

TRUTH AND CONFORMITY ON NETWORKS

AYDIN MOHSENI, COLE RANDAL WILLIAMS

ABSTRACT. Typically, public discussions of questions of social import exhibit two important properties: (1) they are influenced by conformity bias, and (2) the influence of conformity is expressed via social networks. We examine how social learning on networks proceeds under the influence of conformity bias. In our model, heterogeneous agents express public opinions where those expressions are driven by the competing priorities of accuracy and of conformity to one’s peers. Agents learn, by Bayesian conditionalization, from private evidence from nature, and from the public declarations of other agents. Our key findings are that networks that produce configurations of social relationships that sustain a diversity of opinions empower honest communication and reliable acquisition of true beliefs, and that the networks that do this best turn out to be those which are both less centralized and less connected.

1. INTRODUCTION

Epistemology is the study of true, justified, or reliable beliefs. Social epistemology is the study of the effect of social structures and interactions on the emergence and maintenance of such beliefs. Much of the recent work in social epistemology has been in the application of game-theoretic techniques to modeling communities of rational inquirers—epistemic communities—to see what incentive and interaction structures are conducive to the outcomes of accuracy,¹ efficiency,² and equity.³

One of the most important domains of social inquiry is that of broad public discourse. Which social policy will lead to better outcomes? Which political candidate is more qualified for office? Typically, public discussion on such questions of import is influenced by the human tendency of conformity. Individual decisions are informed and influenced by peers; the presence of conformist bias in social discourse is well-studied, and well-supported.⁴

We present a model of social inquiry where it exhibits two properties endemic to matters of public discussion: (1) individuals are subject to varying degrees to conformity bias, and (2) the influence of the pressure to conformity is expressed via social networks. We examine how the structure of social ties in tandem with conformity bias can influence the flow and reliability of information in matters of public opinion.

Date: August 20, 2019.

¹ See Zollman (2007, 2009, 2013), Mayo-Wilson et al. (2013), and Rosenstock et al. (2017).

² See Heesen (2017), Kitcher (1990, 1993), and Strevens (2003, 2013).

³ See O’Connor, Bright, and Bruner (O’Connor et al.), and Bruner and O’Connor (2016)

⁴ See Asch (1955), Bond and Smith (1996), and Morganand and Laland (2012).

In our model, heterogeneous agents express public opinions where those expressions are driven by the competing priorities of accuracy and of conformity. Agents learn, by Bayesian conditionalization, from private evidence from nature, and from the public declarations of other agents.

Several key findings emerge. We see that the most influential public declarations are made by agents when they go against the consensus of their neighbors, but that the most informative declarations, on average, are made by agents when their social influences are balanced. This provides a unifying explanation for our results: networks that produce configurations of social relationships that sustain a diversity of opinions empower honest communication and hence reliable acquisition of the truth.

In related literature on network epistemology (Zollman 2007, 2013), less connected networks are shown, under the right conditions, to increase the reliability of inquiry. In those cases, greater connectivity can cause premature “lock-in” to consensus in epistemic communities dealing with an exploration-exploitation trade-off. We arrive at a similar moral by different means.

We show that networks are differentially conducive to informative communication depending on the degree to which a community is divided in its publicly stated opinions. When communities are most divided, more connected networks, such as complete networks, do best. Whereas, when communities are near consensus, less connected networks exhibiting low degree-centrality, such as circle networks, are optimal.

Across the networks literature, star networks have been shown to possess certain optimality properties: they emerge as the product of various processes of strategic network formation (Goeree et al. 2009; Barrett et al. 2017), can lead to efficient division of cognitive labor (Goyal 2007; Zollman 2013), and can provide optimal conditions for information dissemination (Goeree et al. 2009). In contrast, we find that, in the presence of a modicum of conformity bias, star networks produce to the worst possible conditions for social learning.

In §2, we explain our model. In §3, we present the long run success of learning in the presence of conformist bias. In §4, we present simulations illustrating our central results. In §5, we provide an analysis of the deeper patterns that unify and explain our results. In §6, we conclude.

2. THE MODEL: CAESAR OR POMPEIA?

To animate our model, let us consider an anachronistic allegorical vignette. A community of Roman citizens has come together to discuss which candidate is better qualified for office. The candidates are Caesar and Pompeia. In discussing their beliefs, the citizens are influenced, to varying degrees, by two competing motivations: the motivation to say, honestly, who they believe is the better qualified candidate, and the motivation

to agree with their neighbors, or, more particularly, those with whom they share social or economic ties.

Each citizen varies with respect to the weight she places on each honesty and conformity. On one extreme, we may find Titus the Truth-Teller, who speaks his mind, come what may. Titus has come to believe—both from what is public knowledge, and from his own private information and experiences—that it is Pompeia who is more likely to be a better candidate. And so he declares as much, and he does so without any thought or worry concerning the impact of his declaration on the social regard of his peers.

On the other extreme, we find Cassius the Conformist, for whom harmony with peers is his sole concern. Making his true beliefs known does not enter the picture. Whichever candidate his peers favor, Cassius favors. Now, it happens to be Caesar.

Most of the remaining citizens, however, are not so extreme in their dispositions, but rather fall somewhere between Titus and Cassius. They care about making their true beliefs known, to some degree, and also about harmony with their peers, to some degree. Most make their declarations in a way that is contingent both on the strength of their beliefs, and on the weight of the social pressures around them.

Imagine that there are two states of the world: θ and $-\theta$. We can think of these as corresponding to where one social policy will lead to better outcomes, or one political candidate will be better suited to office.

Agents are interested in learning the true state of the world. This proceeds in two ways: (1) They get private evidence $\sigma \sim f_\theta(\sigma)$ from Nature; we can think of these as hearing some piece of news, or reasoning about an argument. And (2), they observe the public declarations $\mathbf{x}_{-i} \in \{\theta, -\theta\}^{N-1}$ of other agents across the network. Declarations indicate to others the state a declaring agent ostensibly believes to be true.

For each agent i , her payoffs are a convex combination of her *truth-seeking* orientation α_i and desire for *conformity* to her neighbors $(1 - \alpha_i)$. Her payoff for a declaration $x \in \{\theta, -\theta\}$ then is given by

$$U_i(x) = \alpha_i P_i(x) + (1 - \alpha_i) N_i(x)$$

where $P_i(x)$ is agent i 's expectation of the truth of x given her current information, and $N_i(x)$ is the proportion of her neighbors that have also declared x .⁵

We can think of an agent i as engaged in two games simultaneously which determine her payoffs in proportion to her type: a Bayesian learning game that contributes α_i of her payoff, where the data are the agent's private evidence σ and others' public declarations \mathbf{x}_{-i} , and an n_i -player pure coordination game that constitutes the remaining $(1 - \alpha_i)$ of her payoff, where n_i is the count of agent i 's neighbors.

⁵ Note that an agent's truth-seeking payoff for a declaration is based on her expectation that it corresponds to the true state of the world—agents do not know, and do not find out, whether their assessments are accurate.

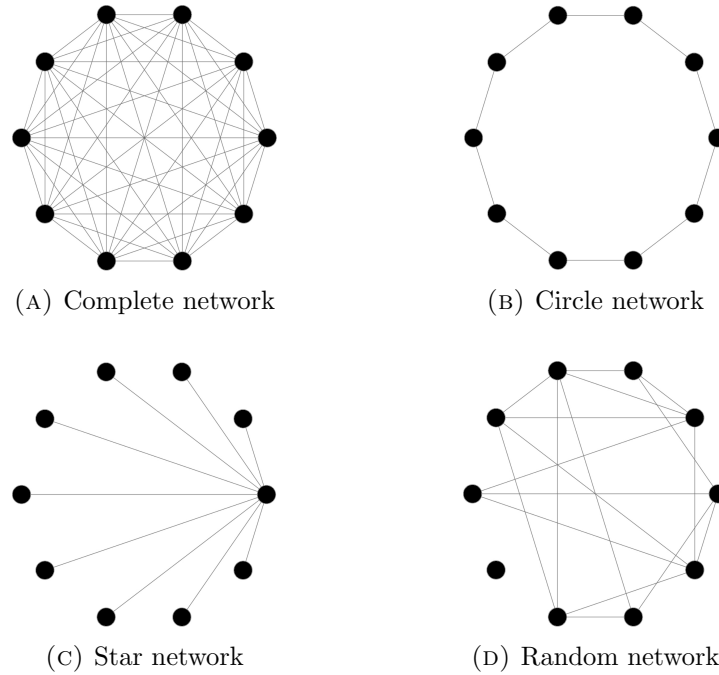


FIGURE 1. Social networks with 10 agents.

Our epistemic community of N agents inhabits a society where their patterns of shared social influence are described by a network. Here, nodes represent agents, and neighbors are connected by edges. Standard networks include complete, circle, star, and random networks (see FIGURE 1).

Networks vary with respect to the patterns of social influence they capture. The complete network describes a social structure in which each agent has social ties with every other. The circle describes a social structure in which each agent shares social ties with exactly two other individuals. Note that complete and circle networks are special cases of regular networks,⁶ where the regular network is of degree $N - 1$ and degree 2, respectively. In contrast, the star network describes a centralized social structure, where one individual (a central agent) has far more connections than the rest (the peripheral agents), who are otherwise socially isolated.

Before the game, agent types (truth-seeking/conformity orientations) are drawn from a continuous distribution:

$$\alpha_1, \dots, \alpha_N \stackrel{iid}{\sim} G \text{ with } \text{supp}(G) = [0, 1]$$

And Nature randomly chooses the state of the world to be θ or $-\theta$. Each state of the world induces a distinct distribution from which evidence $\sigma \sim f_\theta(\sigma)$ may be drawn.

⁶ Regular networks are those in which all nodes are of the same degree, or number of edges. Here, this will correspond to all agents having the same number of neighbors.

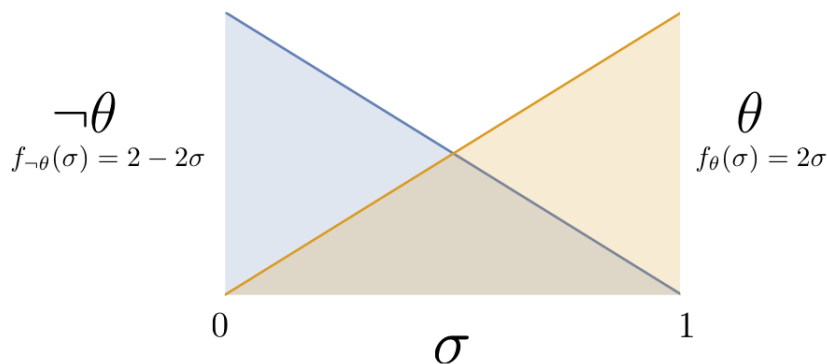


FIGURE 2. Distributions $f_\theta(\sigma)$, $f_{-\theta}(\sigma)$, of evidence σ , for each state of the world θ and $-\theta$.

The distributions $f_\theta(\sigma) = 2\sigma$ and $f_{-\theta}(\sigma) = 2 - 2\sigma$, depicted in FIGURE 2, are used in our simulations due to their convenient functional form. More generally, however, the distributions need only satisfy: *mutual absolute continuity* and *unbounded evidence*. Mutual absolute continuity requires that both distributions agree on what subsets of possible evidence have positive probability, meaning that no single piece of evidence can falsify one or the other. And unbounded evidence give us that evidence has the potential, in principle, to make one arbitrarily (though not completely) confident of either state. We take this to be a reasonable assumption, as we want to allow that, for any degree of belief shy of absolute certainty, there can—in principle—exist some evidence, however unlikely, which is sufficiently compelling to produce that belief.

In each round, an agent is chosen at random to receive private evidence from Nature, and to make a public declaration to be observed by the network. Upon receiving her evidence, an agent updates her beliefs, via conditionalization, about the true state of the world. This is done in the normal way, using Bayes' rule

$$P(\theta|\sigma, \mathbf{h}^t) = \frac{P(\sigma|\theta)P(\theta|\mathbf{h}^t)}{P(\sigma|\theta)P(\theta|\mathbf{h}^t) + P(\sigma|-\theta)P(-\theta|\mathbf{h}^t)}$$

where $P(\sigma|\theta)$ is the likelihood of her new evidence σ given the state θ , and $P(\theta|\mathbf{h}^t)$ is her prior on θ given the history of declarations at that time \mathbf{h}^t . Note that $P(\theta|\mathbf{h}^t)$ is also the public belief at that time—the shared portion of individual beliefs about the true state of the world constituted by the history of learning from public declarations. To simplify exposition, assume the population begins with ignorance priors.⁷

Next, the agent calculates her utilities, given her truth-seeking orientation, chooses her best response as a function of her private evidence and public prior (which, together,

⁷ An *ignorance prior* is a probability distribution assigning equal probability to all possibilities. Our proofs will require only non-degenerate priors, and our simulations will employ a range of priors.

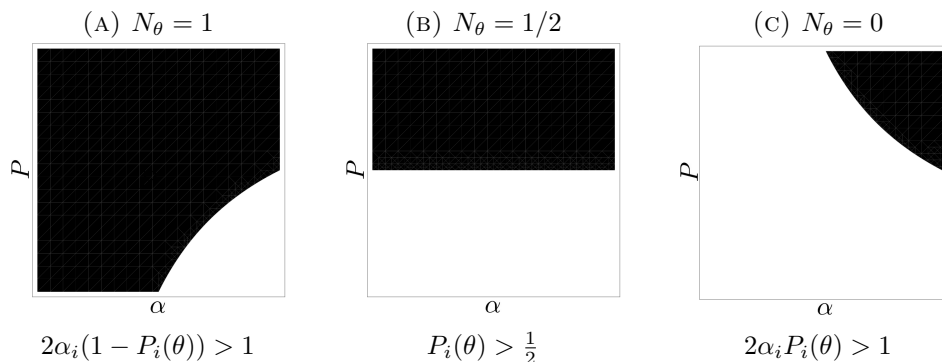


FIGURE 3. What is inferred from agent i 's declaration of θ , as captured by condition (\dagger) , about her posterior belief $P_i(\theta)$ and truth-seeking orientation α_i , when different proportions of her neighbors $N_i(\theta) = 1, 1/2, 0$ are making the same declaration.

form her posterior probability $P(\theta|\sigma, \mathbf{h}^t)$ over θ), and the composition of her neighbors:⁸

$$BR_i(\sigma, N_i(x)) = \arg \max\{U_i(\theta), U_i(-\theta)\}.$$

Following this, the other agents in the network observe her declaration, and update their beliefs. To do so, they must consider the likelihood of her having made her declaration given the composition of her neighbors, her likely evidence, her possible truth-seeking orientations, and their own prior beliefs about the state of the world.

So, what precisely do agents learn from one another's declarations? Well, when agent i declares $x = \theta$, others know that it was her best response to do so. It follows that

$$\begin{aligned} U_i(\theta) &> U_i(-\theta) \\ \alpha_i P_i(\theta) + (1 - \alpha) N_i(\theta) &> \alpha_i(1 - P_i(\theta)) + (1 - \alpha_i)(1 - N_i(\theta)) \\ \alpha_i(2P_i(\theta) - 1) + (1 - \alpha_i)(2N_i(\theta) - 1) &> 0. \end{aligned} \quad (\dagger)$$

We can get an intuitive grasp of this inequality (\dagger) by considering fixed values of the proportion of the declaring agent's neighbors who are also declaring θ (depicted in FIGURE 3.). The shaded area captures the values of agent i 's truth-seeking orientation α_i (on the horizontal axis), and posterior belief P_i (on the vertical axis), that are compatible with her having declared θ . That is, the region in which (\dagger) is satisfied.

Consider the $N_i(\theta) = 1$ case (FIGURE 3A). This is where all of the focal agent's neighbors are also declaring θ . Here, we see that a broad range of beliefs and truth-seeking orientations are compatible with her having declared θ . What can be ruled out (the area in white) is that it was *not* the case that she was both highly truth-seeking and strongly believed in the truth of θ . Here, others do not learn much from observing the focal agent's declaration.

⁸ In the case of payoff ties, the agent chooses among her best responses at random.

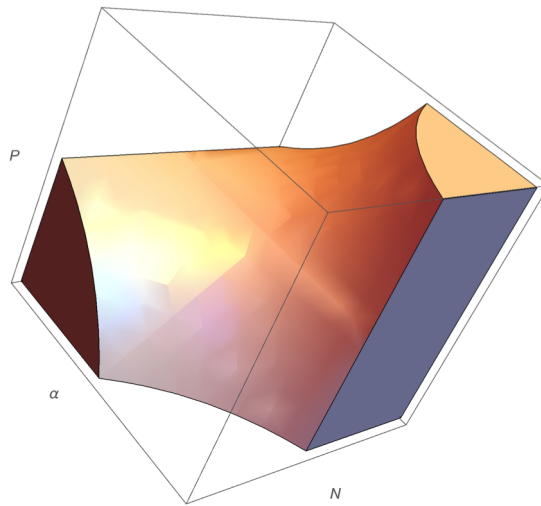


FIGURE 4. The content of an agent’s declaration, visualized as a hyperboloid—corresponding to (†)—of the values of N_θ , α_i , and $P_i(\theta)$ for which she could have made her declaration.

Consider the $N_i(\theta) = 1/2$ case (FIGURE 3B). This is where the focal agent’s neighbors are evenly split; half declaring θ and half $-\theta$. Here, the other agents infer that the focal agent’s social influences are balanced, and so her truth-seeking orientation α_i is no longer relevant. Her declaration is now determined purely by her posterior belief. If $P_i(\theta) > 1/2$, then she would make the declaration she did, if not, she would not. Here, others learn the direction of the focal agent’s belief, but not much about its strength.

Next, consider the $N_i(\theta) = 0$ case (FIGURE 3C). This is where none of the focal agent’s neighbors are declaring θ . Here, we see that only a narrow range of beliefs and truth-seeking orientations are compatible with her declaration of θ . It must have been the case that she was both highly truth-seeking, and possess a strong belief in the truth of θ . Now, others learn both the direction and strength of the focal agent’s belief, and through it about the strength of her evidence.

What happens when we put all this together? The hyperboloid in FIGURE 4 gives us the delimited domain of the likelihood of an agent’s declaration. This captures the reasoning we just covered as to the qualitative inferences agents make about one another’s beliefs from their declarations. From this, the agents in the population update their beliefs about the state of the world, in the normal way, using Bayes’ rule. (See APPENDIX A for the mathematical details.)

In this way, rational agents learn from their own private evidence, the declarations of other agents in the network, and the public belief about the true state evolves through discussion and across the network.

3. TRUTH IN THE LONG RUN

Our primary interest lies in dynamical analysis of the short-to-medium-run behavior of social inquiry under conformist bias. Before we proceed to this analysis, however, it may help us in this to understand the long-run trajectory of social learning under conformity. What we find is that, in the long run, irrespective of social structure or conformist bias, epistemic communities like the ones we have described will converge to believing in, and publicly declaring, the true state of the world.

More precisely, given any social network, unbounded evidence, and the possibility of sufficiently truth-seeking agents $1 \in \text{supp}(\alpha)$, a community of Bayesian learners will, with probability one, converge to knowing and declaring the truth in the long run. This is captured by the following proposition and its corollary.

Proposition 1. *An epistemic community learning about the state of the world will, in the long run, converge in belief to the true state.*⁹

Corollary 2. *For such an epistemic community, converging in belief to the true state implies converging to consensus in declaring the true state.*

Convergence in beliefs follows from the fact that our agents learn via Bayesian conditioning, that the true state is contained in each agent’s hypothesis set, and that agent declarations are always to some degree informative as to the state of the world. Given this, classical convergence results for Bayesian learning¹⁰ guarantee long run acquisition of the truth.

Convergence in declarations follows from the fact that, given convergence in beliefs, the community’s beliefs will inevitably pass a threshold such that a consensus on declaring the true state cannot be escaped. Moreover, with enough time following the passing of this threshold of belief, the population will almost surely traverse a positive probability path to consensus on the true state, whereupon it will never leave this consensus.

It may well be that “in the long run we are all dead,” [Keynes, 1923, p. 80] but it can be helpful to confirm where we are headed. We have seen that our epistemic communities will arrive at the truth in the limit of time, so we turn to short and medium run analysis of social learning for a richer and more pressing picture of inquiry.

4. TRUTH AND CONFORMITY IN THE SHORT AND MEDIUM RUN

What can be said about the short and medium run behavior of learning under conformity? What role does social structure play in the reliable acquisition of true beliefs? To answer these questions, we ran simulations of our model of epistemic communities engaged in social learning and discourse. We recorded and analyzed the resulting behavior

⁹ All proofs can be found in Appendix A.

¹⁰ For an excellent exposition of the classic results, see Smith and Sørensen (2000).

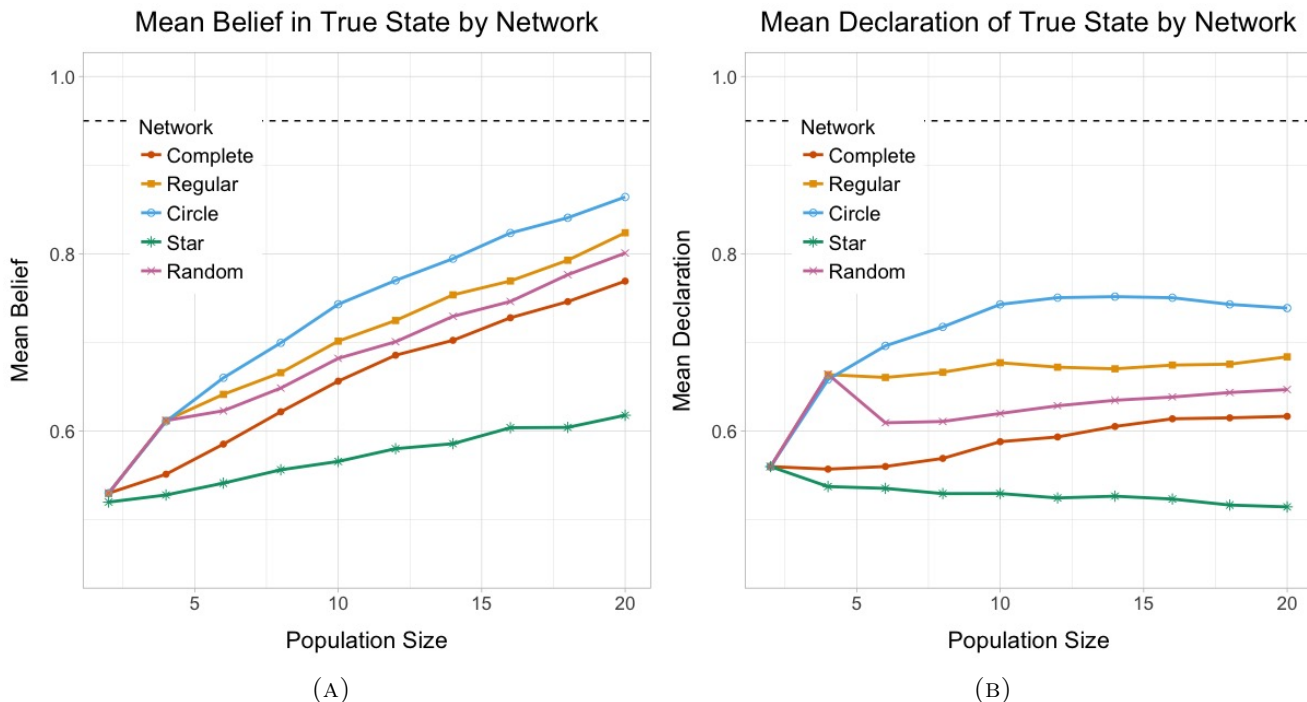


FIGURE 5. Plots of the mean belief in the true state $P(\theta)$ (A), and declaration of the true state θ (B), for each network type, and for network sizes from 2 agents to 20. Note that the networks only become fully distinct at $N = 6$. The dashed line represents performance in total absence of any conformity.

over a parameter sweep of network types, population sizes, initial declarations, prior beliefs, and distributions of the individuals' truth-seeking and conformity orientations.

For the simulations, we varied the structure of social influences by placing our agents on each complete, regular (of degree $N/2$), circle, star, and random (of mean degree $N/2$) networks. We varied the number of agents N in the network from 2 agents (at which all networks are essentially identical) to 20 agents. We considered when the initial declarations of the society were at a consensus on the true state, a consensus on the false state, and an even split. We varied the shared prior beliefs of the population between relative confidence in the true state ($P(\theta) = 0.75$), skepticism toward the true state ($P(\theta) = 0.25$), and ambivalence about the true state ($P(\theta) = 0.5$). Each combination of network structure, population size, initial declarations, and prior beliefs composed one parameter setting.

For each parameter setting we ran 10,000 simulations where each simulation was composed of 100 turns, and where each turn consisted of the following phases: (1) a randomly selected agent receives her private evidence from Nature; (2) the agent updates her private belief in light of this evidence; (3) the agent chooses her best response given her beliefs, her neighbors' declarations, and her truth-seeking/conformity orientation;

(4) the agent makes her declaration to the network; (5) the other agents in the network update their beliefs in light of her declaration.

Three regularities readily emerged from the data (see FIGURES 5A, 5B): (1) In all simulations, the star network performed worse than all other standard networks in terms of generating reliable belief in, and declaration of, the true state. (2) The circle network, on the other hand, performed better than other standard networks on all counts. (3) The other networks—complete, regular (of degree $N/2$), and random (of mean degree $N/2$)—yielded middling performances, neither as good as the circle, nor as poor as the star, with the regular network typically outperforming the random network, and the random network outperforming the complete network.¹¹

To make sense of these regularities in our simulation results, analytic treatment of the model and its dynamics is needed. What should be obvious is that conformity bias muddies the waters with respect to the information content of individuals’ declarations. In the absence of conformity, our epistemic communities would rapidly and reliably acquire the truth, and the underlying network structure would make no difference to this learning.

What we will find is that different networks induce social configurations more or less conducive to honest communication, and that this will also depend on the degree to which the population is divided or unified in their public declarations.

5. INFLUENCE, INFORMATION, AND SOCIAL STRUCTURE

To understand why different social networks are more or less conducive to the reliable acquisition of true beliefs, we first need a measure of informativeness. For this, we introduce the concepts of influence and informativeness of declarations, and show how they are related.

We define the *influence* of a declaration $x \in \{\theta, -\theta\}$ as the difference between the public belief in x before and after its declaration to the network, $q(x|x) - q(x)$, where q is the public belief. Next, we define the *informativeness* of a declaration $x \in \{\theta, -\theta\}$, as the reduction in uncertainty it produces with respect to its corresponding state when starting from a maximal entropy prior, $H(q|q(x) = 1/2) - H(q|x)$, where H is the Shannon entropy function.

We now derive the fact that the informativeness of a declaration is monotonically increasing in its influence on the public belief (see Lemma 9 in Appendix A). This gives us that a declaration will be (minimally) maximally influential just in case it is (minimally) maximally informative. We will use this fact repeatedly to infer the relative informativeness of declarations from their influence.

¹¹ In our simulation plots (FIGURE 5), we mark the performance of learning in the absence of any conformity bias—that is, of unimpeded Bayesian learning—with a dashed line. We will continue to compare our results to this control case, denoting the case of learning in the absence of conformity bias in further plots (FIGURE 6, 7, 8) each time with a dashed line.

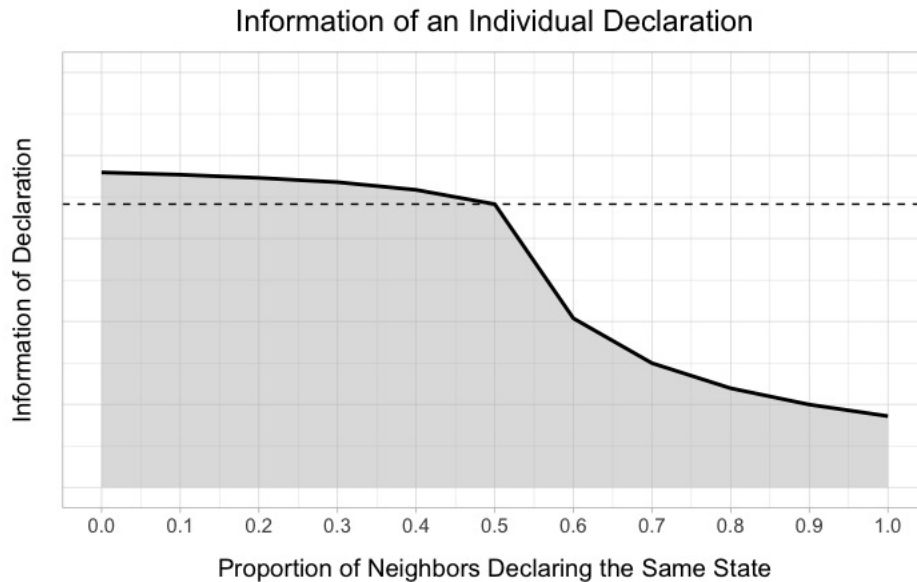


FIGURE 6. The influence and informativeness of an agent’s declaration, as a function of the proportion of her neighbor’s who are declaring the same state.

Optimal Information From Going Against the Grain. Given our measures of influence and informativeness, our first insight follows straightforwardly from our model of agents learning via Bayesian conditioning under uncertainty about one another’s evidence and truth-seeking orientations. It is that the most informative declarations—those that have the most significant effect on the public belief—are those that “go against the grain.” That is, those made by agents exactly when they deviate from the consensus of their neighbors.

This insight is captured by the following proposition:

Proposition 3. *The informativeness of an agent’s declaration is monotonically increasing in proportion of her neighbors who are declaring the opposing state.*

And since the minimum proportion of an agent’s neighbors who may declare in favor of any state is zero, we have the following as an immediate corollary:

Corollary 4. *The most informative declaration in favor of a state is one made by an agent when she goes against the consensus of her neighbors.*

This corresponds to the case in FIGURE 3, where $N_i(\theta) = 0$, and is visualized by the plot of information of declarations in FIGURE 6 where we see the change in belief by the population in response to an agent’s declaration as a function of the proportion of that agent’s neighbors who are declaring the same state.

When an agent deviates from the consensus of her immediate peers, it is inferred by the broader network that she is both likely to be more truth-seeking and that she has

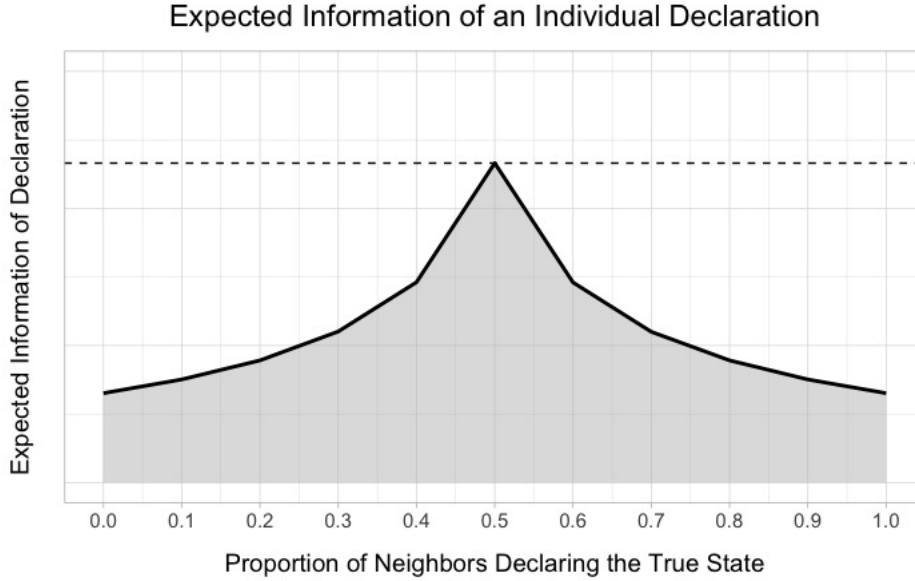


FIGURE 7. The *expected* influence and informativeness of an agent’s declaration, as a function of the proportion of her neighbor’s who are declaring the same state.

received sufficiently strong evidence to justify the loss in social payoffs she incurred. No other declaration is more influential on the public belief.

Optimal Expected Information From Conflicted Neighbors. We have seen what that the most informative declarations occur when an agent goes against the consensus of her peers. But such declarations are rare, as it takes highly truth-seeking agents with good evidence to be willing to make them. We should ask then: under what conditions, *on average*, do we expect to find the most informative declarations?

These turns out to be the obverse of where we find the most influential declaration. The most informative declarations, on average, must come from individuals whose neighbors are perfectly divided in terms of their declarations.

This is captured by the following observation:

Observation 1. *The most influential and informative declaration, in expectation, is that made by an agent when her neighbors are evenly divided in their declarations.*¹²

Observation 2. *The expected information of declarations is convex and increasing for $N_i(\theta) \in (0, 1/2)$ and convex and decreasing for $N_i(\theta) \in (1/2, 1)$.*

¹² Our observations are computationally verified for the following distributions of types and evidence: the distribution of truth-seeking orientations in the population was varied from Beta(1,5) (corresponding to high conformism), to uniform, and Beta(5,1) (corresponding to high truth-seeking). And the distributions of evidence induced by each state of the world were varied between the linear case described before, and Gaussian distributions with means of 1 and -1, and variances of 1, 10, and 100.

This corresponds to the case in FIGURE 3, where $N_i(\theta) = 1/2$, and is visualized by the plot of expected information of declarations in FIGURE 7. In FIGURE 7 we see the expected change in belief of the population in response to an agent's declaration, as a function of the proportion of that agent's neighbors who are declaring the true state. Our propositions make use of these observations.

It is when an agent's social influences equally represent each viable position that she is most free to declare her honest belief, and in such a case others infer that she is most likely doing so.

5.1. Informativeness of Networks. Which networks then are most conducive to the social configurations that yield honest communication? Using the insights developed so far, we extend the concept of expected informativeness to the level of social networks.

Assume that θ is the true state of the world, then *expected influence* of declarations $X = \{\theta, -\theta\}$ for an N -agent network \mathcal{G} is given by

$$E_X[q(\theta|\mathcal{G}) - q(\theta)] \propto \sum_{k=0}^N \sum_{j=1}^{\binom{N}{k}} \sum_{i=1}^N E_X[q(\theta|x_i) - q(\theta)]$$

where the first sum is over the number of the agents in the network declaring the true state, the second sum is over the possible configurations of declarations in the network given the number of agents declaring the true state, and the third sum is over the individuals in the network.¹³ In this way, we infer the informativeness of a network in aggregate as well as for fixed proportions of the community declaring the true state.

With a generalized measure of expected informativeness, we compute the expected informativeness of 10-agent networks for different proportions of the population declaring the true state (see FIGURE 8).

From this, several observations emerge. Denote the proportion of the community declaring θ by N_θ . For all networks, then, the least informative state is that of consensus, $N_\theta = 0$ or 1 , and the most informative state is when there is an even split in declarations $N_\theta = 1/2$. Given Observation 1, it should be clear why this is so. Declarations are expected to be informative in measure to the presence of balanced dissent.

Next, we observe that, when the population is nearly split, the complete network produces the most informative declarations among the networks considered, while the circle network produces the most informative declarations when the population is near consensus. Finally, the star network provides the least informative social configuration no matter the proportion of the population making either declaration.

We may understand these results in terms of our previous insights, and sharpen them by considering large networks. On a star network, when the population is large,

¹³ Note that we suppress the normalizing term from the definition of the influence of a declaration. The reason for this is that both terms are constants, and are therefore irrelevant for determining maxima or minima.

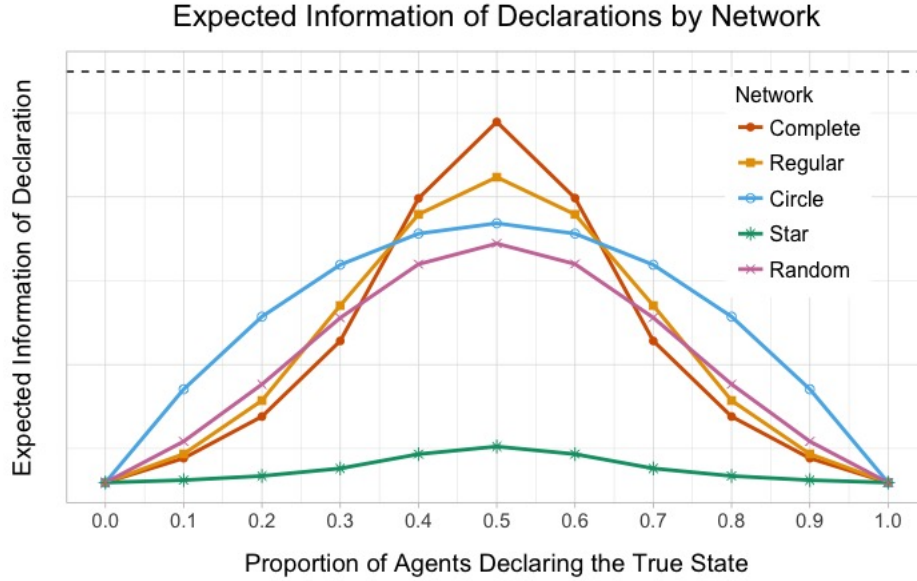


FIGURE 8. The expected informativeness of the next declaration for 10-agent networks as a function of the proportion of the population which is declaring of the true state. The dashed line denotes the expected information in the absence of any conformity.

practically every individual has merely one neighbor. Hence, for any proportion of declarations in the population, the star network will be in the minimally informative state. That is, $\mathbf{I}(\mathcal{G}_{star}|N_\theta) = N_\theta \mathbf{I}(1) + (1 - N_\theta) \mathbf{I}(0) = \mathbf{I}(0)$.¹⁴

Proposition 5. *For large networks, the star network is minimally informative in any state.*

On a complete network, when the population is evenly divided, $N_\theta = 1/2$, each individual is in the optimal position to make informative declarations. When all individuals are neighbors and the population is sufficiently large, the expected informativeness of the network as a whole recapitulates the expected informativeness of individual declarations given in FIGURE 7. That is, $\mathbf{I}(\mathcal{G}_{complete}|N_\theta) = \mathbf{I}(N_\theta)$. Given Observation 1, we show that no network can be more informative in such a state.

Proposition 6. *For large networks, when the population is evenly split in declarations, the complete network is maximally informative.*

On a circle network, when the population is near consensus, a single dissenting individual can make it possible for both her their neighbors to declare their honest beliefs. That is, given that each individual has two neighbors, their neighbors' declarations are binomially distributed with the success parameter given by the population proportion

¹⁴ Given the assumption of symmetry of expected informativeness across $N_\theta = 1/2$, we have that $\mathbf{I}(0) = \mathbf{I}(1)$, and, more generally, that $\mathbf{I}(1/2 - c) = \mathbf{I}(1/2 + c)$ for $c \in [0, 1/2]$.

of declarations, $\mathbf{I}(\mathcal{G}_{circle}|N_\theta) = N_\theta^2 \mathbf{I}(0) + 2N_\theta(1 - N_\theta) \mathbf{I}(1/2) + (1 - N_\theta)^2 \mathbf{I}(1)$. Contrast this with the complete network where, near consensus, every individual faces strong incentives to conform.

Proposition 7. *For large connected networks, for a range of states near consensus in declarations, the circle network is the maximally informative network.*

More generally, we can express the expected informativeness of the declaration of any individual with d connections and proportion N_θ of her neighbors declaring the true state as

$$E_{N_\theta}[\mathbf{I}_d] = \sum_{k=0}^d \binom{d}{k} N_\theta^k (1 - N_\theta)^{d-k} \mathbf{I}\left(\frac{k}{d}\right). \quad (*)$$

From this, we can derive the informativeness of a any network, when we conceive of networks as admixtures of proportions of individuals with different numbers of neighbors.

Given any large network, it can be represented as a distribution $\mu = \langle \mu_d \rangle$ over the degree d of individuals within the network, Thus, the expected informativeness of the network will be $\mathbf{I}(\mathcal{G}_\mu|N_\theta) = \sum_d \mu_d \cdot E_{N_\theta}[\mathbf{I}_d]$. Using this, we provide bounds for the informativeness of epistemic networks near consensus.

Proposition 8. *For large networks, near consensus, any network (including any regular or random network) of minimum degree at least two will be intermediate in informativeness between the circle and complete network.*

6. CONCLUSION

When social learning proceeds under the influence of conformity bias, the structure of social relationships underpinning the epistemic community becomes crucial to reliable acquisition of truth. That disagreement and diversity in publicly held opinions can be optimal for honest communication gives us our key insight into understanding the effects of different social networks. The question as to which social networks lead to reliable beliefs becomes a question as to which social networks produce and sustain optimal patterns of disagreement throughout the process of learning.

In sum, we find that in the presence of even a modicum of conformity bias the star network always provides the worst conditions for informative communication, the complete network provides optimal conditions exactly when the population is evenly divided, the circle network provides optimal conditions near consensus, and that, in such a state, all sufficiently connected networks will be intermediate in informativeness between the circle and complete networks.

This has implications for real-world social networks, which tend to exhibit low average degree and high degree-centrality (Watts and Strogatz 1998). We may conjecture that, when we suspect conformity bias at play in social discourse and decision-making, interventions which reduce the density of connections of a social network while still keeping

it connected, and interventions which decrease its centralization by reducing the relative influence of central individuals, may lead to more informative communication—and so to more reliable beliefs—for the epistemic community as a whole.

REFERENCES

- Asch, S. E. (1955). Opinions and Social Pressure. *Scientific American* 193(5), 31–35.
- Barrett, J., B. Skyrms, and A. Mohseni (2017). Self-Assembling Networks. *British Journal for the Philosophy of Science*.
- Bond, R. and P. B. Smith (1996). Culture and conformity: A meta-analysis of studies using asch’s (1952b, 1956) line judgment task. *Psychological Bulletin* 119(1), 111–137.
- Bruner, J. and C. O’Connor (2016). Power, Bargaining , and Collaboration. In *Scientific Collaboration and Collective Knowledge*, pp. 1–25.
- Goeree, J. K., A. Riedl, and A. Ule (2009). In search of stars: Network formation among heterogeneous agents. *Games and Economic Behavior* 67(2), 445–466.
- Goyal, S. (2007). *Connections: An introduction to the economics of networks*. Number 1.
- Heesen, R. (2017). Communism and the Incentive to Share in Science. *Philosophy of Science* 84(4), 698–716.
- Kitcher, P. (1990). The Division of Cognitive Labor. *The Journal of Philosophy* 87(1), 5.
- Kitcher, P. (1993). *The Advancement of Science: Science Without Legend, Objectivity Without Illusions*. Oxford University Press.
- Mayo-Wilson, C., K. J. S. Zollman, and D. Danks (2013). Wisdom of crowds versus groupthink: Learning in groups and in isolation. *International Journal of Game Theory* 42(3), 695–723.
- Morganand, T. J. and K. N. Laland (2012). The biological bases of conformity.
- O’Connor, C., L. Bright, and J. Bruner. The Evolution of Intersectional Oppression. *Philosophy of Science*.
- Rosenstock, S., J. Bruner, and C. O’Connor (2017). In Epistemic Networks, Is Less Really More? *Philosophy of Science* 84(2), 234–252.
- Smith, L. and P. Sørensen (2000). Pathological outcomes of observational learning. *Econometrica* 68(2), 371–398.
- Strevens, M. (2003). The role of the priority rule in science. *The Journal of Philosophy* 100, 55–79.
- Strevens, M. (2013). Herding and the quest for credit. *Journal of Economic Methodology* 20(1), 19–34.
- Watts, D. J. and S. H. Strogatz (1998). Collective dynamics of ‘small-world’ networks. *Nature* 393(6684), 440–442.
- Zollman, K. J. S. (2007). The Communication Structure of Epistemic Communities. *Philosophy of Science* 74(5), 574–587.

Zollman, K. J. S. (2009). The Epistemic Benefit of Transient Diversity. *Erkenntnis* 72(1), 17–35.

Zollman, K. J. S. (2013). Network epistemology: Communication in epistemic communities. *Philosophy Compass* 8(1), 15–27.

APPENDIX A. MATHEMATICAL APPENDIX

Learning from others' declarations. When agent i declares $x = \theta$, we know that it was her best response. As previously mentioned, this implies that the following condition holds:

$$\alpha_i(2P_i(\theta) - 1) + (1 - \alpha_i)(2N_i(\theta) - 1) > 0. \quad (\dagger)$$

We plug agent i 's (publicly unknown) posterior probability $P(\theta|\sigma)$ into (\dagger) to get the elaborated condition

$$\alpha_i \left(\frac{2}{1 + \frac{1 - \bar{P}}{\bar{P}} \frac{1 - \sigma}{\sigma}} - 1 \right) + (1 - \alpha_i)(2N_i(\theta) - 1) > 0 \quad (\ddagger)$$

where \bar{P} denotes the (publicly known) prior $P(\theta|\mathbf{h}^t)$. We then compute the likelihood of agent i 's declaration θ , given our public prior, as follows.

Let ϕ denote the left-hand term of our elaborated condition (\ddagger) , under which our agent would have declared θ , so that $\mathbb{I}[\phi > 0]$ is its indicator function. We then get the likelihood of the declaration given each possible state of the world,

$$P(x = \theta|\theta, \bar{P}) = \int_A \int_{\Sigma} \mathbb{I}[\phi > 0] dF_{\theta}(\sigma) dG(\alpha),$$

$$P(x = \theta|-\theta, \bar{P}) = \int_A \int_{\Sigma} \mathbb{I}[\phi > 0] dF_{-\theta}(\sigma) dG(\alpha).$$

From these, we obtain the posterior—the belief of the other agents in the network in light of agent i 's declaration of θ —using Bayes' rule as follows

$$P(\theta|x = \theta, \bar{P}) = \left(1 + \frac{\int_A \int_{\Sigma} \mathbb{I}[\phi > 0] dF_{-\theta}(\sigma) dG(\alpha) \frac{1 - \bar{P}}{\bar{P}}}{\int_A \int_{\Sigma} \mathbb{I}[\phi > 0] dF_{\theta}(\sigma) dG(\alpha)} \right)^{-1}$$

which yields the new public belief.

Proof of Proposition 1. There are two states of the world θ and $-\theta$. Without loss of generality, suppose θ to be the true state of the world. Let $q(\mathbf{h}^t) = P(\theta|\mathbf{h}^t)$ be the *public belief* and \mathbf{h}^t the history of declarations up to time t . As is well-known, the likelihood ratio

$$\ell(\mathbf{h}^t) \equiv \frac{1 - q(\mathbf{h}^t)}{q(\mathbf{h}^t)}$$

is a martingale conditional on θ . Let X be the finite set of declarations. For any given declaration $x \in X$,

$$\ell(\mathbf{h}^t, x) = \ell(\mathbf{h}^t) \frac{P(x|\mathbf{h}^t, \neg\theta)}{P(x|\mathbf{h}^t, \theta)}$$

and thus the martingale property follows:

$$E[\ell(\mathbf{h}^{t+1})|\theta] = \sum_{x \in X} \ell(\mathbf{h}^t, x) P(x|\mathbf{h}^t, \theta) = \sum_{x \in X} \ell(\mathbf{h}^t) P(x|\mathbf{h}^t, \theta) = \ell(\mathbf{h}^t).$$

By Theorem 3(b) of Smith and Sørensen (2000), when evidence is unbounded, individuals almost surely converge in belief to the true state. \square

We show that convergence in beliefs implies a convergence in declarations. In particular, we show that convergence in beliefs implies that the community's belief in the true state will be bounded from below over time. We then observe, using simple probabilistic arguments, that given sufficient time the community will almost surely arrive at a consensus state where all individuals are declaring the true state. Finally, we show that, having arrived at such a consensus with individual beliefs in the true state appropriately bounded from below, the community must remain at this consensus forever.

Proof of Corollary 2. Let q and q' denote the public belief before and after hearing a declaration, respectively. Consider a focal agent i having received her evidence from Nature on a given turn. Let P_i denote the focal agent's posterior belief $P(\theta|\sigma, \mathbf{h}^t)$, and suppose that this agent declared $x = -\theta$. It is straightforward to show that if the population could observe the focal agent's posterior, the public belief would be precisely equal to her posterior

$$q'(-\theta, q, N_i(\theta), P_i) = P_i. \quad (*)$$

Let $\Pi(\cdot|-\theta, q, N_\theta)$ be the distribution over the focal agent's posterior belief given her declaration of $-\theta$, q the public belief when she selected her action, and $N_i(\theta)$ the proportion of her neighbors declaring θ . By (*) we can write

$$q'(-\theta, q, N_i(\theta)) = \int_0^1 P_i \, d\Pi(P_i|-\theta, q, N_i(\theta)).$$

We can thus interpret the public belief as the public's expectations of the focal agent's posterior. As the public belief almost surely converges to certainty on the truth, for almost all trajectories of the public belief $\{q_t\}_{t=0}^{+\infty}$, for all $\epsilon > 0$ there exists a time T_ϵ such that, if $t > T_\epsilon$ then $q_t > 1 - \epsilon$. That is, there is a time after which the public belief in θ will always be at least $1 - \epsilon$. Then choose $\epsilon = 1/2$.

With probability 1 at some point along the trajectory after T_ϵ all agents will be declaring θ . To see this, let λ be the probability all N agents choose declarations in sequence, each has an α sufficiently high such that they declare the state they believe to be more likely regardless of their neighbors' declarations, and they receive evidence such that their posterior assigns higher probability on θ . However small the probability

λ might be, it exceeds 0. Hence, the probability that this event does not occur goes to zero as $t \rightarrow +\infty$.

Assume, for the sake of contradiction, that at some point after T_ϵ an agent goes against the consensus and declares $\neg\theta$, then her posterior must satisfy

$$P_i \leq -\frac{1 - \alpha_i}{2\alpha_i} + \frac{1}{2}.$$

But then we get that $E[P_i | \neg\theta, \cdot] \leq 1/2$. That is, her belief in θ was less than 1/2, which contradicts the fact that her belief was bounded from below. Hence, no agent can deviate from the consensus after time T_ϵ , and convergence in belief implies convergence in declaration. \square

Lemma 9 (Monotonicity of Informativeness in Influence). *The informativeness of a declaration about a state is monotonically increasing in its influence on the public belief.*

Proof. Without loss of generality, let the focal agent declare $x = \theta$. We show that the informativeness of her declaration, $H(q|q(\theta) = 1/2) - H(q|x = \theta)$, is monotonically increasing in its influence, $q(\theta|x = \theta) - q(\theta)$.

First, we unpack the definition of informativeness, temporarily omitting the assumption of the maximal entropy prior $q(\theta) = 1/2$, to get

$$\begin{aligned} H(q) - H(q(\theta|x = \theta)) &= E[-\ln(q(\theta|x = \theta))] - E[-\ln(q(\theta))] \\ &= E[\ln(q(\theta)) - \ln(q(\theta|x = \theta))] \\ &= E\left[\ln\left(\frac{q(\theta)}{q(\theta|x = \theta)}\right)\right] \\ &= q(\theta) \cdot \ln\left(\frac{q(\theta)}{q(\theta|x = \theta)}\right) + q(\neg\theta) \cdot \ln\left(\frac{q(\neg\theta)}{q(\neg\theta|x = \theta)}\right) \end{aligned}$$

Now, let $A \equiv q(\theta)$ and $B \equiv q(\theta|x = \theta)$, so that $C \equiv B - A$ denotes the influence of the declaration $x = \theta$. Then we can re-write the preceding expression as

$$A \cdot \ln\left(\frac{A}{A + C}\right) + (1 - A) \cdot \ln\left(\frac{1 - A}{1 - (A + C)}\right)$$

Taking the partial derivative with respect to influence C , and solving for when it is positive—i.e., for when informativeness is increasing—yields

$$A + C - 1 > 0 \quad \text{or} \quad B > 1/2.$$

And when $q(\theta) = 1/2$, we have that $B = q(\theta|x = \theta) \geq 1/2$, and so informativeness is monotonically increasing in influence, as desired. \square

We will show that $q'(\theta, N_i(\theta)') < q'(\theta, N_i(\theta))$ whenever $N_i(\theta)' > N_i(\theta)$. From this it follows straightforwardly that, given $N_i(\theta) \in [0, 1]$, the most influential declaration occurs just when $N_i(\theta) = 0$.

To do so, consider a given focal agent i having received evidence $\sigma \sim f_\theta(\sigma)$ from Nature. Let $r = r(\sigma) \equiv P_i(-\theta|\sigma)$ be one minus her private belief, $G_{-\theta}(r)$ and $G_\theta(r)$ the conditional cdf's for r , and $g(r) \equiv \frac{dG_{-\theta}}{dG_\theta}(r)$ the Radon-Nikodym derivative of $G_{-\theta}$ with respect to G_θ .

Lemma 10. $g(r) = \frac{r}{1-r}$ almost surely.

Proof. If an agent updates her belief after observing r , it will remain unchanged. Thus from Bayes' theorem $r = P_i(-\theta|r) = \frac{g(r)}{g(r)+1}$. \square

Lemma 11. The ratio $\frac{G_{-\theta}}{G_\theta}(r)$ is strictly increasing for r in the common support of G_θ and $G_{-\theta}$.

Proof. Let $r' > r$. From Lemma 10 we have that $g(r)$ is strictly increasing, hence,

$$G_{-\theta}(r) = \int_0^r g(x)dG_\theta(x) < g(r)G_\theta(r)$$

And thus

$$\begin{aligned} G_{-\theta}(r') - G_{-\theta}(r) &= \int_r^{r'} g(x)dG_\theta(x). \\ &> [G_\theta(r') - G_{-\theta}(r)]g(r) \\ &> [G_\theta(r') - G_{-\theta}(r)]\frac{G_{-\theta}(r)}{G_\theta(r)}. \end{aligned}$$

It follows that $\frac{G_{-\theta}(r')}{G_\theta(r')} > \frac{G_{-\theta}(r)}{G_\theta(r)}$. \square

Proof of Proposition 3. Now, we proceed to show that $q'(\theta, N_i(\theta)') < q'(\theta, N_i(\theta))$ whenever $N_i(\theta)' > N_i(\theta)$. Define q' to be the posterior public belief, q the prior public belief, $N_i(\theta)$ the proportion of the focal agent's neighbors declaring θ , and $\Pi(\cdot|x_i, q, N_i(\theta))$ the posterior belief over the declaring agent's truth-seeking orientation $\alpha_i \in [0, 1]$. Then

$$q'(\theta, N_i(\theta)) = \int_0^1 q'(\theta, N_i(\theta), \alpha_i)d\Pi(\alpha_i|\theta, N_i(\theta), q).$$

For a given α_i in the support of $\Pi(\cdot|\theta, N_i(\theta), q)$, there exists a threshold $\bar{r} = \bar{r}(\alpha_i, q, N_i(\theta))$ such that the agent only selects $x_i = \theta$ if $r \leq \bar{r}$. From Bayes' theorem,

$$q'(\theta, N_i(\theta), \alpha_i) = \left(1 + \frac{1-q}{q} \frac{G_{-\theta}(\bar{r})}{G_\theta(\bar{r})}\right)^{-1}.$$

If $\bar{r}(\alpha_i, N_i(\theta)', q) \geq \bar{r}(\alpha_i, N_i(\theta), q)$ holds, and further holds strictly for a subset of α_i with positive posterior probability, then, by Lemma 11, $q'(\theta, N_i(\theta)') < q'(\theta, N_i(\theta))$.

It can be shown that the threshold $\bar{r}(\alpha_i, N_i(\theta), q)$ is strictly increasing in $N_i(\theta)$. This gives us that $q'(\theta, N_i(\theta)', \alpha_i) \leq q'(\theta, N_i(\theta), \alpha_i)$. Furthermore, having assumed that α_i and r take full support in $[0, 1]$, we can find a neighborhood of $\alpha_i = 1$ with positive probability such that $\bar{r}(\alpha_i, N_i(\theta), q) > 0$ for all α_i in this neighborhood. Hence, in this neighborhood $q'(\theta, N_i(\theta)', \alpha_i) < q'(\theta, N_i(\theta), \alpha_i)$. \square

Proof of Corollary 4. We have, from proposition 3, that $q'(\theta, N_i(\theta)') > q'(\theta, N_i(\theta))$ whenever $N_i(\theta)' < N_i(\theta)$. It follows directly that

$$\arg \max_{N_i(\theta) \in [0,1]} q'(\theta, N_i(\theta)) = 0.$$

Thus, the most influential declaration is made just when $N_i(\theta) = 0$. And we have, from Lemma 9, that this is also the most informative declaration. \square

Proof of Proposition 5. On a large star network, proportion 1 of individuals have a single neighbor. Thus for any proportion of the population declaring θ , every individual is in the minimally informative state where either $N_\theta = 0$ or 1. Therefore, for all $N_\theta \in [0, 1]$, and symmetric \mathbf{I} , $\mathbf{I}(\mathcal{G}_{star}) = \mathbf{I}(0) \leq \mathbf{I}(\mathcal{G})$ for any connected network \mathcal{G} . \square

Proof of Proposition 6. On a complete network, every individual is neighbors with every other. Hence, the proportions of an individuals neighbors declaring θ is the same as the proportion of the population declaring θ . The expected informativeness is maximized when an individual's neighbors are equally split $N_i(\theta) = 1/2$. Thus, when exactly half the population is declaring θ , the declaration of every individual in the population is at maximal expected informativeness. Hence, no other network can be more informative in this state. That is, when $N_\theta = 1/2$, $\mathbf{I}(\mathcal{G}_{complete}) = \mathbf{I}(1/2) \geq \mathbf{I}(\mathcal{G})$ for any connected network \mathcal{G} . \square

To show that the circle is maximally informative near consensus, first we show that for regular networks of degree at least 2 informativeness is decreasing in degree near consensus. This implies that any regular network of degree greater than two is less informative than the circle network. We combine this with Proposition 5, which implies that networks of degree 1 are also less informative than the circle network, to show that the circle network is the maximally informative regular network. Next, using the fact that any network can be formulated as an admixture of individuals of various degrees we derive that the circle network is maximally informative near consensus.

Lemma 12. *For regular networks of degree at least 2, informativeness is decreasing in degree near consensus.*

Proof. Take the derivative of the informativeness of any regular network \mathcal{G}_d of degree $d \geq 2$ with respect to the proportion of the population declaring the true state.

$$\frac{d}{dN_\theta} [\mathbf{I}(\mathcal{G}_d)] = \frac{d}{dN_\theta} \left[\sum_{k=0}^d \binom{d}{k} N_\theta^k (1 - N_\theta)^{d-k} \mathbf{I} \left(\frac{k}{d} \right) \right].$$

Let N_θ go to 0. This makes it so only the constant terms of the derivative remain, and the expression simplifies to

$$\lim_{N_\theta \rightarrow 0^+} \frac{d}{dN_\theta} [\mathbf{I}(\mathcal{G}_d)] = d[\mathbf{I}(1/d) - \mathbf{I}(0)].$$

This term corresponds to the slope of the secant line connecting $\mathbf{I}(0)$ and $\mathbf{I}(1/d)$. Since \mathbf{I} is an increasing function, this term must be decreasing in d . Thus, for networks of degree two and greater, informativeness is decreasing in degree near consensus. \square

Lemma 13. *The circle is the maximally informative regular network near consensus.*

Proof. This follows from Lemma 12 and Proposition 5, which state that a regular network of degree 2 (the circle) is more informative than any network of greater degree near consensus, and that a regular network of degree 1 is less informative than any other at any state. Taken together, they imply that, near consensus, regular networks of degree two are maximally informative among regular networks. \square

Proof of Proposition 7. Now, recall that any large connected network \mathcal{G}_μ can be formulated as an admixture $\mu = \langle \mu_d \rangle$ of proportions of individuals of degree $d \geq 1$, where $\sum_d \mu_d = 1$ and $\mu_d \geq 0$. The expected informativeness of any network then is a proportion-weighted sum of the expected informativeness of the individuals of each degree contained in the network. That is, $\mathbf{I}(\mathcal{G}_\mu | N_\theta) = \sum_d \mu_d \cdot E_{N_\theta}[\mathbf{I}_d]$. It follows from Lemma 13 that, near consensus, any network not entirely composed of individuals of degree two is strictly less informative than one which is in fact composed entirely of individuals of degree two. Thus, when $N_\theta = 0$ or 1, $\mathbf{I}(\mathcal{G}_{circle}) > \mathbf{I}(\mathcal{G}_\mu)$ for any \mathcal{G}_μ such that $\mu_0 = 0$ and $\mu_2 \neq 1$. \square

Proof of Proposition 8. It follows directly from Lemma 12 that, near consensus, the maximally and minimally informative regular networks of degree at least two are the circle and complete network, respectively. We combine this with the fact that any large network \mathcal{G}_μ can be formulated as an admixture $\mu = \langle \mu_d \rangle$ of regular networks of degree d , and with the linearity of expected informativeness, to adduce that the informativeness of any network is bounded above by that of the circle network and bounded below by the complete network. That is, when $N_\theta = 0$ or 1, $\mathbf{I}(\mathcal{G}_{circle}) \geq \mathbf{I}(\mathcal{G}_\mu) \geq \mathbf{I}(\mathcal{G}_{complete})$ for any \mathcal{G}_μ such that $\min\{d : \mu_d > 0\} \geq 2$. \square

APPENDIX B. SIMULATION CODE

The source code for our network learning model and simulations can be found at:
<https://amohseni.shinyapps.io/Truth-and-Conformity-on-Networks>

UC IRVINE, IRVINE, CA 92697, USA

E-mail address: amohseni@uci.edu

E-mail address: colerw@uci.edu