Misinformation due to asymmetric information sharing

Berno Buechel[†⋆], Stefan Klößner[†♭], Fanyuan Meng[‡], Anis Nassar[†]

[†]Department of Economics, University of Fribourg, 1700 Fribourg, Switzerland

[♭]Faculty of Educational and Social Sciences, University of Vechta, 49377 Vechta, Germany

[‡]Department of Physics, University of Fribourg, 1700 Fribourg, Switzerland

[⋆]Correspondence: berno.buechel@unifr.ch

May 2021

Abstract

We introduce a model of social learning in which agents share true and false information with different decays and to different networks of people. Our results establish that these asymmetries, thus far largely ignored in theory despite being empirically established, govern the long-run beliefs of a society. We derive a single threshold condition under which misinformation prevails. Misinformation is more likely when false information decays less, and when the false information network is locally denser, as measured by the largest eigenvalue of its adjacency matrix. Under these conditions, all agents guess the wrong state; a result that the literature reaches only in the presence of forceful or stubborn agents. Additionally, we measure speed of convergence and show that agents who are more central in the false information network are more prone to be misinformed. We illustrate our results using numerical simulations that incorporate societies segmented into groups, and derive policy implications that center on human sharing behaviors and network structures.

Keywords: Misinformation, asymmetry, social networks, social learning, opinion dynamics, disinformation, echo chamber

JEL Classification Codes: D83, D85

1 Introduction

Misinformation is thriving. This is problematic because it can lead people to inaccurate beliefs. These beliefs may then generate various socially harmful outcomes, such as lower vaccination coverage or political riots (Pennycook and Rand, 2021; Burki, 2019). Practitioners and academics are thus increasingly voicing their concern and investigating solutions to stop or curb the spread of misinformation (Lazer et al., 2018; European Commission, 2018).

To evaluate adequate policies in the fight against misinformation, the way it spreads and how it competes with true information has to be properly understood. An important dynamic, we argue, is that people do not share true and false information in the same manner. Two significant asymmetries that have been measured in social media sharing behaviours support this idea. First, false information tends to be shared further in a network than true information. Importantly, this asymmetry is not driven by bots but by humans (Vosoughi et al., 2018), even though people of all political orientation agree that accuracy is the most important criteria to consider before sharing information (Pennycook et al., 2021). We refer to this kind of asymmetry as *decay asymmetry*; true and false information need not be shared to the same extent. Second, true and false information is shared more or less heavily in different parts of a given network. Some agents are heavily connected in networks that shares false information, while others are more connected in networks over which true information travels (Del Vicario et al., 2016; Zollo et al., 2017; Johnson et al., 2020). We refer to this kind of asymmetry as *network asymmetry*; true and false information need not be shared with the same people.

In this paper, we propose the first model of social learning that admits such asymmetries. We investigate their consequences on long-run beliefs and on speed of convergence to a belief. Our model builds on the literature on social learning, where agents repeatedly share binary signals (see, e.g. Golub and Sadler, 2016, for a survey). We introduce decay asymmetry and network asymmetry, meaning that agents can share true and false information with different decay and with different people. Our results show that both decay and network asymmetries can cause a society to go from being able to discover the true state to being misinformed in the long run. By distinguishing between networks in which true information and false information are shared, we establish a threshold condition that determines which state a society converges to in the long run. The condition compares the product of decay factor and largest eigenvalue between true and false information sharing networks. Noticeably, the long-run outcome in most cases does not depend on the initial distribution of signals (as long as there is at least one signal of each) and is

thus governed solely by sharing behaviors and network structures. Additionally, we show that the speed of convergence to the long-run belief also depends on the ratio between the products of respective decay factors and largest eigenvalues. The smaller the difference between said products, the larger the half-life to convergence. When the difference gets close to zero, the half-life explodes, implying that the speed of convergence is particularly low when there are only slight asymmetries. We further show that agents can be ordered according to their ratio of eigenvector centralities in the two different networks. Those who are relatively more central in the false information sharing network are more prone to be misinformed. We then extend the model to allow agents to have pair-specific relationships, which accommodates heterogeneous subgroups, directed networks and idiosyncratic learning, and show that the threshold condition, speed of convergence and ordering by eigenvector centralities still hold.

We illustrate the properties of our model using numeric simulations. The simulations highlight the effect of both the number of links in a given network and of the distribution of these links on the long-run state. Additional links in the false information network intuitively always favor reaching the misinformed state. However, for a given number of links, smaller denser groups, or echo chambers, have a (much) stronger influence than sparser larger groups. We thus show the required decay asymmetry favoring true information that is needed to compensate misinformation gets disproportionately large with the existence of misinformation echo chambers.

To counter misinformation, our model suggests two main avenues. The first works on decay asymmetry, the second on network asymmetry. Regarding decay, measures to make true information shared more and false information shared less have to be considered. These can include both changes to the information itself (such as making true information more shareable, easier to understand) or to the behaviors (such as educating people to recognize false information to avoid sharing it further). Regarding the second, measures to make the true information network denser and the false information network sparser have to be considered in order to increase the ratio of largest eigenvalues in favor of the true information sharing network. Particularly, breaking links of highly connected echo chambers over which false information travels will disproportionately favor reaching the true state in the long run.

The remainder of the paper is structured as follows. Section 2 relates it to the literature. We introduce the model in Section 3 and use the special case of symmetry as a benchmark in Section 4. Section 5 presents results and an extension. We illustrate our results using simulations in Section 6. A discussion and policy implications are provided in Section 7.

2   Related Literature

Our paper belongs to the literature on social learning (or opinion dynamics), where agents repeatedly learn from their neighbors in a social network. If agents were fully Bayesian, then perfect information aggregation would occur in any connected network (DeMarzo et al., 2003; Mueller-Frank, 2013). However, Bayesian rationality is very demanding when it comes to learning in a network structure. As experimental studies reveal, actual behaviors are less often consistent with Bayesian learning than they are with simpler updating rules such as repeated averaging (Corazzini et al., 2012; Grimm and Mengel, 2018; Friedkin and Bullo, 2017; Chandrasekhar et al., 2020). The classic model of repeated linear updating (DeGroot, 1974) has been studied intensively and it has been extended in many interesting directions (Friedkin and Johnsen, 1990; DeMarzo et al., 2003; Golub and Jackson, 2010, 2012; Buechel et al., 2015; Grabisch et al., 2019; Banerjee et al., 2019). A main focus is whether the long-run consensus opinion optimally aggregates the initially dispersed pieces of information. This is satisfied, at least asymptotically, if there are no individuals with excessive influence on the others (Golub and Jackson, 2010); or if there is at least one agent who is perfectly Bayesian (Mueller-Frank, 2014); or if agents additionally receive a private signal in *each* period and treat this signal in a Bayesian manner (Jadbabaie et al., 2012). More generally, there are (relatively weak) conditions on non-Bayesian learning which are sufficient for learning the true underlying state in the long run (Molavi et al., 2018).

However, none of these models is able to capture asymmetric treatment of signals. Symmetric treatment of signals is related to so-called "label neutrality," which is an underlying assumption of the models in the literature and in fact a characterizing feature of the DeGroot model (Molavi et al., 2018). We contribute, to our best knowledge, the first model of social learning that relaxes label neutrality. We do so by addressing network or decay asymmetry in signal sharing. As we show, the consequences are drastic as either decay or network asymmetry favoring negative signals can be sufficient for misinformation. That is, not only can the society's long-run belief be a suboptimal aggregation of the initial signals, but all members of the society would guess the wrong state and hence make the wrong decision with high probability. Similarly drastic conclusions are known in the literature only in the presence of forceful (or biased or stubborn) agents who, at some point, do not learn at all but still heavily influence the other agents (Acemoglu et al., 2010; Grabisch et al., 2018; Rusinowska and Taalaibekova, 2019; Azzimonti and Fernandes, 2018; Della Lena, 2019; Sikder et al., 2020). The policy implications that can be drawn from these model rather suggest acting on forceful agents and on fighting

disinformation, which is quite different from ours. Complementary to these insights, we argue that in the absence of forceful agents, the way signals are shared is crucial for avoiding misinformation.

Sharing behavior and signal accumulation in our model is closest to Sikder et al. (2020), in which agents repeatedly share binary signals in a rather naïve way but with Bayesian-style updating. These authors show that the introduction of confirmation bias leads to polarization. While Sikder et al. (2020) mostly focus on regular graphs, we first show for any connected network that misinformation in this baseline model is bounded. We generalize their baseline model to any network structure and by introducing decay and network asymmetries depending on the type of information shared. These asymmetries are closest to Taalaibekova (2020, Chapter 3.3). In one variation of her dissertation, she introduces the concept of "optimists" and "pessimists" who receive more positive signals, respectively negative signals, without being aware of it. Despite receiving signals in a biased manner, these agents still treat different signals in the same manner while our agents treat true and false information asymmetrically.

3   A Model of Asymmetric Signal Sharing

Ingredients.   There are n agents $N = \{1, ..., n\}$ who talk about a binary issue. Classic examples including whether the NASA has landed on the moon, whether there is human induced climate change, or whether vaccines cause autism; new examples are popping up every day. To model uncertainty, nature draws the true state with a commonly known prior probability and then each agent receives a signal. Specifically, the true state $\theta \in \{0, 1\}$ is drawn with prior probability $b^0 = P(\theta = 1) \in (0, 1)$. Each agent i independently receives signal $s_i \in \{0, 1\}$ which matches the true state with probability ρ, i.e. $P(s_i = 1|\theta = 1) = P(s_i = 0|\theta = 0) = \rho \in (\frac{1}{2}, 1)$. Conditional on the state, the signals are independent. As a convention, we let the true state be $\theta = 1$ and call 1 a *positive* signal and 0 as a *negative* signal.[1]

Time is discrete $t = 0, 1, 2, ..., \infty$. At time $t = 0$, the initial signals are received. At each time step $t > 0$, agents communicate with their neighbors in a social network. We model this communication activity in a way that admits asymmetric treatment of positive and negative signals. We introduce decay asymmetry and network asymmetry.

---

[1]This is only known to the modeler and, of course, not to the agents. Agents thus do not consciously recognize a signal to be true or false, but will treat them differently. This might be due to, for example, the difference in emotional content carried by true and false information as suggested by Vosoughi et al. (2018).

When a signal is passed on from one agent to the next, it decays by $\delta^+ \in (0, 1]$ if it is positive and by $\delta^- \in (0, 1]$ if it is negative. Decay can be either (i) due to a sender who only shares a fraction δ^+ of her signals, (ii) due to the communication channel on which a fraction $1 - \delta^+$ gets lost, or (iii) due to the recipient who discounts the received signals by δ^+.[2] If, for any reason, $\delta^+ \neq \delta^-$, there is *decay asymmetry*.

Positive signals are shared in a network (N, A^+), negative signals in a network (N, A^-). A^+ and A^- are both symmetric $n \times n$ matrices with entries 0 or 1, i.e. adjacency matrices representing each an undirected unweighted network. We assume that these networks are connected, which implies that the matrices are irreducible. Of course, these two networks can be highly related to each other.[3] If $A^+ \neq A^-$, there is *network asymmetry*.

Signal Accumulation. Given these ingredients, we can now formulate how signals are shared and accumulated. Let $N_i^+(t)$ denote number of positive signals of node i at time t and let $N^+(t)$ be the $(n \times 1)$ vector. The law of motion for positive signals is

$$N^+(t) = (I + \delta^+ A^+) N^+(t - 1). \tag{1}$$

Hence, the number of agent i's positive signals in a given period is the number of positive signals i held in the previous period plus the number of positive signals that i's neighbors in network A^+ held in the previous period discounted by δ^+. For the negative signals the law of motion and the notation is fully analogous. $N^-(t)$ is the vector of negative signals and the law of motion is $N^-(t) = (I + \delta^- A^-) N^-(t - 1)$. Let $N^+(0)$ and $N^-(0)$ be the vectors of initial signals, i.e. $N_i^+(0) = s_i$ and $N_i^-(0) = 1 - s_i$. Given the initial signals and the law of motion, we can compute the number of signals of any agent at any time. The signal accumulation for positive signals is

$$N^+(t) = (I + \delta^+ A^+)^t N^+(0). \tag{2}$$

Technically, the entries in the matrix $(I + \delta^+ A^+)^t$ can be considered as the number of walks of length t or smaller, when the network A^+ is augmented by self-loops (ones on the diagonal); thereby each walk is discounted by δ^+ at any step, except when using a self-loop. A less technical interpretation is that agents share all the positive signals that they

---

[2]In Appendix C.2, we show that all three interpretations can be explicitly modeled and are captured in a reduced form in our model with one parameter δ^+. In an extension of the model, we will study pair-specific decay factors, which admit heterogeneity in behavior of senders, communication channels, or recipients (Section 5.4).

[3]In several scenarios we investigate, there is a connected network A that is a subnetwork of both A^+ and A^-.

have with their neighbors in the positive signals sharing network; and whenever signals are passed on only a fraction δ^+ fully arrives at the recipient.

Beliefs and Signal Mixes. In this model, beliefs are formed similar to Bayes' rule with a crucial difference: the signals are taken at face value, i.e. treated as independent, ignoring that many accumulated signals are merely repetitions of the same initial signals.[4] Let $k_i(t) := N_i^+(t) - N_i^-(t)$ be the *signal difference* of agent i at time t. Then i's *belief* is

$$b_i(t) = P(\theta = 1|k_i(t)) = \frac{\rho^{k_i(t)} * b^0}{\rho^{k_i(t)} * b^0 + (1 - \rho)^{k_i(t)} * (1 - b^0)}, \tag{3}$$

where b^0 is the prior belief.[5]

The evolution of beliefs $b_i(t)$ is not as easy to study as that of *signal mixes*

$$x_i(t) := \frac{N_i^+(t)}{N_i^+(t) + N_i^-(t)}, \tag{4}$$

i.e. the fraction of positive signals that i holds.[6] Again, we denote vectors by $b(t)$, $k(t)$ and $x(t)$.

To measure misinformation, we assess whether an agent's belief tends to the correct or to the wrong state. We consider an agent i to be *misinformed* if her belief satisfies $b_i(t) < 0.5$, while she is not misinformed if her belief satisfies $b_i(t) > 0.5$. (Agents with $b_i(t) = 0.5$ are considered as something in between.) In most cases, signal mixes determine whether an agent is misinformed, as Lemma 1 assures.

Lemma 1. *Suppose that either $b^0 = 0.5$ or $t \gg 0$ holds (or both), i.e. either the prior belief is one half or sufficient time has elapsed (or both). If an agent's signal mix satisfies $x_i(t) < 0.5$, then she is misinformed, i.e. $b_i(t) < 0.5$. If an agent's signal mix satisfies $x_i(t) > 0.5$, then she is not misinformed, i.e. $b_i(t) > 0.5$.*

Lemma 1 means that, apart from very special cases, it is sufficient to know an agent's signal mix $x_i(t)$ to determine whether she is misinformed. One exception are signal mixes that converge to exactly one half, i.e. $\lim_{t \to \infty} x_i(t) = 0.5$, as they can lead to nasty behavior of beliefs.[7] Based on Lemma 1, we use an agent's signal mix $x_i(t)$ as a proxy for misinformation on the individual level.

---

[4]This kind of behavioral mistake is also known as "persuasion bias" and provides a justification for the DeGroot model (DeMarzo et al., 2003).

[5]This formula can be derived from Bayes' rule. It applies to symmetric binomial signals that are independently drawn.

[6]For instance, $k_i(t)$, which determines the belief, might diverge to $-\infty$, to ∞, converge to 0, converge to some other number, or even not converge at all, while $x_i(t)$ always converges.

[7]Various examples showcasing the behaviour of beliefs and best guesses can be found in Appendix C.3.

4 Benchmark: Misinformation under Symmetry

To assess the effects of decay and network asymmetry, a natural benchmark is, of course, the special case of *symmetry*: $\delta^+ = \delta^- =: \delta$ and $A^+ = A^- =: A$. This is a generalization of the baseline model introduced in Sikder et al. (2020), which our model nests for $\delta = 1$. In the original article by Sikder et al. (2020), the focus is on regular networks (N, A), that are characterized by each node having the same degree.[8] Regular networks happen to minimize the probability of misinformation, as our Proposition 1 below implies. More importantly, this result shows that the probability of misinformation is always bounded under symmetry.

Throughout the analysis, the adjacency spectrum will play an important role. Let λ_1 be the largest eigenvalue of A.[9] Let $c = (c_1, ..., c_n)^\top$ be its corresponding eigenvector, normalized such that its components sum to one, i.e. $A^+ c^+ = \lambda_1^+ c^+$ and $\sum_i c_i^+ = 1$.[10] Entries in the eigenvector c are a measure of *eigenvector centrality* (Bonacich, 1972; Friedkin, 1991).

Proposition 1 (Symmetry). *Under symmetry, the long-run signal mix is a convex combination of the initial signals s_j with weights according to eigenvector centrality, i.e. for all i, $\lim_{t \to \infty} x_i(t) = \sum_{j=1}^n c_j s_j$. Hence, misinformation prevails if $\sum_{j=1}^n c_j s_j < 0.5$ and the probability of misinformation is bounded from above by 0.5.*

The proofs of this proposition and all others are collected in Appendix A. Proposition 1 means that misinformation under symmetry only occurs due to an unlucky distribution of signals. It depends on the relative influence of each agent on the long-run opinions, as measured by her entry in the largest eigenvector. In the best case, all agents, who are by assumption equally well-informed, are equally influential.[11] This is satisfied in regular graphs, in which by definition every agent has the same degree. Then the long-run signal mix of every agent exactly reflects the initial signal distribution, i.e. $\lim_{t \to \infty} x_i(t) = \frac{1}{n} \sum_{j=1}^n s_j$, which is just the mean of the initial signals. Hence, under symmetry and when the network is regular, naïve agents who accumulate the same signals over and over again can fare

---

[8]Sikder et al. (2020) find polarization in an extension of their model, in which they introduce agents with confirmation bias.

[9]More precisely, λ_1 is the largest positive eigenvalue of A and other eigenvalues of A might exist which are as large in absolute value as λ_1. For ease of notation, we usually omit 'positive' when addressing largest positive eigenvalues.

[10]Recall that by assumption the adjacency matrix is symmetric. Hence, there is no need to distinguish between left-hand and right-hand eigenvectors.

[11]More generally, the requirement is a balance between idiosyncratic signal precision and social influence (see, e.g. Buechel et al., 2015).

equally well as Bayesians would (as we discuss in Example C.1 in Appendix C.1). In the worst case, there is a group of agents who are overly influential (as illustrated in Example C.2 in Appendix C.1).

Since misinformation in the worst case can be substantial, the question arises whether this is also true for more realistic networks. We therefore simulate probabilities of misinformation for two classes of random graphs: Erdös-Rényi (ER), in which the probability of every link is fixed (p), and Barabasi-Albert (BA), where in every step of constructing the network, m new links are created. The expected degree distribution of BA networks is scale-free, a feature that many real social networks share (Albert and Barabási, 2002; Strogatz, 2001). Since the degree distribution in ER random graphs is more uniform than in BA random graphs, we expect the probability of misinformation to be higher in the latter, when selecting parameters (p and m) such that the two random networks have the same expected density.

Figure 1 illustrates the simulation results. As expected, misinformation tends to be lower in ER as compared to BA random graphs. The lowest line is the probability of misinformation in a regular graph (from Example C.1) while the highest line is the probability of misinformation in the worst network (from Example C.2), where a small group of agents is extremely influential. Misinformation for the two classes of random graphs is closer to the regular graph's. In particular, misinformation decreases with network size n. The thickness of the two lines represents the variation of misinformation that covers 50% of all simulation runs (from 25th to 75th percentile). For large enough networks, the probability of misinformation is small.
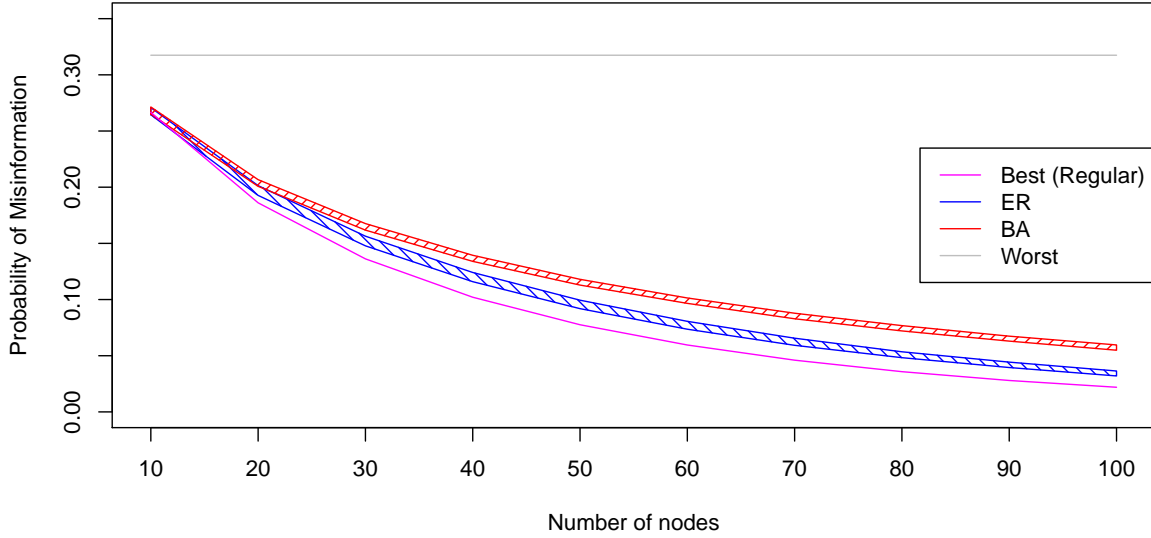
Figure 1: Misinformation under symmetry: comparing different network structures.

Notes: Signal precision $\rho = 0.6$. Number of nodes $n = 10, ..., 100$ on the x-axis. Random graph parameters are set such that asymptotic average degree is 6 in these simulation runs. 1000 simulation runs per class of random network of a given size. Thickness of corresponding line represents variation covering 50% of all outcomes (from 25th to 75th percentile).

The results in this section closely resemble those of the common DeGroot model of naïve learning. In Appendix C.5, we describe the similarities and differences of these two models and their results in detail. Importantly, our model of naïve learning keeps track of all signals, while in the DeGroot model positive and negative signals are mingled into opinions. This has the consequence that asymmetries in signal sharing cannot be introduced into the DeGroot model. Moreover, some measures against misinformation can only be assessed in a model that keeps track of the initial signals.

5 Results

5.1 Key Result

We now study how misinformation in the long run is affected by asymmetric treatment of signals. Decay asymmetry is captured simply by its two parameters δ^+ and δ^-. For

network asymmetry, the adjacency spectrum of the two matrices A^+ and A^- matters. Let λ_1^+ be the largest eigenvalue of A^+. Let $c^+ = (c_1^+, ..., c_n^+)^\top$ be its corresponding eigenvector, normalized such that its components sum to one, i.e. $A^+ c^+ = \lambda_1^+ c^+$ and $\sum_i c_i^+ = 1$. And likewise λ_1^- and c^- are the largest eigenvalue and its normalized eigenvector of A^-, i.e. $A^- c^- = \lambda_1^- c^-$ and $\sum_i c_i^- = 1$. Entries in the eigenvectors c^+, c^- are the respective eigenvector centrality.

Proposition 2 (Key Result). *Suppose that the initial distribution of signals contains at least one positive and at least one negative signal.*

1. *If $\delta^+ \lambda_1^+ < \delta^- \lambda_1^-$, then for all i and large t:*

$$x_i(t) \approx \frac{c_i^+}{c_i^-} \left(\frac{1 + \delta^+ \lambda_1^+}{1 + \delta^- \lambda_1^-} \right)^t \frac{\sum_{k=1}^n (c_k^-)^2}{\sum_{k=1}^n (c_k^+)^2} \frac{\sum_{j=1}^n c_j^+ s_j}{1 - \sum_{j=1}^n c_j^- s_j}$$

   such that $\lim_{t \to \infty} x_i(t) = 0$. Hence, misinformation prevails.

2. *If $\delta^+ \lambda_1^+ > \delta^- \lambda_1^-$, then for all i and large t:*

$$x_i(t) \approx 1 - \frac{c_i^-}{c_i^+} \left(\frac{1 + \delta^- \lambda_1^-}{1 + \delta^+ \lambda_1^+} \right)^t \frac{\sum_{k=1}^n (c_k^+)^2}{\sum_{k=1}^n (c_k^-)^2} \frac{1 - \sum_{j=1}^n c_j^- s_j}{\sum_{j=1}^n c_j^+ s_j}$$

   such that $\lim_{t \to \infty} x_i(t) = 1$. Hence, misinformation vanishes.

3. *If $\delta^+ \lambda_1^+ = \delta^- \lambda_1^-$, then for all i:*

$$\lim_{t \to \infty} x_i(t) = \frac{1}{1 + \dfrac{c_i^-}{c_i^+} \dfrac{\sum_{k=1}^n (c_k^+)^2}{\sum_{k=1}^n (c_k^-)^2} \dfrac{1 - \sum_{j=1}^n c_j^- s_j}{\sum_{j=1}^n c_j^+ s_j}} \in (0, 1).$$

   Hence, long-run misinformation depends on eigenvector centralities and the signal distribution.

Let us first discuss the "big picture". The proposition states that the combination of decay factor and largest eigenvalue is crucial for the level of misinformation in the long run: the condition $\delta^+ \lambda_1^+ \lessgtr \delta^- \lambda_1^-$ determines whether all signal mixes and therefore all beliefs will converge to 0 (Case 1) or whether they will converge to 1 (Case 2). Intuitively, the case with full misinformation, Case 1, is rather reached when the decay factor for positive signals δ^+ is low compared to the decay factor with negative signals δ^-, which means that positive signals are shared to a lower extent; and when the eigenvalue λ_1^+ of the positive signal sharing network A^+ is low compared to λ_1^-, which has the interpretation

that the agents are generally better connected in the negative signals sharing network A^-, as discussed below. The case distinction relies on the product of these two factors: $\delta^+ \lambda_1^+ \lessgtr \delta^- \lambda_1^-$; or $\frac{\delta^+}{\delta^-} \lessgtr \frac{\lambda_1^-}{\lambda_1^+}$. Hence, decay factor and largest eigenvalue can only compensate each other to some extent. Case 3, $\delta^+ \lambda_1^+ = \delta^- \lambda_1^-$, is the knife-edge case, in which these two products coincide. In that case it is possible that some agents are misinformed in the limit while others are not, as we will demonstrate below (in Example 1).

In the limit, the initial distribution of signals does virtually not matter in Cases 1 and 2. Given that there is initially at least one positive and one negative signal, either positive or negative signals fully dominate in the long run. Hence, all beliefs $b_i(t)$ in Case 1 converge to 0 such that all agents in the society are misinformed, independent of the signal distribution and independent of the prior belief b^0. Likewise, all beliefs $b_i(t)$ in Case 2 converge to 1 such that no agent in the society is misinformed. For Case 3, the limit is less trivial. We can first observe that the long-run signal mix $x_i(t)$ is increasing in all s_j, i.e. with increasing number of initial signals being correct, agents' signal mixes and therefore also their beliefs get closer to the truth.

Comparing this key result to the symmetric benchmark draws three main conclusions. First, the symmetric benchmark is nested in Case 3. Second, while misinformation is bounded in the symmetric benchmark, introducing asymmetry induces full misinformation, in which every agent is misinformed, if the conditions of Case 1 are satisfied. Third, while in the symmetric benchmark the common decay factor δ did not affect the long-run opinions, decay factors δ^+ and δ^- are crucial for the case distinction under asymmetry.

Let us now discuss the terms that determine the asymptotic behavior in Cases 1 and 2 of Proposition 2. In Case 1, there are four factors:

$$
x_i(t) \approx \underbrace{\frac{c_i^+}{c_i^-}}_{\text{centrality ratio}} \cdot \underbrace{\left(\frac{1 + \delta^+ \lambda_1^+}{1 + \delta^- \lambda_1^-} \right)^t}_{\text{exponential decay}} \cdot \underbrace{\frac{\sum_{k=1}^n (c_k^-)^2}{\sum_{k=1}^n (c_k^+)^2}}_{\text{normalizing constant}} \cdot \underbrace{\frac{\sum_{j=1}^n c_j^+ s_j}{1 - \sum_{j=1}^n c_j^- s_j}}_{\text{signal averages}} \quad (5)
$$

The first is agent-specific, the three latter factors are common for all agents in a given society. The first factor shows that agent i's characteristics enter her asymptotic signal mix $x_i(t)$ via her ratio of eigenvector centrality in the positive signals sharing network over her centrality in the negative signals sharing network: $\frac{c_i^+}{c_i^-}$. We will refer to this as i's *centrality ratio*. The second factor shows the exponential decay process, which depends on the (information) decay factors δ^+ and δ^-, as well as on the the largest eigenvalues of the two networks. The next factor is the ratio of the sums of squared centralities. This can be considered as a normalizing constant. The last factor is determined by weighted averages of the initial signals, whereas the weights are the network centralities. We can

observe that this factor is increasing in s_j, reflecting the fact that an agent's signal mix $x_i(t)$ becomes closer to 1 when any agent's signal s_j flips from false information ($s_j = 0$) to true information ($s_j = 1$). The centrality ratio will determine opinion diversity, the exponential decay will determine speed of convergence, the weighted signal averages will determine the levels of the signal mixes. The factor decomposition is analogous for Case 2. In Case 3, opinion diversity is also determined by the centrality ratio, whereas speed of convergence is determined differently from Cases 1 and 2, as further discussed below.

5.2   Implications

Through the main result, we can derive implications regarding the state that is reached in the long run, the potential diversity of opinions or consensus and how long it takes to reach the long-run state.

Long-run Misinformation. Suppose there is decay asymmetry, while the networks are symmetric, i.e. $\delta^+ \neq \delta^-$ and $A^+ = A^-$. Then the crucial condition for long-run misinformation, $\delta^+ \lambda_1^+ \lesseqgtr \delta^- \lambda_1^-$ (Proposition 2), becomes $\delta^+ \lesseqgtr \delta^-$. Now, if it is true that negative signals exhibit less decay than positive signals, i.e. $\delta^+ < \delta^-$, then we are in Case 1 of Proposition 2 and misinformation prevails. Hence, already slight asymmetry in decay fully changes the long-run outcome from bounded misinformation, which we have observed in the case of symmetry, to full misinformation. However, it may take time until this long-run consensus on the false opinion is reached.

Suppose there is decay symmetry, but network asymmetry, i.e. $\delta^+ = \delta^-$ and $A^+ \neq A^-$. Then the crucial condition of Proposition 2, $\delta^+ \lambda_1^+ \lesseqgtr \delta^- \lambda_1^-$, becomes $\lambda_1^+ \lesseqgtr \lambda_1^-$. As argued before, the largest eigenvalues capture how well-connected the agents are in some sense. This interpretation becomes more specific in special cases. Suppose the networks A^+ and A^- are the same, apart from some links in A^- that are not present in A^+, e.g. because some agents use these additional channels to share only negative signals. We write $A^+ \subset A^-$ to denote that the positive signals sharing network is a subgraph of the negative signals sharing network (i.e. for each entry, $a_{ij}^+ \leq a_{ij}^-$, and for some entry $a_{ij}^+ < a_{ij}^-$). Then it holds that $\lambda_1^+ < \lambda_1^-$. As a consequence, we are in Case 1 of Proposition 2 and misinformation prevails. Using some links to only share negative signals has hence drastic consequences for long-run misinformation. Likewise, if $A^+ \supset A^-$, then $\lambda_1^+ > \lambda_1^-$ and we are in Case 2 such that misinformation vanishes. In general, however, the positive signals sharing network need not be a subnetwork of the negative signals sharing network, nor the other way around.

Finally, let us admit both decay and network asymmetry, i.e. $\delta^+ \neq \delta^-$ and $A^+ \neq A^-$. Clearly, when both kinds of asymmetry point to the same direction, for instance, when negative signals decay less ($\delta^+ < \delta^-$) and the negative signals sharing network is better connected ($\lambda_1^+ < \lambda_1^-$), then the results follow from Proposition 2: Since $\delta^+ \lambda_1^+ < \delta^- \lambda_1^-$, misinformation prevails. Moreover, this setting boosts the fraction $\log\left(\frac{1+\delta^- \lambda_1^-}{1+\delta^+ \lambda_1^+}\right)$ (i.e. the exponential decay factor of Case 2), which measures the speed of convergence. And analogously for the opposite setting. The more interesting question is: to which extent decay and network asymmetry can compensate each other when they point to opposite directions? We illustrate such cases in Section 6.

Opinion Diversity. Proposition 2 already indicates that individual differences between agent's beliefs are driven by their centrality ratio $\frac{c_i^+}{c_i^-}$. The following corollary establishes this relation, considering two agents' ratios of positive over negative signals, $\frac{N_i^+(t)}{N_i^-(t)} \big/ \frac{N_j^+(t)}{N_j^-(t)}$.[12] Clearly, if an agent's ratio of positive over negative signals is above another agent's ratio, then her signal mix and belief are closer to 1 and hence closer to the truth.

Corollary 1 (Centrality Ratios and Opinion Diversity). *Suppose that the initial distribution of signals contains at least one positive and at least one negative signal. Then the ratio of two agents' ratios of positive over negative signals converges to these agents' ratio of centrality ratios, i.e.*

$$\lim_{t \to \infty} \frac{N_i^+(t)}{N_i^-(t)} \bigg/ \frac{N_j^+(t)}{N_j^-(t)} = \frac{c_i^+}{c_i^-} \bigg/ \frac{c_j^+}{c_j^-}. \tag{6}$$

Hence, an agent i with higher centrality ratio than another agent j has a higher asymptotic signal mix and belief, i.e. if $\frac{c_i^+}{c_i^-} > \frac{c_j^+}{c_j^-}$, then for large t, $x_i(t) > x_j(t)$ and $b_i(t) > b_j(t)$.

Corollary 1 applies to all three cases of Proposition 2. It says that agents with relatively high centrality in the negative signals network are more prone to be misinformed. Moreover, if some agent is misinformed in the long run, then all agents with lower centrality ratios must be as well. Similarly, if an agent is well-informed in the long run, then all agents with higher centrality ratios must be as well.

Minimal opinion diversity is given in all networks where $\frac{c_i^+}{c_i^-}$ is constant across agents, which holds in particular, if $A^+ = A^-$ (which we discuss in the next paragraph). Strong opinion diversity is given in networks where these ratios differ strongly across agents. For example, consider two star networks (N, A^+) and (N, A^-) with different centers. The

---

[12]Notice that agent i's ratio of positive over negative signals, $\frac{N_i^+(t)}{N_i^-(t)}$, can equivalently be written as $\frac{x_i(t)}{1-x_i(t)}$.

ratio of one center is $\frac{c_i^+}{c_i^-} = \frac{\sqrt{n-1}}{1}$, the ratio of the other center $\frac{c_j^+}{c_j^-} = \frac{1}{\sqrt{n-1}}$. Hence, by Corollary 1, $\frac{N_i^+(t)}{N_i^-(t)} \Big/ \frac{N_j^+(t)}{N_j^-(t)}$ converges to $n-1$. That is, agent i has asymptotically $n-1$ times more positive signals over negative signals than agent j. Again, this holds in all of the three cases, even if the absolute differences vanish for large t.

In terms of opinion diversity, the Corollary 1 means that, even if beliefs converge to consensus in Cases 1 and 2, they are ordered by their centrality ratios. In Case 3, this order also holds and opinion diversity even persists in the limit, as Example 1 below shows. We have seen in a special case of Case 3, the symmetry benchmark (Section 4), that consensus may also emerge in Case 3. This was due to the absence of network asymmetry. Indeed, $A^+ = A^-$, implies that an agents i's centrality c_i^+ in the positive signals sharing network equals her centrality c_i^- in the negative signals sharing network. Hence, for every agent i, the ratio of network centralities is $\frac{c_i^+}{c_i^-} = 1$. Therefore, Corollary 1 implies that all agents approach the same signal mix, opinions converge to consensus in the absence of network asymmetry.

Speed of Convergence. The motivation for studying misinformation is that agents make decisions, which may be based on inaccurate information. If the point of decision making is not sufficiently far in the future, then the short or medium term opinion dynamics matter. Speed of convergence can also be measured with the help of Proposition 2. For instance, in Case 1, the speed of convergence is governed by the speed that the exponential decay factor $\left(\frac{1+\delta^+\lambda_1^+}{1+\delta^-\lambda_1^-}\right)^t$ converges to 0 (see also Eq. (5)). Looking for the half-life, as it is standard for exponential decay processes, we define $t_{1/2}$ as the number of periods it takes for this quantity to fall to one half of its initial value. We get $t_{1/2} = \log(0.5) \Big/ \log\left(\frac{1+\delta^+\lambda_1^+}{1+\delta^-\lambda_1^-}\right)$ in Case 1. Analogously, in Case 2 half-time $t_{1/2}$ equals $\log(0.5) \Big/ \log\left(\frac{1+\delta^-\lambda_1^-}{1+\delta^+\lambda_1^+}\right)$. In Case 3, things are more complicated as not only the largest eigenvalues are relevant asymptotically.[13] The case dependent half-life can then be generalized as:

$$t_{1/2} = \frac{\log(0.5)}{\log(\tau)}, \quad \text{with} \tag{7}$$

$$\tau := \begin{cases} \frac{1+\min\{\delta^+\lambda_1^+,\delta^-\lambda_1^-\}}{1+\max\{\delta^+\lambda_1^+,\delta^-\lambda_1^-\}}), & \text{if } \delta^+\lambda_1^+ \neq \delta^-\lambda_1^- \text{ (Cases 1, 2)} \\ \max\left\{\max\left\{\frac{|1+\delta^+\lambda_i^+|}{1+\delta^+\lambda_1^+}, i=2,\ldots,n\right\}, \max\left\{\frac{|1+\delta^-\lambda_i^-|}{1+\delta^-\lambda_1^-}, i=2,\ldots,n\right\}\right\}, & \text{if } \delta^+\lambda_1^+ = \delta^-\lambda_1^- \end{cases}$$

---

[13]The exponential decay factor can be derived from the representations for $N^+(t)$ and $N^-(t)$ given in Equations (A.2) and (A.3).

In the first two cases, half-life will be large when $\delta^+ \lambda_1^+$ and $\delta^- \lambda_1^-$ are close to each other, i.e. when we are close to Case 3. In some sense, Case 3 can be considered as unlikely because it is a special case of the parameter space. Still, it is important to study this case for at least two reasons. First, in the absence of asymmetry – the special case that has been studied in the literature – we are in Case 3, as it was discussed in Section 4. Second, under certain conditions, opinions in the short and medium term are well approximated by the analysis of Case 3, even if this does not hold in the long run.[14]

5.3   Illustration of Key Result and Implications

Example 1 illustrates the case distinction, the signal mix dynamics, the speed of convergence implied by Proposition 2, as well as the opinion diversity implied by Corollary 1.

Example 1. *Consider five agents $N = \{1, 2, 3, 4, 5\}$. Let the prior be $b^0 = 0.5$. The positive signals sharing network (N, A^+) is the complete network, i.e. $a_{ij}^+ = 1$ for all $i \neq j$. The negative signals sharing network (N, A^-) is a star network with agent 1 at the center, i.e. $a_{1j}^- = a_{j1}^- = 1$ for all $j = 2, 3, 4, 5$ and $a_{ij}^- = 0$ else. For illustration purpose, one can think of this example as five colleagues who discuss scientific publications with each other, and discuss conspiracy theories with only one person.*

We fix the decay factor of negative signals to be $\delta^- = 0.8$, whereas we vary the decay factor of the positive signals. Proposition 2 distinguishes three cases based on the condition $\delta^+ \lambda_1^+ \lessgtr \delta^- \lambda_1^-$. In the complete network, we have $\lambda_1^+ = n - 1 = 4$. In the star network, we have $\lambda_1^- = \sqrt{n-1} = 2$. Hence, we are in Case 3 of Proposition 2 if $\delta^+ = 0.4$, in Case 1 if $\delta^+ < 0.4$, and in Case 2 if $\delta^+ > 0.4$ (indeed, $\delta^+ \lambda_1^+ = 0.4 \cdot 4 = 0.8 \cdot 2 = \delta^- \lambda_1^-$ yields Case 3). The corresponding dynamics of signal mixes is illustrated in Figure 2 for the signal distribution $s = (0, 0, 1, 1, 1)$ and for values of parameter δ^+ close to the equality threshold of 0.4. The first four panels belong to Case 1 of Proposition 2, where signal mixes converge to 0 and misinformation prevails. The last four panels belong to Case 2, where signal mixes converge to 1 and misinformation vanishes. The middle panel induces Case 3 by setting $\delta^+ = 0.4$. Interestingly, in that case Agent 1's belief converges to 0 as $x_1(\infty) < 0.5$, whereas the other agent's beliefs converge to 1 as $x_i(\infty) > 0.5$ for $i = 2, 3, 4, 5$.[15] Hence, only the center of this star network, is misinformed in the limit (which holds for any initial distribution of signals that has at least one positive and one negative signal).

---

[14]Conditions are discussed in Appendix C.4.

[15]Indeed, if the signal mix is above 0.5, the difference between positive and negative signals $k_i(t)$ grows
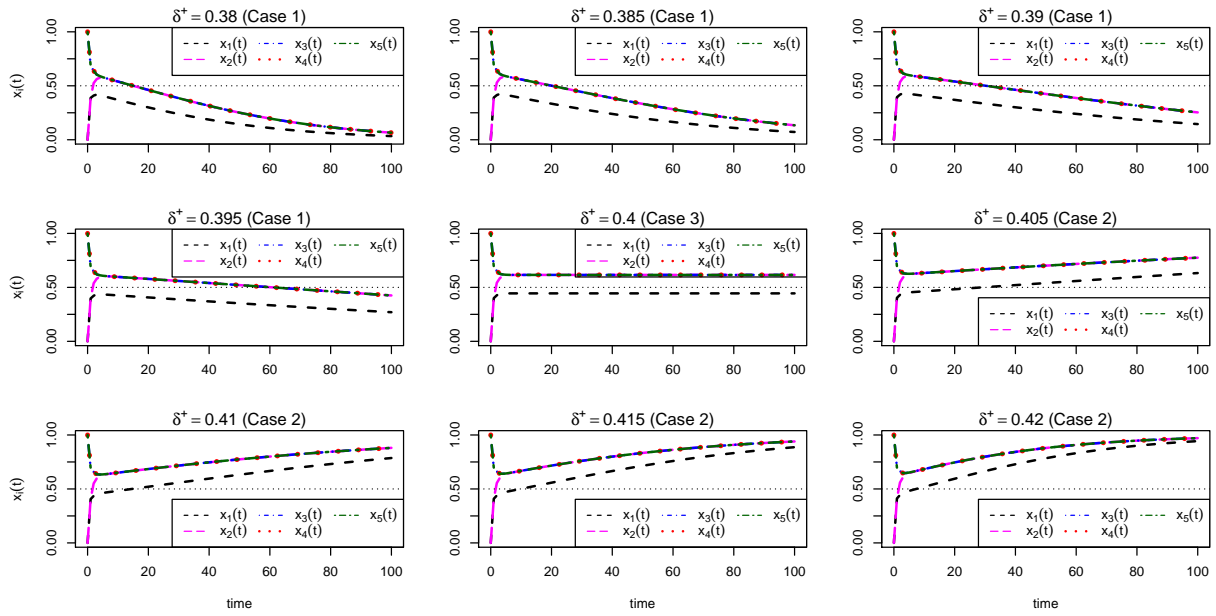
Figure 2: Illustration of key result.

Notes: Dynamics of signal mixes in Example 1 over time. Recall that signal mixes below 0.5 indicate misinformation. The first four panels with $\delta^+ < 0.4$ belong to Case 1 of Proposition 2. The middle panel with $\delta^+ = 0.4$ yields Case 3. The last four panels with $\delta^+ > 0.4$ belong to Case 2.
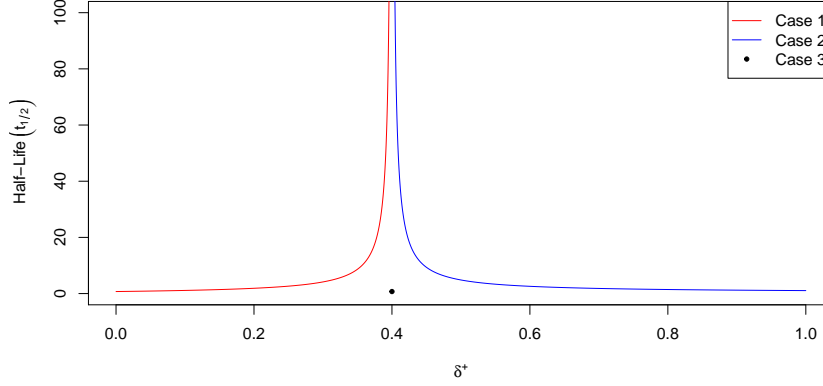
Figure 3: Speed of Convergence in Example 1.

Notes: Half-life in Example 1 for different values of δ^+. Setting $\delta^+ = 0.4$ yields Case 3 such that the black dot represents the relevant half-life. $\delta^+ < 0.4$ induces misinformation in the long-run, with half-life in red. Conversely, with $\delta^+ > 0.4$ misinformation vanishes in the long-run, with half-life in blue. Convergence is mostly fast as half-life is mostly low. Nearby the critical value of $\delta^+ = 0.4$ half-life "explodes", meaning speed of convergence is extremely low.

To illustrate Corollary 1 we compute eigenvector centralities: $c^+_{1,2,3,4,5} = \frac{1}{5}$, $c^-_1 = \frac{1}{3}$ and $c^-_{2,3,4,5} = \frac{1}{6}$. The ratios $\frac{c^+_i}{c^-_i}$ of relative importance in the positive network relative to the negative network for the four agents are hence $(\frac{3}{5}, \frac{6}{5}, \frac{6}{5}, \frac{6}{5}, \frac{6}{5})$. Thus, the first agent has a smaller centrality ratio than all other agents. The reason is that this agent, as the center of the star network, is better connected in the negative signals sharing network than the others, while all agents are equally well-connected in the positive signals sharing network, the complete network. By Corollary 1, this implies that this agent's signal mix will asymptotically always be closer to 0 than those of the other agents: $x_1(t) < x_i(t)$, for $i = 2, 3, 4, 5$. In all panels of Figure 2 this can be observed as Agent 1's signal mix is below the others' signal mix, even if they converge to 1 or 0. Finally, observe that convergence is slower the closer the parameter δ^+ to the critical value that induces Case 3. Figure 3 shows half-life for all the values of parameter δ^+. The red line is half-life in Case 1, the blue line is half-life in Case 2, the black dot is half-life in Case 3, i.e. when $\delta^+ = 0.4$. Half-life "explodes" near this value, but is reasonably low for lower and higher values.

such that the belief converges to 1, as more and more signals are acquired.

5.4 Extension: Heterogeneous Relations

In this section, we extend the model by defining decay factors that are pair-specific. Formally, we consider two weighted graphs (N, M^+) and (N, M^-), where M^+ and M^- are $n \times n$ matrices with entries $m_{ij}^+ = \delta_{ij}^+ \in (0, 1]$ or $m_{ij}^+ = 0$ and $m_{ij}^- = \delta_{ij}^- \in (0, 1]$ or $m_{ij}^- = 0$. Like in the baseline model, a positive entry $\delta_{ij}^+ > 0$ is the decay of information for the pair ij, i.e. the fraction of positive signals that agent i receives when communicating with agent j. Signal accumulation becomes $N^+(t) = (I + M^+)N^+(t-1)$ for positive signals and likewise $N^-(t) = (I + M^-)N^-(t-1)$ for negative signals. In all other aspects, the extended model is defined as the baseline model, introduced in Section 3. In particular, we assume again that both networks are strongly connected.

This extension makes the model more flexible since it can accommodate many forms of heterogeneity. First, it allows for every pair i, j to have a different quality of communication and hence a pair-specific decay factor. Second, networks can be directed such that some agent i receives signals from another agent j, while j does not receive signals of i, i.e. $m_{ij}^+ = \delta_{ij} > 0$ and $m_{ji}^+ = 0$. Third, it allows an agent i to face a stronger decay as a receiver of information, i.e. $\delta_{ij} < \delta_{kj}$ for all j, k linked to i. This can incorporate in particular differences in discounting of received signals. Fourth, it allows for some agent i to face a stronger decay as a sender of information, i.e. $\delta_{ji} < \delta_{jk}$ for and all j, k linked to i. This could incorporate in particular differences in which a fraction of signals is shared. These variations can be applied at the individual level to single agents, or more broadly to groups of agents, as we will discuss in the next section.

Even though this extension carries a rich array of new applications, the key result, Proposition 2, as well as its Corollary 1, neatly generalize. A noteworthy difference is that left and right eigenvectors do not coincide. The right eigenvectors determine the relations between the asymptotic signal mixes of agents, while the left eigenvectors determine the effect of agents' initial signals on the asymptotic signal mix, as Proposition B.1 and Corollary 2 in the appendix demonstrate. The generalization shows that it is not essential that the network is undirected or that discounting is similar for different agents, while symmetry or asymmetry between positive and negative networks is crucial.

6 The Effect of Groups

To explore and illustrate how misinformation depends on the different asymmetries we introduced, we apply our model to group structures. The existence of groups, which are more connected amongst themselves and have similar sharing behaviors is an established

feature of social networks and it has been identified as a source of the spread of mis-information (Zollo et al., 2017; Quattrociocchi et al., 2016). First, we illustrate network asymmetry and how a more connected group can determine the long run state. Second, we show that not only the number of links but also how they are distributed matters, namely that echo chambers generate a disproportional advantage. Third, we put in perspective what network asymmetry would be needed to compensate for a given decay asymmetry.

To this end, we use the following set up. Let a society be partitioned into three groups $N = \{N_0, N_I, N_{II}\}$ of sizes n_0, n_I, n_{II}. Members of N_I are more connected with each other in the positive signal sharing network, while N_{II} members are more connected with each other in the negative signal sharing network. Technically, we first generate a connected network (N, A) with α links. Then, we generate α_I^+ additional links in A^+ between members of group N_I. Finally, we construct A^+ as the union of the links of A and the additionally generated links. Likewise, we generate α_{II}^- additional links in A^- between members of group N_{II}. For illustration, one can think of N_I members as belonging to a social media group dedicated to scientific information (e.g. regarding vaccines), members of N_{II} belonging to a group dedicated to (e.g. vaccine) conspiracies and members of N_0 as the majority that belongs to no such group. The parameter space for such a society can be summarized by $(n_0, n_I, n_{II}, \alpha, \alpha_I^+, \alpha_{II}^-)$. Figure 4 illustrates such a society (with only 30 nodes for ease of readability) with $(n_0, n_I, n_{II}, \alpha, \alpha_I^+, \alpha_{II}^-) = (20, 5, 5, 40, 5, 5)$.
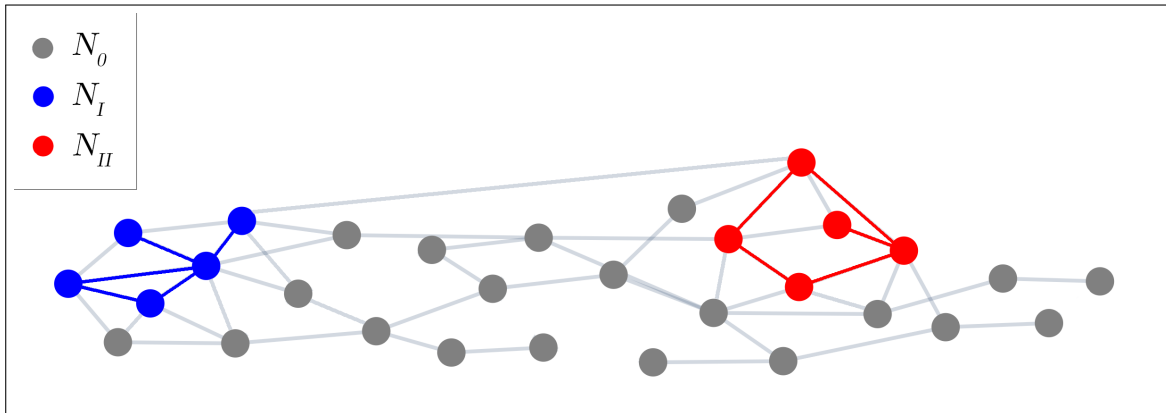

Figure 4: Illustration of a society with groups.

Notes: Members of N_0 and α common links are in grey. Members of N_I and their α_I^+ additional links are blue. Members of N_{II} and their α_{II}^- additional links are red. Positive signals travel over the A^+ network, which consists of the grey and blue links. Negative signals travel over the A^- network, which consists of the grey and red links.

6.1 Network Asymmetry: More Connected Groups

To illustrate the effect of additional links, we use a society of 100 people in a connected Erdős–Rényi random graph. These people are connected with a total of $\alpha = 250$ randomly generated links in the A network, meaning the average degree is 5. We then include the group structure and generate additional random links in the A^+ and A^- with the following parameters: $(n_0, n_I, n_{II}, \alpha, \alpha_I^+, \alpha_{II}^-) = (60, 20, 20, 250, 10, \alpha_{II}^-)$ with $\alpha_{II}^- \in [0, 20]$, i.e. Groups I and II consist of 20 agents each. Group I has 10 additional positive links. Group II has between 0 and 20 additional negative links.

The long-run signal mix as well as the speed of convergence to that signal mix depending on α_{II}^- is illustrated in Figure 5. As one would expect, the probability of this society being misinformed in the long run increases with α_{II}^-, i.e. the more negative links added to Group II. We also see that the larger the difference of number of links between the two networks, the faster the speed of convergence. We observe however that it is possible to have misinformation in the long run with values of $\alpha_{II}^- \leq \alpha_I^+ = 10$. Societies that are more connected in the positive signal sharing network tend to reach the true state but can, with a lower probability, still converge to a misinformed state in the long run. Given the density of the A, A^+ and A^- network is always exactly the same for a fixed set of parameters, it must then be that this fluctuation is a consequence of the randomness of the link generation process. We can also observe this with the violins getting thinner the more links are added. As more links are created, the randomness in the process gets increasingly influential in determining the outcome of the convergence, leading to higher variance around the median. We conclude that not only the number of links in a society, but also how they are distributed affects the long-run belief, as we further investigate in the next section.
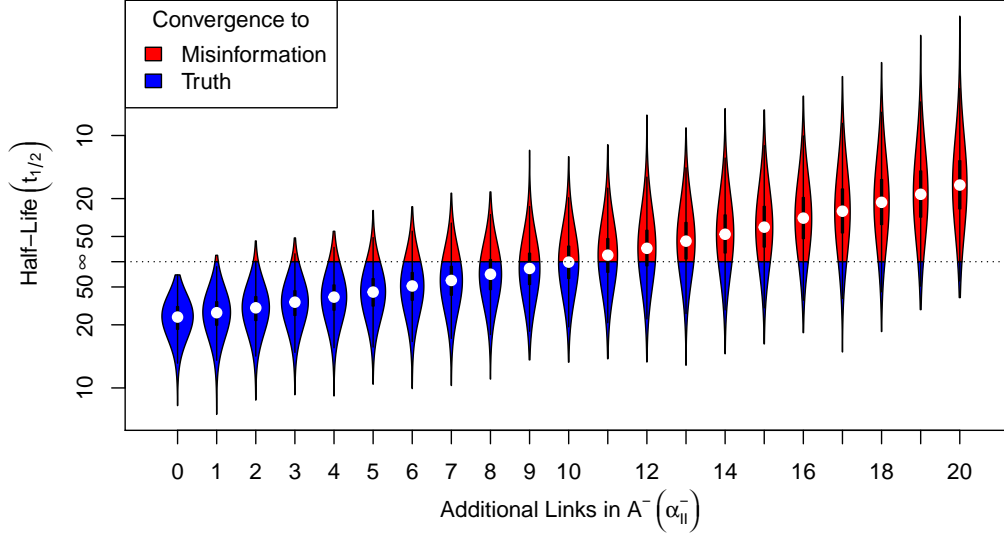
Figure 5: Network asymmetry: the effect of additional links.

Notes: Results based on 10'000 replications per setting. Effect of varying α_{II}^- on misinformation and speed of convergence measured by half-life. α_I^+ is constant and equal to 10. The networks are symmetric in expectations for $\alpha_{II}^- = 10$. For each violin, the red area represents the probability of misinformation for a given set of parameters. White dots represent the median, central black bars are the interquartile range and the thickness of the violin are density estimates.

6.2 Network Asymmetry: Groups as Echo Chambers

To illustrate that a link can have more or less influence on the long-run belief of a society, we start from the same society as in the previous example. 100 people are in a connected Erdős–Rényi random graph with an average degree of 5 in the A network. We then add the group structure with the following parameters: $(n_0, n_I, n_{II}, \alpha, \alpha_I^+, \alpha_{II}^-) = (80 - n_{II}, 20, n_{II}, 250, 10, 10)$ with $n_{II} \in [5, 35]$, i.e. Group I is of size 20, Group II's size ranges from 5 to 35, both groups have 10 additional links, which are added to their respective positive or negative networks. α_I^+ and n_I are fixed. Consequently all properties of the A^+ networks are in expectations the same across the parameter space and across replications. However, despite a fixed α_{II}^-, the local density of A^- varies significantly depending on the size of n_{II}. When $n_{II} = 5$ a small clique, i.e. a fully connected subgroup is generated. We call such a group an echo chamber. When $n_{II} = 35$, the group is larger and the additional

network between them is sparser.

Figure 6 illustrates the effect of the distribution of the links in A^- on misinformation. For a fixed number of links, denser groups (i.e. for smaller values of n_{II}) tend to induce misinformation in the long run. This effect is stronger when generating highly connected groups as the difference in probability of misinformation between $n_{II} = 5$ and $n_{II} = 20$ is much more pronounced than between $n_{II} = 20$ and $n_{II} = 35$. The reason is that high local density, works as an echo chamber, in which the signals of these agents are amplified.
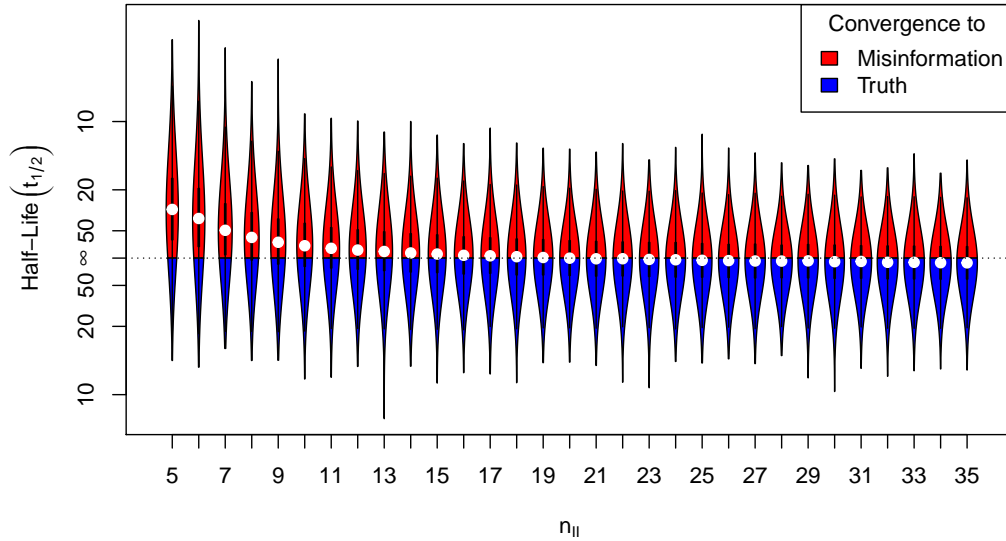

Figure 6: Network asymmetry: the echo chamber effect.

Notes: Results based on 10'000 replications per setting. Effect of varying local density while keeping number of links constant on misinformation and speed of convergence. α_I^+ and α_{II}^- are constant and equal to 10, meaning both group get 10 additional links in respectively the positive and negative network. The smaller n_{II}, the larger the local density, with a complete echo chamber at $n = 5$. The networks are symmetric in expectations for $n_{II} = 20$. For each violin, the red area represents the probability of misinformation for a given set of parameters. White dots represent the median, central black bars are the interquartile range and the thickness of the violin are density estimates.

6.3   Decay Asymmetry vs. Network Asymmetry

Finally, to illustrate how additional links in the positive signal sharing network A^+ can compensate for decay favoring negative signals (or inversely), we start from the same

society as in the two previous cases. 100 people are in a connected Erdős–Rényi random graph with an average degree of 5 in the A network. We then add the group structure with the following parameters, which now include the respective decays δ^+ and δ^-: $(n_0, n_I, n_{II}, \alpha, \alpha_I^+, \alpha_{II}^-, \delta^+, \delta^-) = (100 - n_I, n_I, 0, 250, \alpha_I^+, 0, \delta^+, 1)$. We thus vary the asymmetry between decays, favoring negative signals given $\delta^- = 1$ and $\delta^+ \in [0, 1]$. The asymmetry between networks is also varied, favoring the positive signal sharing network, which is the only one to receive additional links. As we have established, how the additional α_I^+ links are added will influence the state reached by this society. We thus compare two link generation processes: random and echo chamber. In the random link generation process, we first add one link at random in A^+, which generates network asymmetry. We then compute what decay asymmetry it exactly compensates. We then add another link in A^+ at random, and again, compute the new decay asymmetry it compensates, and so on. In the echo chamber process, each additional link is created with the intent of generating an increasingly large echo chamber. The first additional link is generated between two random nodes, the second between these two nodes and a third. The third link will either increase the connectedness of the first three nodes, or, if all links are already present, create a link between a fourth node and the first three, and so on. At each step, we compute what decay asymmetry is compensated.

In Figure 7, we show the ratio of links between A^+ and A^- needed to exactly compensate different levels of decay asymmetry, using the two different link generation processes. For example, if positive signals face twice the decay of negative signals, i.e. $\frac{\delta^+}{\delta^-} = \frac{0.5}{1} = 0.5$, then it takes more than twice as many random links in the positive network than in the negative network to compensate it, as the purple lines are around $\frac{\alpha + \alpha_I^+}{\alpha + \alpha_{II}^-} = \frac{250 + \alpha_I^+}{250} \approx 2.2$. In contrast, for the same compensation it takes only about 20% more links if the new links are arranged in an echo chamber (as the black lines are roughly at ratio of densities $\frac{\alpha + \alpha_I^+}{\alpha + \alpha_{II}^-} = \frac{250 + \alpha_I^+}{250} \approx 1.2$.) Consequently, if the ratio of densities is between these two values and still positive signals face twice the decay of negative signals, there is most likely misinformation when the positive links were added at random, while there is most likely convergence to the truth when these links are arranged as an echo chamber.
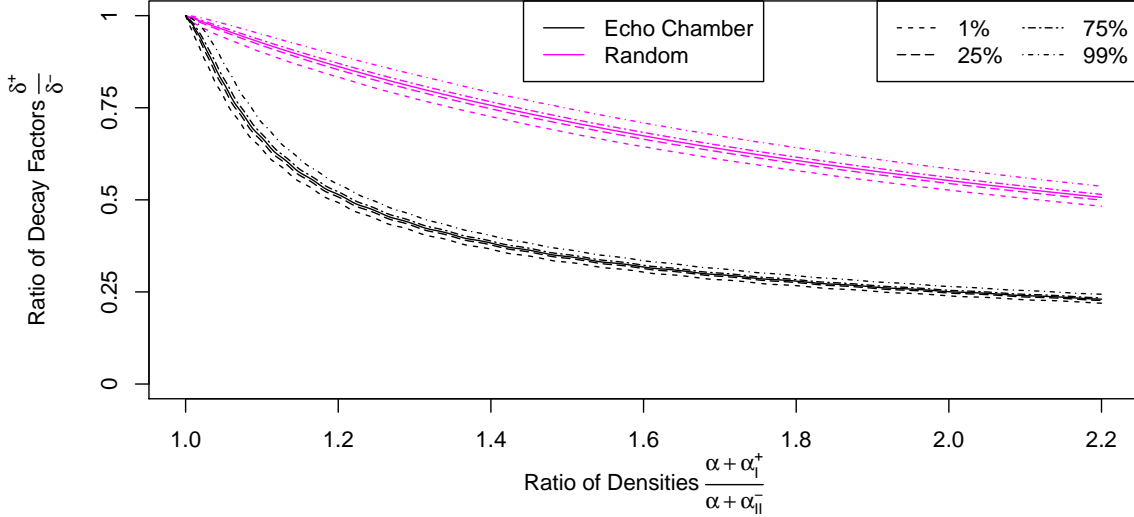
Figure 7: Compensating decay asymmetry with network asymmetry.

Notes: Results based on 10'000 replications per setting. Each line represents a different additional link generation process. The lines draw the median ratios of decay asymmetry that exactly compensate the ratio of links between the A^+ and A^- networks, generating Case 3 of Proposition 2. For parameters at the lower left of a line (with constant link generation process), a society is misinformed. For parameters at the upper right of a line, a society reaches the true state.

7 Discussion

7.1 Limitations

Separating decay and signal type The model assumes decay to be fully dependent on signal type. Agents share all true, and respectively false, signals in the same manner. More realistically, decay would be idiosyncratic to the signal and only partly correlated to signal type. Negative signals might generally be more shareable and hence decay less, but some positive signals might also be highly shareable.

Extending our model to allow for idiosyncratic decay factors would generalize our main condition $\delta^+\lambda_1^+ \lessgtr \delta^-\lambda_1^-$ to considering not the common decay factors, but the maximally realized decay factors in this equation, i.e. the highest decay among all positive signals and the highest decay among all negative signals. While gaining realism, this extension

would not change the comparative-statics of decay asymmetry and network asymmetry.

Accounting for several networks We assume the existence of two networks over which are shared true or false signals exclusively. In reality, there exists a complex overlay of different online networks, messaging apps, live social networks, etc. One can thus be connected with one another through many links. We partly cover this possibility in the extension in Section 5.4 as one can interpret the weighted matrix to reflect the strength of connections between agents across all existing networks collapsed into one number. However, another viable option would be extending the model to allow for multiple links between agents.

7.2   Policy Implications

It is clear from our results that there is no silver bullet in the fight against misinformation. A society reaches the true state or is misinformed as a result of different forces that are opposing each other. We can however conclude that both signal-dependent sharing behaviors and network structures play determining roles and extract counter-measures from these two types of asymmetries. Counter-measures can further be split in two: those aiming to improve the travel of true information and those aiming to decrease the travel of false information.

Improving decay of true information ($\delta^+ \uparrow$) Recently, scholars have reported that the emotional content of true and false information is different. Scientific communication tends to generate more neutral and positive emotional reaction than unfounded conspiracies (Zollo et al., 2015). Similarly, false information in different topics such as business, politics, or technology tend to generate more surprise, fear and disgust than true information on the same topics. Vosoughi et al. (2018) argue that these emotional differences could explain why false information travels further. This implies that we have to consider how the emotional content of true information makes an impact on the receiver and how it will help or hamper its ability to travel. Easily forgettable information will decay faster.

Similarly, the medium has to be considered. The question of how to best get scientific information to the general public is not new (Miller, 2001), and has arguably still to be answered. Most scientific publications still exist only in article form, and are often not accessible to the general public. If they are, their complexity regularly constitutes an impenetrable barrier. In the meantime, false information is created in ways that is easily shareable online via social media or messaging apps such as images, videos and

26

news-like article. Providers of scientific information are thus not always fighting with the appropriate tools and should consider how to tailor their message in a way that allows it to keeps its informative integrity, while being easily shareable.

Weakening decay of false information ($\delta^- \downarrow$) Educating citizens appears to be an important way to accelerate the decay of false information. This education can take different forms. For example, educating to recognize false information, and training to question one's own information sharing behaviors. Mo and Mortensen (2019) have shown that people with higher information literacy scores are significantly better at identifying false information. Information literacy emphasizes the ability to navigate and locate information, to recognize opinion pieces or to search databases (Livingstone et al., 2008). Providing citizens opportunities to develop these skills is thus critical. On top, allowing people to reflect on their existing sharing behavior can prove effective, without even building new skills. Fazio (2020) reports that simply asking participants to pause and consider the veracity of a headline significantly decreases their intention to share false information. Pausing did not affect intentions to share true information. Similarly, priming the importance of accuracy in the mind of participants decreases their intent to share misinformation (Pennycook et al., 2021). Having well equipped citizens, who are able to recognize false information, is thus a necessity to slow down its spread. They should however not be left alone to carry this responsibility, simply due to the sheer amount and complexity of information to verify. Institutions are calling for the support of trusted fact-checkers which should facilitate the identification of false information for the end user (European Commission, 2018).

There is room for social media platforms to integrate these ways to accelerate the decay of false information. May it be by the integration of fact-checkers and subsequent flags in feeds or via introducing mechanisms that nudge users to reflect on what they share. How to best implement it is then still to be determined and tested.

Improving the true information sharing network ($\lambda_1^+ \uparrow$) The intuitive network policy to favor true information would be to increase the density of the true information sharing network. That is, adding links to said network. However, what exactly constitutes a *link* in an online social network is less clear than it seems in the models of social learning. In the context of this paper, what matters are links through which information can travel from one person to another, which happens online mainly through a timeline or feed, depending on which platform is used. The information one receives, is then the result of conscious choices of the user, through the people and pages she links with,

and recommendations from algorithms that are based on previous behaviors and other undisclosed factors. This implies that out of all the possible pieces of information to display based on a user's full network, the algorithm influences which information to display first (or at all). The feed algorithm thus activates or silences existing links. On top, the algorithm can generate new links and display information from nodes that a user was not connected to, e.g. when advertising. Feed algorithms can hence generate new links, and do it. Social media platforms thus have the capacity to both favor displaying trusted sources of information when the link exists, or create it if it does not.

Weakening the false information sharing network $(\lambda_1^- \downarrow)$ Finally, making the network over which false information is shared sparser favors reaching the true state. Our result concur with the theoretical results of (Törnberg, 2018) and the empirical observations of (Johnson et al., 2020); the organization in tight clusters, or echo chambers, of groups that spread misinformation increase their influence on the long-run beliefs. Breaking these echo chambers thus appears as one of the most efficient way to fight misinformation. However, silencing or breaking links that users have generated consciously raises ethical concerns, which are outside of the scope of this paper.

References

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

Albert, R. and Barabási, A.-L. (2002). Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74(1):47.

Azzimonti, M. and Fernandes, M. (2018). Social Media Networks, Fake News, and Polarization. NBER Working Papers 24462, National Bureau of Economic Research, Inc.

Banerjee, A., Breza, E., Chandrasekhar, A. G., and Mobius, M. (2019). Naive Learning with Uninformed Agents. NBER Working Papers 25497, National Bureau of Economic Research, Inc.

Bonacich, P. (1972). Factoring and weighting approaches to status scores and clique identification. *Journal of Mathematical Sociology*, 2(1):113–120.

Bonacich, P. (1987). Power and centrality: A family of measures. *American Journal of Sociology*, 92(5):1170–1182.

Buechel, B., Hellmann, T., and Klner, S. (2015). Opinion dynamics and wisdom under conformity. *Journal of Economic Dynamics and Control*, 52:240 – 257.

Burki, T. (2019). Vaccine misinformation and social media. *The Lancet Digital Health*, 1(6):e258–e259.

Chandrasekhar, A. G., Larreguy, H., and Xandri, J. P. (2020). Testing models of social learning on networks: Evidence from two experiments. *Econometrica*, 88(1):1–32.

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):1276–1288.

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

Del Vicario, M., Bessi, A., Zollo, F., Petroni, F., Scala, A., Caldarelli, G., Stanley, H. E., and Quattrociocchi, W. (2016). The spreading of misinformation online. *Proceedings of the National Academy of Sciences*, 113(3):554–559.

Della Lena, S. (2019). Non-Bayesian Social Learning and the Spread of Misinformation in Networks. Working Papers 2019:09, Department of Economics, University of Venice "Ca' Foscari".

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

European Commission (2018). Tackling online disinformation: a european approach. Communication from the commission to the European parliament, the council, the European economic and social committee and the committee of the regions COM(2018) 236 final, Office of the United Nations High Commissioner for Human Rights.

Fazio, L. (2020). Pausing to consider why a headline is true or false can help reduce the sharing of false news. *Harvard Kennedy School Misinformation Review*, 1(2).

Friedkin, N. E. (1991). Theoretical foundations for centrality measures. *American Journal of Sociology*, 96(6):1478–1504.

Friedkin, N. E. and Bullo, F. (2017). How truth wins in opinion dynamics along issue sequences. *Proceedings of the National Academy of Sciences*, 114(43):11380–11385.

Friedkin, N. E. and Johnsen, E. C. (1990). Social influence and opinions. *Journal of Mathematical Sociology*, 15(3-4):193–206.

Golub, B. and Jackson, M. O. (2010). Naïve learning in social networks and the wisdom of crowds. *American Economic Journal: Microeconomics*, 2(1):112–49.

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

Golub, B. and Sadler, E. D. (2016). Learning in social networks. *Social Science Electronic Publishing*.

Grabisch, M., Mandel, A., Rusinowska, A., and Tanimura, E. (2018). Strategic influence in social networks. *Mathematics of Operations Research*, 43(1):29–50.

Grabisch, M., Poindron, A., and Rusinowska, A. (2019). A model of anonymous influence with anti-conformist agents. *Journal of Economic Dynamics and Control*, 109(C).

Grimm, V. and Mengel, F. (2018). An Experiment on Belief Formation in Networks. *Journal of the European Economic Association*.

Jackson, M. O. (2010). *Social and economic networks*. Princeton University Press.

Jadbabaie, A., Molavi, P., Sandroni, A., and Tahbaz-Salehi, A. (2012). Non-Bayesian social learning. *Games and Economic Behavior*, 76(1):210–225.

Johnson, N. F., Velásquez, N., Restrepo, N. J., Leahy, R., Gabriel, N., El Oud, S., Zheng, M., Manrique, P., Wuchty, S., and Lupu, Y. (2020). The online competition between pro-and anti-vaccination views. *Nature*, pages 1–4.

Lazer, D. M., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., Metzger, M. J., Nyhan, B., Pennycook, G., Rothschild, D., et al. (2018). The science of fake news. *Science*, 359(6380):1094–1096.

Livingstone, S., Van Couvering, E., Thumin, N., Coiro, J., Knobel, M., Lankshear, C., and Leu, D. (2008). Converging traditions of research on media and information literacies. *Handbook of Research on New Literacies*, pages 103–132.

Miller, S. (2001). Public understanding of science at the crossroads. *Public Understanding of Science*, 10(1):115–120.

Mo, J. J. and Mortensen, T. (2019). Does media literacy help identification of fake news? information literacy helps, but other literacies don't. *American Behavioral Scientist*.

Molavi, P., Tahbaz-Salehi, A., and Jadbabaie, A. (2018). A theory of non-bayesian social learning. *Econometrica*, 86(2):445–490.

Mueller-Frank, M. (2013). A general framework for rational learning in social networks. *Theoretical Economics*, 8(1):1–40.

Mueller-Frank, M. (2014). Does one Bayesian make a difference? *Journal of Economic Theory*, 154(C):423–452.

Pennycook, G., Epstein, Z., Mosleh, M., Arechar, A. A., Eckles, D., and Rand, D. G. (2021). Shifting attention to accuracy can reduce misinformation online. *Nature*, 592(7855):590–595.

Pennycook, G. and Rand, D. (2021). Examining false beliefs about voter fraud in the wake of the 2020 presidential election. *The Harvard Kennedy School Misinformation Review*, 2.

Quattrociocchi, W., Scala, A., and Sunstein, C. R. (2016). Echo chambers on facebook. *Available at SSRN 2795110*.

Rusinowska, A. and Taalaibekova, A. (2019). Opinion formation and targeting when persuaders have extreme and centrist opinions. *Journal of Mathematical Economics*, 84(C):9–27.

Sikder, O., Smith, R. E., Vivo, P., and Livan, G. (2020). A minimalistic model of bias, polarization and misinformation in social networks. *Scientific Reports*, 10(1):1–11.

Strogatz, S. H. (2001). Exploring complex networks. *Nature*, 410(6825):268–276.

Taalaibekova, A. (2020). *Diffusion of opinions and innovations among limitedly forward-looking individuals.* PhD thesis, UCL-Université Catholique de Louvain.

Törnberg, P. (2018). Echo chambers and viral misinformation: Modeling fake news as complex contagion. *PLoS ONE*, 13(9):e0203958.

Vosoughi, S., Roy, D., and Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380):1146–1151.

Zollo, F., Bessi, A., Del Vicario, M., Scala, A., Caldarelli, G., Shekhtman, L., Havlin, S., and Quattrociocchi, W. (2017). Debunking in a world of tribes. *PLoS ONE*, 12(7):e0181821.

Zollo, F., Novak, P. K., Del Vicario, M., Bessi, A., Mozetič, I., Scala, A., Caldarelli, G., and Quattrociocchi, W. (2015). Emotional dynamics in the age of misinformation. *PLoS ONE*, 10(9):e0138740.

A Appendix: Proofs

A.1 Proof of Lemma 1

First of all, due to the definitions of $k_i(t) = N_i^+(t) - N_i^-(t)$ and $x_i(t) = \frac{N_i^+(t)}{N_i^+(t)+N_i^-(t)}$, it is easy to see that $k_i(t) = 2\left(x_i(t) - \frac{1}{2}\right)\left(N_i^+(t) + N_i^-(t)\right)$. Thus, Equation (3) can be rewritten as

$$b_i(t) = \frac{1}{1 + \frac{1-b^0}{b^0}\left(\frac{1-\rho}{\rho}\right)^{k_i(t)}} = \frac{1}{1 + \frac{1-b^0}{b^0}\left(\frac{1-\rho}{\rho}\right)^{2\left(x_i(t)-\frac{1}{2}\right)\left(N_i^+(t)+N_i^-(t)\right)}}$$

Thus, $b_i(t)$ is larger (smaller) than 0.5 if and only if $\frac{1-b^0}{b^0}\left(\frac{1-\rho}{\rho}\right)^{2\left(x_i(t)-\frac{1}{2}\right)\left(N_i^+(t)+N_i^-(t)\right)}$ is smaller (larger) than 1. For $b^0 = 0.5$, this is equivalent to $x_i(t) > 0.5$ ($x_i(t) < 0.5$), due to $\frac{1-\rho}{\rho}$ being smaller than 1 and $N_i^+(t) + N_i^-(t)$ being positive. As for $t \gg 0$, $N_i^+(t) + N_i^-(t)$ will be large,[16] $b_i(t)$ will be larger than 0.5 if $x_i(t) > 0.5$, and smaller than 0.5 if $x_i(t)$ is smaller than 0.5.

A.2 Proof of Proposition 1

To prove Proposition 1 we apply Proposition 2, which is proven below. Using the result of Case 3 of Proposition 2 and $c^+ = c^- = c$, we find that

$$\lim_{t\to\infty} x_i(t) = \frac{1}{1 + \frac{c_i^-}{c_i^+}\frac{\sum_{k=1}^n (c_k^+)^2}{\sum_{k=1}^n (c_k^-)^2}\frac{1 - \sum_{j=1}^n c_j^- s_j}{\sum_{j=1}^n c_j^+ s_j}} = \frac{1}{1 + \frac{1 - \sum_{j=1}^n c_j s_j}{\sum_{j=1}^n c_j s_j}}$$

$$= \frac{1}{\frac{\sum_{j=1}^n c_j s_j + 1 - \sum_{j=1}^n c_j s_j}{\sum_{j=1}^n c_j s_j}} = \sum_{j=1}^n c_j s_j.$$

We now show that the probability of misinformation is bounded by 0.5, i.e.

$$P(x_i(\infty) < 0.5) + 0.5 P(x_i(\infty) = 0.5) \leq 0.5 \text{ for all } i.$$

This amounts to showing that

$$P\left(\sum_{j=1}^n c_j s_j < \frac{1}{2}\right) + \frac{1}{2}P\left(\sum_{j=1}^n c_j s_j = \frac{1}{2}\right) \leq \frac{1}{2}, \tag{A.1}$$

as $x_i(\infty) = \sum_{j=1}^n c_j s_j$ for all i. In order to prove Equation (A.1), we define the following quantities: $p_l(\rho) := P\left(\sum_{j=1}^n c_j s_j < \frac{1}{2}\right)$, $p_m(\rho) := P\left(\sum_{j=1}^n c_j s_j = \frac{1}{2}\right)$, and $p_u(\rho) :=$

---

[16]This follows i.a. from the representations for $N^+(t)$ and $N^-(t)$ given in Equations (A.2) and (A.3).

$P\left(\sum_{j=1}^{n} c_j s_j > \frac{1}{2}\right)$, which obviously sum to unity: $p_l(\rho) + p_m(\rho) + p_u(\rho) = 1$ for all ρ. Using these, Equation (A.1) can be restated as $p_l(\rho) + \frac{1}{2} p_m(\rho) \leq \frac{1}{2}$, which is equivalent to $2p_l(\rho) + p_m(\rho) \leq 1 = p_l(\rho) + p_m(\rho) + p_u(\rho)$, which is in turn equivalent to $p_l(\rho) \leq p_u(\rho)$. This part of the proof can therefore be completed by showing that indeed $p_l(\rho) \leq p_u(\rho)$ for all $\rho > \frac{1}{2}$. Due to $\sum_{j=1}^{n} c_j(1 - s_j) = \sum_{j=1}^{n} c_j - \sum_{j=1}^{n} c_j s_j = 1 - \sum_{j=1}^{n} c_j s_j$, the condition $\sum_{j=1}^{n} c_j s_j < \frac{1}{2}$ is equivalent to $\sum_{j=1}^{n} c_j(1 - s_j) > \frac{1}{2}$. As $1 - s_j$ takes the values 0 and 1 with probabilities ρ and $1 - \rho$, respectively, we find that

$$p_l(\rho) = P\left(\sum_{j=1}^{n} c_j s_j < \frac{1}{2}\right) = P\left(\sum_{j=1}^{n} c_j(1 - s_j) > \frac{1}{2}\right) = p_u(1 - \rho) \leq p_u(\rho),$$

where the inequality at the end holds true because $1 - \rho < \rho$, due to $\rho > \frac{1}{2}$. We thus have shown that $p_l(\rho) \leq p_u(\rho)$, concluding the proof.

A.3 Proof of Proposition 2

To begin the proof, we use the eigendecompositions of the real symmetric matrices A^+ and A^-, writing them as $A^+ = Q^+ \Lambda^+ (Q^+)^\top$ and $A^- = Q^- \Lambda^- (Q^-)^\top$, respectively, with Q^+ and Q^- being orthogonal matrices whose columns are eigenvectors of A^+ and A^-, and Λ^+ and Λ^- being diagonal matrices whose entries are the eigenvalues of A^+ and A^-. From this, we have that $I + \delta^+ A^+ = Q^+ (I + \delta^+ \Lambda^+) (Q^+)^\top$ and $I + \delta^- A^- = Q^- (I + \delta^- \Lambda^-) (Q^-)^\top$ as well as $(I + \delta^+ A^+)^t = Q^+ (I + \delta^+ \Lambda^+)^t (Q^+)^\top$ and $(I + \delta^- A^-)^t = Q^- (I + \delta^- \Lambda^-)^t (Q^-)^\top$ for all t. Overall, this delivers

$$\left(I + \delta^+ A^+\right)^t = \sum_{i=1}^{n} \left(1 + \delta^+ \lambda_i^+\right)^t q_i^+ \left(q_i^+\right)^\top, \quad \left(I + \delta^- A^-\right)^t = \sum_{i=1}^{n} \left(1 + \delta^- \lambda_i^-\right)^t q_i^- \left(q_i^-\right)^\top,$$

with q_i^+ and q_i^- $(i = 1, \ldots, n)$ denoting the eigenvectors of A^+ and A^-, respectively. Denoting the vector of initial signals by s, we thus get:

$$N^+(t) = \left(I + \delta^+ A^+\right)^t N^+(0) = \left(I + \delta^+ A^+\right)^t s = \sum_{j=1}^{n} \left(1 + \delta^+ \lambda_j^+\right)^t q_j^+ \left(q_j^+\right)^\top s,$$

$$N^-(t) = \left(I + \delta^- A^-\right)^t N^-(0) = \left(I + \delta^+ A^+\right)^t (\mathbb{1} - s) = \sum_{j=1}^{n} \left(1 + \delta^- \lambda_j^-\right)^t q_j^- \left(q_j^-\right)^\top (\mathbb{1} - s).$$

From this, we get for the numbers of positive signals at time t,

$$N^+(t) = \left(1 + \delta^+ \lambda_1^+\right)^t \left(q_1^+ \left(q_1^+\right)^\top s + \sum_{j=2}^{n} \left(\frac{1 + \delta^+ \lambda_j^+}{1 + \delta^+ \lambda_1^+}\right)^t q_j^+ \left(q_j^+\right)^\top s\right), \qquad \text{(A.2)}$$

34

and for the numbers of negative signals at time t,

$$N^-(t) = \left(1 + \delta^- \lambda_1^-\right)^t \left(q_1^- \left(q_1^-\right)^\top (\mathbb{1} - s) + \sum_{j=1}^n \left(\frac{1 + \delta^- \lambda_j^-}{1 + \delta^- \lambda_1^-}\right)^t q_j^- \left(q_j^-\right)^\top (\mathbb{1} - s)\right). \quad \text{(A.3)}$$

Due to Perron-Frobenius theory, it is clear that $\dfrac{1 + \delta^+ \lambda_j^+}{1 + \delta^+ \lambda_1^+}$ and $\dfrac{1 + \delta^- \lambda_j^-}{1 + \delta^- \lambda_1^-}$ are both smaller than 1 for $j = 2, \ldots, n$, implying

$$\left(1 + \delta^+ \lambda_1^+\right)^{-t} N^+(t) \overset{t \to \infty}{\longrightarrow} q_1^+ \left(q_1^+\right)^\top s, \quad \left(1 + \delta^+ \lambda_1^-\right)^{-t} N^-(t) \overset{t \to \infty}{\longrightarrow} q_1^- \left(q_1^-\right)^\top (\mathbb{1} - s).$$

Now, using that $q_1^+ = \dfrac{c^+}{||c^+||} = \dfrac{c^+}{\sqrt{\sum_{k=1}^n \left(c_k^+\right)^2}}$ and $q_1^- = \dfrac{c^-}{||c^-||} = \dfrac{c^-}{\sqrt{\sum_{k=1}^n \left(c_k^-\right)^2}}$, we finally get:

$$\left(1 + \delta^+ \lambda_1^+\right)^{-t} N^+(t) \overset{t \to \infty}{\longrightarrow} c^+ \frac{\left(c^+\right)^\top s}{\sum_{k=1}^n \left(c_k^+\right)^2} = c^+ \frac{\sum_{j=1}^n c_j^+ s_j}{\sum_{k=1}^n \left(c_k^+\right)^2}, \quad \text{(A.4)}$$

$$\left(1 + \delta^- \lambda_1^-\right)^{-t} N^-(t) \overset{t \to \infty}{\longrightarrow} c^- \frac{\left(c^-\right)^\top (\mathbb{1} - s)}{\sum_{k=1}^n \left(c_k^-\right)^2} = c^- \frac{\sum_{j=1}^n c_j^- (1 - s_j)}{\sum_{k=1}^n \left(c_k^-\right)^2} = c^- \frac{1 - \sum_{j=1}^n c_j^- s_j}{\sum_{k=1}^n \left(c_k^-\right)^2}. \quad \text{(A.5)}$$

With these general results at hand, we can address the three cases considered in Proposition 2.

1. For $\left(\dfrac{1 + \delta^+ \lambda_1^+}{1 + \delta^- \lambda_1^-}\right)^{-t} x_i(t)$, we get when $\delta^+ \lambda_1^+ < \delta^- \lambda_1^-$:

$$\left(\frac{1 + \delta^+ \lambda_1^+}{1 + \delta^- \lambda_1^-}\right)^{-t} x_i(t) = \left(\frac{1 + \delta^+ \lambda_1^+}{1 + \delta^- \lambda_1^-}\right)^{-t} \frac{N_i^+(t)}{N_i^+(t) + N_i^-(t)}$$

$$= \frac{\left(1 + \delta^+ \lambda_1^+\right)^{-t} N_i^+(t)}{\left(1 + \delta^- \lambda_1^-\right)^{-t} \left(N_i^+(t) + N_i^-(t)\right)}$$

$$= \frac{\left(1 + \delta^+ \lambda_1^+\right)^{-t} N_i^+(t)}{\left(\dfrac{1 + \delta^+ \lambda_1^+}{1 + \delta^- \lambda_1^-}\right)^t \left(1 + \delta^+ \lambda_1^+\right)^{-t} N_i^+(t) + \left(1 + \delta^- \lambda_1^-\right)^{-t} N_i^-(t)}$$

$$\overset{t \to \infty}{\longrightarrow} \frac{c_i^+ \dfrac{\sum_{j=1}^n c_j^+ s_j}{\sum_{k=1}^n \left(c_k^+\right)^2}}{c_i^- \dfrac{1 - \sum_{j=1}^n c_j^- s_j}{\sum_{k=1}^n \left(c_k^-\right)^2}} = \frac{c_i^+}{c_i^-} \frac{\sum_{k=1}^n (c_k^-)^2}{\sum_{k=1}^n (c_k^+)^2} \frac{\sum_{j=1}^n c_j^+ s_j}{1 - \sum_{j=1}^n c_j^- s_j}.$$

35

2. For $\left(\dfrac{1+\delta^-\lambda_1^-}{1+\delta^+\lambda_1^+}\right)^{-t}(1-x_i(t))$, we get when $\delta^+\lambda_1^+ > \delta^-\lambda_1^-$:

$$\left(\frac{1+\delta^-\lambda_1^-}{1+\delta^+\lambda_1^+}\right)^{-t}(1-x_i(t)) = \left(\frac{1+\delta^-\lambda_1^-}{1+\delta^+\lambda_1^+}\right)^{-t}\frac{N_i^-(t)}{N_i^+(t)+N_i^-(t)}$$

$$= \frac{\left(1+\delta^-\lambda_1^-\right)^{-t}N_i^-(t)}{\left(1+\delta^+\lambda_1^+\right)^{-t}\left(N_i^+(t)+N_i^-(t)\right)}$$

$$= \frac{\left(1+\delta^-\lambda_1^-\right)^{-t}N_i^-(t)}{\left(1+\delta^+\lambda_1^+\right)^{-t}N_i^+(t)+\left(\dfrac{1+\delta^-\lambda_1^-}{1+\delta^+\lambda_1^+}\right)^{t}\left(1+\delta^-\lambda_1^-\right)^{-t}N_i^-(t)}$$

$$\xrightarrow{t\to\infty} \frac{c_i^-\dfrac{1-\sum_{j=1}^{n}c_j^-s_j}{\sum_{k=1}^{n}\left(c_k^-\right)^2}}{c_i^+\dfrac{\sum_{j=1}^{n}c_j^+s_j}{\sum_{k=1}^{n}\left(c_k^+\right)^2}} = \frac{c_i^-}{c_i^+}\frac{\sum_{k=1}^{n}(c_k^+)^2}{\sum_{k=1}^{n}(c_k^-)^2}\frac{1-\sum_{j=1}^{n}c_j^-s_j}{\sum_{j=1}^{n}c_j^+s_j}.$$

3. Finally, when $\delta^+\lambda_1^+ = \delta^-\lambda_1^-$, we get:

$$x_i(t) = \frac{\left(1+\delta^+\lambda_1^+\right)^{-t}N_i^+(t)}{\left(1+\delta^+\lambda_1^+\right)^{-t}N_i^+(t)+\left(1+\delta^-\lambda_1^-\right)^{-t}N_i^-(t)}$$

$$\xrightarrow{t\to\infty} \frac{c_i^+\dfrac{\sum_{j=1}^{n}c_j^+s_j}{\sum_{k=1}^{n}\left(c_k^+\right)^2}}{c_i^+\dfrac{\sum_{j=1}^{n}c_j^+s_j}{\sum_{k=1}^{n}\left(c_k^+\right)^2}+c_i^-\dfrac{1-\sum_{j=1}^{n}c_j^-s_j}{\sum_{k=1}^{n}\left(c_k^-\right)^2}} = \frac{1}{1+\dfrac{c_i^-}{c_i^+}\dfrac{\sum_{k=1}^{n}(c_k^+)^2}{\sum_{k=1}^{n}(c_k^-)^2}\dfrac{1-\sum_{j=1}^{n}c_j^-s_j}{\sum_{j=1}^{n}c_j^+s_j}}.$$

A.4 Proof of Corollary 1

First of all, notice that $\dfrac{N_i^+(t)}{N_i^-(t)}\Big/\dfrac{N_j^+(t)}{N_j^-(t)}$ can be written as $\dfrac{x_i(t)(1-x_j(t))}{x_j(t)(1-x_i(t))}$. We thus have to prove that $\lim\limits_{t\to\infty}\dfrac{x_i(t)(1-x_j(t))}{x_j(t)(1-x_i(t))} = \dfrac{c_i^+c_j^-}{c_i^-c_j^+}$, which we will do by successively tackling the three cases given in Proposition 2. For Case 1, we know that $\delta^+\lambda_1^+ < \delta^-\lambda_1^-$ and

$$x_i(t)\left(\frac{1+\delta^-\lambda_1^-}{1+\delta^+\lambda_1^+}\right)^{t} \xrightarrow{t\to\infty} \frac{c_i^+}{c_i^-}\frac{\sum_{k=1}^{n}(c_k^-)^2}{\sum_{k=1}^{n}(c_k^+)^2}\frac{\sum_{l=1}^{n}c_l^+s_l}{1-\sum_{l=1}^{n}c_l^-s_l}.$$

36

as well as $x_i(t) \xrightarrow{t\to\infty} 0$ for all i. Trivially, thus, $1 - x_i(t)$ and $1 - x_j(t)$ each converge to 1. For

$$\frac{x_i(t)}{x_j(t)} = \frac{x_i(t)\left(\frac{1+\delta-\lambda_1^-}{1+\delta+\lambda_1^+}\right)^t}{x_j(t)\left(\frac{1+\delta-\lambda_1^-}{1+\delta+\lambda_1^+}\right)^t},$$

we then find that it converges to

$$\frac{\frac{c_i^+}{c_i^-}\frac{\sum_{k=1}^n (c_k^-)^2}{\sum_{k=1}^n (c_k^+)^2}\frac{\sum_{l=1}^n c_l^+ s_l}{1-\sum_{l=1}^n c_l^- s_l}}{\frac{c_j^+}{c_j^-}\frac{\sum_{k=1}^n (c_k^-)^2}{\sum_{k=1}^n (c_k^+)^2}\frac{\sum_{l=1}^n c_l^+ s_l}{1-\sum_{l=1}^n c_l^- s_l}} = \frac{\frac{c_i^+}{c_i^-}}{\frac{c_j^+}{c_j^-}},$$

which proves the assertion for Case 1.

For Case 2, we know that $\delta^+\lambda_1^+ > \delta^-\lambda_1^-$ and

$$(1 - x_i(t))\left(\frac{1+\delta^+\lambda_1^+}{1+\delta^-\lambda_1^-}\right)^t \xrightarrow{t\to\infty} \frac{c_i^-}{c_i^+}\frac{\sum_{k=1}^n (c_k^+)^2}{\sum_{k=1}^n (c_k^-)^2}\frac{\sum_{l=1}^n c_l^- s_l}{1-\sum_{l=1}^n c_l^+ s_l}.$$

as well as $x_i(t) \xrightarrow{t\to\infty} 1$ for all i. Trivially, thus, $x_i(t)$ and $x_j(t)$ each converge to 1. For

$$\frac{1 - x_j(t)}{1 - x_i(t)} = \frac{(1 - x_j(t))\left(\frac{1+\delta^+\lambda_1^+}{1+\delta^-\lambda_1^-}\right)^t}{(1 - x_i(t))\left(\frac{1+\delta^+\lambda_1^+}{1+\delta^-\lambda_1^-}\right)^t},$$

we then find that it converges to

$$\frac{\frac{c_j^-}{c_j^+}\frac{\sum_{k=1}^n (c_k^+)^2}{\sum_{k=1}^n (c_k^-)^2}\frac{\sum_{l=1}^n c_l^- s_l}{1-\sum_{l=1}^n c_l^+ s_l}}{\frac{c_i^-}{c_i^+}\frac{\sum_{k=1}^n (c_k^+)^2}{\sum_{k=1}^n (c_k^-)^2}\frac{\sum_{l=1}^n c_l^- s_l}{1-\sum_{l=1}^n c_l^+ s_l}} = \frac{\frac{c_i^+}{c_i^-}}{\frac{c_j^+}{c_j^-}},$$

which proves the assertion for Case 2.

Finally, for the Case 3, we know that for all i

$$\lim_{t\to\infty} x_i(t) = \frac{1}{1 + \dfrac{c_i^-}{c_i^+}\dfrac{\sum_{k=1}^n (c_k^+)^2}{\sum_{k=1}^n (c_k^-)^2}\dfrac{1-\sum_{l=1}^n c_l^- s_l}{\sum_{l=1}^n c_l^+ s_l}} \in (0,1).$$

This immediately implies that

$$\lim_{t\to\infty} 1 - x_i(t) = \frac{\dfrac{c_i^-}{c_i^+}\dfrac{\sum_{k=1}^n (c_k^+)^2}{\sum_{k=1}^n (c_k^-)^2}\dfrac{1-\sum_{j=1}^n c_j^- s_j}{\sum_{l=1}^n c_l^+ s_l}}{1 + \dfrac{c_i^-}{c_i^+}\dfrac{\sum_{k=1}^n (c_k^+)^2}{\sum_{k=1}^n (c_k^-)^2}\dfrac{1-\sum_{j=1}^n c_j^- s_j}{\sum_{l=1}^n c_l^+ s_l}} \quad \text{as well as}$$

$$\lim_{t\to\infty} \frac{x_i(t)}{1 - x_i(t)} = \frac{c_i^+}{c_i^-}\frac{\sum_{k=1}^n (c_k^-)^2}{\sum_{k=1}^n (c_k^+)^2}\frac{\sum_{l=1}^n c_l^+ s_l}{1-\sum_{l=1}^n c_l^- s_l}$$

for all i, from which it easily follows that $\lim_{t\to\infty} \frac{x_i(t)(1-x_j(t))}{x_j(t)(1-x_i(t))} = \frac{c_i^+ c_j^-}{c_i^- c_j^+}$.

B   Extension: Heterogeneous Relations

B.1   Extended Key Result

Denote by $\lambda_1(M^+)$ the largest eigenvalue of matrix M^+ and denote by c^+, d^+ the corresponding right and left eigenvector, normalized such that $\sum_{j=1}^n c_j^+ = 1 = \sum_{j=1}^n d_j^+.$[17] Likewise, let $\lambda_1(M^-)$ be the largest eigenvalue of matrix M^- and denote by c^-, d^- the corresponding normalized right and left eigenvector. Notice that these eigenvalues and eigenvectors now contain information not only about network asymmetry, but also about decay asymmetries, as the weights δ_{ij}^+ and δ_{ij}^- have already entered the matrices M^+ and M^-. When these matrices are considered as weighted networks, c^+ and c^- are called eigenvector centrality or right-hand eigenvector centrality of M^+ and M^- (Bonacich, 1987), while d^+ and d^- can be called left-hand eigenvector centrality (e.g. Golub and Sadler, 2016).

Proposition B.1 (Extended Key Result). *Suppose that the initial distribution of signals contains at least one positive and at least one negative signal.*

1. *If $\lambda_1(M^+) < \lambda_1(M^-)$, then for all i and large t*

$$x_i(t) \approx \frac{c_i^+}{c_i^-} \left(\frac{1 + \lambda_1(M^+)}{1 + \lambda_1(M^-)} \right)^t \frac{\sum_{k=1}^n c_k^- d_k^-}{\sum_{k=1}^n c_k^+ d_k^+} \frac{\sum_{j=1}^n d_j^+ s_j}{1 - \sum_{j=1}^n d_j^- s_j}$$

   such that $\lim_{t \to \infty} x_i(t) = 0$. Hence, misinformation prevails.

2. *If $\lambda_1(M^+) > \lambda_1(M^-)$, then for all i and large t:*

$$x_i(t) \approx 1 - \frac{c_i^-}{c_i^+} \left(\frac{1 + \lambda_1(M^-)}{1 + \lambda_1(M^+)} \right)^t \frac{\sum_{k=1}^n c_k^+ d_k^+}{\sum_{k=1}^n c_k^- d_k^-} \frac{1 - \sum_{j=1}^n d_j^- s_j}{\sum_{j=1}^n d_j^+ s_j}$$

   such that $\lim_{t \to \infty} x_i(t) = 1$. Hence, misinformation vanishes.

3. *If $\lambda_1(M^+) = \lambda_1(M^-)$, then for all i:*

$$\lim_{t \to \infty} x_i(t) = \frac{1}{1 + \dfrac{c_i^-}{c_i^+} \dfrac{\sum_{k=1}^n c_k^+ d_k^+}{\sum_{k=1}^n c_k^- d_k^-} \dfrac{1 - \sum_{j=1}^n d_j^- s_j}{\sum_{j=1}^n d_j^+ s_j}}.$$

   Hence, long-run misinformation depends on the signal distribution and on the eigenvectors that correspond to the largest eigenvalues.

---

[17]The two eigenvectors c^+ and d^+ coincide in the special case that the matrix M^+ is symmetric.

Proposition B.1 first of all shows that the key result obtained in our baseline model (Proposition 2) is robust to the broad generalization. Second, the crucial condition is now expressed in terms of the largest eigenvalues $\lambda_1(M^+)$ and $\lambda_1(M^-)$. Again there are special cases in which it is clear which of those is larger. For instance, suppose every element of matrix M^+ is weakly smaller than every element of M^-, i.e. for all i, j, $m_{ij}^+ \leq m_{ij}^-$; and for some the inequality is strict. Then $\lambda_1(M^+) < \lambda_1(M^-)$ such that we are in Case 1 and misinformation prevails. Analogously, for the situation that M^+ is element-wise bigger than M^-. Similarly, if $M^+ = M^-$, we are clearly in Case 3 such that misinformation depends on the distribution of initial signals and on the eigenvectors corresponding to the largest eigenvalue. Then, the eigenvectors for the positive and the negative signal sharing networks coincide: $c^+ = c^-$ and $d^+ = d^-$ and long-run misinformation can be expressed in terms of eigenvector centrality with respect to the weighted network $(N, M^+) = (N, M^-)$. Third, the dynamics in the three cases correspond to the dynamics characterized in the baseline model. For instance, the observation in Proposition 2 Case 3 that the signal mix is increasing in s_j translates into the same observation in Case 3 of Proposition B.1. The speed of convergences can also be assessed like in the baseline model. For instance, $\lambda_1(M^+) < \lambda_1(M^-)$ (Case 1), then speed can be measured by $\log\left(\frac{1+\lambda_1(M^-)}{1+\lambda_1(M^+)}\right)$. Finally, the fact that decay and network asymmetry are mingled into single matrices M^+ and M^- makes it more difficult to disentangle their effects.

Corollary 2 (Centrality Ratios and Opinion Diversity). *Suppose that the initial distribution of signals contains at least one positive and at least one negative signal. Then, the ratio of two agents' ratios of positive over negative signals converges to these agents' ratio of centrality ratios, i.e.*

$$\lim_{t\to\infty} \frac{N_i^+(t)}{N_i^-(t)} \Big/ \frac{N_j^+(t)}{N_j^-(t)} = \frac{c_i^+}{c_i^-} \Big/ \frac{c_j^+}{c_j^-}. \tag{B.1}$$

Hence, an agent i with higher centrality ratio than another agent j has a higher asymptotic signal mix and belief, i.e. if $\frac{c_i^+}{c_i^-} > \frac{c_j^+}{c_j^-}$, then for large t, $x_i(t) > x_j(t)$ and $b_i(t) > b_j(t)$.

Eigenvector centralities c and their ratios govern the relations of asymptotic signal mixes between agents.

The results under symmetry (Section 4), do not change in the extended model, i.e. if $\delta_{ij}^+ = \delta_{ij}^-$ and $a_{ij}^+ = a_{ij}^-$ for all i, j, then the results of Proposition 1 stay essentially unchanged. The only difference is that every appearance of (left and right) eigenvector centrality c, in Proposition 1 has to be replaced by d, the right eigenvector centrality. The important point of symmetry thus is symmetry with respect to positive and negative networks, but not symmetry of the matrices A or M: It is not essential that the network

is undirected or that the discounting is symmetric in the sense that $\delta_{ij} = \delta_{ji}$. The only thing that matters is the symmetry between positive and negative networks.

B.2 Proof of Proposition B.1

In order to prove the assertions, we will show that

$$\left(1 + \lambda_1(M^+)\right)^{-t} N^+(t) \overset{t \to \infty}{\longrightarrow} c^+ \frac{(d^+)^\top s}{\sum\limits_{k=1}^n c_k^+ d_k^+} = c^+ \frac{\sum\limits_{j=1}^n d_j^+ s_j}{\sum\limits_{k=1}^n c_k^+ d_k^+}, \tag{B.2}$$

$$\left(1 + \lambda_1(M^-)\right)^{-t} N^-(t) \overset{t \to \infty}{\longrightarrow} c^- \frac{(d^-)^\top (\mathbb{1} - s)}{\sum\limits_{k=1}^n c_k^- d_k^-} = c^- \frac{1 - \sum\limits_{j=1}^n d_j^- s_j}{\sum\limits_{k=1}^n c_k^- d_k^-}. \tag{B.3}$$

With Equations (B.2) and (B.3) at hand, the assertions of Proposition B.1 then follow from exactly the same arguments as those given in the proof of Proposition 2 after Equations (A.4) and (A.5).

As the essential parts of Equations (B.2) and (B.3) coincide, determining the limits of $(1 + \lambda_1(M^+))^{-t} N^+(t)$ is completely analogous to determining the limit of $(1 + \lambda_1(M^-))^{-t} N^-(t)$. Thus, we will do this in one sweep by looking at the limit of $(1 + \lambda_1(M))^{-t} (I + M)^t$, where M and stands for M^+ and M^-, respectively. The proof will thus be complete when showing that

$$\left(1 + \lambda_1(M)\right)^{-t} (I + M)^t \overset{t \to \infty}{\longrightarrow} \frac{c\, d^\top}{c^\top d}, \tag{B.4}$$

where c (d) stands for c^+ and c^- (d^+ and d^-), respectively.[18] In order to prove Equation (B.4), we first rewrite M using its Jordan normal form: $M = SJS^{-1}$, where J is a block diagonal matrix

$$J = \begin{pmatrix} J_1 & & \\ & \ddots & \\ & & J_p \end{pmatrix}$$

formed of Jordan blocks J_i ($i = 1, \ldots, p$), which are either scalars consisting of eigenvalues λ_i of M or have the form

$$J_i = \begin{pmatrix} \lambda_i & 1 & & \\ & \lambda_i & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_i \end{pmatrix}.$$

---

[18] Additionally, Equation (B.4) implies that for Cases 1 and 2, the formulas given in the main text for speed of convergence remain valid in the generalized setting of Proposition B.1.

Due to the network being strongly connected and M containing only non-negative entries, M is irreducible and Perron-Frobenius theory allows to infer that the spectral radius of M is a simple eigenvalue of M. We will assume without loss of generality that this value, $\lambda_1(M)$, corresponds to the matrix J_1. As $\lambda_1(M)$ is the spectral radius of M, we also know that $|\lambda_i| \leq \lambda_1(M)$ for all $i > 1$. From all this, by setting $\widetilde{\lambda}_i := \frac{1+\lambda_i}{1+\lambda_1(M)}$ we find that $\widetilde{M} := \frac{1}{1+\lambda_1(M)}(I + M) = S\widetilde{J}S^{-1}$, with

$$
\widetilde{J} = \begin{pmatrix} 1 & & & \\ & \widetilde{J}_2 & & \\ & & \ddots & \\ & & & \widetilde{J}_p \end{pmatrix}, \quad \widetilde{J}_i = \begin{pmatrix} \widetilde{\lambda}_i & 1 & & \\ & \widetilde{\lambda}_i & \ddots & \\ & & \ddots & 1 \\ & & & \widetilde{\lambda}_i \end{pmatrix} \quad (i = 2, \ldots, p).
$$

Additionally, we know that $\widetilde{\lambda}_i < 1$ due to $|\lambda_i| \leq \lambda_1(M)$ and λ_i being different from $\lambda_1(M)$. Taking all this together, we find that

$$
(1 + \lambda_1(M))^{-t} (I + M)^t = \widetilde{M}^t = S\widetilde{J}^t S^{-1} = S \begin{pmatrix} 1 & & & \\ & \widetilde{J}_2^t & & \\ & & \ddots & \\ & & & \widetilde{J}_p^t \end{pmatrix} S^{-1}.
$$

With respect to the terms $\widetilde{J}_i^t$, it is well known (and easy to prove) that for large t

$$
\widetilde{J}_i^t = \begin{pmatrix} \widetilde{\lambda}_i^t & \binom{t}{1}\widetilde{\lambda}_i^{t-1} & \binom{t}{2}\widetilde{\lambda}_i^{t-2} & \cdots \\ & \widetilde{\lambda}_i^t & \ddots & \vdots \\ & & \ddots & \binom{t}{1}\widetilde{\lambda}_i^{t-1} \\ & & & \widetilde{\lambda}_i^t \end{pmatrix},
$$

implying that all $\widetilde{J}_i^t$ shrink to 0 due to $|\widetilde{\lambda}_i| < 1$, entailing that the rate of convergence of $\widetilde{M}^t$ is essentially being determined by $\max\{\widetilde{\lambda}_i : i = 2, \ldots, p\}$.[19] For the limit of $\widetilde{M}^t$, we thus have: $\widetilde{M}^t \overset{t\to\infty}{\longrightarrow} Se_1 e_1^\top S^{-1}$. Setting $u := Se_1$ and $v := S^{-\top}e_1$, we can rewrite this as $\widetilde{M}^t \overset{t\to\infty}{\longrightarrow} uv^\top$. The following derivations show that u is a right eigenvector of M for $\lambda_1(M)$, while v is a corresponding left eigenvector:

$$
Mu = MSe_1 = SJS^{-1}Se_1 = SJe_1 = S\lambda_1(M)e_1 = \lambda_1(M)u,
$$

$$
v^\top M = \left(S^{-\top}e_1\right)^\top M = e_1^\top S^{-1}M = e_1^\top S^{-1}SJS^{-1} = e_1^\top JS^{-1} = \lambda_1(M)e_1^\top S^{-1}
$$

$$
= \lambda_1(M)\left(S^{-\top}e_1\right)^\top = \lambda_1(M)v^\top.
$$

---

[19]This implies that the formulas for speed of convergence given in the main text still apply to the generalized setting for Case 3.

As the left and right eigenvectors of M for $\lambda_1(M)$ are unique up to multiplying by a constant, $uv^\top$ and $cd^\top$ differ only by a constant: $uv^\top = \alpha cd^\top$ for some constant α. Now, from $uv^\top uv^\top = Se_1 e_1^\top S^{-1} Se_1 e_1^\top S^{-1} = Se_1 e_1^\top e_1 e_1^\top S^{-1} = Se_1 e_1^\top S^{-1} = uv^\top$, we find that $uv^\top uv^\top = \alpha cd^\top \alpha cd^\top$ must equal $uv^\top = \alpha cd^\top$, thus we have $\alpha^2 d^\top c = \alpha$, implying $\alpha = \frac{1}{d^\top c}$ and $\widetilde{M}^t \overset{t\to\infty}{\longrightarrow} \frac{cd^\top}{c^\top d}$, proving Equation (B.4) and concluding the proof.

B.3 Proof of Corollary 2

The proof of this corollary is perfectly analogous to the one of Corollary 1, building on Proposition B.1 instead of Proposition 2.

C Further Appendices

C.1 Additional Examples for the Benchmark Case of Symmetry

To illustrate occurrence of misinformation in the setting of symmetry, we study two extreme examples: A regular graph in Example C.1 and a network with a clique of five in Example C.2.

Example C.1 (Regular network)**.** *Consider network* (N, A) *that is connected and regular of degree* k, *i.e. every agent has exactly* k *links.*

Regularity of degree k *implies that the largest eigenvalue is* $\lambda_1 = k$ *and eigenvector centrality is* $c = (\frac{1}{n}, ..., \frac{1}{n})$. *From Proposition 1,* $\lim_{t\to\infty} x_i(t) = \frac{1}{n} \sum_{j=1}^{n} s_j$, *which is just the mean of the initial signals. This is a remarkable observation: Under symmetry and when the network is regular, the long-run signal mix of every agent exactly reflects the initial signal distribution.*

The only source of misinformation is hence that the initial draw of signals is "unlucky" (i.e. it happens to consist of many negative signals). For instance, let n *be odd. Then the expected fraction of misinformed agents is* $\sum_{r=0}^{\frac{n-1}{2}} \binom{n}{r} \rho^r (1-\rho)^{n-r}$, *which equals the probability that the minority of* n *independent signals is correct. To have concrete numerical examples, let the quality of each initial signal be* $\rho = 0.6$. *Then expected fraction of misinformed agents is* 0.267 *for* $n = 9$ *agents and* 0.022 *for* $n = 99$ *agents. Observe that the probability of misinformation in regular graphs goes to zero for growing* n.

Observe finally the comparison to Bayesian learners. Suppose for a moment that all agents are proper Bayesian learners in the following sense: they account for the repetition of signals and form their beliefs according to Bayes' rule using each independent signal
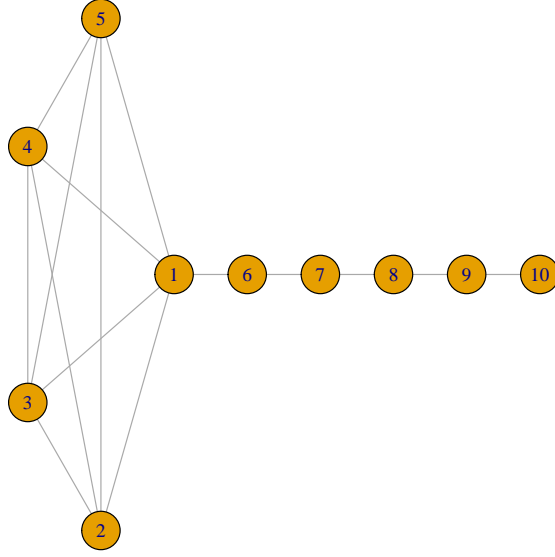
Figure C.1: A network with a clique of five and all other agents arranged in a line. Eigenvector centrality is:

$$c = (19.42\%, 18.41\%, 18.41\%, 18.41\%, 18.41\%, 5.12\%, 1.35\%, 0.36\%, 0.09\%, 0.02\%)$$

only once. In a connected network, these Bayesian learners will update until they have received each initial signal and then form their belief based on exactly the same signal mix as our much more naïve agents form in the long run when the network is regular (see, e.g., DeMarzo et al. (2003), Theorem 3).[20]

Example C.2 (Network with clique of five). *Consider the network* (N, A) *depicted in Figure C.1. This network consists of* $n = 10$ *agents. Five of them,* $1, ..., 5$, *form a clique, i.e. the network restricted to these agents is complete; the others are arranged in a line.*

The normalized eigenvector corresponding to the largest eigenvalue is

$$c = (19.42\%, 18.41\%, 18.41\%, 18.41\%, 18.41\%, 5.12\%, 1.35\%, 0.36\%, 0.09\%, 0.02\%)$$

The nodes are labeled according to their entry in this eigenvector with 1 *having the largest entry and* 10 *the lowest. Observe that the five members of the clique obtain the highest eigenvector entries. In fact, any three of their entries sum up to more than half of all entries. Hence, if it happens that at least three out of the five agents* $1, ..., 5$ *receive the wrong*

---

[20]There is a difference in the belief due to the signal difference $k_i(t)$, which grows in our model and is constant among Bayesian learners, but the signal mix as well as the best guess of each agent is identical in the two models.

signal, we have $\sum_{j=1}^{n} c_j s_j < 0.5$ and hence misinformation prevails (by Proposition 1). The
probability to have such a draw of signals and in fact the probability of misinformation
is $(1 - \rho)^5 + 5\rho(1 - \rho)^4 + 10\rho^2(1 - \rho)^3$, e.g. for $\rho = 0.6$, it is $Ef^{mis}(x(\infty)) = 0.31744$.
There are many more such networks (with the same expected level of misinformation) for
$n = 10$, but there is no network with higher probability of misinformation.

More importantly, we can construct networks with a clique of five and all others ar-
ranged in a line for all $n > 7$. The probability of misinformation is unchanged, as we
checked for n up to 1'000 by using programming language R. The eigenvector centralities
converge to $c_1 = 19.41919\%$ and $c_{2,...,5} = 18.40593\%$ for the members of the clique. Hence,
misinformation still happens when at least three out of these five receive the wrong signal.
Thus, we observe that as the number of nodes grow, the probability of misinformation need
not go to zero, as there are networks with a substantial probability of misinformation.

The example shows that a small group of people who are well-connected among them-
selves may have a disproportional large influence on the long-run signal mixes and hence
can be a cause for misinformation under symmetry. However, under symmetry the prob-
ability of misinformation is always bounded.

C.2   Different Interpretations for the Decay Factor

With respect to the decay factors, there are in fact three different interpretations, as we
will explain in more detail below. In a nutshell, these interpretations are (i) the sender
only shares part of her signals, (ii) the communication channel does not transmit 100%
of the signals, and (iii) the recipient discounts part of the received signals. In order to
improve readability, we will in this subsecion omit all '+' and '-' superscripts, thus A may
stand for A^+ and A^-, respectively, $N(t)$ may denote either the numbers of positive signals
$N^+(t)$ or that of negative signals $N^-(t)$, δ will be either δ^+ or δ^- and so on.

In order to showcase all the explanations given above for the existence of decay factors,
we might consider the following very general model: by $N_{(s)}(t)$, we denote the numbers of
signals that agents send out to their neighbours, and we write $N_{(s)}(t) = \delta_{(s)}N(t)$ to model
that agents do not communicate all their signals to their neighbors, with $\delta_{(s)} \in (0, 1]$
capturing the share of signals that agents are willing to transmit. We then denote by
$N_{(t)}(t)$ the numbers of signals that are transmitted between the agents, and by modeling
$N_{(t)}(t) = \delta_{(t)}AN_{(s)}(t)$, with $\delta_{(t)} \in (0, 1]$ describing the share of signals that are successfully
transmitted by the communication channel. Finally, we use $N_{(p)}(t)$ to denote the numbers
of signals that agents are actually processing when updating their signals from time t to
$t + 1$. Here, by setting $N_{(p)}(t) = \delta_{(p)}N_{(t)}(t)$, the discounting of received signals by agents

would be described by $\delta_{(p)} \in (0,1]$. Taken together and defining $\delta := \delta_{(p)}\delta_{(t)}\delta_{(s)}$, agents process

$$N_{(p)}(t) = \delta_{(p)}N_{(t)}(t) = \delta_{(p)}\delta_{(t)}AN_{(s)}(t) = \delta_{(p)}\delta_{(t)}A\delta_{(s)}N(t) = \delta_{(p)}\delta_{(t)}\delta_{(s)}AN(t) = \delta AN(t),$$
(C.1)

which is exactly the formula that we use in our main model. By doing so, we are able to model any of the three interpretations, by setting two of the three factors to 1 and allowing only one to be smaller than 1: e.g., setting $\delta_{(p)} = 1$, $\delta_{(t)} = 1$, and $\delta_{(s)} < 1$ leads to a model where the decay factor δ captures that agents share only some part of the signals they receive. Furthermore, our model also allows variations where two or even all three effects are at play.

If some of the above phenomena are no longer homogeneous across agents, but agent-specific, we might preserve the general structure, but replace the scalar quantities $\delta_{(p)}$, $\delta_{(t)}$, and $\delta_{(s)}$ by matrices $\Delta_{(p)}$, $\Delta_{(t)}$, and $\Delta_{(s)}$. In this case, the equation describing sharing only parts of available signals becomes $N_{(s)}(t) := \Delta_{(s)}N(t)$, and $\Delta_{(s)}$ will be a diagonal matrix which captures the agent-specific factors describing which share of their signals agents do actually share. Similarly, the generalized equation for the discounting of received signals becomes $N_{(p)}(t) = \Delta_{(p)}N_{(t)}(t)$, with the diagonal matrix $\Delta_{(p)}$ capturing the agent-specific factors used for ignoring some part of the signals transmitted to the agents. Finally, with the non-diagonal matrix $\Delta_{(t)}$ whose entries $\Delta_{(t)_{ij}}$ determine the share of signals that are successfully transmitted from agent j to agent i, the equation for the transmitted signals becomes $N_{(t)}(t) := (\Delta_{(t)} \circ A)N_{(s)}(t)$, with $\Delta_{(t)} \circ A$ denoting the Hadamard product of $\Delta_{(t)}$ and A. Equation (C.1) then generalizes to

$$N_{(p)}(t) = \Delta_{(p)}N_{(t)}(t) = \Delta_{(p)}\left(\Delta_{(t)} \circ A\right)N_{(s)}(t) = \Delta_{(p)}\left(\Delta_{(t)} \circ A\right)\Delta_{(s)}N(t) \qquad \text{(C.2)}$$

Due to properties of diagonal matrices and the Hadamard product, $\Delta_{(p)}\left(\Delta_{(t)} \circ A\right)\Delta_{(s)}$ turns out to be identical to $\left(\Delta_{(p)}\Delta_{(t)}\Delta_{(s)}\right) \circ A$, implying that we finally have $N_{(p)}(t) = (\Delta \circ A)N(t)$, with $\Delta := \Delta_{(p)}\Delta_{(t)}\Delta_{(s)}$, which amounts to the formula we use in our extended model. The components Δ_{ij} thus simultaneously capture agent i's possible discounting $(\Delta_{(p)_i})$, the communication between i and j not working properly $(\Delta_{(t)_{ij}})$, and agent j not sharing all signals $(\Delta_{(s)_j})$. Similar to above, our generalized model therefore also allows for one, two, or even all three of these effects being at play.

C.3   Examples for Special Dynamics

When an agent signal mix $x_i(t)$ converges to $\frac{1}{2}$, then beliefs might behave in a "nasty" way, as we illustrate in the examples below. For all examples, $b^0 = \frac{1}{2}$, $\rho = \frac{3}{5}$ and $\delta^+\lambda_1^+ = \delta^-\lambda_1^-$

(Case 3). In addition to the signal mix $x_i(t)$ and belief $b_i(t)$, we also report the best guess $g_i(t)$, which is the state of nature that agent i considers as most likely to be the true state at time t. When the agent is undecided (i.e. if $b_i(t) = \frac{1}{2}$), we define $g_i(t) = \frac{1}{2}$.

1. Beliefs and best guesses of all agents converge extremely fast to $\frac{1}{2}$:
   Let $A^+ = A^-$ be the complete network on $n = 4$ agents, $\delta^+ = \delta^- = 1$ (symmetry as in Section 4), and initial signal vector $s(0) = (0, 0, 1, 1)$. Then, for all $t > 0$:

$$N^+(t) = \frac{1}{2} 4^t \mathbb{1},$$
$$N^-(t) = \frac{1}{2} 4^t \mathbb{1},$$
$$k(t) = 0,$$
$$b(t) = \frac{1}{2} \mathbb{1},$$
$$g(t) = \frac{1}{2} \mathbb{1}.$$

2. All beliefs converge to $\frac{1}{2}$, best guesses are constant at 0 and 1 for two players each:
   Let $A^+ = A^-$ be the complete network on $n = 4$ agents, $\delta^+ = \delta^- = \frac{1}{2}$ (symmetry as in Section 4), and initial signal vector $s(0) = (0, 0, 1, 1)$. Then, for all $t > 0$:

$$N^+(t) = \frac{1}{2} \left(\frac{5}{2}\right)^t \mathbb{1} + \frac{1}{2} \left(\frac{1}{2}\right)^t (-1, -1, 1, 1)^\top,$$
$$N^-(t) = \frac{1}{2} \left(\frac{5}{2}\right)^t \mathbb{1} - \frac{1}{2} \left(\frac{1}{2}\right)^t (-1, -1, 1, 1)^\top,$$
$$k(t) = \left(\frac{1}{2}\right)^t (-1, -1, 1, 1)^\top \xrightarrow{t \to \infty} 0,$$
$$b(t) \xrightarrow{t \to \infty} \frac{1}{2} \mathbb{1},$$
$$g(t) = (0, 0, 1, 1)^\top.$$

3. Beliefs of all agents alternate, best guess of all agents, but one, alternate:
   Let $A^+ = A^-$ be the star network of $n = 5$ agents, $\delta^+ = \delta^- = 1$ (symmetry as in

Section 4), and initial signal vector $s(0) = (0, 0, 1, 1, 1)$. Then, for all $t > 0$:

$$N^+(t) = \frac{1}{8}3^t\,(6,3,3,3,3)^\top + \frac{1}{4}(0,-3,1,1,1)^\top + \frac{1}{8}(-1)^t\,(-6,3,3,3,3)^\top,$$

$$N^-(t) = \frac{1}{8}3^t\,(6,3,3,3,3)^\top - \frac{1}{4}(0,-3,1,1,1)^\top + \frac{1}{8}(-1)^t\,(2,-1,-1,-1,-1)^\top,$$

$$k(t) = \frac{1}{2}(0,-3,1,1,1)^\top + \frac{1}{2}(-1)^t\,(-2,1,1,1,1)^\top,$$

$$b(t) \text{ alternates between } \frac{1}{5}(2,2,3,3,3)^\top \text{ and } \left(\frac{3}{5},\frac{4}{13},\frac{1}{2},\frac{1}{2},\frac{1}{2}\right)^\top,$$

$$g(t) \text{ alternates between } (0,0,1,1,1)^\top \text{ and } \left(1,0,\frac{1}{2},\frac{1}{2},\frac{1}{2}\right)^\top.$$

Notice that $g_2(t) = 0$ is constant, $g_1(t)$ alternates between 0 and 1, while $g_3(t)$, $g_4(t)$, $g_5(t)$ alternate indefinitely between 1 and $\frac{1}{2}$.

4. Beliefs of all agents converge to $\frac{1}{2}$, best guesses converge for some agents, but alternate for other agents:

Let A^+ be a network on $n = 5$ agents with agents 1 and 2 being connected to everyone and no separate connections between agents 3, 4, and 5, A^- the complete network, $\delta^+ = \frac{4}{5}$, $\delta^- = \frac{3}{5}$ (network and decay asymmetry), and initial signal vector $s(0) = (0, 0, 1, 1, 1)$. Then, for all $t > 0$:

$$N^+(t) = \frac{1}{5}\left(\frac{17}{5}\right)^t (3,3,2,2,2)^\top + \frac{3}{5}\left(-\frac{3}{5}\right)^t (-1,-1,1,1,1)^\top,$$

$$N^-(t) = \frac{2}{5}\left(\frac{17}{5}\right)^t \mathbb{1}^\top + \frac{1}{5}\left(\frac{2}{5}\right)^t (3,3,-2,-2,-2)^\top,$$

$$k(t) = \frac{1}{5}\left(\frac{17}{5}\right)^t (1,1,0,0,0)^\top + \frac{3}{5}\left(-\frac{3}{5}\right)^t (-1,-1,1,1,1)^\top$$
$$\qquad - \frac{1}{5}\left(\frac{2}{5}\right)^t (3,3,-2,-2,-2)^\top \xrightarrow{t\to\infty} (\infty,\infty,0,0,0)^\top,$$

$$b(t) \xrightarrow{t\to\infty} \left(1,1,\frac{1}{2},\frac{1}{2},\frac{1}{2}\right)^\top,$$

$$g(t) \text{ asymptotically alternates between } \mathbb{1}^\top \text{ and } (1,1,0,0,0)^\top.$$

Notice that $g_1(t) = g_2(t)$ converge to 1, while $g_3(t) = g_4(t) = g_5(t)$ alternate indefinitely between 0 and 1.

C.4 Conditions for Case 3 to Approximate Case 1 and 2

The first condition is that parameters are reasonably close to case 3, i.e. $\delta^+\lambda_1^+$ being close to $\delta^-\lambda_1^-$. We can observe this in Example 1 in Figure 2, where values of δ^+ which

are close to the critical value of 0.4 induce similar dynamics. There is however a second condition, which happens to be satisfied in Example 1. For explaining that condition, let us reconsider Equation 7 and define $\tau^{\text{Case 3}} := \max\{\frac{|1+\delta^+\lambda_i^+|}{1+\delta^+\lambda_1^+}, \frac{|1+\delta^-\lambda_i^-|}{1+\delta^-\lambda_1^-}, i = 2, \ldots, n\}$, where λ_i^+ and λ_i^- denote all but the largest eigenvalues of A^+ and A^-, respectively. The second condition is that $\tau^{\text{Case 3}}$ is substantially smaller than $\frac{1+\min\{\delta^+\lambda_1^+,\delta^-\lambda_1^-\}}{1+\max\{\delta^+\lambda_1^+,\delta^-\lambda_1^-\}}$. Intuitively, the first condition assures that the importance of the largest eigenvalues persists sufficiently long (they always matter in Case 3, but vanish in the long run of the two others) and the second condition assures that the importance of all other eigenvalues vanishes sufficiently fast. As a rule of thumb, when slow convergence occurs in Cases 1 or 2, resulting in large values for actual half-life $t_{1/2} = \log(0.5)\big/\log\left(\frac{1+\min\{\delta^+\lambda_1^+,\delta^-\lambda_1^-\}}{1+\max\{\delta^+\lambda_1^+,\delta^-\lambda_1^-\}}\right)$, and at the same time the half-life formula for Case 3, $\frac{\log(0.5)}{\log(\tau^{\text{Case 3}})}$ produces a much lower "pseudo half-life," then the formula given for Case 3 of Proposition 2 may provide a better approximation for the relevant misinformation in the short or medium term than the actual long-term limit of either 0 or 1.

C.5   Comparison with DeGroot Model

In the following, we will discuss the similarities and differences of our model as compared to the classical DeGroot model, with respect to model set-up, conditions for convergence, reaching a consensus, and eigenvector centrality.

With respect to the set-up of our model, the most striking difference to the DeGroot model is that we separate the evolution of positive and of negative signals, while in the DeGroot model this distinction is not possible. However, the evolution of each type of signals (positive or negative) resembles the DeGroot model, in the sense that values at time $t+1$ are linear functions of values at time t. While the weights of these regression-type recursions are restricted to be non-negative in our model as well in the DeGroot model, our model for the signals' evolution does not require convex combinations, i.e. the weights do not have to sum up to unity, in contrast to the DeGroot case. For the special case of symmetry in the sense that $A^+ = A^- =: A$ as well as $\delta^+ = \delta^- =: \delta$ and denoting $I + \delta A$ by W, it is easy to see that $N^+(t) = W^t s$, $N^-(t) = W^t(\mathbb{1} - s)$, and $N(t) = W^t\mathbb{1}$, implying

$$x_i(t) = \frac{e_i^\top W^t s}{e_i^\top W^t \mathbb{1}}, \tag{C.3}$$

with e_i denoting the i-th unit vector. Furthermore, Equation (C.3) implies that the updating of the signal mixes $x(t)$ in our model may be written as

$$x(t) = \widetilde{W}(t)x(t-1) \tag{C.4}$$

48

with the row-stochastic matrices $\widetilde{W}(t)$ having entries $\widetilde{w}(t)_{ij} = w_{ij} \frac{\kappa_j(t-1)}{\kappa_i(t)}$, with w_{ij} denoting the entries of $W = I + \delta A$ and $\kappa(t) := W^t \mathbb{1}$.[21] Equation (C.4) thus is a representation of the updating process in our model as a generalized DeGroot model, where in general the updating matrices $\widetilde{W}(t)$ are not constant, but change over time. Furthermore, if the networks described by A are regular, then these updating matrices will in fact not depend on time, and the updating formula (C.4) will actually become a DeGroot model, in complete analogy to a special case of the model considered by Sikder et al. (2020).

With regard to conditions ensuring convergence, it is well-known that values converge in the DeGroot model if and only if every set of nodes is strongly connected and closure is also aperiodic. This is in fact very similar to our model, where we receive convergence by assuming that the whole society is strongly connected and by observing that the matrices W^+ and W^- are aperiodic. Aperiodicity of the matrices $W^+ = I + \delta^+ A^+$ and $W^- = I + \delta^- A^-$ is guaranteed, as they both involve the identity matrix I.

For the DeGroot model, there emerges a consensus whenever the model converges, where the consensus is a convex combination of the initial values. This is different for our model, which in contrast to the DeGroot model, allows for the signal mixes to converge to the truth (1) or to complete misinformation (0), depending on the networks' eigenvalues and the decay parameters. In addition, in our model, signal mixes will converge to consensus if $\delta^+ \lambda^+ = \delta^- \lambda^-$ and $\frac{c_i^-}{c_i^+}$ does not depend on i: the latter condition is equivalent to $c^- = c^+$, i.e. in our main as well as in the extended model (Section 5.4), the networks must have identical centralities. This actually can happen even though A^+ and A^- are different, with an example being the case of both networks being regular, but of different degrees.

With respect to eigenvector centrality in DeGroot models, the left-hand eigenvector determines the weights for the asymptotically emerging consensus (Jackson, 2010; Golub and Sadler, 2016). In our model, however, centralities only play this role when the relation $\delta^+ \lambda_1^+ = \delta^- \lambda_1^-$ holds (i.e. in case 3). In this case, eigenvector centralities play two roles (in our main model): they influence the long-run signal mix of all agents commonly through $\frac{1 - \sum_{j=1}^n c_j^- s_j}{\sum_{j=1}^n c_j^+ s_j}$, reflecting the initial signals' impact, and they influence the agents' individual long-run signal mixes through the centrality ratios $\frac{c_i^-}{c_i^+}$. In our extended model, the former role (the part common to all agents) is played by the left-hand eigenvectors d^+ and d^-, while the latter role (the individual part) is still played by the (right-hand) eigenvector centralities c^+ and c^-.

---

[21]A similar representation of $x(t)$ also holds for the extended model, it can be developed by simply replacing all appearances of δA by M.

49