# Confirmation Bias in Social Networks

Working Paper 2019-05

MARCOS FERNANDES

November, 2019

Stony Brook University

# CONFIRMATION BIAS IN SOCIAL NETWORKS

MARCOS FERNANDES

*This version: November, 2019*

Abstract. I propose a social learning model that investigates how confirmatory bias affects public opinion when agents exchange information over a social network. For that, besides exchanging opinions with friends, individuals observe a public sequence of potentially ambiguous signals and they interpret it according to a rule that accounts for confirmation bias. I first show that, regardless the level of ambiguity and both in the case of a single individual or of a networked society, only two types of opinions might be formed and both are biased. One opinion type, however, is necessarily less biased (more efficient) than the other depending on the state of the world. The size of both biases depends on the ambiguity level and the relative magnitude of the state and confirmatory biases. In this context, long-run learning is not attained even when individuals interpret ambiguity impartially. Finally, since it is not trivial to ascertain analytically the probability of emergence of the efficient consensus when individuals are connected through a social network and have different priors, I use simulations to analyze its determinants. Three main results derived from this exercise are that, in expected terms, i) some network topologies are more conducive to consensus efficiency, ii) some degree of partisanship enhances consensus efficiency even under confirmatory bias and iii) open-mindedness, i.e. when partisans agree to exchange opinions with other partisans with polar opposite beliefs, might harm efficiency in some cases.

*JEL Classification:* C11, D83.

*Keywords:* Social Networks, Social Learning, Misinformation, Confirmation Bias, Folly of crowds.

---

*E-mail address*: marcos.fernandes@stonybrook.edu.

I thank Sandro Brusco, Pradeep Dubey, Ben Golub, Laura Karpuska, Ehud Lehrer, Ting Liu, Mihai Manea, Alejandro Melo, Ron Peretz, Max Silva, Eilon Solan, Troy Tassier and Yair Tauman for helpful comments and suggestions. I also thank all participants of the 29th International Conference on Game Theory (Stony Brook), 23rd Coalition Theory Network workshop (Maastricht), 4th Annual Conference on Network Science and Economics (Nashville), 44th Eastern Economic Association conference (Boston) and the seminar participants at UFABC, Fordham University and Stony Brook Center for Behavioral Political Economy. I gratefully acknowledge the Department of Economics of Stony Brook University (specially Juan Carlos Conesa), Stony Brook Graduate Students Organization and The Network Science in Economics conference series (specially Myrna Wooders) for financial support.

## 1. Introduction

Individuals form opinions over a myriad of economic, political and social issues based on information they get from both media and trustworthy acquaintances such as friends, coworkers, professors, family members, etc. This process of information acquisition usually takes place when the issue being discussed has no clear-cut *right/wrong* or *true/false* distinction or when the set of information available is not easily or readily understood by individuals. In this case, consulting friends opinions has its appeal since it is an easy way to gather information. For that, social networks appear as a primary tool for many people to get informed and debate their world views. In view of this, it is important to understand how beliefs depend on the way agents perceive and process the information and on the social network structure. With this regard, I examine one potential aspect of learning in social networks: how public opinion is affected by confirmation bias?

Confirmation bias, as the term is typically employed in the psychology literature, connotes the interpretation of evidence in ways that are in line with existing beliefs. In this sense, an individual is said to suffer from confirmatory bias if he tends to interpret ambiguous evidence as confirming his current belief. This can be done in different ways, like restricting attention to favored hypothesis, disregarding evidence that could falsify the current world view or overweighting positive confirmatory instances, etc. In all cases, individuals restrict attention to a single hypothesis and, for that, fail to give appropriate consideration to alternative hypotheses and this process creates a friction in belief formation.

As per the social psychology literature, individuals show the tendency to interpret ambiguous evidences as supporting their initial impressions. For instance, banks and companies may misinterpret Central Banks' stance toward high inflation after ambiguous statements, professors may misinterpret the quality of students after ambiguous performance, people may misinterpret scientists after ambiguous announcements, etc. In this view, while relying on friends might help individuals to aggregate the information in some cases, in others it might lead individuals to expose themselves to other individuals that rely on their own world view to derive information from ambiguous evidence. In these cases, efficient aggregation of information is not guaranteed and I investigate how social opinion is affected by that.

To analyze such phenomenon, I consider a society where agents are interested in learning some underlying state $\theta \in \Theta = [0, 1]$. For example, this underlying state $\theta$ might represent the degree (from 0 to 1, say) in which the anthropic activity causes global warming. All agents have a initial prior belief about it and observe a sequence of public signals, one at each date $t$. Signals are either i) informative, ii) uninformative or iii) ambiguous. Signals in the class i) are simply

binary variables that indicate 1 if right states are more likely and 0 if left states are more likely. Therefore, even though there is no noise in the signal, agents can only learn the state (the right proportion of 1's and 0's) asymptotically. Signals in the class ii) are simply disregarded for not being informative and, for that, prior beliefs are kept unchanged. Finally, signals in the class iii), the ambiguous one, are open to interpretation. In this case, I allow agents to interpret them using a fairly general randomization rule proposed by Fryer, Harms, and Jackson (2018) that accounts for confirmation bias. As per this rule, the interpretation of the ambiguous signal received at time $t$ is influenced, to a greater or lesser extent, by the likelihood of 0 and 1 at time $t-1$ (more details below). Ambiguity in this context has nothing to do with the noise of a signal in standard information theory or with the classical notion of ambiguity aversion as defined by Ellsberg (1961). Instead, is just a way of capturing the situations in which people feel compelled to give meaning to ambiguous evidences about the issue or subject analyzed.

As in Jadbabaie, Molavi, Sandroni, and Tahbaz-Salehi (2012), information exchange between agents result from multilateral communication. Before the beginning of each period, all agents meet their social media friends and observe their opinions and precisions about the subject of interest. At the beginning of every period $t$, the public signal is realized. Thus, each agent first *interprets* ambiguous signals (if the case) using the randomization rule, *stores* it and *computes* his Bayesian posterior opinion and precision. After that, every agent sets his *final* opinions and precisions to be a linear combination of the Bayesian posterior opinion and precision computed with the interpreted signal and the opinions and precisions of neighbors met in the period before. The social connectivity among agents is fixed over time and strong connectivity is assumed, i.e. all agents are exposed to all the other agents either through a directed or undirected path in the social networks.

I show that, regardless the level of ambiguity and both in the case of a single individual or a connected society, only two types of opinions can emerge and both are biased, one left-biased and the other right-biased. One opinion type, however, is more efficient (less biased) than the other depending on the magnitude of the state. Opinion efficiency, in this case, is only guaranteed under a "favorable" combination of "low" ambiguity and "sufficiently pronounced" state. If this condition holds, I show that the efficient consensus is attained with probability 1, otherwise, efficient consensus is reached with some probability. Moreover, long-run learning is not attained even if individuals are impartial when interpreting ambiguous signals. Those results contrast with some results presented by Rabin and Schrag (1999) and Fryer et al. (2018), where long-run learning takes place with a positive probability and impartiality helps learning the state. Furthermore, the network effect presented here, together with signals realizations, reinforces the interpreting

"*tug-of-war*" since individuals might have their own biases confirmed (or mitigated) by other agents.

Finally, since it is not trivial to derive analytically the probability of emergence of the most efficient (less biased) consensus, I use graphs simulations to show its determinants. I show that the presence of partisan agents in societies who suffer from confirmatory bias has a double effect on the expected consensus efficiency: i) it helps to countervail the misinterpretation of initial signals when there degree of partisanship is low and for that it increases expected efficiency; and ii) exacerbates misinterpretation of signals when the degree of partisanship is high, reducing expected consensus efficiency. Moreover, I also show that open-mindedness of partisan agents, i.e. when partisans agree to exchange opinions with partisans with polar opposite beliefs, might reduce expected consensus efficiency in some social topologies.

This work, even though it does not generalize results for other conjugate families, seems sufficiently general to capture some relevant real-world situations. The public signals realized every period and observed by all agents might represent the information reported by media outlets, such as TV channels, Radio, Youtube, Twitter, etc. The level of ambiguity of the information content reported by such outlets, measured by a parameter $\mu$, represents the fraction of instances where a signal conveys two polar opposite meanings at the same time and agents feel impelled to interpret them. In this regard, agents can be more or less biased when interpreting signals. They can be biased and conform the interpretation with their prior to some extent, be impartial and choose an interpretation uniformly at random (say 0 or 1'with the same probability $\frac{1}{2}$ each) or even go against their world-view. The interpretation behavior of every agent is dictated by the parameters of the signal interpretation function.

The structure of this work is as follows. Section 2 provides a brief literature review. Section 3 describes a framework for updating beliefs when agents communicate over a social network and evidences (signals) are potentially open to interpretation and present the main theoretical results. Section 4 describes a simulation exercise when there is priors heterogeneity. Section 5 concludes. There are four appendices. Appendices A and B contain the primitives of the Beta-Bernoulli conjugate family employed in this work. Appendix C contains the proofs of auxiliary results, whereas Appendix D presents the proofs main results.

## 2. Literature review and contribution

A great deal of empirical evidence on social psychology supports the idea that the confirmation bias is extensive and that it appears in many ways. Nickerson (1998) argues that most studies in the field confirm the human tendency of casting doubt on information that conflicts with preexisting

beliefs and to be more likely to see ambiguous information to be confirming of preexisting beliefs. This selectivity in the acquisition and use of evidence, however, takes place without intending to treat evidence in a biased way or even being aware of doing so. Molden and Higgins (2004, 2008) identify that both *vagueness* (when the evidence is weak) and *ambiguity* (when the evidence is conflicting) influence the interpretation of an uncertain evidence, whereas Furnham and Ribchester (1995) and Furnham and Marks (2013) find evidences that the way individuals perceive and process information about ambiguous situations is related to their degree of ambiguity tolerance.[1]

In this context, confirmation bias can be seen as an information process that departs from standard Bayesian updating because agents scrutinize ambiguous signals in line with their world views. Some examples of decision-making models that account for Bayesian updating deviation are Hellman and Cover (1970), Rabin and Schrag (1999), Wilson (2014) and Fryer et al. (2018). In Rabin and Schrag (1999), for instance, signals believed to be less likely are misinterpreted with an exogenous probability, whereas in Fryer et al. (2018), in its simplest version with binary states, ambiguous signals are produced with certain probability and agents interpret those before performing the Bayesian update. To interpret such signals individuals employ three methods that simply differ in the intensity with which agents conform their interpretation with their current world-view.

In this regard, Rabin and Schrag (1999) and Fryer et al. (2018) are the closest references to this work, both in spirit and results. In this work, however, I move away from the binary state space case and allow states to be continuously distributed over the unit interval according to a Beta distribution. Binary signals are drawn from a Bernoulli distribution, ambiguous signals appear with some positive probability and I allow agents to use the interpretation strategies proposed in Fryer et al. (2018). Two important implication of such modeling strategy are that an impartial interpretation strategy (flipping a fair coin to interpret ambiguous signals, for instance) is not sufficient to overcome confirmatory bias and that long-run learning is not attained even when ambiguity level is "sufficiently" low. Moreover, I introduce a network structure among agents and allow them to set their final beliefs to be a linear combination of the Bayesian posterior and the opinions of their neighbors as in Jadbabaie et al. (2012). In this case, since there are network externalities, is not immediately clear how opinions will evolve as interpretations are influenced by both the realization of signals and the confirmatory biases of friends. One important implication of network externalities is that some network topologies induce less biased consensus when there is heterogeneity of initial priors.

[1]More recently the concept of tolerance of ambiguity has been conceived by part of scholars to reflect the contemporary definition of ambiguity proposed by Ellsberg (1961). For a good coverage of the classical literature on ambiguity aversion, see Gilboa and Schmeidler (1989) and Gilboa and Schmeidler (1993), Epstein and Schneider (2007).

This work is also related to the literature of *biased assimilation* in networks. In a nutshell, this literature focus on models of social learning in which agents have tendency to overweight the opinion of friends with similar beliefs. Some examples are Hegselmann and Krause (2002), Hegselmann and Krause (2005), Dandekar, Goel, and Lee (2013) and Mao, Bolouki, and Akyol (2018). While bias assimilation has to do with the tendency to *conform* with the majority or leading individuals, confirmatory bias, as argued above, is some sort of failure in the Bayesian updating process. From this perspective, modeling confirmatory bias as either a biased assimilation or a failure in the Bayesian update has different consequences. On the one hand, bias assimilation presumes that the connections between agents are broken (or temporarily interrupted) according to how "far" opinions are and, therefore, it implies in non trivial changes in the topology of the network. Thus, a natural result found in this literature is that long run polarization takes place when there is biased assimilation. Thus, polarization is a natural product of the initial heterophily of opinions in the system and the eventual permanent deletion of some links. On the other hand, modeling confirmatory bias as a Bayesian update failure, like this work, is inconsequential to the network topology and under the strong connectivity assumption leads to a bias (misinformation) that can be analytically studied.

Finally, there exists a great deal of works on *social learning*, both assuming bounded and fully rationality. The Bayesian social learning literature (fully rational agents) mainly focuses on formulating stylized games with incomplete information and characterizing its equilibria. More specifically, rather than considering complex and repeated interactions, most part of the works focuses on environments where agents are myopic or interact only once. Some works of reference are Banerjee (1992), Bala and Goyal (1998), Bala and Goyal (2001), Banerjee and Fudenberg (2004), Acemoglu, Dahleh, Lobel, and Ozdaglar (2011).[2]

On the other hand, the non-bayesian learning (bounded rational agents) literature focus on studying generalizations or departures of the seminal DeGroot (1974) model. For instance, DeMarzo, Vayanos, and Zwiebel (2003) show that the classical consensus result does not rely on the social weighting matrix being a stationary matrix, Acemoglu, Ozdaglar, and ParandehGheibi (2010) consider a random meeting (Poisson) model and characterize how the presence of forceful agents, i.e. agents who influence others disproportionately and hardly revise their beliefs, interferes with information aggregation, whereas Golub and Jackson (2010) show that convergence holds if (and only if) the influence of the most influential agent vanishes as the society grows unboundedly. Jadbabaie et al. (2012) is the first work to consider the possibility of constant arrival of informative

[2]For an overview of recent research on belief and opinion dynamics in social networks, see Acemoglu and Ozdaglar (2011).

signals every period of time in networked environments. The novelty in the paper is that the update rule that sets the final belief to be a linear combination of the Bayesian posterior and the opinions of her neighbors is an efficient alternative to the complicated task of implementing Bayesian update in networks. Lastly, Azzimonti and Fernandes (2018), similar to this work in modeling strategy, investigate how the structure of social networks and the presence of fake news affect the degree of polarization and misinformation. The two major differences with respect to this paper are that i) their model consider the presence of stubborn agents called *Internet bots* whose sole purpose is to deceive other agents, whereas the main source of bias in my model derives from confirmatory bias; and ii) that the connectivity among all agents evolves stochastically, whereas it is fixed in this paper. Those two features together are the main drivers of misinformation and polarization cycles in a dynamic system that does not reach convergence, whereas my model focus on understanding how misinformation depends on both the structure of the network and the way agents interpret ambiguous signals.

## 3. The model

**Notation:** All vectors are viewed as column vectors, unless stated otherwise. Given a vector $v \in \mathbb{R}^n$, I denote by $v_i$ its $i$-th entry. When $v_i \geq 0$ for all entries, I write $v \geq 0$. Moreover, I define $v^\top$ as the transpose of the vector $x$ and for that, the inner (scalar) product of two vectors $x, y \in \mathbb{R}^n$ is denoted by $x^\top y$. I denote by $\mathbf{1}$ the vector with all entries equal to 1. A matrix $W$ is said to have size $m \times n$ whenever $W$ has exactly $m$ rows and $n$ columns. Moreover, whenever $m = n$, $W$ is called a square matrix of size $n$. The identity matrix of size $n$ is denoted by $\mathbb{I}$. For a matrix $W$, I write $W_{ij}$ to denote the entry in the $i$-th row and $j$-th column. The notation $W_{ij}^k$ is used to denote the entry in the $i$-th row and $j$-th column of the matrix $W^k$, i.e. the matrix $W$ raised to the power $k$. Finally, a vector $v$ is said to be a stochastic vector when $v \geq 0$ and $\sum_i v_i = 1$. A square matrix $W$ is said to be a (row) stochastic matrix when each row of $W$ is a stochastic vector.

### 3.1. Network structure.

The connectivity among agents in a network is described by a directed graph $G = (N, g)$, where $N = \{1, 2, \ldots, n\}$ is the set of agents, fixed over time, and $g$ is a real-valued $n \times n$ adjacency (or incidence) matrix, also fixed over time. Each element $g_{ij}$ in the directed-graph represents the connection between agents $i$ and $j$. More precisely, $g_{ij} = 1$ if individual $i$ is paying attention to (e.g. receiving information from) individual $j$, and 0 otherwise. Since the graph is directed, it is possible that some agents pay attention to others who are not necessarily reciprocating, i.e. $g_{ij} \neq g_{ji}$. The out-neighborhood of any agent $i$ represents the set of agents that $i$ is receiving information from (e.g. $i$'s references), and is denoted by $N_i^{out}(g) = \{j : g_{ij} = 1\}$.

Similarly, the in-neighborhood of any agent $i$, denoted by $N_i^{in}(g)$, represents the set of agents that are receiving information from $i$ (i.e. $i$'s followers), $N_i^{in}(g) = \{j : g_{ji} = 1\}$. We define a directed path in $G$ from agent $i$ to agent $j$ as a sequence of agents starting with $i$ and ending with $j$ such that each agent is a neighbor of the next agent in the sequence. We say that a social network is *strongly connected* if there exists a directed path from each agent to any other agent.

3.2. **Initial beliefs and signals.** Let $\Theta = [0, 1]$ to denote the set of possible states of the world. For instance, one may find useful to interpret $\Theta$ as the degree to which the anthropic activity might cause global warming, such that a state close to 0 means that human activity has no impact on global warming, whereas a state close to 1 means that human activity is fully responsible for the global warming. I assume that each agent $i$ in this society starts with an initial belief about an underlying state $f_{i,0}(\theta) \in \Delta\Theta$, represented by a Beta probability distribution over the set $\Theta$ with shape parameters $\alpha_{i,0}, \beta_{i,0} > 0$. Given prior beliefs, I denote by *opinion* of agent $i$ at time $t$ by $y_{i,t}$

$$y_{i,t} = \mathbb{E}[\theta] = \frac{\alpha_{i,t}}{\alpha_{i,t} + \beta_{i,t}}.^{3}$$

Conditional on the state of the world $\theta$, every agent observes a sequence of public signals $s_t$, one at each date $t \in \{1, 2, \dots\}$. Public signals lie in the set $S = \{1, 0, a, \emptyset\}$. As per the example of global warming given above, a signal 1 is evidence that human activity is the main responsible for global warming, a signal 0 is evidence on the contrary (no responsibility), the signal $\emptyset$ contains no information and a signal $a$ is *ambiguous* and open to idiosyncratic interpretation (section 3.3 explains how agents deal with those signals).

Signals are independent over time, conditional on the state. With probability $\delta \in D = [0, 1]$, independent of the state, the no informational signal $\emptyset$ is observed and with probability $(1 - \delta)$ some signal is observed. Conditional on observing a signal, the probability that the new signal is ambiguous is $\mu \in M = (0, 1]$. In this case, the signal conveys informational aspects that could lead one to interpret as either 1 or 0. With the remaining probability $1 - \mu$ the information provided by the signal is clear. In any state $\theta \in \Theta$, the probability that an unambiguous signal is 1 is $\theta \in (0, 1)$ and 0 with probability $1 - \theta$. The signal structure is depicted in the Figure 1.

[3]See Appendix A for the primitives of the Beta distribution. For tractability, the opinion is intended to be a real number that summarizes "well" the whole belief. For that, one can understand the opinion of an agent as the Bayesian estimator of $\theta$ that minimizes some sort of mean squared error or absolute error. Since the mean, mode and median of the beta distribution are asymptotically equivalent, as shown in Appendix B, the functional form is believed to be irrelevant for the results.

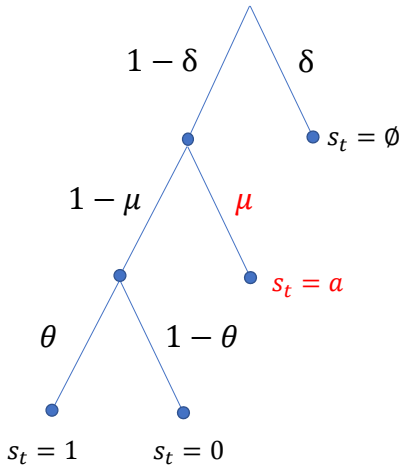



FIGURE 1. Signal structure

3.3. **Interpretation of ambiguous signals.** Although ambiguous signals are as uninformative about the state as the signal $\emptyset$ and, therefore, should also be disregarded from a pure Bayesian point of view, agents are constrained to make interpretations of ambiguous signals. This constraint captures the idea that, in some instances, people react to ambiguous pieces of information. They fail to perceive the unformativeness of such signal and end up using their prior world view to derive information from such signals.

Furthermore, one may also think that people might interpret ambiguous information in different ways. A potential one is to use their prior assessment of the subject to categorize the ambiguous signal, i.e. the interpretation of an ambiguous signal $s_t = a$ as 0 or 1 could depend on how the agent perceives the state, i.e. if $\theta = 0$ is more likely than $\theta = 1$ (likelihood ratio greater than one), then the agent could be prone to store $s_t = a$ as 0 and vice-versa. Conversely, the agent could interpret $s_t = a$ as 0 or 1 depending on the *mode* of his belief, i.e. if the mode is greater (less) than 0.5, then agent stores the ambiguous signal as 1 (as 0). In this sense, the interpretation depends on the intensity of the confirmatory bias intensity.[4]

For the interpretation of ambiguous signals, I use a randomization rule proposed in Fryer et al. (2018), adapted here for some technical idiosyncrasies, which says that with probability $\gamma_{i,t} \in [0, 1]$ the agent $i$ conforms with his posterior at time $t$ and with probability $1 - \gamma_{i,t}$ goes against it. In other words, with probability

$$\psi_{i,t} = \gamma_{i,t} \, \mathbb{1}\{y_{i,t-1} \geq 0.5\} + (1 - \gamma_{i,t}) \, \mathbb{1}\{y_{i,t-1} < 0.5\} \tag{1}$$

[4]In this paper, I move away from the decision-theoretic aspects of such problem and assume this randomization rule as given.

agent $i$ interprets the ambiguous signals as 1 and with the remaining probability $(1 - \psi_{i,t})$ interprets the ambiguous signals as 0 at time $t$.[5]

Therefore, the parameter $\gamma_{i,t}$ represents the intensity of the confirmatory bias experienced by any individual $i$ at any time $t$ and its distribution over time can be very general. I only assume $\gamma_{i,t}$ to be independent of opinion $y_{i,t}$ for any $i \in N$, history of opinions of all individuals and all other parameters in this model. From this randomization rule, there are three definitions of interest.

**Definition 1.** *An individual $i \in N$*

(1) *has **confirmatory tendency** if $\frac{1}{2} < \gamma_{i,t} \leq 1$ for all $t$,*
(2) *is **biased** if $\gamma_{i,t} = 1$ for all $t$.*
(3) *is **impartial** if $\bar{\gamma}_i = \mathbb{E}_t[\gamma_{i,t}] = \frac{1}{2}$. Two distinctions apply:*
   - ***always impartial** if impartial and $\gamma_{i,t} = \frac{1}{2}$ for all $t$,*
   - ***moderately impartial** if impartial and $\gamma_{i,k} \neq \frac{1}{2}$ for some $k$.*

That said, the signal interpretation functions, $s_{i,t}^{(0)}$ and $s_{i,t}^{(1)}$, for each individual at any point in time can be generally defined as

$$s_{i,t}^{(0)} = \mathbb{1}\{s_t = 0\} + \mathbb{1}\{s_t = a\}\mathbb{1}\{u_t > \psi_{i,t}\} \tag{2}$$

$$s_{i,t}^{(1)} = \mathbb{1}\{s_t = 1\} + \mathbb{1}\{s_t = a\}\mathbb{1}\{u_t \leq \psi_{i,t}\}, \tag{3}$$

where $\psi_{i,t}$ is as defined in Equation (1), $s_t$ is the publicly observed signal and $u_t$ is the realization of a continuous $U[0, 1]$ random variable at time $t$ simply used to break the tie. The draws $\{u_t\}$ are independent across time and also independent of all other random variables in this model. In words, the signal interpretation functions are basically transforming the observed signals $\{s_t\}_{t=1}^{\infty}$ into binary interpretations. When the realized public signal is $s_t = 1$ ($s_t = 0$), all agents undoubtedly interpret it as 1 (as 0) and set $s_{i,t}^{(0)} = 0$ and $s_{i,t}^{(1)} = 1$ (set $s_{i,t}^{(0)} = 1$ and $s_{i,t}^{(1)} = 0$). However, when the realized public signal is ambiguous, i.e. $s_t = a$, agents use their their prior information (summarized by $y_{i,t-1}$) to categorize the signal as either 0 or 1, as per Equation (1). For a more detailed description of the signals likelihood function, see Appendix A.

3.4. **Belief evolution.** We assume that agents update their beliefs based on public signals $s_t \in \{1, 0, a, \emptyset\}$ and on the influence of friends in their social clique. Before the beginning of each

[5]From Appendix B, notice that since mean and mode of the Beta distribution are very close for different compositions of parameters $(\alpha, \beta)$ and are also asymptotically equivalent, Equation (1) uses $y_{i,t-1}$ (mean and note mode) to interpret public signals. I believe this is neutral to all results even though I have not checked it.

period, agent $i$ meets individuals in his neighborhood $N_i^{out}(g)$. These neighbors share their world-views, summarized by $\alpha_{j,t}$ and $\beta_{j,t}$ for all $j \in N_i^{out}(g_t)$.[6]

At the beginning of period $t$, a signal profile is realized and the signal $s_{i,t}$ is privately observed by agent $i$. After observing the public signal $s_t$, agent $i$ computes his posterior in a standard Bayesian fashion. Following Jadbabaie et al. (2012), I assume that the final parameters $\alpha$ and $\beta$ will be a convex combination between the parameters $\alpha$ and $\beta$ of his Bayesian posterior and the weighted average of the his neighbors parameters.[7]

In mathematical terms, the update rule is as follows

$$\alpha_{i,t+1} = b\left[\alpha_{i,t} + s_{i,t+1}^{(1)}\right] + (1-b)\sum_j \hat{g}_{ij}\alpha_{j,t} \tag{4}$$

$$\beta_{i,t+1} = b\left[\beta_i + s_{i,t+1}^{(0)}\right] + (1-b)\sum_j \hat{g}_{ij}\beta_{j,t}. \tag{5}$$

Notice that when $b = 1$, agents fully rely on the signals and behave like a standard Bayesian agent. As $b$ approaches zero, agents are more influenced by the network, as more weight is given to his neighbors' opinions. Moreover, let $\alpha_t = (\alpha_{1,t}, \alpha_{2,t}, \ldots, \alpha_{n,t})^\top$ and $\beta_t = (\beta_{1,t}, \beta_{2,t}, \ldots, \beta_{n,t})^\top$ denote the column vectors of length $n$ of agents beliefs parameters at time $t$, $\mathbb{I}$ be an identity matrix of dimension $n$ and $B = \text{diag}(b, b, \ldots, b)$ be the diagonal Bayesian (or self-reliance) matrix. We can rewrite equation (4) as

$$\begin{aligned}\alpha_{t+1} &= B(\alpha_t + s_{t+1}^{(1)}) + (\mathbb{I} - B)\hat{g}\alpha_t \\ &= \left(B + (\mathbb{I} - B)\hat{g}\right)\alpha_t + Bs_{t+1}^{(1)} \\ &= W\alpha_t + Bs_{t+1}^{(1)},\end{aligned} \tag{6}$$

and equation (5) as

$$\beta_{t+1} = W\beta_t + Bs_{t+1}^{(0)}, \tag{7}$$

[6]Moreover, it is assumed they do that in such a way that the final posterior remains in the same conjugate family as the prior. i.e. since the initial prior is represented by a Beta distribution, I will assume the posterior will always be a Beta distribution. This is done by assuming that agents share the real-valued parameters rather than sharing the whole belief (distribution). This assumption is neutral to all results and asymptotically equivalent to the case in which agents update their beliefs as a linear combination of Beta distributions. The benefit of doing it is that the both algebra and intuition get clearer. Moreover, in the spirit of the bounded rational assumption, it is arguable that agents find easier to handle the mental computation involved in this process when dealing with real numbers than with the whole distribution.

[7]One may also think of agents sharing opinions (mean) and precisions (variance) with each other rather than sharing distribution parameters. Those are equivalent modeling strategies, we only need to use the relationships $y = \frac{\alpha}{\alpha+\beta}$ and $\sigma^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$ to fully determine $\alpha$ and $\beta$. Algebraic manipulation yields $\alpha = -\frac{y(\sigma^2+y^2-y)}{\sigma^2}$ and $\beta = \frac{(\sigma^2+y^2-y)(y-1)}{\sigma^2}$.

where, $W = B + (\mathbb{I} - B)\hat{g}$ is a homogeneous row-stochastic matrix. Notice that since the graph $G$ induced by the Adjacency matrix $g$ is assumed to be strongly connected, the graph induced by $W$ is trivially strongly connected as well.

## 4. Theoretical results

Before illustrating the network effects over the public opinion when agents are exposed to ambiguity, I first focus on explaining what is expected to happen in the case of a single individual. For that, one might imagine that this is a special case of the environment introduced in the previous section when the parameter $b = 1$. In this regard, the following result shows that only two types of opinions might emerge when agents interpret ambiguity under confirmatory bias.

**Proposition 1** (Polarization)**.** *If individual $i$ randomizes interpretation of ambiguous signals according to Equation* (1)*, fully relies on signals and disregards people's opinions (no network effect), then his opinion converges to either $(1-\mu)\theta + \mu\bar{\gamma}_i$ (right-biased) or $(1-\mu)\theta + \mu(1-\bar{\gamma}_i)$ (left-biased) almost surely, where $\bar{\gamma}_i = \mathbb{E}_t[\gamma_{i,t}]$. The result holds for any initial belief.*

The first thing to notice is that both left and right biased opinions may be formed with some positive probability and that both are biased since any limiting opinion is a weighted average that places weight $\mu$ on the confirmatory bias and weight $1-\mu$ in the true state $\theta$. Thus, if the fraction of ambiguous signals realized is zero, i.e. $\mu = 0$, then agents would learn the state asymptotically and no bias is observed. In this regard, it is clear that confirmatory bias is the main source of misinformation. Moreover, individuals may exhibit polarized opinions even if individuals with different confirmatory bias observe a common stream of evidence. In this case, the degree of polarization naturally depends on the relative bias and the initial priors. Some agents would naturally have a right bias and others a left bias and the intensity of the resulting polarization depends on the mass of these groups.

Moreover, depending on the sizes of $\theta$ and $\mu$, we can guarantee which type of opinion will emerge. For that, it is convenient to partition the space $\Theta \times M = [0,1]^2$ (unit square) into three regions: region $L$ characterized by both state $\theta$ and ambiguity $\mu$ "sufficiently" low, region $R$ characterized by the combination of "sufficiently" high state $\theta$ and "sufficiently" low ambiguity $\mu$ whereas region $\mathcal{W}$ is the complement of the union of $L$ and $R$. If the pair $(\theta, \mu)$ falls in to the region L, then we can say that with probability 1 just the left-biased opinion emerges. Conversely, if the pair falls in to the region $R$, then with probability 1 the right-biased opinion is formed with probability 1. If the pair falls in to the area $\mathcal{W}$, then we can not tell which opinion type will

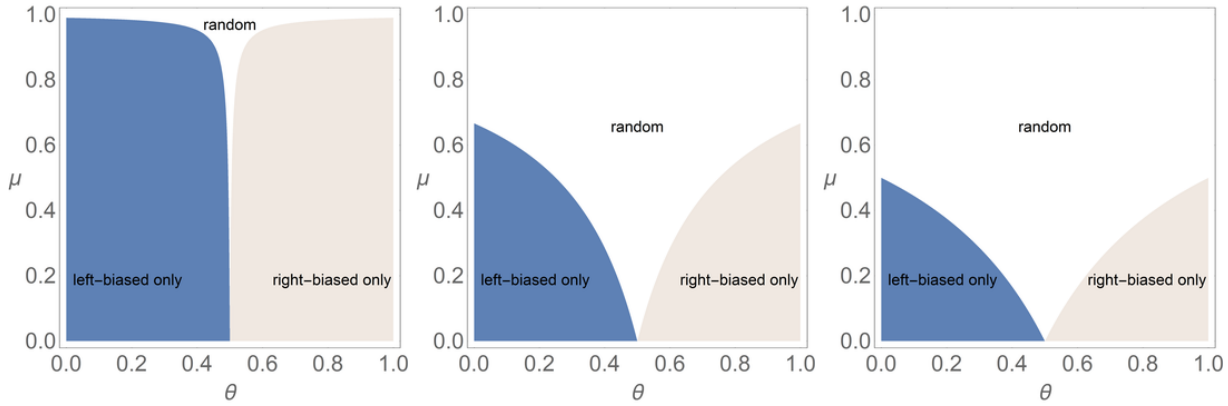
be formed as both might emerge with some positive probability. This remark is generalized as follows.

**Proposition 2** (Opinion types likelihood)**.** *For any individual with confirmatory tendency, right-biased opinion emerges with probability 1 when the frequency of ambiguity is sufficiently low and the state is sufficiently high (i.e.* $(\theta, \mu) \in R$*), whereas left-biased opinion emerges with probability 1 when both the state and the frequency of ambiguity are sufficiently low (i.e.* $(\theta, \mu) \in L$*). In all other cases (i.e.* $(\theta, \mu) \in \mathcal{W}$*), opinion type is a Bernoulli random variable which takes the value 1 (right-biased) with probability* $p$ *and the value 0 (left-biased) with probability* $(1 - p)$*. The result holds regardless his initial beliefs and the observed sequence of signals.*

The proof relies on figuring which combinations of $\theta$ and $\mu$ are sufficient to let both types of opinion to fall in the same side of the 0-1 spectrum and which combinations lead opinions to diverge in location (one above 0.5 and the other below 0.5). In mathematical terms, we have that those partitions are

$$R = \left\{(\theta, \mu) | \frac{1}{2} < \theta \leq 1 \text{ and } 0 \leq \mu < \frac{\theta - 0.5}{\bar{\gamma}_i + \theta - 1}\right\},$$
$$L = \left\{(\theta, \mu) | 0 \leq \theta < \frac{1}{2} \text{ and } 0 \leq \mu < \frac{\theta - 0.5}{\theta - \bar{\gamma}_i}\right\},$$
$$\mathcal{W} = [0, 1]^2 \setminus \{R \cup L\}.$$

The intuition of Proposition 2 is depicted in Figure 2 for three cases of confirmatory bias. In case 1, when the agent is roughly impartial, case 2 when the agent has and intermediary level of confirmatory bias and case 3 when agent is biased.




(A) case 1: $\bar{\gamma}_i = 0.505$ (B) case 2: $\bar{\gamma}_i = 0.75$ (C) case 3: $\bar{\gamma}_i = 1$

FIGURE 2. Parameter space and emergence of different types of consensus

The lightly shaded areas on the right represent the set of parameters $\mu$ (vertical axis) and $\theta$ (horizontal axis) that ensures the emergence of right-biased opinion, whereas the darkly shaded areas represent the set of parameters that ensure the emergence of left-biased opinion. In both areas, for a given level of confirmatory bias, the left (right) biased opinion emerge with probability 1 if there is a favorable combination of low frequency of ambiguous signals and a low state, i.e. below 0.5 (high state, above 0.5). The white area, on the other hand, represents the combinations of $\theta$ and $\mu$ such that the opinion type becomes a random variable, i.e. both types might emerge with positive probability. In order to finish conveying the intuition of the results presented so far and to introduce the next result, first consider the following example.

**Example 1.** *Suppose that a biased individual ($\bar{\gamma}_i = 1$) faces a "low" frequency of ambiguity, say $\mu = 0.20$ (i.e. 20% of the non-empty signals are ambiguous) and consider three possible states $\theta$: low, middle and high, e.g. $\theta_L = 0.1$, $\theta_M = 0.5$ and $\theta_R = 0.9$, respectively. In this case, Propositions 1 and 2 show that under state*

$$\theta_L, \text{ the } \begin{cases} \text{right-biased opinion } (1-0.2)\times 0.1 + 0.2 \times 1 = 0.28 \text{ is formed with prob. } 0, \\ \text{left-biased opinion } (1-0.2)\times 0.1 + 0.2 \times (1-1) = 0.08 \text{ is formed with prob. } 1, \end{cases}$$

*under state*

$$\theta_M, \text{ the } \begin{cases} \text{right-biased opinion } (1-0.2)\times 0.5 + 0.2 \times 1 = 0.60 \text{ is formed with prob. } p \in (0,1), \\ \text{left-biased opinion } (1-0.2)\times 0.5 + 0.2 \times (1-1) = 0.40 \text{ is formed with prob. } 1-p, \end{cases}$$

*and under state*

$$\theta_H, \text{ the } \begin{cases} \text{right-biased opinion } (1-0.2)\times 0.9 + 0.2 \times 1 = 0.92 \text{ is formed with prob. } 1, \\ \text{left-biased opinion } (1-0.2)\times 0.9 + 0.2 \times (1-1) = 0.72 \text{ is formed with prob. } 0. \end{cases}$$

Based on the results shown so far and on the numerical example above, one can see clearly that for an individual with confirmatory tendency, one opinion type is less biased than the other depending on the magnitude of the state $\theta$. Moreover, in regions $L$ and $R$ the opinion type formed, even though biased, is the most efficient with probability 1. This generalizes as follows.

**Corollary 1** (Efficiency)**.** *For any individual with confirmatory tendency and for any ambiguity level, the right-biased (left-biased) opinion is less biased than the left-biased (right-biased) opinion if $\theta > \frac{1}{2}$ $\left(\theta < \frac{1}{2}\right)$. Conversely, both right and left biased opinions are equally biased when $\theta = \frac{1}{2}$.*

The intuition is that these two opinions are not symmetric around $\theta$ as shown in the Example 1 above. This happens because the bias of each one depends on the relative size of $\theta$ and $\bar{\gamma}_i$. Since we are restricting attention to the case in which individual has confirmatory tendency, i.e. $\bar{\gamma}_i > \frac{1}{2}$,

it is the case that individuals make less mistakes when they are in the correct side of the spectrum. For that to happen, the ambiguity has to be low enough to not mislead individuals and the state has to be high (or low) enough to nudge individuals opinions to the correct side.

Besides that, it is not trivial to ascertain analytically the functional form of the probability of emergence of right-biased opinion when $\theta > \frac{1}{2}$ and the probability of emergence of left-biased opinion when $\theta < \frac{1}{2}$. It is possible to see that all parameters related to the priors $(\alpha_0, \beta_0)$, to the signal structure $(\delta, \mu, \theta)$ and to the confirmatory bias intensity $(\gamma)$ influence it but it is not clear how to generalize the result. I return to this discussion on Section (5) when I discuss the welfare consequences of different network topologies.

Another particular case of interest is the one in which the agent is impartial. In this case, it can be shown that bias is not overcome.

**Corollary 2** (Bias from impartiality)**.** *If an individual is impartial, then his limiting opinion is* $(1-\mu)\,\theta + \mu\,\frac{1}{2}$ *almost surely, regardless his initial prior and the sequence of observed signals.*

The reason why impartiality does not overcome bias is because it forces individuals to set a disproportionate probability mass in the center of the spectrum $(0, 1)$. Thus, impartiality make agents excessively centrists instead making them neutral towards the possible states. One can show that this phenomenon is a direct consequence of the Beta-Bernoulli conjugate family employed here that would not take place in the case of a binary state space $\Theta = \{0, 1\}$.

Moreover, under impartiality, for any mass of ambiguity $\mu > 0$, if the true state is located in the left side of the 0-1 spectrum $(\theta < \frac{1}{2})$, then consensus has a positive bias and lies in $\left(\theta, \frac{1}{2}\right)$. Conversely, if $\theta > \frac{1}{2}$, then consensus has a negative bias and lies in $\left(\frac{1}{2}, \theta\right)$. The only instance when the individual learn the state is when $\theta = \frac{1}{2}$, an almost anywhere event. The results presented so far both extend the intuition and contrast with Propositions 4 and 5 (i) in Rabin and Schrag (1999) and with Propositions 2 and 3 in Fryer et al. (2018). It extends the intuition to the case in which the state is continuously distributed over the interval 0-1, and contrasts because impartiality no longer can help an individual to overcome bias, as per the result above. Moreover, it also contrasts with previous results as long-run learning is an event with probability zero as it happens almost anywhere in the full parameter space. The next result elaborates this argument.

**Corollary 3** (No long-run learning)**.** *For any individual with confirmatory tendency, long-run learning is an event with probability 0, regardless his initial prior and the sequence of observed signals.*

The intuition of Corollary (3), together with Proposition 1, is that confirmatory bias invariably nudges opinions and lead to bias. Under interpretation of ambiguity, an individual can learn

the state only in some very particular situations that almost never happen. Finally, at the other extreme, one could ask under what conditions an individual would reach an extreme opinion, i.e. either opinion 0 (extreme left) or opinion 1 (extreme right). The next result shows that those cases can only be sustained under two extreme conditions: i) the fraction of ambiguous signals to be maximal ($\mu = 1$) and, ii) individual to be biased ($\bar{\gamma}_i = 1$).

**Proposition 3** (Extreme opinions)**.** *An individual $i$ has extreme opinion (0 or 1) only if he is biased and the mass of ambiguity is maximal ($\mu = 1$). The result holds for any state $\theta$ and regardless his initial beliefs and the observed sequence of signals.*

As argued before in the case of long-run learning, this is also considered to be an event that happens almost anywhere in the parameter space. With all the intuition of what happens in the single agent case, one may ask what happens if besides learning from signals, agents also learn from their friends. This case, at first, seems to impose an extra challenge because the interpretation of ambiguity not only depends on the initial realization of signals, but also depends on the influence of friends that potentially interpret ambiguity in different ways. The "*tug-of-war*" played between left and right biases has one driver more, the network externalities. Before discussing the implications of a network structure, I define the concept of consensus and state an instrumental Lemma.

**Definition 2.** *Society reaches a consensus almost surely for any initial beliefs if there is a $y$ such that for a small $\epsilon > 0$*

$$P\left(\lim_{t\to\infty} |y_{i,t} - y| < \epsilon\right) = 1$$

*for any $i \in N$.*

The auxiliary result below illustrates, in terms of ergodicity of a Markov chain, the social influence of agents derived from the reliance weight matrix $W$ presented in equations (6) and (7). The proof of such statement can be found in Appendix C.

**Lemma 1.** *The $t$-th power of matrix $W$, $W^t$, converges to a unique row-stochastic matrix with unit rank (all rows the same) as $t$ tends to infinity, i.e.*

$$\lim_{t\to\infty} W^t = W^\infty = \mathbf{1}\pi^\top = \Pi,$$

*where the invariant distribution $\pi$ is the normalized left eigenvector of the matrix $W$ associated to the unit eigenvalue, i.e. $\pi^\top W = \pi^\top$ and $\sum_i \pi_i = 1$.*

Therefore, I show next that under the assumption of strong connectivity, consensus is reached in this dynamic system and it has the same functional form of the individual limiting opinion derived in Proposition 1.

A first case of interest is the limiting case in which individuals exclusively pay attention to friends. This case can be constructed in two different (but equivalent) ways: i) setting $b = 0$, for any $\delta$ or ii) setting $\delta = 1$, for any $b$. The first one represents the situation where agents disregard signals completely and are pure conformists, whereas the second represents the case where no signals enter in the network and agents (forcefully) pay exclusive attention to friends. In both cases, consensus reached is the same and differently than the classical DeGroot case, limiting opinion is not properly a weighted average of the initial opinions, even though is still very close to it. The discrepancy has to do with the fact that agents are exchanging opinions and precisions (parameters $\alpha$ and $\beta$ of each individual prior). this is stated as follows.

**Proposition 4** (Pure DeGroot: Consensus)**.** *If the social network $G = (N, g)$ is strongly connected, individuals randomize interpretation of ambiguous signals according to Equation* (1)*, and the Bayesian parameter is set at $b = 0$, for any $\delta$ (or equivalently $\delta = 1$, for any $b$), then society reaches consensus $y = \frac{\sum_j \Pi_{ij}\alpha_{j,0}}{\sum_j \Pi_{ij}(\alpha_{j,0}+\beta_{j,0})}$, for any $i \in N$.*

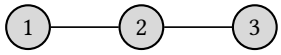
The intuition of this result is that if signals $1, 0$ and $a$ are not observed very often ($\delta$ close to one) by connected agents, then this will force their priors to converge to a common one. The longer it takes this society to observe some non-empty signal, the more agents interact and the more likely they will find a common ground. The implications of this result are highlighted in the Section 5 when I explore the effects of priors heterogeneity on the probability of attaining consensus efficiency. Next, I introduce the result when there are network externalities and non-empty signals are observed with positive probability ($\delta < 1$).

**Proposition 5** (Network effect)**.** *With network externalities, the sequences $\{y_{i,t}\}_{t=1}^{\infty}$ generated by the update rule converge almost surely to either right-biased consensus $(1-\mu)\theta + \mu\tilde{\bar{\gamma}}$ or left-biased consensus $(1-\mu)\theta + \mu(1-\tilde{\bar{\gamma}})$, for all $i \in N$ and where $\Pi$ is the invariant distribution matrix, $\tilde{\bar{\gamma}} = \sum_j \Pi_{ij}\bar{\gamma}_j$ and $\bar{\gamma}_j = \mathbb{E}_t\left[\gamma_{j,t}\right]$.*

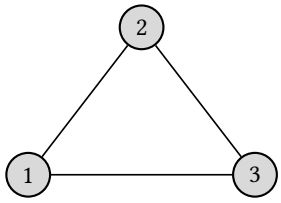
Again, the result basically shows that consensus again takes the form of a weighted average between the true state $\theta$ and the social confirmatory bias $\tilde{\bar{\gamma}}$, where the mass of ambiguity $\mu$ is the weight of the later. If $\mu = 0$, then there is no consensus bias and agent would aggregate information efficiently. Moreover, the consensus (of any type) does not depend on the parameter $\delta$, i.e. the consensus does not depend on the frequency with which the network receives signals.

Thus, neither a system that remains "quiet" for a long time (high $\delta$) nor a system that receives information all the time (low $\delta$) can influence consensus. Additionally, the parameter $b$ does impact the vector of social influence (i.e. the invariant distribution $\pi$ of the matrix $W$ (see Lemma (1)) and therefore does impact consensus. As $b \to 0$, the social influence is basically dictated by the normalized adjacency matrix $\hat{g}$ and is somehow directly proportional to the degree centrality of the agents. As $b \to 1$, agents tend to almost disregard in full friends' opinions and the social weight of each individual converges to $\frac{1}{n}$.

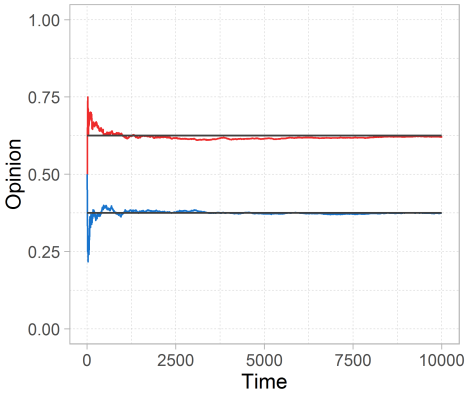
Moreover, the results above show that the consensus type in this dynamic system is also a tail event, i.e. right-biased consensus will either almost surely emerge as the stable equilibrium or almost surely not emerge. If it does not emerge as an equilibrium of this system, then it is surely the case that the left-biased consensus has emerged as the equilibrium (and vice-versa). As an illustration of this result, Figures (3c) and (3d) show the typical opinion sample path of any agent in the line and wheel networks, respectively, and the convergence to different consensus types (horizontal lines) in different simulations.



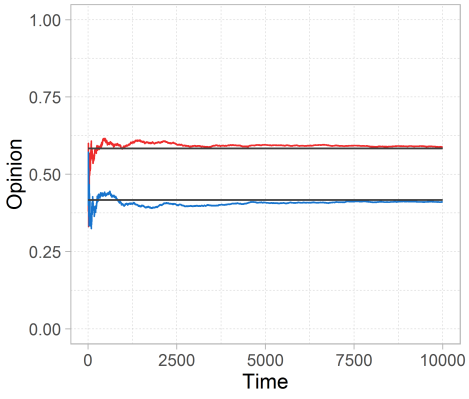

(A) Line network




(B) wheel network




(C) Typical opinion path in line network and type 1 and type 2 theoretical consensus




(D) Typical opinion path in wheel network and type 1 and type 2 theoretical consensus

FIGURE 3. Simulation with parameters $T = 10,000$, $n = 3$, $\delta = \mu = \theta = b = 0.5$, $\alpha_{i,0} = \beta_{i,0} = 1$ for all $i \in N$ (so $y_{i,0} = 0.5$ for any $i$) and $\gamma_t = (\gamma_{1,t}, \gamma_{2,t}, \gamma_{3,t}) = (0.8, 1, 0.2)$ for all $t \in T$.

Given the importance of the probability of emergence of the efficient consensus in terms of social efficiency, the next section is devoted to numerically characterizes it. The exercise is not trivial and for that I will rely on simulations of the learning process described in Section (3) for selected classical network topologies and for different sets of parameters of interest.

## 5. Random graph simulation: determining consensus type

Ascertaining $p$ analytically depends on many circumstances. One particular example is when the initial priors at time $t = 0$ differ. They can either be skewed to the right or to the left (partisans) and in different proportions (degree of partisanship). Thus, under this circumstance, the society is prone to interpret ambiguous signals as 0 or 1 not only according to the initial realization of the signals, but also according to the initial partisanship. Besides, the heterogeneity of priors has the consequence of creating centrality heterogeneity, i.e. partisan individuals might be located at more or less central nodes. This is a challenging instance because partisans might influence other agents disproportionately and this can amplify the underlying interpreting dispute in the network. Another example is that agents might differ in the intensity of confirmatory bias they suffer, i.e. agents with polar opposite bias (say, $\gamma_{i,0} = 0$ and $\gamma_{j,0} = 1$, for some $i, j \in N$) might be directly connected or not and this might play a role on how much this heterogeneity affects the interpreting conflict. This issue gets particularly hard if one allows different partisan individuals to suffer from different confirmatory bias. A final example is the association of partisanship unbalance with the frequency with which the society receives signals (i.e. low/high $\delta$). In this case, the less often signals enter in the network (large $\delta$), the more society behaves as purely DeGrootian agents and consequently the initial heterogeneity of priors might lose importance. This happens because even though partisans start with very different beliefs, agents communicate very often and might converge to a common prior even before some ambiguous signal enters the network.

Those are just some examples to illustrate the challenging nature of computing $p$ analytically. That said, I proceed with the analysis of $p$ in two general cases: i) when agents are biased but have common centrist prior and ii) biased agents with heterogeneous priors. Next I explain the strategy employed to control the inherent multi-dimensionality of this exercise.

### 5.1. **Common prior.**

The common prior case can be represented by the situation in which priors parameters have the same configuration, i.e. $\alpha_{i,0} = \bar{\alpha} \in \mathbb{R}_+$ and $\beta_{i,0} = \bar{\beta} \in \mathbb{R}_+$ for all $i \in N$. In particular, when $\bar{\alpha} = \bar{\beta} = 1$, all agents hold a uniform common prior over the unit interval. For any other value, say $\bar{\alpha} = \bar{\beta} = k > 1$, agents hold a symmetric "bell-shaped" common prior over the unit interval, centered at 0.5. Moreover, as $k \to \infty$, the bell-shaped priors collapse to the

point $\frac{1}{2}$, i.e. the precision of the prior diverges and all opinions are $y_{i,0} = \frac{1}{2}$. Those cases represent the situation in which agents start as *"centrists"* and the subsequent asymmetry of interpretation stems from the signals realizations.[8] On the other hand, when $\alpha_{i,0} = \bar{\alpha}$ and $\beta_{i,0} = \bar{\beta}$ for all $i \in N$ and $\bar{\alpha} > \bar{\beta}$ $\left(\bar{\beta} > \bar{\alpha}\right)$, the society holds a rightist (leftist) common prior, i.e. $y_{i,0} = y_R > \frac{1}{2}$ $(y_{i,0} = y_L < \frac{1}{2})$ and as $\frac{\bar{\alpha}}{\bar{\beta}} \to \infty$ $(\to 0)$, the bell-shaped priors collapse to the point 1 (0), i.e. the precision of the prior diverges and all opinions become extreme.

5.2. **Heterogeneous priors.** The heterogeneous prior case refers to the situation in which there are three types of agents in the society at time $t = 0$: centrists ($\mathcal{C}_0$) and two partisans, *leftists* ($\mathcal{L}_0$) and rightists ($\mathcal{R}_0$). To make a distinction between those agents, first let's consider two parameters that intend to measure the degree of *partisanship* of such agents $\tau_l, \tau_r \in \mathbb{N}$. With that, I define such groups as follows: centrists, $\mathcal{C}_0 = \{i \in N \,|\, \alpha_{i,0} = 1 \text{ and } \beta_{i,0} = 1\}$, left-partisan, $\mathcal{L}_0 = \{i \in N \,|\, \alpha_{i,0} = 1 \text{ and } \beta_{i,0} = 1 + \tau_l\}$ and right-partisan, $\mathcal{R}_0 = \{i \in N \,|\, \alpha_{i,0} = 1 + \tau_r \text{ and } \beta_{i,0} = 1\}$. Notice that the definition implies that initial opinions and precisions $y_{i,0} = \alpha_{i,0}(\alpha_{i,0} + \beta_{i,0})^{-1}$ and $\sigma_{i,0}^{-2} = (\alpha_{i,0}\beta_{i,0})^{-1}(\alpha_{i,0} + \beta_{i,0})^2(\alpha_{i,0} + \beta_{i,0} + 1)$, respectively.

$$y_{i,0} = \begin{cases} \dfrac{1}{2 + \tau_l}, & \text{if } i \in \mathcal{L}_0 \\ \dfrac{1}{2}, & \text{if } i \in \mathcal{C}_0 \\ \dfrac{1 + \tau_r}{2 + \tau_r}, & \text{if } i \in \mathcal{R}_0 \end{cases}$$

and

$$\sigma_{i,0}^{-2} = \begin{cases} \dfrac{6 + 5\tau_l + \tau_l^2}{1 + \tau_l}, & \text{if } i \in \mathcal{L}_0 \\ 12, & \text{if } i \in \mathcal{C}_0 \\ \dfrac{6 + 5\tau_r + \tau_r^2}{1 + \tau_r}, & \text{if } i \in \mathcal{R}_0 \end{cases}$$

[8]To be more precise, interpretation neutrality does not exist as there is a non-neutral tie-break rule described by Equation (1). Thus, if the very first signal happens to be ambiguous, then it will be interpreted as 1 by all agents, as per the tie-break rule. This is without loss of generality for the results presented in this work. The tie-break rule could have been defined in a way that the initial interpretation would benefit the left-interpretation and intuition and conclusions would remain the same. Finally, if the tie-break rule is neutral towards the initial interpretation, that would require a more intricate update rule that would force agents to keep the prior unchanged when opinions are 0.5 whenever they face an ambiguous signal. In such situation, agents would have to draw another signal until some non-ambiguous realization takes place. I conjecture that results would not change in this case as well, since the neutral tie-break, even though does not benefit any state, could also trap individuals into a wrong state and the nature of the problem would remain the same. The modeling effort, however, could change significantly.

Notice that $\lim_{\tau\to\infty} y_{i,0}$ is 0, $\frac{1}{2}$ and 1, whereas $\lim_{\tau\to\infty} \sigma_{i,0}^{-2}$ is $+\infty$, 12 and $+\infty$ for left-partisan, centrists and right-partisan, respectively.

5.3. **Simulation.** In order to compute the empirical frequency with which the efficient consensus emerges as an equilibrium in the system $(\hat{p})$ when the pair $(\theta, \mu) \in \mathcal{W}$ I simulate the learning process described on Section (3) many times for a sufficiently long period in all selected classical networks shown in Figure 4. The number of simulations is described by $S \in \mathbb{N}$ and the agents maximal interaction time is $t \in \mathbb{N}$.

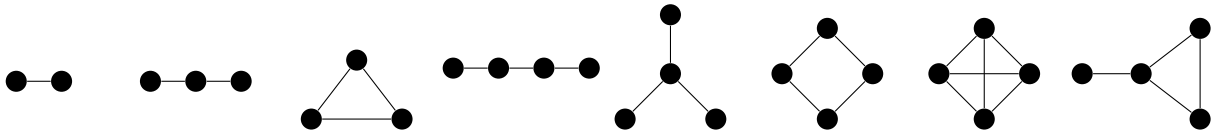



FIGURE 4. Selected network topologies

For each simulation I allow some parameters to vary (see details below) so I can capture changes in $\hat{p}$ due to changes in such parameters. However, the choice of the parameters is the same for all networks in each simulation so I can properly isolate effects on $\hat{p}$ due to parameters discrepancies. Moreover, for a small $\epsilon > 0$ and given the chosen parameters in each simulation $S$, the simulated frequency of efficient consensus in a network $G$ is computed as

$$\hat{p}_G = \frac{1}{S}\sum_{S} \mathbb{1}\Bigg\{ \left| \lim_{t\to\infty} y_{i,t}^{S,G} - \left((1-\mu_S)\theta_S + \mu_S \tilde{\tilde{\gamma}}_S\right)\right| < \epsilon \text{ and } \theta_S > \frac{1}{2} \text{ , or} \\ \left| \lim_{t\to\infty} y_{i,t}^{S,G} - \left((1-\mu_S)\theta_S + \mu_S (1-\tilde{\tilde{\gamma}}_S)\right)\right| < \epsilon \text{ and } \theta_S < \frac{1}{2}\Bigg\}. \tag{8}$$

The description of each exercise follows below.

5.3.1. *Exercise 1: common (centrist) prior vs. heterogeneous (balanced) priors.* The purpose of this exercise is to understand how heterogeneous priors affect the probability of emergence of the efficient consensus. For that, I keep the number of partisans in its minimum, one left and one right-partisan so the proportion is balanced, and place them uniformly at random in the available nodes in each simulation. In terms of degree of partisanship, when $\tau_l = \tau_r = \tau = 0$ agents have a common centrist prior (uniform distribution over the unit interval) and there are no partisan agents, whereas when $\tau_l = \tau_r = \tau > 0$ represents the case of heterogeneous priors in which the degree of partisanship of both partisans is positive and equally balanced.

Moreover, in order to avoid an extra layer of heterogeneity I allow all agents to be biased $\tilde{\tilde{\gamma}} = \gamma_{i,t} = 1$ for all $i, t$. Finally, for every simulation, I fix some selected parameters and draw

uniformly at random other parameters from the following sets in a way that each network in each simulations has the same parameters. The description is summarized below and the summary statistics of the simulations can be found in Appendix E. The simulated probability of emergence of the efficient consensus $\hat{p}$ are reported in the Table 1. .

- **Fixed parameters** (for all simulations)
  - **Learning**: $\tilde{\bar{\gamma}} = \gamma_{i,t} = 1$ for all $i$ and $t$ and $b = 0.5$.
  - **Duration**: $t = 700$.
  - **Initial conditions**: $(\alpha_{i,0}, \beta_{i,0}) = (1, 1)$ for all $i \in N$, $|\mathcal{L}_0| = |\mathcal{R}_0| = 1$.
- **Variable parameters** (for each simulation $S$)
  - **Information**:

    $\delta_S \in \Delta = \{0.05, 0.20, 0.35, \ldots, 0.95\}$,

    $\theta_S \in \Delta \setminus \{0.5\}$ and

    $\mu_S \in \tilde{M}$ is such that $(\theta_S, \mu_S) \in \mathcal{W}$, $|\Delta| = |\tilde{M}|$ and $\sup \tilde{M} = \sup \Delta = 0.95$.
  - **Initial conditions**:

    $\tau_l = \tau_r = \tau_S$ such that $\tau_S \in \mathcal{T} = \{0, 1, 10, 30\}$.

Based on the simulation statistics derived from the Exercise 1 simulations, I present some results of interest. First, as per the data from simulations with common centrist prior ($\tau = 0$, no partisanship), we can see that topology seems to be innocuous to the probability $\hat{p}$. This evidence is stated as the following result.

**Result 1** (Topology neutrality)**.** *If all agents are biased and centrists, then network topology has no impact on consensus efficiency.*

The intuition of this result relies on the fact that since signals are public and all agents share the same bias intensity $\bar{\gamma}_i = 1$, there is no interpretation diversity regardless signals realization. If agents start observing signal 1, then all agents will become more rightists and network externalities can not countervail this effect anyhow. The same argument applies to all other signals, including the ambiguous one. Therefore, this is identical to the case of a single individual learning from signals. Moreover, based on the data from simulations with common prior ($\tau = 0$, no partisanship) and low priors heterogeneity ($\tau = 1$, low partisanship), there seems to have a non-negative effect of partisanship on consensus efficiency.

**Result 2** (Efficiency of low partisanship)**.** *In expected terms, a biased society with low partisanship ($\tau = 1$) is at least as able to reach the efficient consensus as the same biased society with no partisanship at all ($\tau = 0$).*

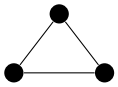
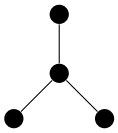
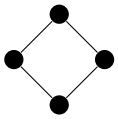
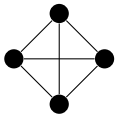
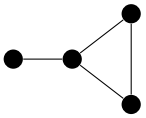
| Size | Network Topology | Type | Label | $\hat{p}$ ($\tau$=0) | $\hat{p}$ ($\tau$=1) | $\hat{p}$ ($\tau$=10) | $\hat{p}$ ($\tau$=30) |
|---|---|---|---|---|---|---|---|
| $n = 1$ |  | single agent | (SA) | 0.688 | - | - | - |
| $n = 2$ |  | line (complete) | (A) | 0.688 | 0.716 | 0.698 | 0.678 |
| $n = 3$ |  | line | (B) | 0.688 | 0.707 | 0.608 | 0.590 |
| |  | wheel (complete) | (C) | 0.688 | 0.730 | 0.726 | 0.717 |
| $n = 4$ |  | line | (D) | 0.688 | 0.766 | 0.680 | 0.648 |
| |  | star | (E) | 0.696 | 0.709 | 0.630 | 0.602 |
| |  | wheel | (F) | 0.688 | 0.766 | 0.787 | 0.794 |
| |  | complete | (G) | 0.695 | 0.725 | 0.719 | 0.718 |
| |  | paw | (H) | 0.694 | 0.733 | 0.633 | 0.559 |
| $S$ | | | | 11,539 | 4,604 | 4,680 | 4,554 |

TABLE 1. Simulated frequency of the emergence of efficient consensus $\hat{p}$.

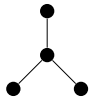
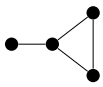
Partisanship acts in a way that countervails the effect of initial ambiguous signals. Under no influence of partisans, centrists interpretations are solely dependent on observed signals, therefore the realization of the initial signals are crucial to determine what bias opinions will have and, for that, it is determinant to consensus efficiency. On the other hand, when some partisan agents are present, priors parameters $\alpha$'s and $\beta$'s are shifted up, by right and left-partisans respectively, which makes opinions more robust to the initial signals realizations. However, it seems that there is some sort of "optimal" level of partisanship since high partisanship, for most topologies, has a non-monotonic effect over the probability of emergence of the efficient consensus. This result is generalized as follows.

**Result 3** (Inefficiency of high partisanship)**.** *In expected terms, a biased society with low partisanship ($\tau = 1$) is at least as able to reach the efficient consensus as the same biased society with high partisanship ($\tau = 30$). One exception is the wheel network with four agents (network (F)) in which efficiency seems to increase monotonically with partisanship.*

The intuition is that if there is disproportionately partisanship, then partisan individuals can create an unbalance similar to the one created by the realization of the first signals. More explicitly, one can imagine that a partisan agent with high degree of partisanship will almost never interpret ambiguous evidences in a way that disagrees with his beliefs and a similar effect applies to his neighbors. However, partisan agents might be more or less connected and even connected to each other. The later situation is defined as *open-mindedness* and its effect is particularly important to consensus efficiency. The definition of open-mindedness stated below is very similar in nature to the one of *heterophily* already established in the social and economic networks literature and both reflect the tendency that different individuals have to connect with each other.

**Definition 3** (Open/Narrow-mindedness)**.** *For any given network induced by some adjacency matrix $g$, a partisan agent $i \in N$ is said to be open-minded if, for some other partisan agent $j \in N$ with opposite belief, we have that $j \in N_i^{out}(g)$. Conversely, $i$ is narrow-minded if $j \notin N_i^{out}(g)$.*

Naturally, in networks A, C and G partisan agents are invariably open-minded because those networks are complete, i.e. all individuals are connected with every other individual in the network, regardless their types. In this sense, it makes sense to analyze the effect of open/narrow-mindedness in networks B, D, E, F and H since it is not always that those agents are connected. The simulated probability $\hat{p}$ in those cases are reported in Table **??** and the next result is stated immediately.

| Network | Partisans | $\hat{p}$ ($\tau$=0) | $\hat{p}$ ($\tau$=1) | $\hat{p}$ ($\tau$=10) | $\hat{p}$ ($\tau$=30) |
|---|---|---|---|---|---|
|  | pooled | 0.688 | 0.707 | 0.608 | 0.59 |
| | open-minded | 0.692 | 0.703 | 0.555 | 0.523 |
| | narrow-minded | 0.681 | 0.715 | 0.714 | 0.722 |
|  | pooled | 0.688 | 0.766 | 0.680 | 0.648 |
| | open-minded | 0.687 | 0.762 | 0.662 | 0.641 |
| | narrow-minded | 0.689 | 0.770 | 0.699 | 0.656 |
|  | pooled | 0.696 | 0.709 | 0.630 | 0.602 |
| | open-minded | 0.694 | 0.700 | 0.545 | 0.494 |
| | narrow-minded | 0.697 | 0.719 | 0.713 | 0.717 |
|  | pooled | 0.688 | 0.766 | 0.787 | 0.794 |
| | open-minded | 0.682 | 0.779 | 0.827 | 0.833 |
| | narrow-minded | 0.701 | 0.739 | 0.706 | 0.714 |
|  | pooled | 0.694 | 0.733 | 0.633 | 0.559 |
| | open-minded | 0.694 | 0.736 | 0.629 | 0.575 |
| | narrow-minded | 0.693 | 0.728 | 0.641 | 0.529 |

**Result 4** (Open-minded partisans)**.** *In expected terms, for biased individuals connected through*

- *networks F and H, open-mindedness of partisan agents induces consensus efficiency,*
- *networks B and E, narrow-mindedness of partisan agents induces consensus efficiency,*
- *network D, open-mindedness of partisan agents is neutral in terms of inducing consensus efficiency.*

Moreover, for the case of complete networks we have that population size seems to have no influence whatsoever on efficiency, for any .

**Result 5** (Complete networks)**.** *For biased individuals connected through a sufficiently large complete network ($n \geq 3$), the size of the network is neutral to the expected consensus efficiency, regardless the degree of partisanship.*

A final case of interest is the one in which agents are connected through a line.

**Result 6** (Line networks)**.** *In expected terms, biased individuals connected through any sufficiently long line network ($n \geq 3$), high partisanship ($\tau = 30$) reduces the chance of reaching the efficient consensus. Moreover, for any given level of partisanship $\tau > 0$, a shorter line (lower $n$) reduces the chance of reaching the efficient consensus.*

## 6. Conclusions

Confirmatory bias is one of the most notorious cognitive biases documented and it appears in many ways. Since it is a systematic deviation from rationality in judgment, it is expected to have a significant influence in the process of belief formation. In this sense, since social networks appear as a primary tool for many people to get informed and debate their world views, one could expect confirmatory bias to have some influence on the public opinion formation. To date, however, there has been little understanding of how such phenomenon influences public opinion. To shed some light on this topic, I consider a social learning model in which a fraction of signals, external to the social network, is ambiguous and open idiosyncratic interpretation. The interpretation, however, is affected by individuals' confirmatory biases. Moreover, I also allow agents to be influenced by friends in their social clique and to set their beliefs to be a linear combination of the (biased) Bayesian posterior and the (also biased) friends' posteriors.

It follows directly from my model that biased individuals connected through a social network can only reach two types of consensus and both are biased, one to the left and the other to the right. One consensus type, however, is more efficient (less biased) than the other depending on the state. Moreover, I show that long-run learning is not attained even if individuals are impartial when interpreting ambiguous signals. Those results contrast with Rabin and Schrag (1999) and Fryer

et al. (2018) in which long-run learning takes place with a positive probability and impartiality helps learning the state. Furthermore, the network effect presented here, together with signals realizations, reinforces the interpreting "*tug-of-war*" since individuals might have their own biases confirmed (or mitigated) by other agents.

Finally, since it is not trivial to derive analytically the probability of emergence of the most efficient (less biased) consensus, I use graphs simulations to show its determinants. I show that the presence of partisan agents in societies who suffer from confirmatory bias has a double effect on the expected consensus efficiency: i) it helps to countervail the misinterpretation of initial signals when there degree of partisanship is low and for that it increases expected efficiency; and ii) exacerbates misinterpretation of signals when the degree of partisanship is high, reducing expected consensus efficiency. Moreover, I also show that open-mindedness of partisan agents, i.e. when partisans agree to exchange opinions with partisans with polar opposite beliefs, might reduce expected consensus efficiency in some social topologies. These results suggest that policies designed to mitigate partisanship and confirmatory bias effects in social networks have to consider also the positive network externalities induced by them.

## References

Acemoglu, D., K. Bimpikis, and A. Ozdaglar (2014): "Dynamics of information exchange in endogenous social networks," *Theoretical Economics*, 9, 41–97.

Acemoglu, D., M. A. Dahleh, I. Lobel, and A. Ozdaglar (2011): "Bayesian Learning in Social Networks," *Review of Economic Studies*, 1017, 35–1201.

Acemoglu, D. and A. Ozdaglar (2011): "Opinion Dynamics and Learning in Social Networks," *Dynamic Games and Applications*, 1, 3–49.

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

Allahverdyan, A. E. and A. Galstyan (2014): "Opinion Dynamics with Confirmation Bias," *PLOS*.

Andreoni, J. and T. Mylovanov (2012): "Diverging opinions," *American Economic Journal: Microeconomics*, 4, 209–232.

Azzimonti, M. and M. Fernandes (2018): "Social Media Networks, Fake News, and Polarization," *NBER Working Paper*, 24462, 65.

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

——— (2001): "Conformism and diversity under social learning," *Economic Theory*, 17, 101–120.

Baliga, S., E. Hanany, and P. Klibanoff (2013): "Polarization and Ambiguity," *American Economic Review*, 103, 3071–3083.

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

Banerjee, A. V. (1992): "A Simple Model of Herd Behavior," *The Quarterly Journal of Economics*, 107, 797–817.

——— (1993): "The Economics of Rumours," *Review of Economic Studies*, 60, 309–327.

Banerjee, A. V., A. Chandrasekhar, E. Duflo, and M. O. Jackson (2017): "Gossip: Identifying Central Individuals in a Social Network," *SSRN Electronic Journal*.

Dandekar, P., A. Goel, and D. T. Lee (2013): "Biased assimilation, homophily, and the dynamics of polarization." *PNAS*, 110, 5791–6.

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

DeGroot, M. H. and M. J. Schervish (2012): *Probability and Statistics*, Pearson.

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

Ellison, G. and D. Fudenberg (1993): "Rules of Thumb for Social Learning," *Journal of Political Economy*, 101, 612–643.

Ellsberg, D. (1961): "Risk, Ambiguity, and the Savage Axioms," *The Quarterly Journal of Economics*, 75, 643–669.

Epstein, L. G., J. Noor, and A. Sandroni (2010): "Non-Bayesian Learning," *Journal of Theoretical Economics Advances The B.E. Journal of Theoretical Economics*, 10.

Epstein, L. G. and M. Schneider (2007): "Learning Under Ambiguity," *Review of Economic Studies*, 74, 1275–1303.

Fryer, R. G., P. Harms, and M. O. Jackson (2018): "Updating Beliefs with Ambiguous Evidence: Implications for Polarization," *SSRN Electronic Journal*.

Fryer, R. G. and M. O. Jackson (2008): "A Categorical Model of Cognition and Biased Decision-Making," *The B.E. Journal of Theoretical Economics*.

Furnham, A. and J. Marks (2013): "Tolerance of Ambiguity: A Review of the Recent Literature," *Psychology*, 4, 717–728.

Furnham, A. and T. Ribchester (1995): "Tolerance of ambiguity: A review of the concept, its measurement and applications," *Current Psychology*, 14, 179–199.

Gale, D. and S. Kariv (2003): "Bayesian learning in social networks," *Games and Economic Behavior*, 45, 329–346.

Gennaioli, N. and A. Shleifer (2010): "What Comes to Mind," *Quarterly Journal of Economics*, 125.

Gilboa, I. and D. Schmeidler (1989): "Maxmin expected utility with non-unique prior," *Journal of Mathematical Economics*, 18, 141–153.

——— (1993): "Updating Ambiguous Beliefs," *Journal of Economic Theory*, 59, 33–49.

Glaeser, E. and C. Sunstein (2013): "Why Does Balanced News Produce Unbalanced Views?" *NBER Working Paper*.

Golub, B. and M. O. Jackson (2010): "Naïve Learning in Social Networks and the wisdom of the crowds," *American Economic Journal: Microeconomics*, 2, 112–149.

Hegselmann, R. and U. Krause (2002): "Opinion Dynamics and Bounded Confidence Models, Analysis and Simulation," *Journal of Artifical Societies and Social Simulation*, 5.

——— (2005): "Opinion Dynamics Driven by Various Ways of Averaging," *Computational Economics*, 381–405.

Hellman, M. E. and T. M. Cover (1970): "Learning with Finite Memory," *The Annals of Mathematical Statistics*, 41, 765–782.

Jackson, M. O., E. Kalai, and R. Smorodinsky (1999): "Bayesian Representation of Stochastic Processes Under Learning: De Finetti Revisited," *Econometrica*, 67, 875–893.

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

Kalai, E. and E. Lehrer (1994): "Weak and strong merging of opinions," *Journal of Mathematical Economics*, 23, 73–86.

Lord, C. G., L. Ross, and M. R. Lepper (1979): "Biased Assimilation and Attitude Polarization: The Effects of Prior Theories on Subsequently Considered Evidence," *Journal of Personality and Social Psychology*, 37, 2098–2109.

Mahler, R. P. (1995): "Combining ambiguous evidence with respect to ambiguous a priori knowledge. Part II: Fuzzy logic," *Fuzzy Sets and Systems*, 75, 319–354.

Mao, Y., S. Bolouki, and E. Akyol (2018): "Spread of Information with Confirmation Bias in Cyber-Social Networks," *arXiv:1803.06377*.

Mercier, H. and D. Sperber (2011): "Why do humans reason? Arguments for an argumentative theory," *Behavioral and Brain Sciences*, 34, 57–74.

Meyer, C. D. (2000): "Matrix analysis and applied linear algebra," *Society for Industrial and Applied Mathematics*.

Molavi, P., A. Tahbaz-Salehi, and A. Jadbabaie (2018): "A Theory of Non-Bayesian Social Learning," *Econometrica*, 86, 445–490.

Molden, D. C. and E. T. Higgins (2004): "Categorization Under Uncertainty: Resolving Vagueness and Ambiguity With Eager Versus Vigilant Strategies," *Social Cognition*, 22, 248–277.

——— (2008): "How Preferences For Eager Versus Vigilant Judgment Strategies Affect Self-Serving Conclusions." *Journal of experimental social psychology*, 44, 1219–1228.

Mullainathan, S. (2002): "A Memory-Based Model of Bounded Rationality," *The Quarterly Journal of Economics*, 117, 735–774.

Nickerson, R. S. (1998): "Confirmation Bias: A Ubiquitous Phenomenon in Many Guises," Tech. Rep. 2.

Rabin, M. and J. L. Schrag (1999): "First Impressions Matter: A Model of Confirmatory Bias," *The Quarterly Journal of Economics*, 114, 37–82.

Sherman, D. K. and G. L. Cohen (2006): "The Psychology of Selffifftdefense: SelffifftAffirmation Theory," *Advances in Experimental Social Psychology*, 38, 183–242.

Wilson, A. (2014): "Bounded Memory and Biases in Information Processing," *Econometrica*, 82, 2257–2294.

## Appendix A. Beta-Bernoulli model and likelihood function of interpreted signals

At any time $t$, the belief of agent $i$ is represented by the Beta probability distribution with parameters $\alpha_{i,t}$ and $\beta_{i,t}$

$$f_{i,t}(\theta) = \begin{cases} \dfrac{\Gamma(\alpha_{i,t}+\beta_{i,t})}{\Gamma(\alpha_{i,t})\,\Gamma(\beta_{i,t})}\,\theta^{\alpha_{i,t}-1}(1-\theta)^{\beta_{i,t}-1} & \text{, for } 0<\theta<1 \\ 0 & \text{, otherwise,} \end{cases} \tag{9}$$

where $\Gamma(\cdot)$ is a Gamma function and the ratio of Gamma functions in the expression above is a normalization constant that ensures that the total probability integrates to 1. In this sense,

$$f_{i,t}(\theta) \propto \theta^{\alpha_{i,t}-1}(1-\theta)^{\beta_{i,t}-1}.$$

The idiosyncratic likelihood induced by the agent $i$'s interpretation of the public signal $s_{t+1}$ is

$$\ell_i(s_{t+1}|\theta) = \theta^{s^{(1)}_{i,t+1}}(1-\theta)^{s^{(0)}_{i,t+1}}$$

and, therefore, the standard Bayesian posterior is computed as

$$f_{i,t+1}(\theta|s_{t+1}) = \frac{\ell_i(s_{t+1}|\theta)\; f_{i,t}(\theta)}{\displaystyle\int_\Theta \ell_i(s_{t+1}|\theta)\; f_{i,t}(\theta)\, d\theta}.$$

Since the denominator of the expression above is just a normalizing constant, the posterior distribution is said to be proportional to the product of the prior distribution and the likelihood function as

$$\begin{aligned} f_{i,t+1}(\theta|s_{t+1}) &\propto \ell_i(s_{t+1}|\theta)\; f_{i,t}(\theta) \\ &\propto \theta^{\alpha_{i,t}+s^{(1)}_{i,t+1}-1}\,(1-\theta)^{\beta_{i,t}+s^{(0)}_{i,t+1}-1}. \end{aligned}$$

Therefore, the posterior distribution is

$$f_{i,t+1}(\theta) = \begin{cases} \dfrac{\Gamma(\alpha_{i,t+1}+\beta_{i,t+1})}{\Gamma(\alpha_{i,t+1})\,\Gamma(\beta_{i,t+1})}\,\theta^{\alpha_{i,t+1}-1}(1-\theta)^{\beta_{i,t+1}-1} & \text{, for } 0<\theta<1 \\ 0 & \text{, otherwise,} \end{cases}$$

where $\alpha_{i,t+1} = \alpha_{i,t} + s^{(1)}_{i,t+1}$ and $\beta_{i,t+1} = \beta_{i,t} + s^{(0)}_{i,t+1}$.

## Appendix B. Beta Distribution: Mode, Mean, Median

**Mode.** The mode of a random variable beta-distributed is the value that appears most often. It is the value $\theta$ at which its probability density function takes its maximum value. As per Equation (9), the mode $\theta^{mod}_{i,t}$, for any $i$ at any point in time $t$, is the $\arg\max_\theta f_{i,t}(\theta)$. Computed as

$$\frac{df_{i,t}}{d\theta} = \frac{\Gamma(\alpha_{i,t}+\beta_{i,t})}{\Gamma(\alpha_{i,t})\,\Gamma(\beta_{i,t})}\left[(\alpha_{i,t}-1)\theta^{\alpha_{i,t}-2}(1-\theta)^{\beta_{i,t}-1} - \theta^{\alpha_{i,t}-1}(\beta_{i,t}-1)(1-\theta)^{\beta_{i,t}-2}\right] = 0.$$

Implying that

$$(\alpha_{i,t}-1)\theta^{\alpha_{i,t}-2}(1-\theta)^{\beta_{i,t}-1} - \theta^{\alpha_{i,t}-1}(\beta_{i,t}-1)(1-\theta)^{\beta_{i,t}-2} = 0,$$

and therefore

$$\theta_{i,t}^{mod} = \begin{cases} \dfrac{\alpha_{i,t}-1}{\alpha_{i,t}+\beta_{i,t}-2} & \text{, for } \alpha_{i,t},\beta_{i,t} > 1 \\ 0 & \text{, for } \alpha_{i,t}=1,\beta_{i,t}>1 \\ 1 & \text{, for } \alpha_{i,t}>1,\beta_{i,t}=1 \\ \text{any value in } (0,1) & \text{, for } \alpha_{i,t},\beta_{i,t}=1 \end{cases} \tag{10}$$

**Mean.** The mean of a random variable Beta-distributed, denoted by $\theta_{i,t}^{mean}$ for any $i$ and $t$, is computed as follows

$$\begin{aligned} \theta_{i,t}^{mean} &= \int_0^1 \theta \frac{\Gamma(\alpha_{i,t}+\beta_{i,t})}{\Gamma(\alpha_{i,t})\,\Gamma(\beta_{i,t})}\theta^{\alpha_{i,t}-1}(1-\theta)^{\beta_{i,t}-1}d\theta \\ &= \frac{\Gamma(\alpha_{i,t}+\beta_{i,t})}{\Gamma(\alpha_{i,t})\,\Gamma(\beta_{i,t})}\int_0^1 \theta^{(\alpha_{i,t}+1)-1}(1-\theta)^{\beta_{i,t}-1}d\theta \\ &= \frac{\Gamma(\alpha_{i,t}+\beta_{i,t})}{\Gamma(\alpha_{i,t})\,\Gamma(\beta_{i,t})}\frac{\Gamma(\alpha_{i,t}+1)\,\Gamma(\beta_{i,t})}{\Gamma(\alpha_{i,t}+\beta_{i,t}+1)} \\ &= \frac{\Gamma(\alpha_{i,t}+\beta_{i,t})}{\Gamma(\alpha_{i,t})\,\Gamma(\beta_{i,t})}\frac{\alpha_{i,t}\Gamma(\alpha_{i,t})\,\Gamma(\beta_{i,t})}{(\alpha_{i,t}+\beta_{i,t})\Gamma(\alpha_{i,t}+\beta_{i,t})} = \frac{\alpha_{i,t}}{\alpha_{i,t}+\beta_{i,t}}. \end{aligned} \tag{11}$$

**Median.** There is no general closed formula for the median of the beta distribution for arbitrary values of the parameter $\alpha_{i,t}$ and $\beta_{i,t}$. The median, denoted by $\theta_{i,t}^{med}$, is the function that satisfies

$$\frac{\Gamma(\alpha_{i,t}+\beta_{i,t})}{\Gamma(\alpha_{i,t})\,\Gamma(\beta_{i,t})}\int_0^{\theta_{i,t}^{med}} \theta^{\alpha_{i,t}-1}(1-\theta)^{\beta_{i,t}-1} = \frac{1}{2}.$$

An accurate approximation of the value of the median of the beta distribution, for both $\alpha_{i,t}, \beta_{i,t} \geq 1$, is given by

$$\theta_{i,t}^{med} = \frac{\alpha_{i,t}-\frac{1}{3}}{\alpha_{i,t}+\beta_{i,t}-\frac{2}{3}}.^{9} \tag{12}$$

Therefore, if $1 < \alpha_{i,t} < \beta_{i,t}$, then $\theta_{i,t}^{mod} < \theta_{i,t}^{med} < \theta_{i,t}^{mean}$. If $1 < \beta_{i,t} < \alpha_{i,t}$, then the order of the inequalities is reversed. Finally, it is trivial to see that those three statistical measures are asymptotically equal as $\alpha_{i,t}, \beta_{i,t} \to \infty$.

[9]With relative error of less than 4%, rapidly decreasing to zero as both shape parameters increase.

## Appendix C. Auxiliary Lemmas

**Proof of Lemma 1.** In order to see how $W^t$ behaves as $t$ grows large, I rewrite $W$ using its diagonal decomposition. In particular, let $v$ be the squared matrix of left-hand eigenvectors of $W$ and $D = (d_1, d_2, \ldots, d_n)^\top$ the eigenvector of size $n$ associated to the unity eigenvalue $\lambda_1 = 1$. Without loss of generality, we assume the following normalization $\mathbf{1}^\top D = 1$. Therefore, $W = v^{-1}\Lambda v$, where $\Lambda = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$ is the squared matrix with eigenvalues on its diagonal, ranked in terms of absolute values, i.e. $|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_n|$. More generally, for any time $t$ we write

$$W^t = v^{-1}\Lambda^t v.$$

Since $v^{-1}$ has ones in all entries of its first column, it follows that

$$W^t_{ij} = d_j + \sum_r \lambda_r^t v^{-1}_{ir} v_{rj},$$

for each $r$, where $\lambda_r$ is the $r$-th largest eigenvalue of $W$. Therefore, $\lim_{t\to\infty} W^t_{ij} = D\mathbf{1}^\top$, i.e. each row of $W^t$ for all $t \geq \bar{t}$ converge to $D$, which coincides with the stationary distribution. Moreover, if the eigenvalues are ordered the way we have assumed, then $\|W^t - D\mathbf{1}^\top\| = o(|\lambda_2|^t)$, i.e. the convergence rate will be dictated by the second largest eigenvalue, as the others converge to zero more quickly as $t$ grows. ■

**Lemma 2.** *The opinion of every agent $i$ in any point in time $t$, $y_{i,t}$, can be written as*

$$y_{i,t} = \frac{\sum_{j=1}^n W^t_{ij}\alpha_{j,0} + bK(i,t)}{\sum_{j=1}^n W^t_{ij}\left(\alpha_{j,0} + \beta_{j,0}\right) + bL(i,t)},$$

*where* $K(i,t) = \sum_{k=0}^{t-1}\sum_{j=1}^{n} W^k_{ij} s^{(1)}_{j,t-k}$ *and* $L(i,t) = \sum_{k=0}^{t-1}\sum_{j=1}^{n} W^k_{ij}\left(s^{(0)}_{j,t-k} + s^{(1)}_{j,t-k}\right)$.

*Proof.* The update process of both parameters described by the equations (6) and (7) can be solved iteratively for any period $t$ as

$$\alpha_t = W^t\alpha_0 + \sum_{k=0}^{t-1} W^k B s^{(1)}_{t-k} \tag{13}$$

$$\beta_t = W^t\beta_0 + \sum_{k=0}^{t-1} W^k B s^{(0)}_{t-k}. \tag{14}$$

In agebraic formulation, we have that each entry of the vector in equation (13) can be written as

$$\begin{aligned}\alpha_{i,t} &= \sum_{j=1}^{n} W_{ij}^{t}\alpha_{j,0} + \sum_{k=0}^{t-1}\sum_{j=1}^{n} W_{ij}^{k} b s_{j,t-k}^{(1)} \\ &= \sum_{j=1}^{n} W_{ij}^{t}\alpha_{j,0} + b\sum_{k=0}^{t-1}\sum_{j=1}^{n} W_{ij}^{k} s_{j,t-k}^{(1)} \\ &= \sum_{j=1}^{n} W_{ij}^{t}\alpha_{j,0} + b\sum_{k=0}^{t-1}\sum_{j=1}^{n} W_{ij}^{k} s_{j,t-k}^{(1)} \\ &= \sum_{j=1}^{n} W_{ij}^{t}\alpha_{j,0} + b\, K(i,t). \end{aligned} \tag{15}$$

Similarly for the expression $\alpha_{i,t} + \beta_{i,t}$ using both equations (13) and (14) as follows

$$\begin{aligned}\alpha_{i,t} + \beta_{i,t} &= \sum_{j=1}^{n} W_{ij}^{t}\left(\alpha_{j,0} + \beta_{j,0}\right) + b\sum_{k=0}^{t-1}\sum_{j=1}^{n} W_{ij}^{k}\left(s_{j,t-k}^{(0)} + s_{j,t-k}^{(1)}\right) \\ &= \sum_{j=1}^{n} W_{ij}^{t}\left(\alpha_{j,0} + \beta_{j,0}\right) + b\, L(i,t). \end{aligned} \tag{16}$$

Therefore, from the definition of opinion we have that $y_{i,t} = \frac{\alpha_{i,t}}{\alpha_{i,t}+\beta_{i,t}}$ and the statement is proven. ■

**Lemma 3.** *Let $k \in [0,1]$, $X_1, X_2, \ldots, X_t$ be a sequence of i.n.i.d. random variables such that $\mathbb{P}(X_t \geq x) = p$ and $u_1, u_2, \ldots, u_t$ be i.i.d. $U[0,1]$ random variables. Moreover, assume that the pair $(X_t, u_t)$ is independent, for any $t$. In this case, the expressions $\mathbb{E}\left[\mathbb{1}\{u_t \leq \mathbb{1}\{X_t \geq x\}k\}\right]$ and $\mathbb{E}\left[\mathbb{1}\{u_t \leq \mathbb{E}\left[\mathbb{1}\{X_t \geq x\}\right]k\}\right]$ are equal.*

*Proof.* The first expression can be written as

$$\mathbb{E}\left[\mathbb{1}\{u_t \leq \mathbb{1}\{X_t \geq x\}k\}\right] = (1-p)\mathbb{E}\left[\mathbb{1}\{u_t \leq 0\}\right] + p\mathbb{E}\left[\mathbb{1}\{u_t \leq k\}\right] = pF_u(k) = pk.$$

The second expression simplifies to

$$\mathbb{E}\left[\mathbb{1}\{u_t \leq \mathbb{E}\left[\mathbb{1}\{X_t \geq x\}\right]k\}\right] = \mathbb{E}\left[\mathbb{1}\{u_t \leq (1-p)0 + pk\}\right] = \mathbb{E}\left[\mathbb{1}\{u_t \leq pk\}\right] = pk.$$

■

## Appendix D. Proofs of main propositions and corollaries

**Proof of Proposition 1.**

$$\begin{aligned}
\lim_{t\to\infty} y_{i,t} &= \lim_{t\to\infty} \frac{\alpha_{i,0} + \sum_{k=1}^{t} s_{i,k}^{(1)}}{\alpha_{i,0} + \sum_{k=1}^{t} s_{i,k}^{(1)} + \beta_{i,0} + \sum_{k=1}^{t} s_{i,k}^{(0)}} \\
&= \lim_{t\to\infty} \frac{\alpha_{i,0} + \sum_{k=1}^{t} \left(\mathbb{1}\{s_k = 1\} + \mathbb{1}\{s_k = a\}\mathbb{1}\{u_k \leq \psi_{i,k}\}\right)}{\alpha_{i,0} + \beta_{i,0} + \sum_{k=1}^{t} \left(\mathbb{1}\{s_k = 1\} + \mathbb{1}\{s_k = 0\} + \mathbb{1}\{s_k = a\}\right)} \\
&= \lim_{t\to\infty} \frac{\frac{\alpha_{i,0}}{t} + \frac{1}{t}\sum_{k=1}^{t} \left(\mathbb{1}\{s_k = 1\} + \mathbb{1}\{s_k = a\}\mathbb{1}\{u_k \leq \psi_{i,k}\}\right)}{\frac{\alpha_{i,0}+\beta_{i,0}}{t} + \frac{1}{t}\sum_{k=1}^{t} \left(\mathbb{1}\{s_k = 1\} + \mathbb{1}\{s_k = 0\} + \mathbb{1}\{s_k = a\}\right)} \\
&= \frac{\mathbb{E}_t\left[\mathbb{1}\{s_t = 1\}\right] + \mathbb{E}_t\left[\mathbb{1}\{s_t = a\}\right] \lim_{t\to\infty} \frac{1}{t}\sum_{k=1}^{t} \left(\mathbb{1}\{u_k \leq \psi_{i,k}\}\right)}{\mathbb{E}_t\left[\left(\mathbb{1}\{s_t = 1\}\right] + \mathbb{E}_t\left[\mathbb{1}\{s_t = 0\}\right] + \mathbb{E}_t\left[\mathbb{1}\{s_t = a\}\right]\right)} \\
&= \frac{(1-\delta)(1-\mu)\theta + (1-\delta)\mu \lim_{t\to\infty} \frac{1}{t}\sum_{k=1}^{t} \left(\mathbb{1}\{u_k \leq \psi_{i,k}\}\right)}{(1-\delta)} \\
&= (1-\mu)\theta + \mu \lim_{t\to\infty} \frac{1}{t} \sum_{k=1}^{t} \left(\mathbb{1}\{u_k \leq \gamma_{i,k}\, \mathbb{1}\{y_{i,k-1} \geq 0.5\} + (1-\gamma_{i,k})\, \mathbb{1}\{y_{i,k-1} < 0.5\}\}\right) \\
&= (1-\mu)\theta + \mu\mathbb{E}_t\left[\mathbb{1}\{u_t \leq \mathbb{E}_t\left[\gamma_{i,t}\, \mathbb{1}\{y_{i,t-1} \geq 0.5\} + (1-\gamma_{i,t})\, \mathbb{1}\{y_{i,t-1} < 0.5\}\}\right]\right] \\
&= (1-\mu)\theta + \mu\mathbb{E}_t\left[\mathbb{1}\{u_t \leq \mathbb{E}_t\left[\mathbb{1}\{y_{i,t-1} \geq 0.5\}\right] \mathbb{E}_t\left[2\gamma_{i,t} - 1\right] + 1 - \mathbb{E}_t\left[\gamma_{i,t}\right]\}\right]
\end{aligned}$$

The expression above might take two distinct forms because both rational learning and the repeated average process are Martingales. Thus, convergence is expected to be attained and random variable $\mathbb{E}_t\left[\mathbb{1}\{y_{i,t-1} \geq 0.5\}\right] = \mathbb{P}\left(y_{i,\infty} \geq 0.5\right)$ takes either value 1 with some positive probability $p \in (0,1)$ or 0 with probability $1-p$. For simplicity, say the realization 1 of such R.V. is called A and B otherwise. Therefore,

$$\begin{aligned}
\lim_{t\to\infty} y_{i,t} &= \begin{cases} (1-\mu)\theta + \mu\mathbb{E}_t\left[\mathbb{1}\{u_t \leq \mathbb{E}_t\left[2\gamma_{i,t} - 1\right] + 1 - \mathbb{E}_t\left[\gamma_{i,t}\right]\}\right] & , \text{if } A \\ (1-\mu)\theta + \mu\mathbb{E}_t\left[\mathbb{1}\{u_t \leq 1 - \mathbb{E}_t\left[\gamma_{i,t}\right]\}\right] & , \text{if } B \end{cases} \\
&= \begin{cases} (1-\mu)\theta + \mu\mathbb{E}_t\left[\mathbb{1}\{u_t \leq \bar{\gamma}_i\}\right] & , \text{if } A \\ (1-\mu)\theta + \mu\mathbb{E}_t\left[\mathbb{1}\{u_t \leq 1 - \bar{\gamma}_i\}\right] & , \text{if } B \end{cases} \\
&= \begin{cases} (1-\mu)\theta + \mu F_u\left(\bar{\gamma}_i\right) & , \text{if } A \\ (1-\mu)\theta + \mu F_u\left(1 - \bar{\gamma}_i\right) & , \text{if } B \end{cases} \\
&= \begin{cases} (1-\mu)\theta + \mu\bar{\gamma}_i & , \text{if } A \\ (1-\mu)\theta + \mu\left(1 - \bar{\gamma}_i\right) & , \text{if } B \end{cases}
\end{aligned} \tag{17}$$

■

**Proof of Proposition 2.** The claim is supported by the solution of two system of inequalities $S_1$ (for right-biased opinion) and $S_2$ (for left-biased opinion) below.

$$S_1 = \begin{cases} (1-\mu)\theta + \mu\bar{\gamma}_i > \frac{1}{2} \\ (1-\mu)\theta + \mu(1-\bar{\gamma}_i) > \frac{1}{2} \\ 0 < \mu \leq 1 \\ 0 \leq \theta \leq 1 \\ \frac{1}{2} < \bar{\gamma}_i \leq 1 \end{cases} \qquad S_2 = \begin{cases} (1-\mu)\theta + \mu\bar{\gamma}_i < \frac{1}{2} \\ (1-\mu)\theta + \mu(1-\bar{\gamma}_i) < \frac{1}{2} \\ 0 < \mu \leq 1 \\ 0 \leq \theta \leq 1 \\ \frac{1}{2} < \bar{\gamma}_i \leq 1 \end{cases}$$

The solution of those systems, together with the equation (17) in Proof of proposition 1 ensure the uniqueness of opinion types in the parameter spaces defined in the statement. ■

**Proof of Corollary 1.** From Proposition 1, we can write both right-biased and left-biased opinions as $\theta + \mu(\bar{\gamma}_i - \theta)$ and $\theta + \mu(1 - \bar{\gamma}_i - \theta)$, respectively, where the second term in each expression represents their respective biases. From those expressions, we can see that both sign and magnitude of those biases naturally depend on the relative size of $\theta$ and $\bar{\gamma}_i$. For both biases to be positive, we need $\theta < \min\{\bar{\gamma}_i, 1 - \bar{\gamma}_i\} = 1 - \bar{\gamma}_i$, since $\bar{\gamma}_i > \frac{1}{2}$. For both biases to be negative, we need $\theta > \max\{\bar{\gamma}_i, 1 - \bar{\gamma}_i\} = \bar{\gamma}_i$, since $\bar{\gamma}_i > \frac{1}{2}$. For the right-bias to be positive and the left-bias to be negative, we need $1 - \bar{\gamma}_i < \theta < \bar{\gamma}_i$ to hold. The case in which the right bias is negative while the right-bias is positive never holds, since we assume $\bar{\gamma}_i > \frac{1}{2}$. Therefore, we have the following summary.

(1) if $\theta < 1 - \bar{\gamma}_i$, then both biases are strictly positive
(2) if $1 - \bar{\gamma}_i < \theta < \bar{\gamma}_i$, then right-bias is strictly positive and left-bias is strictly negative
(3) if $\theta > \bar{\gamma}_i$, then both biases are strictly negative.

In the case (1) listed above, we say that the right-bias is less than the left bias whenever $\mu(\bar{\gamma}_i - \theta) < \mu(1 - \bar{\gamma}_i - \theta)$, meaning that $\bar{\gamma}_i < \frac{1}{2}$. However, this contradicts the assumption that individual is confirmatory and we can conclude that whenever $\theta < 1 - \bar{\gamma}_i$, the left-biased opinion is less biased than the right-biased one. In the case (3), we say that the right-bias is less than the left bias whenever $\mu(\theta - \bar{\gamma}_i) < \mu(\bar{\gamma}_i + \theta - 1)$, meaning that the statement is true if $\bar{\gamma}_i > \frac{1}{2}$. Therefore, if $\theta > \bar{\gamma}_i$, the right-biased opinion is less biased than the left-biased one. Finally, in the case (2), we say that the right-bias is less than the left bias whenever $\mu(\bar{\gamma}_i - \theta) < \mu(\bar{\gamma}_i + \theta - 1)$, meaning that it can only be true when $\theta > \frac{1}{2}$. These three arguments together prove the statement and we conclude that the right-bias is less than the left bias whenever $\theta > \frac{1}{2}$ (and vice-versa).

Finally, when $\theta = \frac{1}{2}$, the biases are equal since $|\bar{\gamma}_i - \frac{1}{2}| = |\frac{1}{2} - \bar{\gamma}_i|$ for any $\bar{\gamma}_i$. ■

**Proof of Corollary 2.** When an individual $j$ is always impartial, we have that

$$\begin{aligned}\psi_{j,t} &= \frac{1}{2}\,\mathbb{1}\{y_{j,t-1} \geq 0.5\} + \frac{1}{2}\,\mathbb{1}\{y_{j,t-1} < 0.5\} \\ &= \frac{1}{2}\,\mathbb{1}\{y_{j,t-1} \geq 0.5\} + \frac{1}{2}\,\left(1 - \mathbb{1}\{y_{j,t-1} \geq 0.5\}\right) \\ &= \frac{1}{2}, \end{aligned} \tag{18}$$

for all $t$. Since $u_t$ is a continuous $U\,[0,1]$ random variable in every period $t$, we have that

$$\mathbb{E}_t\left[\mathbb{1}\left\{u_t \leq \frac{1}{2}\right\}\right] = \mathbb{P}\left(u_t \leq \frac{1}{2}\right) = F_u\left(\frac{1}{2}\right) = \frac{\frac{1}{2} - 0}{1 - 0} = \frac{1}{2}, \tag{19}$$

where $F_u(\cdot)$ is the cumulative distribution function of $U\,[0,1]$. Thus, equations (17) and (19) together prove the statement when agents are impartial (both always impartial and moderately impartial). ■

**Proof of Proposition 3.** Say extreme opinion 1 (i.e. $y_{i,\infty} = 1$) is formed, then as per Propositions 1 and 2 we know this is the right-biased opinion and therefore it should be the case that $(1-\mu)\theta + \mu\bar{\gamma}_i = 1$. Conversely, say extreme opinion 0 (i.e. $y_{i,\infty} = 0$) is formed. Then, we know this is the left-biased opinion and it should be that $(1-\mu)\theta + \mu(1-\bar{\gamma}_i) = 0$. These two conditions together imply that $\mu(2\bar{\gamma}_i - 1) = 1$. If we generally consider that $0 \leq \mu \leq 1$ and $0 \leq \bar{\gamma}_i \leq 1$, then the relation $\mu(2\bar{\gamma}_i - 1) = 1$ is only met when $\mu = \bar{\gamma}_i = 1$. ■

**Proof of Corollary 3.** In measure theory, loosely speaking, a property is said to hold *almost everywhere* if, in a technical sense, the set for which the property holds takes up nearly all possibilities. The concept of *almost anywhere* can be thought of as the polar opposite case. In this sense, as per Proposition 2, if $(\theta, \mu) \in R$, then long-run learning implies that $(1-\mu)\theta + \mu\bar{\gamma}_i = \theta$. Therefore, since $\mu > 0$, the equality only holds when $\bar{\gamma}_i = \theta$ (almost anywhere). Likewise, if $(\theta, \mu) \in L$, then long-run learning implies that $(1-\mu)\theta + \mu(1-\bar{\gamma}_i) = \theta$ and the equality holds only when $\bar{\gamma}_i = 1 - \theta$ (almost anywhere). Conversely, if $(\theta, \mu) \in \mathcal{W}$ learning is a more stringent event because opinion type is a random variable and long-run learning is not a deterministic event even if the two previous conditions are met. In all three cases, since parameters are continuously distributed, there is opinion bias almost everywhere.

**Proof of Proposition 5.** As per Lemma 2 in the Appendix C, the limiting opinion of any agent $i$ can be written as

$$\lim_{t\to\infty} y_{i,t} = \lim_{t\to\infty} \frac{\frac{1}{t}\sum_{j=1}^{n} W_{ij}^{t}\alpha_{j,0} + b\frac{1}{t}K(i,t)}{\frac{1}{t}\sum_{j=1}^{n} W_{ij}^{t}\left(\alpha_{j,0}+\beta_{j,0}\right) + b\frac{1}{t}L(i,t)}$$
$$= \lim_{t\to\infty} \frac{\frac{1}{t}\sum_{k=0}^{t-1}\sum_{j=1}^{n} W_{ij}^{k} s_{j,t-k}^{(1)}}{\frac{1}{t}\sum_{k=0}^{t-1}\sum_{j=1}^{n} W_{ij}^{k}\left(s_{j,t-k}^{(0)} + s_{j,t-k}^{(1)}\right)}.$$

By Lemma 1 we can split both series in the numerator and denominator in two parts

$$\lim_{t\to\infty} y_{i,t} = \lim_{t\to\infty} \frac{\frac{1}{t}\left(\sum_{k=0}^{t_{\text{mix}}}\sum_{j=1}^{n} W_{ij}^{k} s_{j,t-k}^{(1)} + \sum_{k=t_{\text{mix}}+1}^{t-1}\sum_{j=1}^{n} W_{ij}^{k} s_{j,t-k}^{(1)}\right)}{\frac{1}{t}\left(\sum_{k=0}^{t_{\text{mix}}}\sum_{j=1}^{n} W_{ij}^{k}\left(s_{j,t-k}^{(0)} + s_{j,t-k}^{(1)}\right) + \sum_{k=t_{\text{mix}}+1}^{t-1}\sum_{j=1}^{n} W_{ij}^{k}\left(s_{j,t-k}^{(0)} + s_{j,t-k}^{(1)}\right)\right)}$$
$$= \lim_{t\to\infty} \frac{\frac{1}{t}\sum_{k=t_{\text{mix}}+1}^{t-1}\sum_{j=1}^{n} W_{ij}^{k} s_{j,t-k}^{(1)}}{\frac{1}{t}\sum_{k=t_{\text{mix}}+1}^{t-1}\sum_{j=1}^{n} W_{ij}^{k}\left(s_{j,t-k}^{(0)} + s_{j,t-k}^{(1)}\right)}.$$

Since the subindex $k$ spans from $t_{mix}$ onwards (i.e. when the chain is already mixed), we can use the invariant distribution matrix in the previous expression. Therefore the limiting opinion

becomes

$$\begin{aligned}
\lim_{t\to\infty} y_{i,t} &= \lim_{t\to\infty} \frac{\sum_{j=1}^{n} \Pi_{ij} \frac{1}{t} \sum_{k=t_{\text{mix}}+1}^{t-1} s_{j,t-k}^{(1)}}{\sum_{j=1}^{n} \Pi_{ij} \frac{1}{t} \sum_{k=t_{\text{mix}}+1}^{t-1} \left(s_{j,t-k}^{(0)} + s_{j,t-k}^{(1)}\right)} \\
&= \frac{\sum_{j=1}^{n} \Pi_{ij} \lim_{t\to\infty} \frac{t-1-t_{\text{mix}}}{t} \frac{1}{t-1-t_{\text{mix}}} \sum_{k=t_{\text{mix}}+1}^{t-1} s_{j,t-k}^{(1)}}{\sum_{j=1}^{n} \Pi_{ij} \lim_{t\to\infty} \frac{t-1-t_{\text{mix}}}{t} \frac{1}{t-1-t_{\text{mix}}} \sum_{k=t_{\text{mix}}+1}^{t-1} \left(s_{j,t-k}^{(0)} + s_{j,t-k}^{(1)}\right)} \\
&= \frac{\sum_{j=1}^{n} \Pi_{ij} \lim_{t\to\infty} \frac{1}{t-1-t_{\text{mix}}} \sum_{k=t_{\text{mix}}+1}^{t-1} \left(\mathbb{1}\{s_{t-k}=1\} + \mathbb{1}\{s_{t-k}=a\}\mathbb{1}\{u_{t-k} \le \psi_{j,t-k}\}\right)}{\sum_{j=1}^{n} \Pi_{ij} \lim_{t\to\infty} \frac{1}{t-1-t_{\text{mix}}} \sum_{k=t_{\text{mix}}+1}^{t-1} \left(\mathbb{1}\{s_{t-k}=0\} + \mathbb{1}\{s_{t-k}=1\} + \mathbb{1}\{s_{t-k}=a\}\right)} \\
&= \frac{\sum_j \Pi_{ij} \mathbb{E}_t \left[\mathbb{1}\{s_t=1\} + \mathbb{1}\{s_t=a\}\mathbb{1}\{u_t \le \psi_{j,t}\}\right]}{\sum_j \Pi_{ij} \mathbb{E}_t \left[\mathbb{1}\{s_t=0\} + \mathbb{1}\{s_t=1\} + \mathbb{1}\{s_t=a\}\right]} \\
&= \frac{(1-\delta)(1-\mu)\theta + (1-\delta)\mu \sum_j \Pi_{ij} \mathbb{E}_t \left[\mathbb{1}\{u_t \le \psi_{j,t}\}\right]}{(1-\delta)},
\end{aligned}$$

where the term $\mathbb{E}_t \left[\mathbb{1}\{u_t \le \psi_{j,t}\}\right]$ is as in Proposition 1, implying that the limiting consensus is

$$\lim_{t\to\infty} y_{i,t} = \begin{cases} (1-\mu)\theta + \mu \sum_j \Pi_{ij} \bar{\gamma}_j & \text{, if } A \\ (1-\mu)\theta + \mu \sum_j \Pi_{ij} (1-\bar{\gamma}_j) & \text{, if } B \end{cases}$$

■

**Proof of Proposition 4.** From Equation (15) in Appendix C, we know that $\alpha_{i,t}$, for any $i$, can be iterated forwardly as

$$\alpha_{i,t} = \sum_{j=1}^{n} W_{ij}^t \alpha_{j,0} + b \sum_{k=0}^{t-1} \sum_{j=1}^{n} W_{ij}^k s_{j,t-k}^{(1)}.$$

.

Similarly, the expression $\alpha_{i,t} + \beta_{i,t}$ in Equation (16) can be written as

$$\alpha_{i,t} + \beta_{i,t} = \sum_{j=1}^{n} W_{ij}^t \left(\alpha_{j,0} + \beta_{j,0}\right) + b \sum_{k=0}^{t-1} \sum_{j=1}^{n} W_{ij}^k \left(s_{j,t-k}^{(0)} + s_{j,t-k}^{(1)}\right).$$

Thus, if $b = 0$ (i.e. agents do not pay attention to signals) and for any $\delta$, the opinion of any agent $i \in N$ at any time $t$ boils down to

$$y_{i,t} = \frac{\sum_{j=1}^{n} W_{ij}^{t} \alpha_{j,0}}{\sum_{j=1}^{n} W_{ij}^{t} \left(\alpha_{j,0} + \beta_{j,0}\right)}$$

and therefore

$$\lim_{t\to\infty} y_{i,t} = y = \frac{\sum_{j=1}^{n} \Pi_{ij} \alpha_{j,0}}{\sum_{j=1}^{n} \Pi_{ij} \left(\alpha_{j,0} + \beta_{j,0}\right)}$$

for any $i$. Equivalently, if $\delta = 1$ (i.e. no signal enters into the network) and for any $b$, we have that $s_{i,t-k}^{(0)} = s_{i,t-k}^{(1)} = 0$ for any $i$ and $t$, since $\mathbb{1}\{s_t = \emptyset\} = 1$ for all $t$ as per Equations (2) and (3). In this case, the limiting opinion of any agent $i$ can be written as in the case when $b = 0$ shown above. ■

## Appendix E. Simulations statistics

E.1. **Tests concerning differences among $k$ proportions.** To decide whether observed differences among sample proportions are significant or whether they can be attributed to chance we must use tests concerning differences among proportions. For that, suppose that $x_1, x_2, \ldots, x_k$ are observed values of $k$ independent random variables $X_1, X_2, ..., X_k$ having binomial distributions with the parameters $n_1$ and $\theta_1$, $n_2$ and $\theta_2, \ldots, n_k$ and $\theta_k$. If the sample sizes are sufficiently large, we can approximate the distributions of the independent random variables

$$Z_i = \frac{X_i - n_i\theta_i}{\sqrt{n_i\theta_i(1-\theta_i)}} \qquad \text{for } i = 1, 2, \ldots, k$$

with standard normal distributions. Therefore, we know that we can look upon the test-statistic

$$\chi^2 = \sum_{i=1}^{k} Z_i^2 = \sum_{i=1}^{k} \frac{(x_i - n_i\theta_i)^2}{n_i\theta_i(1-\theta_i)}$$

as a value of a random variable having chi-square distribution with $k$ degrees of freedom. When the null hypothesis $H_0$ is $\theta_1 = \theta_2 = \cdots = \theta_k$ and the alternative hypothesis is that at least one of the $\theta$'s is different, we can use the *pooled estimate*

$$\hat{\theta} = \frac{\sum_{i=1}^{k} x_i}{\sum_{i=1}^{k} n_i}$$

and the test statistic becomes

$$\chi^2 = \sum_{i=1}^{k} \frac{(x_i - n_i\hat{\theta})^2}{n_i\hat{\theta}(1-\hat{\theta})}$$

a random variable whose value has chi-square distribution with $k - 1$ degrees of freedom because an estimate is substituted for the unknown parameter $\theta$.

TABLE 2. Summary statistics - simulated $\hat{p}$ and parameters ($\tau = 0$)

| Statistic | N | Mean | St. Dev. | Min | Pctl(25) | Median | Pctl(75) | Max |
|---|---|---|---|---|---|---|---|---|
| $\hat{p}$ in network (A) | 11,539 | 0.688 | 0.463 | 0 | 0 | 1 | 1 | 1 |
| $\hat{p}$ in network (B) | 11,539 | 0.688 | 0.463 | 0 | 0 | 1 | 1 | 1 |
| $\hat{p}$ in network (C) | 11,539 | 0.688 | 0.463 | 0 | 0 | 1 | 1 | 1 |
| $\hat{p}$ in network (D) | 11,539 | 0.688 | 0.463 | 0 | 0 | 1 | 1 | 1 |
| $\hat{p}$ in network (E) | 11,539 | 0.696 | 0.460 | 0 | 0 | 1 | 1 | 1 |
| $\hat{p}$ in network (F) | 11,539 | 0.688 | 0.463 | 0 | 0 | 1 | 1 | 1 |
| $\hat{p}$ in network (G) | 11,539 | 0.695 | 0.461 | 0 | 0 | 1 | 1 | 1 |
| $\hat{p}$ in network (H) | 11,539 | 0.694 | 0.461 | 0 | 0 | 1 | 1 | 1 |
| $\delta$ | 11,539 | 0.499 | 0.302 | 0.050 | 0.200 | 0.500 | 0.800 | 0.950 |
| $\mu$ | 11,539 | 0.652 | 0.205 | 0.232 | 0.475 | 0.663 | 0.830 | 0.950 |
| $\theta$ | 11,539 | 0.500 | 0.324 | 0.050 | 0.200 | 0.650 | 0.800 | 0.950 |
| $\mathcal{R}_0$ degree advantage in (B) | 11,539 | 1.164 | 0.625 | 0 | 0.5 | 1 | 2 | 2 |
| $\mathcal{R}_0$ degree advantage in (D) | 11,539 | 1.161 | 0.626 | 0 | 0.5 | 1 | 2 | 2 |
| $\mathcal{R}_0$ degree advantage in (E) | 11,539 | 1.321 | 0.998 | 0 | 0.3 | 1 | 1 | 3 |
| $\mathcal{R}_0$ degree advantage in (H) | 11,539 | 1.218 | 0.765 | 0 | 0.7 | 1 | 1.5 | 3 |
| $\mathcal{R}_0$ and $\mathcal{L}_0$ neighbors in (B) | 11,539 | 0.670 | 0.470 | 0 | 0 | 1 | 1 | 1 |
| $\mathcal{R}_0$ and $\mathcal{L}_0$ neighbors in (D) | 11,539 | 0.500 | 0.500 | 0 | 0 | 0 | 1 | 1 |
| $\mathcal{R}_0$ and $\mathcal{L}_0$ neighbors in (E) | 11,539 | 0.504 | 0.500 | 0 | 0 | 1 | 1 | 1 |
| $\mathcal{R}_0$ and $\mathcal{L}_0$ neighbors in (H) | 11,539 | 0.671 | 0.470 | 0 | 0 | 1 | 1 | 1 |

TABLE 3. Summary statistics - simulated $\hat{p}$ and parameters ($\tau = 1$)

| Statistic | N | Mean | St. Dev. | Min | Pctl(25) | Median | Pctl(75) | Max |
|---|---|---|---|---|---|---|---|---|
| $\hat{p}$ in network (A) | 4,604 | 0.716 | 0.451 | 0 | 0 | 1 | 1 | 1 |
| $\hat{p}$ in network (B) | 4,604 | 0.707 | 0.455 | 0 | 0 | 1 | 1 | 1 |
| $\hat{p}$ in network (C) | 4,604 | 0.730 | 0.444 | 0 | 0 | 1 | 1 | 1 |
| $\hat{p}$ in network (D) | 4,604 | 0.766 | 0.423 | 0 | 1 | 1 | 1 | 1 |
| $\hat{p}$ in network (E) | 4,604 | 0.709 | 0.454 | 0 | 0 | 1 | 1 | 1 |
| $\hat{p}$ in network (F) | 4,604 | 0.766 | 0.423 | 0 | 1 | 1 | 1 | 1 |
| $\hat{p}$ in network (G) | 4,604 | 0.725 | 0.447 | 0 | 0 | 1 | 1 | 1 |
| $\hat{p}$ in network (H) | 4,604 | 0.733 | 0.442 | 0 | 0 | 1 | 1 | 1 |
| $\delta$ | 4,604 | 0.501 | 0.302 | 0.050 | 0.200 | 0.500 | 0.800 | 0.950 |
| $\mu$ | 4,604 | 0.659 | 0.208 | 0.232 | 0.475 | 0.663 | 0.836 | 0.950 |
| $\theta$ | 4,604 | 0.503 | 0.324 | 0.050 | 0.200 | 0.650 | 0.800 | 0.950 |
| $\mathcal{R}_0$ degree advantage in (B) | 4,604 | 1.169 | 0.621 | 0 | 0.5 | 1 | 2 | 2 |
| $\mathcal{R}_0$ degree advantage in (D) | 4,604 | 1.178 | 0.622 | 0.500 | 0.500 | 1.000 | 2.000 | 2.000 |
| $\mathcal{R}_0$ degree advantage in (E) | 4,604 | 1.338 | 1.007 | 0 | 0.3 | 1 | 3 | 3 |
| $\mathcal{R}_0$ degree advantage in (H) | 4,604 | 1.217 | 0.772 | 0.333 | 0.500 | 1.000 | 1.500 | 3.000 |
| $\mathcal{R}_0$ and $\mathcal{L}_0$ neighbors in (B) | 4,604 | 0.658 | 0.474 | 0 | 0 | 1 | 1 | 1 |
| $\mathcal{R}_0$ and $\mathcal{L}_0$ neighbors in (D) | 4,604 | 0.498 | 0.500 | 0 | 0 | 0 | 1 | 1 |
| $\mathcal{R}_0$ and $\mathcal{L}_0$ neighbors in (E) | 4,604 | 0.507 | 0.500 | 0 | 0 | 1 | 1 | 1 |
| $\mathcal{R}_0$ and $\mathcal{L}_0$ neighbors in (H) | 4,604 | 0.664 | 0.472 | 0 | 0 | 1 | 1 | 1 |

TABLE 4. Summary statistics - simulated $\hat{p}$ and parameters ($\tau = 10$)

| Statistic | N | Mean | St. Dev. | Min | Pctl(25) | Median | Pctl(75) | Max |
|---|---|---|---|---|---|---|---|---|
| $\hat{p}$ in network (A) | 4,680 | 0.698 | 0.459 | 0 | 0 | 1 | 1 | 1 |
| $\hat{p}$ in network (B) | 4,680 | 0.608 | 0.488 | 0 | 0 | 1 | 1 | 1 |
| $\hat{p}$ in network (C) | 4,680 | 0.726 | 0.446 | 0 | 0 | 1 | 1 | 1 |
| $\hat{p}$ in network (D) | 4,680 | 0.680 | 0.466 | 0 | 0 | 1 | 1 | 1 |
| $\hat{p}$ in network (E) | 4,680 | 0.630 | 0.483 | 0 | 0 | 1 | 1 | 1 |
| $\hat{p}$ in network (F) | 4,680 | 0.787 | 0.410 | 0 | 1 | 1 | 1 | 1 |
| $\hat{p}$ in network (G) | 4,680 | 0.719 | 0.449 | 0 | 0 | 1 | 1 | 1 |
| $\hat{p}$ in network (H) | 4,680 | 0.633 | 0.482 | 0 | 0 | 1 | 1 | 1 |
| $\delta$ | 4,680 | 0.504 | 0.301 | 0.050 | 0.200 | 0.500 | 0.800 | 0.950 |
| $\mu$ | 4,680 | 0.658 | 0.206 | 0.232 | 0.475 | 0.663 | 0.830 | 0.950 |
| $\theta$ | 4,680 | 0.490 | 0.323 | 0.050 | 0.200 | 0.350 | 0.800 | 0.950 |
| $\mathcal{R}_0$ degree advantage in (B) | 4,680 | 1.164 | 0.624 | 0.500 | 0.500 | 1.000 | 2.000 | 2.000 |
| $\mathcal{R}_0$ degree advantage in (D) | 4,680 | 1.172 | 0.623 | 0 | 0.5 | 1 | 2 | 2 |
| $\mathcal{R}_0$ degree advantage in (E) | 4,680 | 1.327 | 0.995 | 0 | 1 | 1 | 1 | 3 |
| $\mathcal{R}_0$ degree advantage in (H) | 4,680 | 1.208 | 0.753 | 0.333 | 0.667 | 1.000 | 1.500 | 3.000 |
| $\mathcal{R}_0$ and $\mathcal{L}_0$ neighbors in (B) | 4,680 | 0.668 | 0.471 | 0 | 0 | 1 | 1 | 1 |
| $\mathcal{R}_0$ and $\mathcal{L}_0$ neighbors in (D) | 4,680 | 0.504 | 0.500 | 0 | 0 | 1 | 1 | 1 |
| $\mathcal{R}_0$ and $\mathcal{L}_0$ neighbors in (E) | 4,680 | 0.495 | 0.500 | 0 | 0 | 0 | 1 | 1 |
| $\mathcal{R}_0$ and $\mathcal{L}_0$ neighbors in (H) | 4,680 | 0.671 | 0.470 | 0 | 0 | 1 | 1 | 1 |

TABLE 5. Summary statistics - simulated $\hat{p}$ and parameters ($\tau = 30$)

| Statistic | N | Mean | St. Dev. | Min | Pctl(25) | Median | Pctl(75) | Max |
|---|---|---|---|---|---|---|---|---|
| $\hat{p}$ in network (A) | 4,554 | 0.678 | 0.467 | 0 | 0 | 1 | 1 | 1 |
| $\hat{p}$ in network (B) | 4,554 | 0.590 | 0.492 | 0 | 0 | 1 | 1 | 1 |
| $\hat{p}$ in network (C) | 4,554 | 0.717 | 0.451 | 0 | 0 | 1 | 1 | 1 |
| $\hat{p}$ in network (D) | 4,554 | 0.648 | 0.478 | 0 | 0 | 1 | 1 | 1 |
| $\hat{p}$ in network (E) | 4,554 | 0.602 | 0.489 | 0 | 0 | 1 | 1 | 1 |
| $\hat{p}$ in network (F) | 4,554 | 0.794 | 0.405 | 0 | 1 | 1 | 1 | 1 |
| $\hat{p}$ in network (G) | 4,554 | 0.718 | 0.450 | 0 | 0 | 1 | 1 | 1 |
| $\hat{p}$ in network (H) | 4,554 | 0.559 | 0.497 | 0 | 0 | 1 | 1 | 1 |
| $\delta$ | 4,554 | 0.499 | 0.299 | 0.050 | 0.200 | 0.500 | 0.800 | 0.950 |
| $\mu$ | 4,554 | 0.656 | 0.207 | 0.232 | 0.475 | 0.663 | 0.830 | 0.950 |
| $\theta$ | 4,554 | 0.501 | 0.324 | 0.050 | 0.200 | 0.650 | 0.800 | 0.950 |
| $\mathcal{R}_0$ degree advantage in (B) | 4,554 | 1.168 | 0.623 | 0.500 | 0.500 | 1.000 | 2.000 | 2.000 |
| $\mathcal{R}_0$ degree advantage in (D) | 4,554 | 1.142 | 0.618 | 0 | 0.5 | 1 | 2 | 2 |
| $\mathcal{R}_0$ degree advantage in (E) | 4,554 | 1.357 | 1.017 | 0.333 | 0.333 | 1.000 | 3.000 | 3.000 |
| $\mathcal{R}_0$ degree advantage in (H) | 4,554 | 1.214 | 0.761 | 0.333 | 0.500 | 1.000 | 2.000 | 3.000 |
| $\mathcal{R}_0$ and $\mathcal{L}_0$ neighbors in (B) | 4,554 | 0.665 | 0.472 | 0 | 0 | 1 | 1 | 1 |
| $\mathcal{R}_0$ and $\mathcal{L}_0$ neighbors in (D) | 4,554 | 0.502 | 0.500 | 0 | 0 | 1 | 1 | 1 |
| $\mathcal{R}_0$ and $\mathcal{L}_0$ neighbors in (E) | 4,554 | 0.515 | 0.500 | 0 | 0 | 1 | 1 | 1 |
| $\mathcal{R}_0$ and $\mathcal{L}_0$ neighbors in (H) | 4,554 | 0.658 | 0.474 | 0 | 0 | 1 | 1 | 1 |

TABLE 6. Two Population Proportions - Result 6

| $i$ | $j$ | $c_i$ | $c_j$ | $\hat{p}_i(c_i)$ | $\hat{p}_j(c_j)$ | $CI_{5\%}$ | $CI_{95\%}$ | $\chi^2$ | p-value |
|---|---|---|---|---|---|---|---|---|---|
| (A) | (A) | $\tau = 0$ | $\tau = 30$ | 0.688 | 0.678 | -0.006 | 0.026 | 1.641 | 0.2 |
| (B) | (B) | $\tau = 0$ | $\tau = 30$ | 0.688 | 0.59 | 0.082 | 0.115 | 140.542 | 0 |
| (D) | (D) | $\tau = 0$ | $\tau = 30$ | 0.688 | 0.648 | 0.024 | 0.056 | 23.694 | 0 |
| (B) | (D) | $\tau = 1$ | $\tau = 1$ | 0.707 | 0.766 | -0.077 | -0.041 | 41.405 | 0 |
| (B) | (D) | $\tau = 30$ | $\tau = 30$ | 0.59 | 0.648 | -0.078 | -0.039 | 32.948 | 0 |

TABLE 7. Correlation Matrix - Probability of efficient consensus ($\tau = 0$)

| | (A) | (B) | (C) | (D) | (E) | (F) | (G) | (H) |
|---|---|---|---|---|---|---|---|---|
| (A) | 1 | 1 | 1 | 1 | 0.981 | 1 | 0.979 | 0.985 |
| (B) | 1 | 1 | 1 | 1 | 0.981 | 1 | 0.979 | 0.985 |
| (C) | 1 | 1 | 1 | 1 | 0.981 | 1 | 0.979 | 0.985 |
| (D) | 1 | 1 | 1 | 1 | 0.981 | 1 | 0.979 | 0.985 |
| (E) | 0.981 | 0.981 | 0.981 | 0.981 | 1 | 0.981 | 0.981 | 0.995 |
| (F) | 1 | 1 | 1 | 1 | 0.981 | 1 | 0.979 | 0.985 |
| (G) | 0.979 | 0.979 | 0.979 | 0.979 | 0.981 | 0.979 | 1 | 0.983 |
| (H) | 0.985 | 0.985 | 0.985 | 0.985 | 0.995 | 0.985 | 0.983 | 1 |

TABLE 8. Correlation Matrix - Probability of efficient consensus ($\tau = 1$)

| | (A) | (B) | (C) | (D) | (E) | (F) | (G) | (H) |
|---|---|---|---|---|---|---|---|---|
| (A) | 1 | 0.555 | 0.855 | 0.540 | 0.601 | 0.672 | 0.803 | 0.571 |
| (B) | 0.555 | 1 | 0.578 | 0.464 | 0.509 | 0.513 | 0.567 | 0.481 |
| (C) | 0.855 | 0.578 | 1 | 0.610 | 0.635 | 0.784 | 0.892 | 0.624 |
| (D) | 0.540 | 0.464 | 0.610 | 1 | 0.480 | 0.625 | 0.611 | 0.523 |
| (E) | 0.601 | 0.509 | 0.635 | 0.480 | 1 | 0.565 | 0.633 | 0.496 |
| (F) | 0.672 | 0.513 | 0.784 | 0.625 | 0.565 | 1 | 0.774 | 0.612 |
| (G) | 0.803 | 0.567 | 0.892 | 0.611 | 0.633 | 0.774 | 1 | 0.619 |
| (H) | 0.571 | 0.481 | 0.624 | 0.523 | 0.496 | 0.612 | 0.619 | 1 |

TABLE 9. Correlation Matrix - Probability of efficient consensus ($\tau = 10$)

| | (A) | (B) | (C) | (D) | (E) | (F) | (G) | (H) |
|---|---|---|---|---|---|---|---|---|
| (A) | 1 | 0.304 | 0.811 | 0.194 | 0.361 | 0.607 | 0.780 | 0.252 |
| (B) | 0.304 | 1 | 0.339 | 0.130 | 0.214 | 0.276 | 0.341 | 0.151 |
| (C) | 0.811 | 0.339 | 1 | 0.203 | 0.410 | 0.702 | 0.899 | 0.261 |
| (D) | 0.194 | 0.130 | 0.203 | 1 | 0.121 | 0.245 | 0.191 | 0.139 |
| (E) | 0.361 | 0.214 | 0.410 | 0.121 | 1 | 0.314 | 0.433 | 0.155 |
| (F) | 0.607 | 0.276 | 0.702 | 0.245 | 0.314 | 1 | 0.690 | 0.230 |
| (G) | 0.780 | 0.341 | 0.899 | 0.191 | 0.433 | 0.690 | 1 | 0.267 |
| (H) | 0.252 | 0.151 | 0.261 | 0.139 | 0.155 | 0.230 | 0.267 | 1 |

TABLE 10. Correlation Matrix - Probability of efficient consensus ($\tau = 30$)

| | (A) | (B) | (C) | (D) | (E) | (F) | (G) | (H) |
|---|---|---|---|---|---|---|---|---|
| (A) | 1 | 0.259 | 0.801 | 0.130 | 0.297 | 0.581 | 0.703 | 0.135 |
| (B) | 0.259 | 1 | 0.296 | 0.045 | 0.137 | 0.220 | 0.288 | 0.071 |
| (C) | 0.801 | 0.296 | 1 | 0.141 | 0.338 | 0.668 | 0.828 | 0.152 |
| (D) | 0.130 | 0.045 | 0.141 | 1 | 0.089 | 0.172 | 0.140 | 0.036 |
| (E) | 0.297 | 0.137 | 0.338 | 0.089 | 1 | 0.256 | 0.368 | 0.043 |
| (F) | 0.581 | 0.220 | 0.668 | 0.172 | 0.256 | 1 | 0.654 | 0.132 |
| (G) | 0.703 | 0.288 | 0.828 | 0.140 | 0.368 | 0.654 | 1 | 0.159 |
| (H) | 0.135 | 0.071 | 0.152 | 0.036 | 0.043 | 0.132 | 0.159 | 1 |

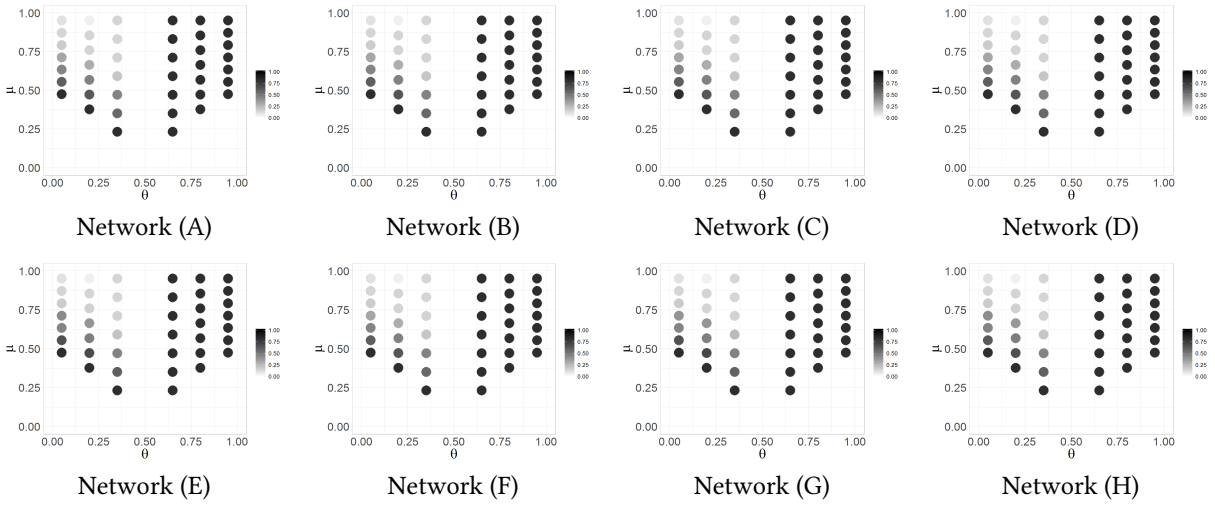



FIGURE 5. Simulated frequency $\hat{p}$ in $\Theta \times M$: common prior case ($\tau = 0$)

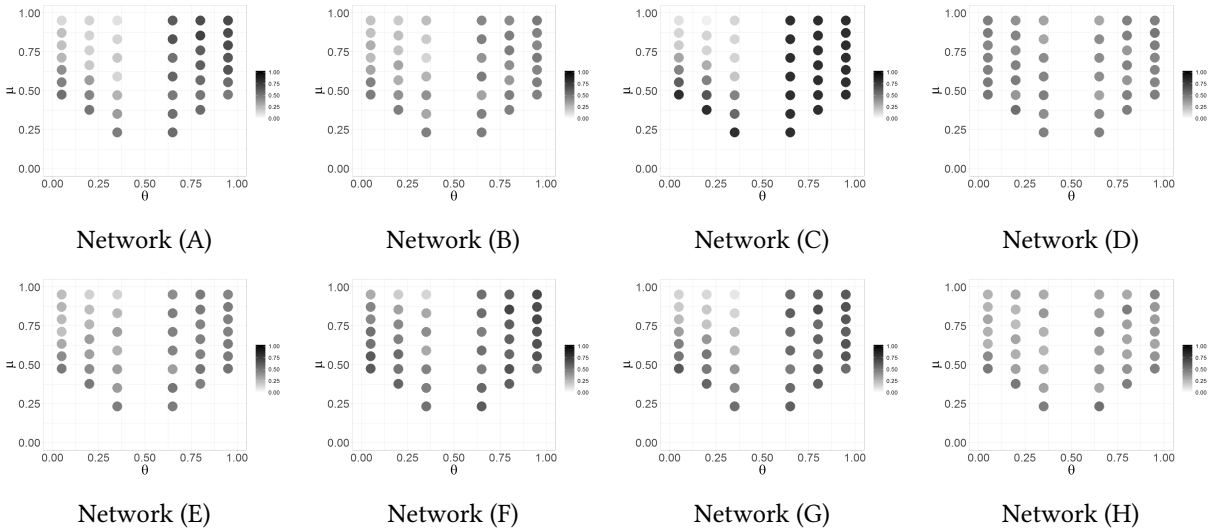



FIGURE 6. Simulated frequency $\hat{p}$ in $\Theta \times M$: heterogeneous priors case ($\tau = 30$)