mathcal{D}_{m^*}$, then for any agent $h \in H_i$, γ_{hm^*} increases, while γ_{hm} decreases for $m \neq m^*$, and $\bar{\gamma}_{hs}$ decreases for all s.

(ii) *If $j \in \mathcal{S}$, then for any agent $h \in H_i$, $\bar{\gamma}_{hj}$ increases, while $\bar{\gamma}_{hs}$ decreases for $s \neq j$, and γ_{hm} decreases for all m.*

(iii) *If $j \in \mathcal{R}$, then γ_{hm} changes iff $h \in H_i$ and m is such that $\gamma_{im} \neq \gamma_{jm}$. In this case, γ_{hm} increases if $\gamma_{im} < \gamma_{jm}$, and γ_{hm} decreases if $\gamma_{im} > \gamma_{jm}$. Similarly, $\bar{\gamma}_{hs}$ changes iff $h \in H_i$ and s is such that $\bar{\gamma}_{is} \neq \bar{\gamma}_{js}$. In this case, $\bar{\gamma}_{hs}$ increases if $\bar{\gamma}_{is} < \bar{\gamma}_{js}$, and $\bar{\gamma}_{hs}$ decreases if $\bar{\gamma}_{is} > \gamma_{js}$.*

Proof. See Appendix B.8. ■

Proposition 10, similarly to Proposition 4, implies that an increase in conformity of agent i to agent j affects not only i, but all agents who are affected by j's belief via i. If j belongs to the communication class $\mathcal{D}_{m^*}$, then the external impact of this class m^* on each affected agent increases, while the external impacts of all other classes (as well as stubborn agents) on these agents decrease. The same is true if j is a stubborn agent: the less agent i wants to disagree with j, the higher the relative weight is of the stubborn agent j on the long-run beliefs of i and other agents from H_i.

If both i and j belong to $\mathcal{R}$, then the change in external impact of some class m on agents from H_i depends on how the external impacts of m on i and j are related. As the sensitivity to dissonance that i experiences from disagreement with j increases, the external impacts of those classes which had higher impacts on j increase, while the external impacts of those classes which had higher impacts on i decrease. Analogous results also hold for the external impacts of all stubborn agents.

We end this section with an example that illustrates Proposition 10.

Example 5. Consider the society with $N = 4$ agents, where agents 1 and 2 are stubborn, and agents 3 and 4 are associates of each other with $d_{34} = d_{43} = 1$. Suppose that agent 3 experiences dissonance to agent 2 with sensitivity $d_{32} = a$, and agent 4 experiences dissonance to agent 1 with sensitivity $d_{41} = d$. The resulting **T**- and **P**-networks are illustrated in Fig. 5. The stubborn agents 1 and 2 always keep their initial beliefs $b_1(0)$ and $b_2(0)$. The long-run beliefs of agents 3 and 4 can be easily computed using (14) or (15):

$$b_3^* = \frac{d}{ad+a+d} b_1(0) + \frac{ad+a}{ad+a+d} b_2(0), \qquad b_4^* = \frac{d(1+a)}{ad+a+d} b_1(0) + \frac{a}{ad+a+d} b_2(0).$$

External impacts of stubborn agent 1 on both agents 3 and 4, $\gamma_{31} = d/(ad+a+d)$ and $\gamma_{41} = (ad+d)/(ad+a+d)$, are increasing in d and decreasing in a. In accordance

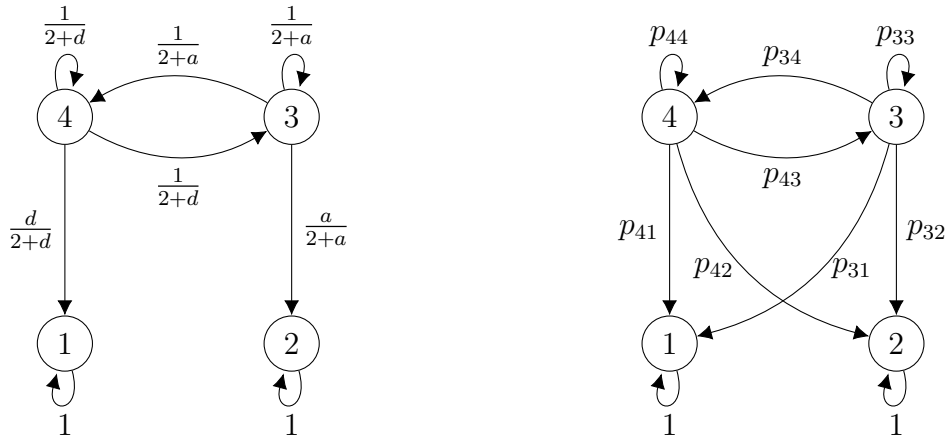

Figure 5: **T**-network (left) and **P**-network (right) in Example 5.

with Proposition 10, an increase in conformity of agent 4 wrt agent 1 leads to a higher weight of agent 1's belief in the long-run beliefs of both agent 4 and their associate agent 3. Note also that $\gamma_{41} > \gamma_{31}$, which is also natural, since agent 4 is closer to agent 1 in the network and directly affected by agent 1's belief.

An interesting application of this result concerns the ability of a stubborn agent to manipulate beliefs in the society. Suppose that agent 1 wants to convince remaining agents 3 and 4 of something. To do this, agent 1 can use one of two alternative approaches: either change their own belief or invest in increasing the sensitivity of agent 4 to disagreement with them. Clearly, if agent 1 chooses to change $b_1(0)$, the resulting change in b_3^* and b_4^* is proportional to the external impacts γ_{31} and γ_{41} respectively:

$$\Delta b_3^* = \frac{d}{ad+a+d}\Delta b_1(0), \qquad \Delta b_4^* = \frac{d(1+a)}{ad+a+d}\Delta b_1(0).$$

However, if agent 1 is somehow able to increase d, then the change in b_3^* and b_4^* is proportional to the difference between $b_1(0)$ and the rival belief $b_2(0)$:

$$\Delta b_3^* = \frac{a}{(ad+a+d)^2}\left(b_1(0)-b_2(0)\right)\Delta d, \qquad \Delta b_4^* = \frac{a(1+a)}{(ad+a+d)^2}\left(b_1(0)-b_2(0)\right)\Delta d.$$

The reason for this is that the change in $b_1(0)$ does not affect the impact of $b_2(0)$ on the remaining agents' beliefs. At the same time, an increase in d leads to not only an increase in the external impacts of agent 1 on both remaining agents, but also a decrease in the external impacts of agent 2.

Therefore, it might be the case that, for a more efficient manipulation, the stubborn agent should, instead of changing their belief, increase the sensitivity to disagreement of the remaining agents. As the above example shows, this conclusion is especially relevant when the difference between the two rival beliefs is large.

5 On endogenous network

Finally, we turn to the model with an endogenous change of network motivated by dissonance minimization, which is the crucial feature of AEW.[25] We describe the model, characterize stationary states and provide a simple example of dynamics.

Suppose that, in each period, agents minimize their dissonance firstly by changing their beliefs and secondly by changing their networks; namely, by replacing those of their associates whose beliefs are far from their own. Let the network in period t be given by beliefs $\mathbf{b}(t-1)$ and the dissonance matrix $\mathbf{D}(t-1) = \{d_{ij}(t-1)\}_{i,j=1}^{N}$, which can be non-stationary. First, agents choose their statements and beliefs as described in Section 2, solving problems (4) and (5) where the dissonance sensitivities are given by $\mathbf{D}(t-1)$. In each period, statements coincide with beliefs, and the new beliefs of agents satisfy the following equation: $\mathbf{b}(t-1) = (\mathbf{I}_N + \operatorname{diag}\{\mathbf{D}(t-1)\mathbb{1}_N\} - \mathbf{D}(t-1))\,\mathbf{b}(t)$.

Second, given their new beliefs $\mathbf{b}(t)$, agents revise their sets of associates. Agent i is given the opportunity to replace one of their associates with another agent, who can be either a current associate (in which case agent i swaps the corresponding sensitivities) or a new associate (in which case agent i drops the link with their previous associate and forms a new link with a new agent with the same sensitivity). Formally, after the beliefs are chosen, the instant dissonance of agent i is given by

$$(b_i(t) - b_i(t-1))^2 + \sum\nolimits_{j=1}^{N} d_{ij}(t-1)(b_i(t) - b_j(t))^2\,. \tag{20}$$

All $b_j(t)$ in (20) are already known, so the only decision variables are elements in row i of $\mathbf{D}(t-1)$. Agent i minimizes the instant dissonance by swapping any two elements in row i of $\mathbf{D}(t-1)$. The swap will take place if, in the new network, dissonance is reduced by at least the switching cost $\varepsilon \geq 0$. After all agents have made their choices, the new dissonance matrix $\mathbf{D}(t)$ emerges.[26] The network in period $t+1$ is then given

[25] See also Bolletta and Pin (2020) for the DeGroot model with endogenous network formation.

[26] Clearly, $\mathbf{D}(t)$ does not depend on the order in which agents replace their associates, and may differ from $\mathbf{D}(t-1)$ by no more than 2 elements in each of the N rows.

by beliefs $\mathbf{b}(t)$ and dissonance matrix $\mathbf{D}(t)$, and the described procedure repeats.

We call a network *stable* if no agent has incentives to change either their beliefs or their associates: $\mathbf{b}(t) = \mathbf{b}(t-1) \equiv \mathbf{b}^*$ and $\mathbf{D}(t) = \mathbf{D}(t-1) \equiv \mathbf{D}^*$. In other words, stable networks are stationary states of the described dynamics. Beliefs in a stable network are given by a right eigenvector of the matrix $\mathbf{P}^* = (\mathbf{I}_N + \operatorname{diag}\{\mathbf{D}^* \mathbb{1}_N\} - \mathbf{D}^*)^{-1}$, and hence a stable network is fully defined by a dissonance matrix $\mathbf{D}^*$.

The stability of a network generically depends on the value of the switching cost — in particular, for a sufficiently high ε, any dissonance matrix $\mathbf{D}^*$ defines a stable network.[27] However, it follows from (20) that, for a stubborn agent or any agent from a communication class, instant dissonance is equal to zero in the long run. Therefore, we come to the following

Proposition 11 (**Stable networks**). *If $\mathbf{D}^*$ consists of S stubborn agents and M communication classes with different beliefs, then the network is stable for any $\varepsilon \geq 0$.*

In particular, if $\mathbf{D}^*$ is strongly connected, then the network is always stable.

In order to provide further insights into the model with endogenous network, let us describe a simple example where we analytically derive the joint dynamics of beliefs and network, and study the convergence to a stable network. For the details, see Appendix C.1. We assume that there are two groups of agents with different initial beliefs, that all agents have the same dissonance sensitivities wrt any associate, and that agents from the same group initially have identically composed sets of associates.

Suppose that society initially consists of two groups of agents, $\mathcal{H} = \{i \in \mathcal{N} \mid b_i(0) = b_h(0)\}$ and $\mathcal{L} = \{i \in \mathcal{N} \mid b_i(0) = b_l(0)\}$, with high $b_h(0)$ and low $b_l(0)$ initial beliefs, respectively. The difference in beliefs between the two groups in period 0 is positive, $\Delta(0) = b_h(0) - b_l(0) > 0$.

Each agent i has A associates; that is, agents j for whom $d_{ij} > 0$, and $d_{ij} = d$ for all i. We assume that initially the network is strongly connected, and that the composition of the sets of associates is identical for each agent in the same group. Formally, in period 0, each agent from $\mathcal{H}$ has $\bar{A}_h(0)$ associates from the opposite group $\mathcal{L}$, and each agent from $\mathcal{L}$ has $\bar{A}_l(0)$ associates from $\mathcal{H}$.[28] Thus the initial conditions of the model are essentially given by initial beliefs within groups, $\{b_h(0), b_l(0)\}$, and the numbers of associates from the opposite group in the network, $\{\bar{A}_h(0), \bar{A}_l(0)\}$.

[27] In our model, unlike in typical models of network formation (Jackson and Wolinsky, 1996; Bala and Goyal, 2000), agents do not benefit from forming a link, so switching cost plays a decisive role.

[28] The number of associates from the own group is $N - \bar{A}_h(0)$ for agents from $\mathcal{H}$, and $N - \bar{A}_l(0)$ for agents from $\mathcal{L}$. Hence the values $\{\bar{A}_h(0), \bar{A}_l(0)\}$ fully determine the network in period 0.

Consider period t when the agents' beliefs are $\{b_h(t), b_l(t)\}$, and sets of associates are $\{\bar{A}_h(t), \bar{A}_l(t)\}$. First, given $\{\bar{A}_h(t), \bar{A}_l(t)\}$, agents make statements and revise their beliefs. Since agents from the same group are homogeneous, in period $t+1$, agents from the same group have identical beliefs given by

$$b_h(t+1) = b_h(t) - \frac{\Delta(t)\bar{A}_h(t)d}{1+[\bar{A}_h(t)+\bar{A}_l(t)]d}, \qquad b_l(t+1) = b_l(t) + \frac{\Delta(t)\bar{A}_l(t)d}{1+[\bar{A}_h(t)+\bar{A}_l(t)]d},$$

where $\Delta(t) = b_h(t) - b_l(t)$. Therefore, in each period is it sufficient to keep track only of the group's beliefs.

The difference in beliefs of groups $\mathcal{H}$ and $\mathcal{L}$ is positive and decreases over time, at the (decreasing) rate which depends on the composition of the sets of associates and on the dissonance sensitivity d:

$$\Delta(t+1) = \frac{1}{1+[\bar{A}_h(t)+\bar{A}_l(t)]d}\Delta(t).$$

Note that the more that agents conform to each other (the higher is d), the faster the beliefs of the two groups converge (the lower is $\Delta(t+1)/\Delta(t)$).

Second, given $\{b_h(t+1), b_l(t+1)\}$, agents revise their sets of associates. An agent replaces a current associate with a new associate if, in the new network, the instant dissonance is reduced by at least ε. Clearly, to decrease the dissonance, an agent has to replace their associate from the opposite group with a new associate from their own group. Therefore, agents from the same group again make identical decisions, and hence it is sufficient to keep track only of the group variables.

Let $T' = \min\{\bar{A}_h(0), \bar{A}_l(0)\}$, $T'' = \max\{\bar{A}_h(0), \bar{A}_l(0)\}$, and define the value

$$\varepsilon^* = \prod_{t=0}^{T'} \frac{(\Delta(0))^2 d}{(1+[\bar{A}_h(0)+\bar{A}_l(0)-2t]d)^2} \cdot \prod_{t=T'+1}^{T''} \frac{1}{(1+[\bar{A}_h(0)+\bar{A}_l(0)-(T'+t)]d)^2}. \quad (21)$$

We show that, depending on the parameters, the model with endogenous network may exhibit different long-run outcomes.

Proposition 12 (**Endogenous network**). *Consider the model with endogenous network described above, and let ε^* be given by* (21).

(i) If $\varepsilon > \varepsilon^*$, then the society reaches a consensus. The resulting stable network is strongly connected, and agents from different groups converge to the same long-run belief.

(ii) If $\varepsilon \leq \varepsilon^*$, then the society is polarized in the long run. The resulting stable network has two communication classes, and agents from different groups converge to different long-run beliefs.

Proof. See Appendix C.2. ■

Thus, depending on whether the switching cost is higher or lower than the threshold ε^*, the society may either reach a consensus or split into two non-interacting classes. We can characterize the properties of the threshold in terms of the model parameters.

Proposition 13 (**Switching cost threshold**). *Let ε^* be given by* (21).

(i) ε^* is increasing in $\Delta(0)$, and decreasing in both $\bar{A}_h(0)$ and $\bar{A}_l(0)$.

(ii) There is $d^* > 0$ such that ε^* is increasing in d for $d \leq d^*$, and ε^* is decreasing in d for $d > d^*$.

Proof. See Appendix C.3. ■

Thus, the higher the initial polarization of the society (initial difference in beliefs between the groups), the more likely the society is to split into two non-interacting classes. Moreover, when agents initially have more associates from the opposite group, the society is more likely to reach a consensus, which is quite natural.

More interesting and ambiguous is the dependence of the threshold value on the sensitivity d, as there are two counteracting effects. First, the higher the sensitivity, the more agents want to replace their associates, which reinforces the tendency towards polarization. Second, the higher the sensitivity, the more agents conform to each other and the lower the difference between the beliefs of the two groups, which reduces polarization.

Proposition 13 shows that the first effect dominates for small values of d, while the second effect takes over for large d. The general form of the threshold ε^* as a function of sensitivity d is illustrated in Fig. 6. When the sensitivity is small (in societies with a low level of conformity), then the higher the sensitivity, the more likely it is that society splits into non-interacting classes. However, when the sensitivity is sufficiently large, then the more that agents conform to each other, the more likely it is that society reaches a consensus.

It follows that *when agents revise both their beliefs and their networks, the dissonance minimization may not fully reduce the polarization.* In our example of the model with endogenous network, a strongly connected society with two different initial

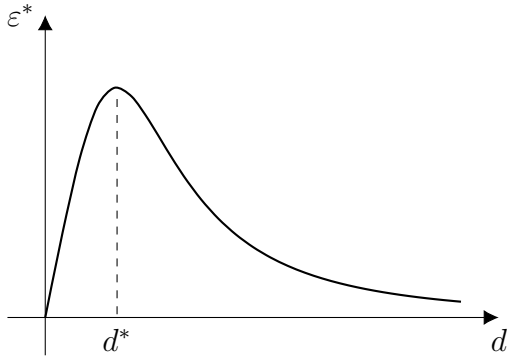

Figure 6: General form of ε^* as a function of d.

beliefs does not necessarily reach a consensus. When the switching cost is sufficiently low, a society splits into two communication classes with different long-run beliefs. The higher the initial difference in beliefs, and the fewer associates from the opposite group that agents initially have, the more likely it is that society is polarized in the long run.[29]

6 Conclusion

Any model studying belief dynamics in social networks explicitly or implicitly assumes that belief is a state of mind, and hence it becomes known only through a conversation between individuals. However, most models do not take into account the fact that people in conversation may not express sincere views in conversation, when confronted with beliefs that contradict their own.

In this paper we argue that social networks are sources of both belief formation and dissonance arousal, and we study a model of social interaction in the spirit of Arifovic et al. (2015), where the belief dynamics are driven by both dissonance minimization and conversation. We derive an analytical solution to the model, and show that it provides microfoundations for the DeGroot learning model. We fully characterize the long-run beliefs in the model, providing necessary and sufficient conditions for a society to reach a consensus. Moreover, we disentangle the effects of conversation and dissonance minimization by comparing the outcomes of the models with and without

[29]This example can be easily generalized: in each period, each agent can replace $W > 1$ potential associates; or there can be more initial groups in the society; or sensitivities of agents from different groups can differ. In all these cases our results remain qualitatively the same: for a sufficiently low switching cost, an initially strongly connected society splits into several non-interacting classes.

conversation. We show that conversation redistributes agents' social influences within communication classes in favor of agents with higher self-confidence, and leaves the external impacts of those classes on the remaining agents unaffected.

We also provide some insights for a model with endogenous network where agents minimize dissonance by revising both their beliefs and associates. While in the model with a fixed network a strongly connected society eventually reaches a consensus, this is not necessarily true for the model with a changing network, where even a strongly connected society may remain polarized in the long run. This result suggests that the phenomenon of homophily generally prevents a society from reaching a consensus.

References

Acemoglu, D., Dahleh, M. A., Lobel, I., and Ozdaglar, A. (2011). Bayesian Learning in Social Networks. *Review of Economic Studies*, **78** (4), pp. 1201–1236.

Anufriev, M., Borissov, K., and Pakhnin, M. (2021). Hypocrisy in a Simple Social Interaction Model. Mimeo.

Arifovic, J., Eaton, B. C., and Walker, G. (2015). The Coevolution of Beliefs and Networks. *Journal of Economic Behavior and Organization*, **120**, pp. 46–63.

Bala, V. and Goyal, S. (1998). Learning from Neighbours. *Review of Economic Studies*, **65** (3), pp. 595–621.

Bala, V. and Goyal, S. (2000). A Noncooperative Model of Network Formation. *Econometrica*, **68** (5), pp. 1181–1229.

Berman, A. and Plemmons, R. J. (1979). *Nonnegative Matrices in the Mathematical Sciences*. Academic Press, New York.

Bolletta, U. and Pin, P. (2020). Polarization When People Choose Their Peers. Working Paper 3245800, SSRN.

Buechel, B., Hellmann, T., and Klößner, S. (2015). Opinion Dynamics and Wisdom under Conformity. *Journal of Economic Dynamics and Control*, **52**, pp. 240–257.

Buechel, B., Hellmann, T., and Pichler, M. M. (2014). The Dynamics of Continuous Cultural Traits in Social Networks. *Journal of Economic Theory*, **154**, pp. 274–309.

Chandrasekhar, A. G., Larreguy, H., and Xandri, J. P. (2020). Testing Models of Social Learning on Networks: Evidence from Two Experiments. *Econometrica*, **88** (1), pp. 1–32.

Choi, S., Gale, D., and Kariv, S. (2008). Sequential Equilibrium in Monotone Games: A Theory-based Analysis of Experimental Data. *Journal of Economic Theory*, **143** (1), pp. 302–330.

Corazzini, L., Pavesi, F., Petrovich, B., and Stanca, L. (2012). Influential Listeners: An Experiment on Persuasion Bias in Social Networks. *European Economic Review*, **56** (6), pp. 1276–1288.

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

Della Lena, S. (2019). The Spread of Misinformation in Networks with Individual and Social Learning. Working Paper 3511080, SSRN.

DeMarzo, P. M., Vayanos, D., and Zwiebel, J. (2003). Persuasion Bias, Social Influence, and Unidimensional Opinions. *Quarterly Journal of Economics*, **118** (3), pp. 909–968.

Dietzenbacher, E. (1990). Perturbations of the Perron Vector: Applications to Finite Markov Chains and Demographic Population Models. *Environment and Planning A*, **22** (6), pp. 747–761.

Echterhoff, G., Higgins, E. T., and Groll, S. (2005). Audience-tuning Effects on Memory: The Role of Shared Reality. *Journal of Personality and Social Psychology*, **89** (3), pp. 257.

Eyster, E. and Rabin, M. (2014). Extensive Imitation is Irrational and Harmful. *Quarterly Journal of Economics*, **129** (4), pp. 1861–1898.

Festinger, L. (1957). *A Theory of Cognitive Dissonance*. Stanford University Press.

Golub, B. and Jackson, M. O. (2010). Naive Learning in Social Networks and the Wisdom of Crowds. *American Economic Journal: Microeconomics*, **2** (1), pp. 112–149.

Golub, B. and Jackson, M. O. (2012). How Homophily Affects the Speed of Learning and Best-response Dynamics. *Quarterly Journal of Economics*, **127** (3), pp. 1287–1338.

Golub, B. and Sadler, E. (2016). Learning in Social Networks. In Bramoulle, Y., Galeotti, A., and Rogers, B. W., editors, *The Oxford Handbook of the Economics of Networks*. Oxford University Press, Oxford.

Groeber, P., Lorenz, J., and Schweitzer, F. (2014). Dissonance Minimization as a Microfoundation of Social Influence in Models of Opinion Formation. *The Journal of Mathematical Sociology*, **38** (3), pp. 147–174.

Harmon-Jones, E., editor (2019). *Cognitive Dissonance: Reexamining a Pivotal Theory in Psychology*. American Psychological Association.

Higgins, E. T. (1999). "Saying Is Believing" Effects: When Sharing Reality about Something Biases Knowledge and Evaluations. In Thompson, L. L., Levine, J. M., and Messick, D. M., editors, *Shared Cognition in Organizations: The Management of Knowledge*, volume 1, pp. 33–49. Lawrence Erlbaum Associates Publishers, Mahwah, NJ, US.

Horn, R. A. and Johnson, C. R. (2013). *Matrix Analysis*. Cambridge University Press.

Jackson, M. O. (2008). *Social and Economic Networks*. Princeton University Press.

Jackson, M. O. and Wolinsky, A. (1996). A Strategic Model of Social and Economic Networks. *Journal of Economic Theory*, **71** (1), pp. 44–74.

Johnson, C. R. and Smith, R. L. (2011). Inverse M-matrices, II. *Linear Algebra and its Applications*, **435** (5), pp. 953–983.

Matz, D. C. and Wood, W. (2005). Cognitive Dissonance in Groups: The Consequences of Disagreement. *Journal of Personality and Social Psychology*, **88** (1), pp. 22–37.

McGrath, A. (2017). Dealing with Dissonance: A Review of Cognitive Dissonance Reduction. *Social and Personality Psychology Compass*, **11** (12), pp. e12362.

McKimmie, B. M. (2015). Cognitive Dissonance in Groups. *Social and Personality Psychology Compass*, **9** (4), pp. 202–212.

Merlone, U. and Radi, D. (2014). Reaching Consensus on Rumors. *Physica A: Statistical Mechanics and its Applications*, **406**, pp. 260–271.

Olcina, G., Panebianco, F., and Zenou, Y. (2017). Conformism, Social Norms and the Dynamics of Assimilation. Discussion Paper 12166, Centre for Economic Policy Research.

Panebianco, F. and Verdier, T. (2017). Paternalism, Homophily and Cultural Transmission in Random Networks. *Games and Economic Behavior*, **105**, pp. 155–176.

Ushchev, P. and Zenou, Y. (2020). Social Norms in Networks. *Journal of Economic Theory*, **185** (104969).

Appendix

A The AEW computational model

Arifovic et al. (2015) consider a numerical model of social learning where N agents are embedded in a directed network and in each period choose their beliefs and associates by minimizing dissonance. They assume that agents are homogeneous in their dissonance sensitivities d, and have the same number of associates A in each period; i.e., the network is always *regular*. AEW study the case of $d \in [0, 1]$. In each period t, the dissonance matrix in AEW has in every row exactly A non-zero elements, all equal to d. The positions of these non-zero elements are either always the same (for a fixed network) or may change over time (for an endogenous network).

AEW generate initial beliefs by averaging a number of initial private clues about an issue for every agent. Those clues are taken as independent draws from a beta distribution, encompassing both uni- and bimodal distributions. At the beginning of each period t, agents engage in conversation, as described in Section 2, and revise their beliefs motivated by a desire to minimize private and social dissonance.

Further, agents revise their sets of associates motivated by a desire to minimize the dissonance between conversations. Each agent is given a random set of A associates where $A \ll N$. At the end of each period t, after the beliefs of all agents have been revised, each agent meets in sequence $W \leq A$ randomly chosen agents who are not their associates. During each such encounter, the agent considers an option to swap the associate whose belief is furthermost from their own belief with a new

acquaintance. The agent will do so iff this swap reduces the dissonance by at least the switching cost, $\varepsilon > 0$.

AEW simulate the model, keeping the total size of the population constant, $N = 200$. They define the simulated path as "converged" when two conditions are met: (a) during the last 20 periods the network did not change, and (b) the sum of belief adjustments over all N agents and over the last 20 periods did not exceed a small number 0.001. AEW find that, for a fixed network ($W = 0$), beliefs converge to the consensus, but convergence may fail when the network is endogenous ($W = 1$ with the same values of other parameters). When this happens, the population becomes visibly divided into two groups with no connections (communication classes), with each group converging to the consensus. This polarization results in an upward bias of average beliefs, indicating a failure of aggregation of social information.[30]

AEW present their results by showing how, after the simulations converge, several statistics depend on the key parameters of the model: the size of sets of associates A (which takes values 4 and 8); the dissonance sensitivity d (which has 10 different values between 0.01 and 1); and the number of possibilities for changing associates per period W (for an endogenous network, it takes values 1, 2, and 4).

The simulations suggest the following results. First, different outcomes are possible. Society may converge to the same opinion with no bias in beliefs. However, under a different parameterization, society may be divided into as many as 6 different classes and may exhibit an opinion divergence (measured by the standard deviation of beliefs after the simulation converges). It may also have a substantial bias after convergence (measured by the difference between the mean belief and the "true" belief).[31]

Second, for every considered combination of parameters, the "enlightenment" index (based on the sum of absolute deviations of beliefs from the truth) gets closer to the "true" average value, whereas the bias of beliefs, if it exists, is always positive.

Third, comparative statics reveals a strong impact of the sensitivity d on the long-run outcome.[32] When people hold stronger dissonance considerations, the result is a society with a smaller number of classes, a smaller standard deviation of beliefs, and

[30]The initialisation is somewhat special, with 60% of the population having the belief close to 1 and 40% of the population having the belief close to 0. For a fixed network, the average belief does not change much over time and is about 0.6 after 60 periods. For an endogenous network, the average belief is about 0.7 after 60 periods, when convergence is declared. See Figs. 4 and 5 in AEW.

[31]Since the population is finite, true belief is defined as the mean of the initial beliefs. If the population were to grow indefinitely, this would converge to the mean of the beta distribution.

[32]The impact of an increase in d is the same as for an increase in the local network size (A), and for a decrease in the number of possibilities to revise information (W) for all statistics, except the bias.

a higher enlightenment. In other words, information is better aggregated and society is less divided when people are more sensitive to cognitive dissonance and thus tune their messages more strongly to the messages of the others.

B Proofs of the main results

B.1 Proof of Proposition 1

Let $\mathbf{H} = (\mathbf{I}_N + \operatorname{diag}\{\mathbf{D}\mathbb{1}_N\})^{-1}$. Consider the matrix

$$\mathbf{HD} = \begin{pmatrix} 0 & \frac{d_{12}}{1+\sum_k d_{1k}} & \cdots & \frac{d_{1N}}{1+\sum_k d_{1k}} \\ \frac{d_{21}}{1+\sum_k d_{2k}} & 0 & \cdots & \frac{d_{2N}}{1+\sum_k d_{2k}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{d_{N1}}{1+\sum_k d_{Nk}} & \frac{d_{N2}}{1+\sum_k d_{Nk}} & \cdots & 0 \end{pmatrix}. \tag{B.1}$$

Since each row sum of $\mathbf{HD}$ is less than 1, its spectral radius is less than 1. Hence $\lim_{t\to\infty}(\mathbf{HD})^t = 0$, and $\mathbf{I}_N - \mathbf{HD}$ is invertible: $(\mathbf{I}_N - \mathbf{HD})^{-1} = \sum_{t=0}^{\infty}(\mathbf{HD})^t$. Matrix $\mathbf{P}$ can be written as $\mathbf{P} = (\mathbf{I}_N + \operatorname{diag}\{\mathbf{D}\mathbb{1}_N\} - \mathbf{D})^{-1} = (\mathbf{H}^{-1} - \mathbf{D})^{-1} = (\mathbf{H}^{-1}(\mathbf{I}_N - \mathbf{HD}))^{-1} = (\mathbf{I}_N - \mathbf{HD})^{-1}\mathbf{H}$, and therefore $\mathbf{P}$ is non-negative, because

$$\mathbf{P} = \left(\mathbf{I}_N + \sum\nolimits_{t=1}^{\infty}(\mathbf{HD})^t\right)\mathbf{H}. \tag{B.2}$$

Since $(\mathbf{HD} + \mathbf{H})\mathbb{1}_N = \mathbb{1}_N$, we have $\mathbf{H}\mathbb{1}_N = (\mathbf{I}_N - \mathbf{HD})\mathbb{1}_N$, and it immediately follows that $\mathbf{P}\mathbb{1}_N = (\mathbf{I}_N - \mathbf{HD})^{-1}\mathbf{H}\mathbb{1}_N = \mathbb{1}_N$, i.e., $\mathbf{P}$ is row-stochastic.

B.2 Proof of Proposition 2

The dynamics of statements in the fast timescale can be written as

$$\mathbf{s}^{\tau}(t) = \mathbf{Hb}(t-1) + \mathbf{HDs}^{\tau-1}(t), \tag{B.3}$$

where $\mathbf{H} = (\mathbf{I}_N + \operatorname{diag}\{\mathbf{D}\mathbb{1}_N\})^{-1}$ and $\mathbf{HD}$ is given by (B.1). Iterating (B.3), we get

$$\mathbf{s}^{\tau}(t) = (\mathbf{HD})^{\tau}\mathbf{s}^{0}(t) + (\mathbf{I}_N + \mathbf{HD} + \ldots + (\mathbf{HD})^{\tau})\mathbf{Hb}(t-1).$$

Since $\lim_{\tau\to\infty}(\mathbf{HD})^\tau = 0$, the dynamic system (B.3) is stable, and statements in the fast timescale converge to some $\mathbf{s}^*(t)$ given by

$$\mathbf{s}^*(t) = \lim_{\tau\to\infty}\mathbf{s}^\tau(t) = \left(\mathbf{I}_N + \sum\nolimits_{\tau=1}^{\infty}(\mathbf{HD})^\tau\right)\mathbf{Hb}(t-1) = \mathbf{Pb}(t-1),$$

where the last equality follows from (B.2). Comparing it with (7), we get $\mathbf{s}^*(t) = \mathbf{s}(t)$.

B.3 Proof of Proposition 3 and Corollary 1

Let $\mathbf{H} = (\mathbf{I}_N + \operatorname{diag}\{\mathbf{D}\mathbb{1}_N\})^{-1}$, and $\mathbf{HD}$ be given by (B.1). It is easily seen from (B.1) that the (i,j)-entry in $(\mathbf{HD})^t$ is determined by all paths in $\mathbf{D}$ from i to j of length t. Therefore, it follows from (B.2) that for $i \neq j$,

$$p_{ij} = \frac{1}{1+\sum_k d_{jk}}\frac{d_{ij}}{1+\sum_k d_{ik}} + \sum_{t=2}^{\infty}\sum_{n_1,n_2,\ldots,n_{t-1}=1}^{N}\frac{1}{1+\sum_k d_{jk}}\cdot \\ \cdot\frac{d_{in_1}}{1+\sum_k d_{ik}}\frac{d_{n_1n_2}}{1+\sum_k d_{n_1k}}\cdots\frac{d_{n_{t-2}n_{t-1}}}{1+\sum_k d_{n_{t-2}k}}\frac{d_{n_{t-1}j}}{1+\sum_k d_{n_{t-1}k}}.$$

Each term in p_{ij} contains the product of the form $d_{in_1}\cdot d_{n_1n_2}\cdots d_{n_{t-1}j}$. Therefore, $p_{ij} = 0$ iff there is no path in $\mathbf{D}$ from i to j.

It follows from (B.2) that $\mathbf{P} = (\mathbf{HD})\,\mathbf{P} + \mathbf{H}$, and hence for all i,

$$p_{ii} \;=\; \frac{1}{1+\sum_k d_{ik}} + \sum\nolimits_{j\neq i}\frac{d_{ij}}{1+\sum_k d_{ik}}p_{ji} \;\geq\; \frac{1}{1+\sum_k d_{ik}} \;=\; t_{ii} \;>0.$$

Moreover, the matrix $\mathbf{P}^{-1} = (\mathbf{I}_N + \operatorname{diag}\{\mathbf{D}\mathbb{1}_N\} - \mathbf{D})$ is diagonally dominant of its rows; i.e., its diagonal elements are greater than the sum of the moduli of the corresponding off-diagonal elements by row (for all i, we have $1+\sum_{k\neq i}d_{ik} > \sum_{k\neq i}d_{ik}$). Therefore (see, e.g., Theorem 1.3 in Johnson and Smith, 2011), $\mathbf{P}$ is diagonally dominant of its column entries; i.e., $p_{ii} > p_{hi}$ for all $h \neq i$.

If $d_{hi} = 0$ for all h, then there is no path in $\mathbf{D}$ ending in i. Hence $p_{hi} = 0$ for $h \neq i$, and $p_{ii} = 1/(1+\sum_k d_{ik})$. If $d_{ij} = 0$ for all j, then there is no path in $\mathbf{D}$ starting from i. Hence $p_{ij} = 0$ for $j \neq i$, and $p_{ii} = 1$ because $\mathbf{P}$ is row-stochastic.

B.4 Proof of Proposition 4

Let $\tilde{\mathbf{D}}$ be a perturbation of $\mathbf{D}$ such that $\tilde{d}_{ij} > d_{ij}$, while all other elements in the two matrices are the same. Let also $\tilde{\mathbf{H}} = (\mathbf{I}_N + \operatorname{diag}\{\tilde{\mathbf{D}}\mathbb{1}_N\})^{-1}$. Consider the matrix

$\tilde{\mathbf{P}} = (\mathbf{I}_N - \tilde{\mathbf{H}}\tilde{\mathbf{D}})^{-1}\tilde{\mathbf{H}}$, and note that we have $\tilde{\mathbf{P}} = (\tilde{\mathbf{H}}\tilde{\mathbf{D}})\tilde{\mathbf{P}} + \tilde{\mathbf{H}}$.

Denote for brevity $\tilde{\mathbf{A}} \equiv \tilde{\mathbf{H}}\tilde{\mathbf{D}}$ and $\tilde{\mathbf{C}} \equiv \tilde{\mathbf{H}}$, and let similarly $\mathbf{A} \equiv \mathbf{HD}$ and $\mathbf{C} \equiv \mathbf{H}$. Since $\tilde{d}_{ij} > d_{ij}$, $\tilde{\mathbf{A}}$ and $\tilde{\mathbf{C}}$ differ from $\mathbf{A}$ and $\mathbf{C}$ respectively only in rows i. In particular, $\tilde{a}_{ij} > a_{ij}$ while $\tilde{a}_{ik} \leq a_{ik}$ for all $k \neq j$, and $\tilde{c}_{ik} \leq c_{ik}$ for all k.

Let $\{\tilde{\mathbf{P}}(t)\}_{t=0}^{\infty}$ be recursively defined by $\tilde{\mathbf{P}}(t) = \tilde{\mathbf{A}}\tilde{\mathbf{P}}(t-1) + \tilde{\mathbf{C}}$, for $\tilde{\mathbf{P}}(0) = \mathbf{P}$. Consider $\tilde{\mathbf{P}}(1) = \tilde{\mathbf{A}}\mathbf{P} + \tilde{\mathbf{C}}$. Since $\tilde{\mathbf{A}}$ and $\tilde{\mathbf{C}}$ are perturbed only in row i, $\tilde{p}_{h\ell}(1) = p_{h\ell}$ for all $h \neq i$ and for all ℓ. At the same time, for all ℓ,

$$\tilde{p}_{i\ell}(1) \;=\; \sum\nolimits_{k\neq j} \tilde{a}_{ik}p_{k\ell} + \tilde{a}_{ij}p_{j\ell} + \tilde{c}_{i\ell} \;=\; \tilde{a}_{ij}p_{j\ell} + (1-\tilde{a}_{i\ell})\frac{\sum_{k\neq j}\tilde{a}_{ik}p_{k\ell} + \tilde{c}_{i\ell}}{\sum_{k\neq j}\tilde{a}_{ik} + \sum_k \tilde{c}_{ik}}.$$

Similarly,

$$p_{i\ell} \;=\; a_{ij}p_{j\ell} + (1-a_{ij})\frac{\sum_{k\neq j}a_{ik}p_{k\ell} + c_{i\ell}}{\sum_{k\neq j}a_{ik} + \sum_k c_{ik}} \;=\; a_{ij}p_{j\ell} + (1-a_{ij})\frac{\sum_{k\neq j}\tilde{a}_{ik}p_{k\ell} + \tilde{c}_{i\ell}}{\sum_{k\neq j}\tilde{a}_{ik} + \sum_k \tilde{c}_{ik}},$$

where the last equality uses the fact that $\tilde{a}_{ik}/a_{ik} = \tilde{c}_{ir}/c_{ir}$ for all $k \neq j$ and all r.

Suppose that ℓ is such that $p_{j\ell} > p_{i\ell}$. Then

$$p_{j\ell} \;>\; p_{i\ell} \;>\; \frac{\sum_{k\neq j}\tilde{a}_{ik}p_{k\ell} + \tilde{c}_{i\ell}}{\sum_{k\neq j}\tilde{a}_{ik} + \sum_k \tilde{c}_{ik}},$$

and since $\tilde{a}_{ij} > a_{ij}$, we have $\tilde{p}_{i\ell}(1) > p_{i\ell}$.

Let us show that in this case, for all $t \geq 2$, $\tilde{p}_{h\ell}(t) \geq p_{h\ell}$ for all h; and, furthermore, for sufficiently large t, the inequality is strict for $h \in H_i$. We will use mathematical induction. Consider $t = 2$. For all $h \neq i$, we have

$$\tilde{p}_{h\ell}(2) \;=\; \sum\nolimits_k \tilde{a}_{hk}\tilde{p}_{k\ell}(1) + \tilde{c}_{h\ell} \;\geq\; \sum\nolimits_k a_{hk}p_{k\ell} + c_{h\ell} \;=\; p_{h\ell},$$

and the inequality is strict if $a_{hi} > 0$ (which is true iff $d_{hi} > 0$). For $h = i$, we have

$$\tilde{p}_{i\ell}(2) \;=\; \sum\nolimits_k \tilde{a}_{ik}\tilde{p}_{k\ell}(1) + \tilde{c}_{i\ell} \;\geq\; \sum\nolimits_k \tilde{a}_{ik}p_{k\ell} + \tilde{c}_{i\ell} \;=\; \tilde{p}_{i\ell}(1) \;>\; p_{i\ell}.$$

Suppose that we have proved our inequalities for $t = T$. Let us prove them for $t = T + 1$. For all $h \neq i$, we have

$$\tilde{p}_{h\ell}(T+1) \;=\; \sum\nolimits_k \tilde{a}_{hk}\tilde{p}_{k\ell}(T) + \tilde{c}_{h\ell} \;\geq\; \sum\nolimits_k a_{hk}p_{k\ell} + c_{h\ell} \;=\; p_{h\ell},$$

and the inequality is strict if $a_{hk} > 0$ for k, for whom strict inequality was shown at some previous step (i.e., there is a path in $\mathbf{D}$ from h to i). For $h = i$, we have

$$\tilde{p}_{i\ell}(T+1) \;=\; \sum\nolimits_k \tilde{a}_{ik}\tilde{p}_{k\ell}(T) + \tilde{c}_{i\ell} \;\geq\; \sum\nolimits_k \tilde{a}_{ik}p_{k\ell} + \tilde{c}_{i\ell} \;=\; \tilde{p}_{i\ell}(1) \;>\; p_{i\ell}.$$

It is easily seen that the sequence $\{\tilde{\mathbf{P}}(t)\}_{t=0}^{\infty}$ converges to $\tilde{\mathbf{P}}$. In the limit, for all ℓ such that $p_{j\ell} > p_{i\ell}$, we get $\tilde{p}_{h\ell} \geq p_{h\ell}$ for all h. Furthermore, the inequality is strict for $h \in H_i$. The case where $p_{j\ell} < p_{i\ell}$ can be considered analogously.

B.5 Proof of Proposition 7

Since social influences satisfy the system of equations (16), $\boldsymbol{\pi}$ is the positive unit left eigenvector of the non-negative and strongly connected matrix

$$\mathbf{Q} = \begin{pmatrix} 0 & \frac{d_{12}}{\sum_k d_{2k}} & \cdots & \frac{d_{1N}}{\sum_k d_{Nk}} \\ \frac{d_{21}}{\sum_k d_{1k}} & 0 & \cdots & \frac{d_{2N}}{\sum_k d_{Nk}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{d_{N1}}{\sum_k d_{1k}} & \frac{d_{N2}}{\sum_k d_{2k}} & \cdots & 0 \end{pmatrix}.$$

Let $\tilde{\mathbf{Q}}$ be a perturbation of $\mathbf{Q}$ such that $\tilde{d}_{ij} > d_{ij}$, while all other elements in the two matrices are the same. Clearly, in $\tilde{\mathbf{Q}}$ column j has increased and column i has decreased, while all other columns remain intact. Let $\tilde{\boldsymbol{\pi}}$ be the positive unit left eigenvector of $\tilde{\mathbf{Q}}$. It follows from Lemma 1 in Dietzenbacher (1990) that for all ℓ, $\tilde{\pi}_j/\pi_j \geq \tilde{\pi}_\ell/\pi_\ell \geq \tilde{\pi}_i/\pi_i$, and therefore

$$\frac{\pi_\ell}{\pi_j} \;\geq\; \frac{\tilde{\pi}_\ell}{\tilde{\pi}_j}, \qquad \text{and} \qquad \frac{\pi_\ell}{\pi_i} \;\leq\; \frac{\tilde{\pi}_\ell}{\tilde{\pi}_i}. \tag{B.4}$$

Since $\mathbf{Q}$ and $\tilde{\mathbf{Q}}$ do not commute, $\boldsymbol{\pi}$ is not proportional to $\tilde{\boldsymbol{\pi}}$. Therefore, for some ℓ, the first inequality in (B.4) is strict. Summing up the first inequalities in (B.4) over ℓ and taking into account that both $\boldsymbol{\pi}$ and $\tilde{\boldsymbol{\pi}}$ are normalized to one, we get $1/\pi_j = \sum_\ell \pi_\ell/\pi_j > \sum_\ell \tilde{\pi}_\ell/\tilde{\pi}_j = 1/\tilde{\pi}_j$, and hence $\tilde{\pi}_j > \pi_j$. Applying the same argument to the second inequality in (B.4), we also get that $\tilde{\pi}_i < \pi_i$.

Using (16) and (17), we get $\tilde{\theta}_\ell/\theta_\ell = \tilde{\pi}_\ell/\pi_\ell \cdot (1 + \sum_k d_{\ell k})/(1 + \sum_k \tilde{d}_{\ell k})$ for all ℓ. Hence $\tilde{\theta}_j/\theta_j \geq \tilde{\theta}_\ell/\theta_\ell \geq \tilde{\theta}_i/\theta_i$, and by the same argument as above, $\tilde{\theta}_j > \theta_j$ and $\tilde{\theta}_i < \theta_i$.

B.6 Proof of Proposition 8

Let $\mathbf{G} = \{g_{ij}\}_{i,j=1}^N$ be such that $g_{ij} = 1$ when $d_{ij} > 0$ and $g_{ij} = 0$ when $d_{ij} = 0$. Then Eq. (16) takes the form $\pi_i d_i \sum_k g_{ik} = \sum_{h:g_{hi}=1} \pi_h d_h$. It is directly checked that the solution to this system is given by $\pi_i = \lambda_i \prod_{j \neq i} d_j$, where $\{\lambda_i\}_{i=1}^N$ satisfies the system of equations $\lambda_i \sum_k g_{ik} = \sum_{h:g_{hi}=1} \lambda_h$, which depends only on the adjacency matrix (i.e., the network structure) and does not depend on the sensitivities (i.e., the values d_i). Since the sum of social influences is normalized to one, we finally have

$$\pi_i \;=\; \frac{\lambda_i \prod_{j \neq i} d_j}{\sum_k \lambda_k \prod_{\ell \neq k} d_\ell}, \qquad \text{for any } i\,.$$

Thus, for each i, π_i is decreasing in d_i and increasing in every d_j for $j \neq i$.

B.7 Proof of Proposition 9

When all agents have the same number of associates A and identical sensitivity parameters d, we have $\mathbf{H} = (\mathbf{I}_N + \mathrm{diag}\,\{\mathbf{D}\mathbb{1}_N\})^{-1} = \frac{1}{1+Ad}\mathbf{I}_N$, and it is clear that $\mathbf{H}$ commutes with $\mathbf{HD}$. It is well known (see Theorem 2.4.8.1 in Horn and Johnson, 2013) that when two matrices commute, the eigenvalues of the sum (resp., the product) of two matrices are the sum (resp., product) of their eigenvalues. Since $\mathbf{T} = \mathbf{H} + \mathbf{HD}$, its eigenvalues satisfy $\lambda_i(\mathbf{T}) = \lambda_i(\mathbf{H}) + \lambda_j(\mathbf{HD}) = \frac{1+\lambda_j(\mathbf{D})}{1+Ad}$. Since $\mathbf{P}$ is given by (B.2), its eigenvalues satisfy $\lambda_i(\mathbf{P}) = (1 + \sum_{t=1}^\infty (\lambda_j(\mathbf{D}))^t)\,\frac{1}{1+Ad} = \frac{1}{1+Ad-\lambda_j(\mathbf{D})}$. It is easily seen that $|\lambda_i(\mathbf{T})|$ and $|\lambda_i(\mathbf{P})|$ are increasing in the real part of $\lambda_j(\mathbf{D})$, and hence the second-largest eigenvalues of $\mathbf{T}$ and $\mathbf{P}$ correspond to the same $\lambda_2(\mathbf{D})$.

For the complete network, $A = N - 1$ and $\lambda_2(\mathbf{D}) = -d$. Then

$$|\lambda_2(\mathbf{T})| \lessgtr |\lambda_2(\mathbf{P})| \;\Leftrightarrow\; |1-d| \lessgtr 1 - \frac{d}{1+Nd}.$$

Clearly, when $0 < d \leq 1$, $|\lambda_2(\mathbf{T})| < |\lambda_2(\mathbf{P})|$. When $d > 1$, the above condition means

$$|\lambda_2(\mathbf{T})| \lessgtr |\lambda_2(\mathbf{P})| \;\Leftrightarrow\; Nd^2 - (2N-2)d - 2 \lessgtr 0.$$

The positive solution to the equation $Nd^2 - (2N-2)d - 2 = 0$ is given by $d^* = 1 - \frac{1}{N} + \sqrt{1 + \frac{1}{N^2}}$. Thus $|\lambda_2(\mathbf{T})| < |\lambda_2(\mathbf{P})|$ for $d < d^*$, and *vice versa*.

B.8 Proof of Proposition 10

In the proof of this proposition, any stubborn agent may be considered a separate communication class consisting of a single agent. Therefore, without loss of generality, we put $S = 0$. It then follows from (15) that the matrix of external impacts $\mathbf{\Gamma}$ satisfies $\mathbf{\Gamma} = \mathbf{A}\mathbf{\Gamma} + \mathbf{C}$, where we have denoted $\mathbf{A} = \mathbf{T}_{\mathcal{R}\mathcal{R}}$ and $\mathbf{C} = (\mathbf{T}_{\mathcal{R}1}\mathbb{1}_{D_1}, \ldots, \mathbf{T}_{\mathcal{R}M}\mathbb{1}_{D_M})$. Note that since $\mathbf{A}\mathbb{1}_R + \mathbf{C}\mathbb{1}_M = \mathbb{1}_R$, we have $\mathbf{C}\mathbb{1}_M = (\mathbf{I}_R - \mathbf{A})\,\mathbb{1}_R$, and hence $\mathbf{\Gamma}\mathbb{1}_M = (\mathbf{I}_R - \mathbf{A})^{-1}\,\mathbf{C}\mathbb{1}_M = \mathbb{1}_R$, so $\mathbf{\Gamma}$ is row-stochastic.

Suppose that agent $i \in \mathcal{R}$ increases their dissonance sensitivity wrt some agent $j \in \mathcal{D}_{m^*}$. Let $\tilde{\mathbf{D}}$ be a perturbation of $\mathbf{D}$ such that $\tilde{d}_{ij} > d_{ij}$, while all other elements in the two matrices are the same. Let also $\tilde{\mathbf{A}} = (\mathbf{I}_R + \mathrm{diag}\{\tilde{\mathbf{D}}_{\mathcal{R}\mathcal{R}}\mathbb{1}_R\})^{-1}(\mathbf{I}_R + \tilde{\mathbf{D}}_{\mathcal{R}\mathcal{R}})$, and $\tilde{\mathbf{C}} = \{\tilde{c}_{rm}\}$ be such that $\tilde{c}_{rm} = \frac{1}{1+\sum_{\ell\in\mathcal{N}}\tilde{d}_{r\ell}}\sum_{\ell\in\mathcal{D}_m}\tilde{d}_{r\ell}$. Then the perturbed matrix of external impacts satisfies $\tilde{\mathbf{\Gamma}} = \tilde{\mathbf{A}}\tilde{\mathbf{\Gamma}} + \tilde{\mathbf{C}}$.

Note that $\tilde{\mathbf{A}}$ and $\tilde{\mathbf{C}}$ differ from $\mathbf{A}$ and $\mathbf{C}$ respectively only in rows i. In particular, $\tilde{a}_{i\ell} \le a_{i\ell}$ for all ℓ, and $\tilde{a}_{ii} < a_{ii}$, while $\tilde{c}_{im^*} > c_{im^*}$, and $\tilde{c}_{im} \le c_{im}$ for all $m \neq m^*$.

Let $\{\tilde{\mathbf{\Gamma}}(t)\}_{t=0}^{\infty}$ be recursively defined by $\tilde{\mathbf{\Gamma}}(t) = \tilde{\mathbf{A}}\tilde{\mathbf{\Gamma}}(t-1) + \tilde{\mathbf{C}}$, for $\tilde{\mathbf{\Gamma}}(0) = \mathbf{\Gamma}$. Consider $\tilde{\mathbf{\Gamma}}(1) = \tilde{\mathbf{A}}\mathbf{\Gamma} + \tilde{\mathbf{C}}$. Since $\tilde{\mathbf{A}}$ and $\tilde{\mathbf{C}}$ are perturbed only in row i, $\tilde{\gamma}_{rm}(1) = \gamma_{rm}$ for all $r \neq i$, and for all $m = 1, \ldots, M$. At the same time, for all $m \neq m^*$,

$$\tilde{\gamma}_{im}(1) \;=\; \sum_{\ell\in\mathcal{R}} \tilde{a}_{i\ell}\gamma_{\ell m} + \tilde{c}_{im} \;<\; \sum_{\ell\in\mathcal{R}} a_{i\ell}\gamma_{\ell m} + c_{im} \;=\; \gamma_{im^*}.$$

Using the same argument as in the proof of Proposition 4 (see Appendix B.4), it can be easily seen that for all $r \in \mathcal{R}$ and $m \neq m^*$, $\tilde{\gamma}_{rm}(t) \le \gamma_{rm}$ for all $t \ge 2$. Furthermore, for sufficiently large t, this inequality is strict for $r = i$ and r such that there is a path in $\mathbf{D}_{\mathcal{R}\mathcal{R}}$ from r to i. Since $\{\tilde{\mathbf{\Gamma}}(t)\}_{t=0}^{\infty}$ converges to $\tilde{\mathbf{\Gamma}}$, in the limit we have $\tilde{\gamma}_{rm} \le \gamma_{rm}$. Furthermore, the inequality is strict for $r = i$ and r such that there is a path in $\mathbf{D}_{\mathcal{R}\mathcal{R}}$ from r to i. Since $\tilde{\mathbf{\Gamma}}$ is row-stochastic, the opposite inequalities hold for $m = m^*$, which proves parts (*i*) and (*ii*).

Suppose now that agent $i \in \mathcal{R}$ increases sensitivity d_{ij} wrt agent $j \in \mathcal{R}$. Let $\tilde{\mathbf{D}}$ be a perturbation of $\mathbf{D}$ such that $\tilde{d}_{ij} > d_{ij}$, while all other elements in the two matrices are the same. Let $\tilde{\mathbf{A}}$ and $\tilde{\mathbf{C}}$ be the corresponding perturbations of $\mathbf{A}$ and $\mathbf{C}$ respectively, as defined above. Again, $\tilde{\mathbf{A}}$ and $\tilde{\mathbf{C}}$ differ from $\mathbf{A}$ and $\mathbf{C}$ respectively only in rows i. In this case $\tilde{a}_{ij} > a_{ij}$ while $\tilde{a}_{i\ell} \le a_{i\ell}$ for all $\ell \neq j$, and $\tilde{c}_{im} \le c_{im}$ for all m. To prove part (*iii*), it remains only to repeat the argument from the proof of Proposition 4 (see Appendix B.4) for the matrix $\tilde{\mathbf{\Gamma}}$ defined by $\tilde{\mathbf{\Gamma}} = \tilde{\mathbf{A}}\tilde{\mathbf{\Gamma}} + \tilde{\mathbf{C}}$.

C Model with endogenous network

C.1 Description of the model

Let us characterize the dynamics of the example of the endogenous network model. Suppose that, in period t, agents' beliefs $\{b_h(t), b_l(t)\}$ and networks $\{\bar{A}_h(t), \bar{A}_l(t)\}$ are given. First, given networks $\{\bar{A}_h(t), \bar{A}_l(t)\}$, agents make statements and revise their beliefs. Since agents from the same group are homogeneous, new statements in period $t+1$ within each group are the same (equilibrium is symmetric). Beliefs always coincide with statements, and hence in period $t+1$ agents from the same group have identical beliefs. Thus in each period we need to keep track only of the group beliefs: $b_h(t+1)$ for agents from $\mathcal{H}$ and $b_l(t+1)$ for agents from $\mathcal{L}$. By the first order condition (6) and the fact that $\mathbf{b}(t) = \mathbf{s}(t)$, we get

$$b_h(t+1) = \frac{1}{1+Ad} b_h(t) + \frac{d}{1+Ad}(A - \bar{A}_h(t)) b_h(t+1) + \frac{d}{1+Ad} \bar{A}_h(t) b_l(t+1),$$
$$b_l(t+1) = \frac{1}{1+Ad} b_l(t) + \frac{d}{1+Ad} \bar{A}_l(t) b_h(t+1) + \frac{d}{1+Ad}(A - \bar{A}_l(t)) b_l(t+1),$$

and hence, denoting $\Delta(t) = b_h(t) - b_l(t)$, we have

$$b_h(t+1) = b_h(t) - \frac{\Delta(t)\bar{A}_h(t)d}{1 + [\bar{A}_h(t) + \bar{A}_l(t)]d}, \qquad b_l(t+1) = b_l(t) + \frac{\Delta(t)\bar{A}_l(t)d}{1 + [\bar{A}_h(t) + \bar{A}_l(t)]d}.$$

Second, given the new beliefs $\{b_h(t+1), b_l(t+1)\}$, agents revise their sets of associates. An agent replaces a current associate whose belief is furthermost from their own with a new associate if, in the new network, the dissonance is reduced by at least the switching cost $\varepsilon > 0$. After the beliefs are formed, the instant dissonance of the agent from $\mathcal{H}$ is given by $(b_h(t+1) - b_h(t))^2 + \bar{A}_h(t) d \left(\Delta(t+1)\right)^2$. In this expression, $b_h(t+1)$ and $\Delta(t+1)$ are already chosen, so the only decision variable is $\bar{A}_h(t)$. To decrease the dissonance, an agent has to decrease $\bar{A}_h(t)$; i.e., to replace an associate from the opposite group $\mathcal{L}$ with a new associate from their group $\mathcal{H}$. Similarly, any agent from $\mathcal{L}$ has to decrease the number of their associates from the opposite group $\mathcal{H}$. It again follows that agents from the same group make identical decisions, and it is sufficient to keep track only of the group variables.

If an agent from any group replaces their current associate from the opposite group with a new associate from the own group, the dissonance reduces by $d\left(\Delta(t+1)\right)^2$. Therefore, in each period there are two cases. First, $d\left(\Delta(t+1)\right)^2 < \varepsilon$. Then no

agent replaces an associate, the network does not change, and the sets of associates of all agents remain the same: $\bar{A}_h(t+1) = \bar{A}_h(t)$ and $\bar{A}_l(t+1) = \bar{A}_l(t)$. Second, $d\left(\Delta(t+1)\right)^2 \geq \varepsilon$. Here any agent from $\mathcal{H}$ replaces an associate from $\mathcal{L}$ (with a new associate from $\mathcal{H}$), and vice versa. However, this is possible only as long as there are associates in the opposite group. Hence, before the next period, we have

$$\bar{A}_i(t+1) = \max\{\bar{A}_i(t) - 1, 0\} = \max\{\bar{A}_i(0) - (t+1), 0\}, \qquad i = h, l.$$

If agents continue to replace their associates, there are two periods in which the dynamics of beliefs and the rate of decrease of $\Delta(t)$ change. In period $T' = \min\{\bar{A}_h(0), \bar{A}_l(0)\}$, either agents from $\mathcal{H}$ do not have any associates from $\mathcal{L}$, or agents from $\mathcal{L}$ do not have any associates from $\mathcal{H}$. Then for $t \geq T' + 1$, in the former case we have $\bar{A}_h(t) = 0$ and $b_h(t) = b_h(T' + 1)$, while in the latter case $\bar{A}_l(t) = 0$ and $b_l(t) = b_l(T' + 1)$. In period $T'' = \max\{\bar{A}_h(0), \bar{A}_l(0)\}$, agents from the less influential group will also remove all their associates from the opposite group: for $t \geq T'' + 1$, $\bar{A}_h(t) = \bar{A}_l(t) = 0$, and $b_h(t) = b_h(T'' + 1)$, $b_l(t) = b_l(T'' + 1)$.

C.2 Proof of Proposition 12

When $\varepsilon > \varepsilon^*$, it is costly for agents to fully replace their associates from the opposite group. There is period T ($0 \leq T \leq T''$) in which agents from both groups do not replace their associates. Starting from T, the sets of associates do not change, and agents from at least one group have at least one associate from the opposite group (either $\bar{A}_h(T) > 0$, or $\bar{A}_l(T) > 0$, or both). Hence, the network remains strongly connected, and agents from both groups converge to the same long-run belief: $\lim_{t\to\infty} b_h(t) = b_h^* = b_l^* = \lim_{t\to\infty} b_l(t)$.

Since for all $t \geq T$, $\bar{A}_h(t) = \bar{A}_h(T)$ and $\bar{A}_l(t) = \bar{A}_l(T)$, it is easily seen that $b_h^* = b_l^* = \frac{\bar{A}_l(T)}{\bar{A}_h(T)+\bar{A}_l(T)} b_h(T) + \frac{\bar{A}_h(T)}{\bar{A}_h(T)+\bar{A}_l(T)} b_l(T)$. The consensus belief in the society is a weighted average of the beliefs in both groups in the period when the network becomes fixed, and the weight of each group is proportional to the number of associates from that group that are eventually linked to agents from the opposite group.

When $\varepsilon \leq \varepsilon^*$, after period T'' the initially strongly connected society splits into two classes with different beliefs. Agents from $\mathcal{H}$ ($\mathcal{L}$) are linked only to agents from $\mathcal{H}$ ($\mathcal{L}$). Therefore, for $t \geq T'' + 1$, agents do not change their beliefs or associates: $\bar{A}_h(t) = \bar{A}_l(t) = 0$, and $b_h(t) = b_h(T'' + 1) = b_h^*$, while $b_l(t) = b_l(T'' + 1) = b_l^*$. Since $b_h^* - b_l^* = \Delta(T'' + 1) > 0$, the consensus is never reached.

C.3 Proof of Proposition 13

Clearly, ε^* is increasing in $\Delta(0)$. The higher $\bar{A}_h(0)$ (resp., $\bar{A}_l(0)$) is, the lower the terms of the form $1/(1+[\bar{A}_h(0)+\bar{A}_l(0)-2t]d)^2$ are, and the higher T' is, so ε^* has more terms of this form (which are less than 1), and thus ε^* is decreasing in $\bar{A}_h(0)$ and $\bar{A}_l(0)$. The dependence of ε^* on d can be written as $F(d) = \frac{d}{C_n d^n + C_{n-1} d^{n-1} + \ldots + C_2 d^2 + C_1 d + 1}$, where $n = 2T''$ and $C_i > 0$ for all i. Therefore,

$$F'(d) = \frac{1 - (n-1)C_n d^n - (n-2)C_{n-1} d^{n-1} - \ldots - C_2 d^2}{(C_n d^n + C_{n-1} d^{n-1} + \ldots + C_2 d^2 + C_1 d + 1)^2}.$$

Since $F'(0) = 1$ and its numerator is decreasing in d, there exists $d^* > 0$ such that $F'(d) \geq 0$ for $d \leq d^*$, while $F'(d) < 0$ for $d > d^*$.