Culture and Communication*

Rajiv Sethi[†] Muhamet Yildiz[‡]

October 8, 2018

Abstract

A defining feature of culture is similarity in the manner in which information about the world is interpreted. This makes it easier to extract information from the beliefs of those within one's own group. But this information may be of low quality if better informed sources lie elsewhere. Furthermore, observing individuals outside one's group deepens our understanding not only of those individuals, but also of their culture. We model this process, using unobservable, heterogeneous priors to represent fundamental belief differences across individuals; these priors are correlated within but not across groups. We characterize long run communication patterns as follows. When uncertainty about the priors is low, there is a merging of cultures and individuals seek information wherever it is most precise. Otherwise, extreme homophily arises with positive probability, and with certainty when priors are highly correlated within groups. At moderate levels of correlation, individuals in each group can be partitioned into two categories: some individuals exhibit extreme homophily, rarely if ever stepping outside group boundaries, while others exhibit baseline homophily and seek information wherever it is most precise. The degree of homophily can vary non-monotonically with the level of correlation in priors, and small groups can exhibit heterophily at intermediate levels of correlation.

*We thank David Budescu, Ben Golub, and Debraj Ray for helpful comments.

[†]Department of Economics, Barnard College, Columbia University and the Santa Fe Institute.

[‡]Department of Economics, MIT.

1 Introduction

A defining feature of culture is similarity in the manner in which information about the world is interpreted. Two individuals who share a common culture—defined by ethnicity, religion, or even politics—will tend to have correlated mental models of the world that affect the way they process information. We call such mental models *perspectives*. Correlated perspectives facilitate communication, since it is easier to extract the informational content of a statement when the listener has a better understanding of the speaker's frame of reference. In seeking information, therefore, people will often turn to those whose perspectives they understand. This is a force for informational homophily.

But no single group has a monopoly on information, and informational homophily therefore comes at a price. Those who are willing to seek information from beyond the boundaries of their own group will have access to a richer information pool, even if this information is sometimes harder to extract. This is a force against informational homophily.

The trade-off between these two forces changes over time, based on an individual's observational history. Previously observed sources of information become better understood and hence more likely to be consulted again. But the degree to which an individual's understanding of another deepens through observation depends on how well-informed the observer herself happens to be in the period of observation. This has a symmetry-breaking effect, leading to divergence over time in the behavior of initially identical individuals. Furthermore, those who repeatedly observe individuals outside their own group learn not only about the perspectives of their targets, but also about the groups to which those targets belong—learning about a person teaches us both about the person and about their culture. As a consequence, such individuals become more likely to step across group boundaries when seeking information in the future.

We model this process, using unobservable, heterogeneous priors to represent perspectives. There are two social groups, and priors are correlated within but not across groups. That is, individuals initially have more precise beliefs about the perspectives of those within their culture than those outside it. There is a sequence of periods, and in each period there is a state of the world about which individuals would like to have precise beliefs. We call these states *issues*. In each period, each individual receives an informative private signal about the current issue. The precision of these signals varies stochastically across individuals and over time, and those with more precise signals in a given period are said to have higher *expertise* on that period's issue. We assume that expertise is publicly observable. In each period, each person selects a target individual and observes the target's posterior belief about the current issue; we call this posterior the target's *opinion*. The opinion is informative about the current issue, the target's perspective, and the perspectives of all others who belong to the target's group. The observer accordingly updates her beliefs about the current state, and about the perspectives of others in the population.

Within this framework, culture affects communication through two channels. First, initially, each individual learns something about the perspectives of others in her group based on knowledge of her own perspective. This creates a bias towards homophily in the first period of observation. Second, there is ongoing and indirect learning, based on inferences from observed opinions, about the culture of each chosen target. This indirect learning plays an important role in shaping long run communication structures.

We show that when initial uncertainty about the perspectives of those outside one's group is small relative to the range of possible expertise, cultures merge in the long run: all individuals eventually seek information from wherever it is most precise and group boundaries effectively dissolve. When initial uncertainty about perspectives is large, however, an extreme form of homophily—involving homogeneity and insularity in observational patterns—emerges with positive probability at all correlation levels.

We obtain a lower bound on the probability of this event, and show that extreme homophily arises with high probability in large groups—or more generally in groups where one is likely to find very well-informed individuals.

Extreme homophily is possible at all levels of correlation, but happens with certainty when correlation is sufficiently high (and initial beliefs about perspectives sufficiently uncertain). At lower levels of correlation, heterogeneity in behavior both within and across groups can arise, even if individuals in each group are identical at the outset. Nevertheless, observational patterns exhibit considerable structure. Under moderately high levels of correlation, individuals in each group can be partitioned into two categories. One of these exhibits considerable homophily, rarely if ever stepping outside group boundaries, while the other is unbiased and seeks information wherever it is most precise.

This bimodality in observational choices is a key testable prediction of the model, and arises because of indirect learning. If an individual observes some members of another group repeatedly, their perspectives will eventually be learned with high precision. This provides considerable information about those in the other group who have *not* been observed, making it likely that they too will be consulted when endowed with sufficiently high expertise. And this means that their perspectives too will be eventually learned. This gives rise to a tipping phenomenon—those who step outside group boundaries often in early periods cease to exhibit own-group bias, while others remain parochial and insular.

There is a limit to power of indirect learning when correlation in perspectives is low, and heterogeneity in perspectives within groups is accordingly considerable. In this case a form of opinion leadership arises with high probability. We obtain a stochastic upper bound on the number of individuals one observes in the long run, and show that this number is likely to be small. That is, in the long run, each individual has a small set of "long-run experts" and targets the best-informed within this set, even when there are better informed individuals outside it.

In addition, we show using a family of examples that the extent of homophily exhibited by a group can vary non-monotonically with the degree of correlation in perspectives, reaching its lowest values when this correlation is neither too low nor too high. This too is a consequence of indirect learning. When correlation is low, an individual is initially more likely to step outside her group when seeking information, but what she learns about her target tells her relatively little about the target's culture. As correlation rises, one learns more about a group from observing a single member, and this can reduce the likelihood of homophily in the long run. In fact, we show that small groups can exhibit *heterophily* at certain intermediate levels of correlation in perspectives. This too is consistent with the available empirical evidence.

2 Related Literature

The idea that culture affects cognition has been explored extensively in anthropology, social psychology, and law. Kahan and Braman (2006), building on prior work by Douglas and Wildavsky (1982), argue that "were indeterminacy or inaccessibility of scientific knowledge the source of public disagreement, we would expect beliefs on discrete issues to be uncorrelated with each other." And yet, on questions such as the effects on crime of gun control, the effects on health of abortion, and the effects on the climate of fossil fuel combustion, there is a high degree of correlation in opinion: "factual beliefs on these and many other seemingly unrelated issues do cohere." Their proposed explanation relies on the concept of *cultural cognition*:

> Essentially, cultural commitments are *prior* to factual beliefs on highly charged political issues. Culture is prior to facts, moreover, not just in the evaluative sense that citizens might care more about how gun control, the death penalty, environmental regulation and the like cohere with their cultural values than they care about the consequences of those policies. Rather, culture is prior to facts in the cognitive sense that what citizens believe about the empirical consequences of those policies *derives* from their cultural worldviews.

Along similar lines, *cultural consensus theory* is based on the premise that individuals within a culture have correlated beliefs, centered around some representative or consensus belief about any particular issue (Romney et al., 1986; Batchelder and Anders, 2012). This consensus need not correspond to any notion of ground truth, and idiosyncratic differences in cultural competence result in heterogeneity of beliefs within groups. The theory allows for the joint estimation of both consensus beliefs and individual levels of cultural competence using survey data. Our model has a similar flavor, with imperfectly correlated prior beliefs within groups, and uncorrelated beliefs across groups. But we are also concerned with beliefs about these prior beliefs, and these evolve over time as individuals receive information about the world both directly from signals and indirectly through observing the opinions of selected others.

We examine these questions using a framework previously developed in Sethi and Yildiz (2012, 2016), where heterogeneous priors represent fundamental belief differences and signals of varying precision represent information. Although priors are initially unobserved, they are drawn from a commonly known distribution, so individuals can reason and update their beliefs about these as time unfolds and information is received. In Sethi and Yildiz (2012) we focused on sequential public belief announcements about a state as in Geanakoplos and Polemarchakis (1982), and identified conditions under which distributed information is fully aggregated. In Sethi and Yildiz (2016) we studied the endogenous formation of information networks in a population with independent priors, private posteriors, and a sequence of states.

In the present work, we build on this latter paper by allowing for distinct identity groups, with a particular correlation structure on the distribution from which priors are drawn. Although we use techniques developed previously, the questions explored and the economic insights obtained are quite different. In Sethi and Yildiz (2016), opinion leadership plays a prominent role, and one of the main findings is that expert sets tend to be small. In contrast, our main concern here is with homophily, and indirect learning can lead to large expert sets and bimodal observational patterns when correlation in perspectives is sufficiently high. (Our earlier result on small expert sets generalizes to the current setup when correlation is low.)

In related and complementary work, Liang and Mu (2018) have examined a model of learning from multiple sources with correlated biases or confounds. Agents in their model learn about a single state through repeated observation over time, rather than a sequence of states as we consider here. But what their work shares with ours is the idea that learning about a state also teaches us about the source, as well as other sources with correlated confounds. This can give rise to learning traps in their model, just as it gives rise to homophily and within-group bimodality in ours.

Lazarsfeld and Merton (1954) are credited with coining the term homophily, and associating it with the proverb "birds of a feather flock together" (McPherson et al., 2001). Homophily can arise along multiple dimensions of affiliation, and prior theoretical work has focused on social interactions and friendship. For instance, Currarini et al. (2009) have developed a model in which homophily arises through a process of costly search, when people have a preference for own-group social affiliates. Those in larger groups have greater incentives to search, since they encounter own-group members more frequently. This gives rise to a friendship gradient, with larger groups having more connections on average. Currarini et al. (2009) also provide empirical evidence for heterophily in small groups using data from the National Longitudinal Survey of Adolescent Health.[1] Such heterophily arises endogenously in our model for intermediate values of correlation in perspectives.

Taking a different approach, Kets and Sandroni (2015) have examined the role of strategic uncertainty in generating homophily. Individuals in their framework are characterized by an impulse to play a particular action in a coordination game, and these impulses are correlated within but not across groups. Each player finds it rational to follow her impulse when interacting with members of her own group, since she expects her counterpart to have the same impulse with high likelihood, and to follow it in equilibrium. This reduces strategic uncertainty and makes interactions with own-group members more desirable. Cultural similarity in their work serves as a mechanism for equilibrium selection rather than effective information extraction.

[1]See, especially, Figure 2 in their paper, which shows that heterophily is quite common in groups that constitute a small share of the overall school population.

Gentzkow and Shapiro (2011) have examined informational homophily empirically, in the context of ideological identity (conservative and liberal) in the United States. They find that ideological homophily in access to online news sources is greater than that in access to offline news, though considerably smaller than that in face-to-face interactions in neighborhoods, workplaces and voluntary associations. Here news sources are themselves are placed on an ideological spectrum based on the distribution of their users across political identity groups.

The idea that individuals can extract information more easily from those with whom they share a culture is the basis for a branch of the statistical discrimination literature descended from Phelps (1972); see especially Aigner and Cain (1977) and Cornell and Welch (1996).[2] Our contribution here may be viewed as providing firmer foundations for this approach. While our starting point is a greater capacity for individuals to interpret the opinions of those in their own group, this capacity evolves over time in ways that can generate substantial within-group heterogeneity.

3 The Model

Consider a population $N = \{1, \ldots, n\}$ partitioned into two sets N_1 and N_2, each of which corresponds to a distinct identity group. We refer to these as group 1 and group 2 respectively. Let $n_k \geq 3$ denote the size of group k.

There is a sequence of periods $t = 0, 1, ...$ in each of which there is a state $\theta_t \in \mathbb{R}$ about which individuals would like to have precise beliefs. Each individual holds an idiosyncratic prior belief about the distribution from which θ_t is drawn. Specifically, according to the prior belief of individual i, θ_t is normally distributed with mean μ_i and unit variance:

$$\theta_t \sim_i N(\mu_i, 1).$$

We refer to the prior mean μ_i as the *perspective* of i. The interpretation is that the perspective governs the manner in which information regarding a broad range of issues is filtered, with the state in each period corresponding to a distinct issue.

An individual's perspective is not observable by others, but it is commonly believed

[2]In contrast, models of statistical discrimination such as Arrow (1973) and Coate and Loury (1993) involve ex ante identical groups.

that the perspectives $\mu = (\mu_1, ..., \mu_n)$ are jointly distributed according to

$$\mu \sim N(\overline{\mu}, \Sigma),$$

where $\overline{\mu} = (\overline{\mu}_1, \ldots, \overline{\mu}_n)$ is the mean and Σ is a variance-covariance matrix with typical element σ_{ij}. Here, $\overline{\mu}_i$ represents observable attributes of i that affect her perspective in a known manner, while Σ describes the residual uncertainty about perspectives. The variance-covariance matrix is a central component of our analysis, while the means $\overline{\mu}_i$ do not play any role.

We assume that perspectives are correlated within but uncorrelated across groups:

$$\Sigma = \sigma_0^2 \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix}.$$

Here the submatrices Σ_k have diagonal elements 1 and off-diagonal elements $\rho_k \in (0, 1]$. That is, for any $i \neq j$, $\sigma_{ij} = \sigma_0^2 \rho_k$ if i and j are both members of group k, and $\sigma_{ij} = 0$ otherwise. We can think of a group with high correlation ρ_k as being relatively homogeneous or tight-knit. The parameter σ_0^2 reflects the degree to which an individual's beliefs about the perspectives of others are imprecise, before any information contained in one's own prior has been taken into account. Since priors are correlated within groups, this information will reduce uncertainty about the perspectives of own-group members, as we discuss below.

All individuals receive private, informative signals about states. Specifically, in period t, each individual i observes the signal

$$x_{it} = \theta_t + \varepsilon_{it}, \tag{1}$$

where $\varepsilon_{it} \sim N(0, \tau_{it}^2)$. The signal variance τ_{it}^2 captures the degree to which i is well-informed; the lower this variance the greater the *expertise* of i about the period t state.

While signal realizations are privately observed, the signal variances $(\tau_{1t}^2, \ldots, \tau_{nt}^2)$ are public information. That is, at the start of each period, each individual knows who is well-informed about the current issue and who is not, but does not know the content of anyone else's information.

Having observed x_{it}, individual i updates her belief about the period t state in accordance with Bayes' rule, resulting in the posterior:

$$\theta_t \sim_i N\left(y_{it}, \frac{\tau_{it}^2}{1 + \tau_{it}^2}\right). \tag{2}$$

We refer to y_{it} as i's *opinion* in period t. This is simply a weighted average of an individual's prior and her signal, with weights determined by the precision of her information:

$$y_{it} = \frac{\tau_{it}^2}{1+\tau_{it}^2}\mu_i + \frac{1}{1+\tau_{it}^2}x_{it}. \tag{3}$$

We assume that the variances are uniformly bounded: $\underline{\tau}^2 \leq \tau_{ij}^2 \leq \overline{\tau}^2$ for some strictly positive real numbers $\underline{\tau}^2$ and $\overline{\tau}^2$ with $\underline{\tau}^2 < \overline{\tau}^2$. That is, no individual is ever perfectly informed, but all signals carry some information. We allow serial and cross-individual correlation in the distributions of expertise, but assume that they satisfy the following: for each open neighborhood V of each $\tau^2 \in \{\underline{\tau}^2, \overline{\tau}^2\}^N$, there is a strictly positive uniform lower bound on the probability that $(\tau_{1t}^2, ..., \tau_{nt}^2)$ is in V across all histories. For example, the probability of $\tau_{it}^2 < \underline{\tau}^2 + \epsilon$ and $\tau_{it}^2 > \overline{\tau}^2 - \epsilon$ is bounded away from 0 across all histories, although it may be vanishingly small as $\epsilon \to 0$.

In each period, each individual can observe one other individual's opinion, but cannot separately observe the constituent priors and signals. That is, in each period t, after the realized expertise levels $(\tau_{1t}^2, \ldots, \tau_{nt}^2)$ have been publicly observed, each i chooses a target $\lambda_t(i) \in N \setminus \{i\}$ and observes the opinion $y_{\lambda_t(i)t}$. In choosing targets, individuals seek the most informative opinion. That is, individuals are myopic with payoff at t equal to the negative of the variance of their belief about θ_t, once this belief has been updated based on their own private signal and the opinion they choose to observe.

Certain aspects of the model are worth emphasizing. First, this is a model of listening, in that individuals do not actively communicate or misrepresent their opinions, and targets are the only endogenously chosen source of information. Second, the interaction is not strategic: the payoff to an individual depends only on her own choices. And third, the model extends that in Sethi and Yildiz (2016), which corresponds to the special case of $\rho_1 = \rho_2 = 0$. Here we introduce distinct groups, defined by correlated perspectives, and study the effects of such correlation on patterns of communication.

4 Preliminaries

In this section we describe how people choose targets and make inferences about perspectives, and how this affects patterns of subsequent communication.

Beliefs about the perspectives of others change through the observation of posteriors.

Let $\Sigma(i,t)$ denote the variance-covariance matrix of μ at time t as perceived by individual i. Here $\Sigma(i,t)$ has entries $\sigma_{jj'}(i,t)$ with variances $\sigma_{jj}(i,t) \equiv \sigma_j^2(i,t)$ on the diagonal. These reflect i's uncertainty about each individual j's perspective. For all i and t, $\sigma_{jj'}(i,t) = 0$ whenever j and j' belong to distinct groups.[3]

Initially, an individual's uncertainty about the perspectives of those who are not in her own group is given by σ_0^2. That is, $\sigma_j^2(i,0) = \sigma_0^2$ if i and j belong to different groups. However, if i and j both belong to group k, initial beliefs are updated as follows. Observing μ_i, i believes that μ_j is distributed normally with some mean $E_i[\mu_j|\mu_i]$ and variance

$$\sigma_j^2(i,0) = Var_i(\mu_j|\mu_i) = \sigma_0^2(1-\rho_k^2). \tag{4}$$

If $\rho_k = 1$, all individuals in group k have the same perspective, and this common perspective is known to all members of the group. If $\rho_k < 1$, there is subjective uncertainty about all perspectives other than one's own. But uncertainty is smaller for own-group members than out-group members, and this greater initial understanding of one's own culture is a key channel through which group membership influences communication in the model.

4.1 Choosing Targets

Each individual can observe exactly one other opinion in each period, and makes this choice with the objective of having the most precise beliefs about the current state. But the most informative opinion need not come from the best-informed source.

If i observes j's opinion in period t, then from (1) and (3), she obtains the following signal for θ_t:

$$(1+\tau_{it}^2)y_{jt} = \theta_t + \varepsilon_{jt} + \tau_{jt}^2\mu_j.$$

This signal is noisy for two reasons: j's information is not perfect, and j's perspective is not perfectly known to i. The variance of the noise in the signal is

$$\gamma_{ij}(t) = \tau_{jt}^2 + \tau_{jt}^4\sigma_j^2(i,t). \tag{5}$$

This expression reveals clearly that in choosing a target j, an individual i has to trade-off noise in the information of j against uncertainty in i's understanding of j's perspective.

[3]Furthermore, since i's own perspective is known to her, the terms $\sigma_{ij}(i,t)$ and $\sigma_{ji}(i,t)$ are all identically zero for all t and all $j \in N$.

This is the trade-off between sources who are well-informed, and those who are well-understood by the observer. Initially, individuals within one's own group are better understood relative to those outside, but this can change over time as information is realized and perspectives become better understood.

The assignment of individuals to targets in period t may be represented by a function $\lambda_t : N \to N$, where

$$\lambda_t(i) \in \arg\min_j \gamma_{ij}(t). \tag{6}$$

That is each individual in each period chooses a target whose opinion is the most informative about the state. In case of ties we assume simply that i chooses the target with the smallest label. This tie-breaking rule does not play a role in our results.[4]

4.2 Learning Perspectives

Suppose that i observes the opinion y_{lt} of individual l in period t; i.e., $\lambda_t(i) = l$. This opinion has been formed in accordance with (2-3), and hence provides the following signal for μ_l:

$$\left(\frac{1+\tau_{lt}^2}{\tau_{lt}^2}\right) y_{lt} = \mu_l + \frac{1}{\tau_{lt}^2}\left(\theta_t + \varepsilon_{lt}\right).$$

The signal contains an additive noise term with variance

$$\alpha(\tau_{it}^2, \tau_{lt}^2) = \frac{1}{\tau_{lt}^4}\left(\frac{\tau_{it}^2}{1+\tau_{it}^2} + \tau_{lt}^2\right). \tag{7}$$

Note that the variance of the noise is bounded above and below:

$$\alpha(\underline{\tau}^2, \overline{\tau}^2) \leq \alpha(\tau_{it}^2, \tau_{jt}^2) \leq \alpha(\overline{\tau}^2, \underline{\tau}^2).$$

After observing her target's opinion, i's beliefs about the target's perspective become more precise in accordance with:[5]

$$1/\sigma_l^2(i, t+1) = 1/\sigma_l^2(i, t) + 1/\alpha(\tau_{it}^2, \tau_{lt}^2). \tag{8}$$

That is, the precision of i's belief about her target l's perspective increases by an amount that depends on the expertise levels of both i and l. Note that α is decreasing in τ_{lt}^2 and

[4]In cases with discrete signal variance distributions ties arise with positive probability. For simulations dealing with such cases, we assume that ties are broken uniformly at random.

[5]This expression follows from standard formulas, and can also be obtained by replacing both j and j' with l in equation (9) below.

increasing in τ_{it}^2. Hence, other things equal, if i happens to observe l during a period in which l is very precisely informed about the state, then i learns very little about l's perspective. This is because l's opinion largely reflects her signal and is therefore relatively uninformative about her prior. And if i is very well informed when observing l, the opposite effect arises, and i learns a great deal about l's perspective. Having good information about the state also means that i has good information about the distribution of l's signal, and is able to make a sharper inference about l's perspective based on the observed opinion.

Since perspectives are correlated, i also learns something about other members of the group to which her target belongs. We call this *indirect learning*. Specifically, observing the opinion of her target, i updates her beliefs about $(\mu_1, \ldots, \mu_n)$, resulting in a new variance-covariance matrix $\Sigma(i, t+1)$ with entries

$$\sigma_{jj'}(i, t+1) = \sigma_{jj'}(i, t) - \frac{\sigma_{jl}(i, t)\sigma_{j'l}(i, t)}{\alpha(\tau_{it}^2, \tau_{lt}^2) + \sigma_l^2(i, t)} \tag{9}$$

for each pair $j, j' \in N$. Since perspectives are correlated within groups, i updates her beliefs about all those in the group to which her target belongs, even though these individuals are not directly observed by i in t. That is, $\sigma_{jj}(i, t+1) < \sigma_{jj}(i, t)$ for each j from the target's group. Note that i does not update her beliefs about the perspectives of those in the group to which her target does *not* belong.

That is, observing a target is informative about the current state, the target's perspective, and the perspectives of all others in the target's group. Given the distribution governing expertise realizations, the dynamics of belief updating define a Markov process where the period t state consists of the variance-covariance matrices $\Sigma(i, t)$ for $i \in N$. These matrices, together with the expertise realizations in t, fully determine the pattern of observation that will arise in each period.

4.3 Free and Broken Links

Define the threshold variance

$$\overline{\sigma}^2 = \frac{\overline{\tau}^2 - \underline{\tau}^2}{\underline{\tau}^4}. \tag{10}$$

By (5) an individual i is indifferent between a target j with maximally precise signal and $\sigma_j^2(i, t) = \overline{\sigma}^2$ and a target j' with minimally precise signal and $\sigma_{j'}^2(i, t) = 0$. Hence, if $\sigma_j^2(i, t) < \overline{\sigma}^2$ at some period t, then i links to j when j has very high expertise and every

$j' \neq j$ has very low expertise. This is a positive probability event in period t, and remains so in all subsequent periods. Accordingly, if $\sigma_j^2(i,t) < \overline{\sigma}^2$ we say that the link from i to j is *free*. In this case i will observe j infinitely often almost surely, learning the perspective of j in the long run.

Next define the mapping $\beta : (\overline{\sigma}^2, \infty) \to \mathbb{R}$ by

$$\beta(s^2) = \frac{\underline{\tau}^4}{\overline{\tau}^4}(s^2 - \overline{\sigma}^2). \tag{11}$$

Here, $\beta(s^2)$ is defined by equality $\underline{\tau}^2 + \underline{\tau}^4 s^2 = \overline{\tau}^2 + \overline{\tau}^4 \beta\left(s^2\right)$, so that by (5), an individual i is indifferent between a target j with maximally precise signal and $\sigma_j^2(i,t) = s^2$ and a target j' with minimally precise signal and $\sigma_{j'}^2(i,t) = \beta\left(s^2\right)$. Hence, if $\sigma_{j'}^2(i,t) < \beta(\sigma_j^2(i,t))$, then i will not link to j at period t, regardless of expertise realizations.

If priors were uncorrelated, as in Sethi and Yildiz (2016), i would never link to j again once this condition is met. But with correlated priors, i may learn about j's perspective by observing the opinions of other members of the group to which j belongs, and this may induce i to link to j in subsequent periods. However, if there is j' from a group k' with

$$\sigma_{j'}^2(i,t) < \min_{j \in N_k} \beta(\sigma_j^2(i,t)), \tag{12}$$

then i will never link to any j in group k thereafter, and we say that all links from i to those in group k are *broken*. More generally, we say that a link from i to j is broken if the probability that it will form in any future period is zero.

If $\sigma_0^2 < \overline{\sigma}^2$, then all links are free at the outset. In this case (as we show below) targets will be chosen on the basis of expertise alone, regardless of group membership, in the long run. If, instead, we have $\sigma_0^2 > \overline{\sigma}^2$, then observational patterns even in the long run will depend on the degree to which perspectives are correlated within each group.

4.4 Long Run Experts and Homophily

Over time, each individual sharpens her understanding of her targets and, to a lesser extent, also her understanding of those who share a culture with these targets. As in Sethi and Yildiz (2016), after a finite number of periods, each potential link j either becomes free or breaks. For each i, therefore, there exists some (history-dependent) set $J_i \subseteq N \setminus \{i\}$ of long-run experts who are observed infinitely often. The perspectives of these long-run experts are learned to an arbitrarily high degree of precision, and hence i eventually links with high likelihood to the most informed individual in J_i in each period. Formally, for each infinite history and $\epsilon > 0$, there exists some $\bar{t}$ such that, for each $t > \bar{t}$,

$$\lambda_t(i) \in J_i \text{ and } \tau^2_{\lambda_t(i)t} \leq \min_{j \in J_i} \tau^2_{jt} + \epsilon.$$

Our main concern in this paper is the manner in which culture affects these sets of long run experts, with particular focus on the degree of homophily.

In the context of information gathering, homophily refers to the extent to which individuals exhibit a preference for own-group members in the process of observing opinions. To measure this, for each individual i in group k, we define a (history-dependent) index of homophily as follows:

$$\eta_i = \frac{|J_i \cap N_k|}{|J_i|}$$

This is the proportion of i's long run experts who belong to i's own group. The index lies in the unit interval and equals 1 if i's long run experts all lie in her group, and equals 0 if they all lie outside it; we shall refer to these cases as extreme homophily and extreme heterophily respectively.

This measure of homophily does not adjust for group size, so even if all individuals were completely unbiased in their choice of targets, members of larger groups will exhibit greater values of the homophily index. We say that an individual i satisfies *baseline homophily* in the long run if $\eta_i = \eta_i^*$ where

$$\eta_i^* = \frac{n_k - 1}{n - 1}.$$

An individual who eventually chooses targets based only on their expertise levels will exhibit baseline homophily.[6] If $\eta_i > \eta_i^*$ then i is said to exhibit *inbreeding homophily*, and if $\eta_i < \eta_i^*$ then she exhibits *heterophily*.

To measure the degree to which an individual i in group k exhibits inbreeding homophily (or heterophily) we follow Currarini et al. (2009) and define

$$\zeta_i = \frac{\eta_i - \eta_k^*}{1 - \eta_k^*}.$$

This index equals zero at baseline homophily, and is negative if i exhibits heterophily. It takes its maximum value of one when i consults only individuals from her own group.

[6]So will an individual who chooses targets entirely at random, but such choices will not be consistent with the assumed decision rule.

These indexes can be applied to groups rather than individuals by simple aggregation. As we shall see, a group may exhibit moderate levels of homophily even as its individual members exhibit either extreme or baseline homophily.

5 Observational Networks

Although our primary interest is in characterizing long run observational patterns for general levels of correlation, we start with a static model that corresponds to the initial period of our dynamic model. We then consider two extreme cases—perfect correlation and independent perspectives—as benchmarks for the multi-period setting. Finally we turn to the general case, present our main results, and explore some additional features of the model using a family of examples.

5.1 The Initial Period

Consider the patterns that can arise in a version of the model with a single period. Given myopic choices, this is equivalent to the first period of the dynamic model.

Since there is no observational history, the perspectives of others in one's group are learned only through introspection, and those of individuals outside one's group are not learned at all. For any i and j both belonging to the same group k, i updates her beliefs about the perspective of j based on her own perspective, resulting in a more precise belief about μ_j in accordance with (4). The variance of i's beliefs about the perspectives of those not in group k remains at σ_0^2. Thus individuals have a better understanding of the perspectives of their fellow group members and can better extract information from their opinions in the initial period.

It is clear from this that a group containing the best-informed person in the population as a whole will exhibit inbreeding homophily. In fact, there are only two possible types of observational pattern in the static model: either both groups exhibit inbreeding homophily, or one exhibits extreme heterophily.

To see why, let $j_1 \in \arg\min_{j \in N_1} \tau_{j0}^2$ and $j_2 \in \arg\min_{j \in N_2} \tau_{j0}^2$ denote the best-informed members of their respective groups. As long as these two individuals have similar levels of expertise, the initial affinity for own-group targets will ensure that we have inbreeding

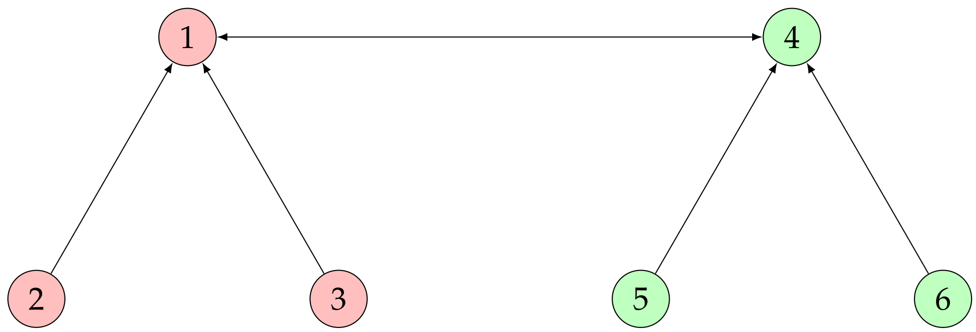

Figure 1: Inbreeding homophily in both groups

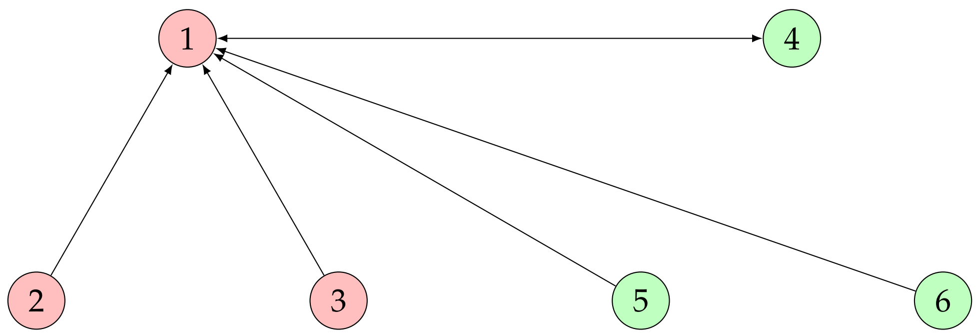

Figure 2: Inbreeding homophily in one group and heterophily in the other

homophily in both groups:

$$\lambda_0(i_1) = j_1 \text{ and } \lambda_0(i_2) = j_2 \text{ for all } i_1 \in N_1 \setminus \{j_1\} \text{ and } i_2 \in N_2 \setminus \{j_2\}$$

This outcome is illustrated in Figure 1.

When the expertise levels of j_1 and j_2 differ sharply, a different outcome can arise, with the globally best-informed individual being targeted by all others in the population. This outcome is illustrated in Figure 2, where we see inbreeding homophily in one group and extreme heterophily in the other.

As the correlation in perspectives gets larger, homophily in both groups becomes more likely. If this correlation is high enough then homophily is ensured, as long as $\sigma_0^2 > \overline{\sigma}^2$. This is because $\sigma_j^2(i,0) < \beta(\sigma_0^2)$ whenever $\rho_k > \overline{\rho}$, where $\overline{\rho}$ is the solution to $\sigma_0^2(1-\rho^2) = \beta(\sigma_0^2)$, and given by:

$$\overline{\rho} = \sqrt{1 - (\underline{\tau}^4/\overline{\tau}^4)(1 - \overline{\sigma}^2/\sigma_0^2)}. \tag{13}$$

To summarize, inbreeding homophily in both groups arises with positive probability for all parameter values. If there are large differences in expertise between the best-informed individuals in the two groups, and the correlation in perspectives is not too great in either group, then the best-informed individual in the population as a whole will attract all observers. But if $\sigma_0^2 > \overline{\sigma}^2$, then any group with highly correlated perspectives will exhibit inbreeding homophily. The degree to which members of such groups understand each other will overwhelm any informational disadvantage that might arise.

This applies to the static model in which all individuals within a group are symmetrically placed with respect to each other. In the dynamic model, this symmetry is broken over time and more complex patterns can arise.

5.2 Two Benchmarks

Recall that for each individual i, there exists a history-dependent set $J_i \subseteq N \setminus \{i\}$ of long-run experts, such that, in each period, i targets the most informed individual in J_i for that period in the long run. We now investigate what form these long-run expert sets must take, starting with two benchmarks: perfectly correlated priors and uncorrelated priors.

Suppose first that perspectives are fully determined by group membership:

$$\rho_1 = \rho_2 = 1.$$

In this case, for any given group k, all members have a common perspective, and this perspective is known to all members of the group. Suppose, first, that we have high cross-cultural uncertainty:

$$\sigma_0^2 > \overline{\sigma}^2.$$

Since $\rho_k = 1 > \overline{\rho}$ for each k, each i in each group k will initially target the most informed individual (other then herself) in her own group. Since individuals already know the common perspective of their own group, and learn nothing about the common perspective of the other group, their beliefs about perspectives remain unchanged. Hence, at $t = 1$, they repeat the this behavior, and keep repeating it thereafter:

$$\lambda_t(i) \in \arg\min_{j \in N_k \setminus \{i\}} \tau_{jt}^2 \qquad (\forall t, k, \forall i \in N_k).$$

Given the initial high uncertainty about each other, groups remain isolated throughout and there is no cross-cultural communication.

Next consider the case of low cross-cultural uncertainty:

$$\sigma_0^2 < \overline{\sigma}^2.$$

As saw in Section 5.1, the pattern of communication in the initial period now depends on expertise realizations. If the best-informed individuals in the two groups have similar expertise levels, we will see extreme homophily to begin with. However, if the difference between the expertise levels of the best informed individuals in two groups is large, one of the groups will exhibit heterophily. Since there is a positive lower bound for the probability of cross-cultural communication at every period, all individuals will have an opportunity to learn about the other culture infinitely many times, almost surely, in the long run. Moreover, there is a positive lower bound for the reduction in uncertainty about perspectives, and hence every individual will eventually learn every perspective. Cultures will ”merge” in the long run, and everybody will target the most informed individual—regardless of group identity—in the population at large.

To summarize, when perspectives are perfectly correlated within each group, there is either a merging of cultures—with all individuals eventually targeting the best informed person in the population—or complete segregation in the long run. These relative simple communication structures preclude any within-group heterogeneity.

As a second benchmark, consider the case of uncorrelated priors in each group:

$$\rho_1 = \rho_2 = 0.$$

This case has been studied extensively in Sethi and Yildiz (2016), and we summarize some of the most relevant findings here. With uncorrelated priors, since the perspective of one individual tells us nothing about the perspectives of others, one can view this case as n cultural groups with a single member in each.

When $\sigma_0^2 < \overline{\sigma}^2$, as in the case of perfect correlation, individuals eventually learn all other perspectives and target the most informed individual in the long run. When $\sigma_0^2 > \overline{\sigma}^2$, an extreme form of informational leadership arises with positive probability: for some $i^* \in N$, every individual $i \neq i^*$ targets i^* at every period t. Moreover, such extreme individual leadership is the only possibility when uncertainty about perspectives is sufficiently high. Although a wide range of observational patterns emerge with positive probability for intermediate values of σ_0^2 in the long run, a weaker form of opinion leadership arises with high probability: each individual focuses on a small set of long-run experts in the long run. Formally, for $\sigma_0^2 > \overline{\sigma}^2$, Sethi and Yildiz (2016) obtain a uniform upper bound for $\Pr\left(|J_i| > m\right)$ where the upper bound is exponentially decreasing in m. We show below that a similar "law of the few" also holds in the more general case of correlated priors.

In both these benchmarks, when cross-cultural uncertainty is low (i.e. $\sigma_0^2 < \overline{\sigma}^2$), there is a merging of cultures, and each individual observes the most informed source in each period in the long run. Our first result below shows that this holds more generally, with arbitrary levels of within-group correlation in perspectives.

5.3 Merging of Cultures

When initial uncertainty about perspectives is low, individuals get to observe others and eventually learn their perspectives, listening to the most informed individuals in the long run. In particular, group identity eventually becomes irrelevant.

Proposition 1. *If $\sigma_0^2 < \overline{\sigma}^2$, then $J_i = N \setminus \{i\}$ for all i almost surely; all individuals exhibit baseline homophily in the long run.*

When $\sigma_0^2 < \overline{\sigma}^2$, all individuals exhibit baseline homophily except on histories that arise with zero probability. The reason is that an individual will always prefer to observe a target from the other group if the latter is sufficiently well informed, provided that the best-informed in her own-group is sufficiently poorly informed. This is a positive probability event regardless of history. The implication is that all perspectives are learned to a high degree of accuracy in the long run, and all medium run effects arising from the dependence on history of observational choices are washed away.

The remainder of this section is focused on the case of high cross-cultural uncertainty (i.e. $\sigma_0^2 > \overline{\sigma}^2$). Here the two benchmark models exhibited very different long run characteristics: extreme homophily under perfect correlation ($\rho_k = 1$) and opinion leadership with positive probability under independent perspectives ($\rho_k = 0$). For more general correlations, we will show that there must be extreme homophily when ρ_k is sufficiently high. Rich patterns of behavior—including extreme homophily—can emerge otherwise. We will provide a lower bound on the probability of extreme homophily, where this bound increases to 1 as ρ_k rises.

For intermediate values of ρ_k, we will show that individuals will either listen to a limited number of individuals from a culture or open up to the culture completely, listening to any member of the group when she is more informed than others in the long run. This is due to the indirect learning of perspectives. When ρ_k is low, indirect learning is not sufficient, and our model inherits the properties of the benchmark case with independent perspectives. For that case, we will show that each individual is likely to listen to only a few individuals in the long run.

5.4 Homophily

If $\sigma_0^2 > \overline{\sigma}^2$, then long run observational patterns depend on the correlation in perspectives, and individuals may exhibit extreme levels of homophily:

Proposition 2. *When $\sigma_0^2 > \overline{\sigma}^2$, there is a positive probability that each individual exhibits extreme homophily:*

$$\lambda_t(i) \in N_k \qquad (\forall t, \forall k, \forall i \in N_k).$$

In addition, if $\rho_k > \overline{\rho}$, then all individuals in group k exhibit extreme homophily almost surely.

Hence, unless the sufficient condition for baseline homophily holds, each individual exhibits extreme homophily with positive probability. In groups with sufficiently high correlation in perspectives, extreme homophily is ensured. The reason is as follows. For high enough correlation, all individuals observe within group members in the first period regardless of expertise realizations. But this only increases understanding of within group perspectives, without raising knowledge of any perspectives outside the group. As a result, the likelihood of extreme homophily cannot decline, and remains at 1. Even if correlation in perspectives is low, extreme homophily in the first period is a positive probability event, and a repetition of this first period network for some finite number of periods is also a positive probability event. If this number of periods is large enough, then each individual in the group develops so great an understanding of their initial target's perspective that no other target is ever subsequently observed, regardless of the expertise realizations that may later arise.

We next present a lower bound on the probability that a given individual in a given group will exhibit homophily in the long run. As the group size gets large, this probability approaches 1 under any $\rho_k > 0$.[7] For simplicity, we use a binary expertise distribution:

$$\tau_{it}^2 = \begin{cases} \underline{\tau}^2 & \text{with probability } q, \\ \overline{\tau}^2 & \text{with probability } 1-q. \end{cases} \tag{14}$$

Our result can be extended easily to general distributions at the expense of clarity.

Proposition 3. *Assume that expertise levels are independently and identically distributed across individuals and over time according to the binomial distribution in (14). Then, for any $\sigma_0^2 >$*

[7] Our bound is derived using the techniques in Sethi and Yildiz (2016).

$\overline{\sigma}^2$, any group k and any $i \in N_k$, individual i exhibits extreme homophily in the long run with probability

$$\Pr\left(J_i \subset N_k\right) \geq \left(\frac{q+(1-q)^{n-1}}{q+(1-q)^{n_k-1}}\right)^{\max\{\lceil\kappa\rceil,0\}} \equiv p^*$$

where

$$\kappa = \frac{\alpha(\overline{\tau}^2,\underline{\tau}^2)}{\sigma_0^2}\left[\frac{1}{1-\overline{\rho}^2}-\frac{1}{1-\rho_k^2}\right]. \tag{15}$$

Proposition 3 provides a lower bound p^* on the probability of extreme homophily under the binomial distribution. This bound has two elements. First, the term in brackets measures the relative likelihood of listening to the most informed in-group member, rather than crossing the boundary and listening to someone from the other group. Second, the exponent $\max\{\lceil\kappa\rceil, 0\}$ measures the number of repeated observations of an in-group member that is sufficient to break all the links to the other group.[8] That is, links to the other group are broken permanently after $\lceil\kappa\rceil$ observations of an in-group member's opinion if nobody from the other group has been observed in the meantime.

The bound p^* satisfies several intuitive comparative statics. First, it is increasing in group size n_k, and it approaches to 1 as $n_k \to \infty$. That is, as the group size gets large, extreme homophily becomes virtually certain. This is intuitive. The likelihood of extreme homophily in the initial period is increasing in group size; the larger the group the more likely it will be that the group contains an individual who is the globally best-informed. If this happens repeatedly for some initial set of periods, all links to out-group members break and expertise realizations in subsequent periods become irrelevant. The number of needed repetitions may be large, but is finite for any $\rho_k > 0$. The probability of this event can be made arbitrarily close to 1 by increasing group size.[9] In the same vein, the lower bound is increasing in probability q of high expertise and approaches 1 as $q \to 1$.

Second, the bound p^* is increasing in ρ_k and approaches 1 as ρ_k approaches the cutoff $\overline{\rho}$, after which there is extreme homophily with probability 1 (by Proposition 2). That is, the more tight-knit the group, the stronger is the force for homophily. Under the binomial distribution, when $\overline{\rho} > \rho_k > 0$, ρ_k does not affect the probability of homophily in the first period. It affects the probability of homophily in the long run by weakening the links to

[8]The operators $\lceil\cdot\rceil$ and $\lfloor\cdot\rfloor$ respectively round up and round down to the nearest integer.

[9]Note that this argument is quite general, and applies even if expertise is continuously distributed. When a group is large, there is a high likelihood that it will contain someone with expertise close to that of the globally best informed. This is true even if the other group is even larger. With correlated perspectives, we again get homophily with high probability.

the other group, so that the links to the other group break faster as ρ_k increases. Finally, it is increasing in initial uncertainty σ_0^2 about the perspectives of the other group. The lower bound approaches 0 as σ_0^2 goes down to $\overline{\sigma}^2$ (recall that individuals exhibit baseline homophily in the long run when $\sigma_0^2 < \overline{\sigma}^2$ by Proposition 1).

5.5 A Bang-Bang Result

More complex and interesting observational patterns can arise if $\sigma_0^2 > \overline{\sigma}^2$, and $\rho_k < \overline{\rho}$ for at least one group k. In this case neither universal unbiasedness nor universal extreme homophily are ensured, and substantial within-group behavioral diversity can arise. To explore this case, for each individual i and group k, let

$$m_{ik} = |(J_i \cup \{i\}) \cap N_k|$$

be the number of individuals from group k whose opinions i observes infinitely often; this set includes herself if she belongs to group k. Using information on these m_k perspectives alone, i can update her beliefs about the perspectives of all other individuals in group k, even if none of these has ever been observed. The variance of these updated beliefs is $\sigma_0^2/\phi\left(m_{ik}, \rho_k\right)$ where

$$\phi\left(m, \rho\right) = \frac{1 - \rho/\left(m\rho + 1\right)}{1 - \rho} \geq 1. \tag{16}$$

That is, i's uncertainty about the perspectives of all individuals in group k shrinks by a factor of (at least) $\phi\left(m_{ik}, \rho_k\right)$ if she observes m_k members of the group infinitely often. This is the *indirect learning* that we discussed in the introduction. Culture affects the dynamics of communication patterns through this channel, as we establish now.

Note that the factor $\phi\left(m, \rho\right)$ is increasing in both m and ρ. It is increasing in ρ because it is easier to learn about others indirectly in more homogenous groups. It is increasing in m simply because m is the size of the data. Importantly for our paper, there is a limit to indirect learning, and the initial sampling accounts for a large fraction of indirect learning. Formally, as $m \to \infty$, $\phi\left(m, \rho\right)$ approaches to $1/\left(1 - \rho\right)$, and indirect learning can at most reduces the variance by a factor of $1 - \rho$. In contrast, the first observation reduces the uncertainty by a factor of

$$1/\phi(1, \rho) = 1 - \rho^2.$$

Hence, the initial sampling accounts for $1/\left(1 + \rho\right)$ fraction of infinite sampling, accounting for more than half of the reduction in uncertainty.

If the variance $\sigma_0^2/\phi(m_{ik},\rho_k)$ of the updated beliefs falls below the cutoff $\overline{\sigma}^2$, then i learns the perspectives of the other members of group k so well that she links to each member of group k infinitely often. In this case we must have $m_{ik} = n_k$. When ρ_k is below the cutoff

$$\underline{\rho} = 1 - \overline{\sigma}^2/\sigma_0^2, \tag{17}$$

the variance $\sigma_0^2/\phi(m_{ik},\rho_k)$ never falls below $\overline{\sigma}^2$, and indirect learning is never sufficient to free a link by itself, regardless how many perspectives one learns from group k. On the other hand, when the correlation ρ_k for group k is above the cutoff $\sqrt{\underline{\rho}}$, the variance $\sigma_0^2/\phi(m_{ik},\rho_k)$ falls below $\overline{\sigma}^2$ at $m_{ik} = 1$, and all links within group k become free at the beginning as they learn about the perspectives of the other group members from their own perspectives. In the long run, they link to a group member whenever that group member is more informed than the others. These two cutoffs will play important role in our analyses below, as the long-run communication patterns will dramatically depend on where the correlations ρ_k falls with respect to these cutoffs.

To identify when $\sigma_0^2/\phi(m_{ik},\rho_k)$ falls below $\overline{\sigma}^2$ more generally, suppose that $\rho > \underline{\rho}$ and define

$$\overline{m}\left(\sigma_0^2,\rho\right) = \left(\frac{1-\rho}{\rho}\right)\frac{\sigma_0^2 - \overline{\sigma}^2}{\overline{\sigma}^2 - (1-\rho)\sigma_0^2}. \tag{18}$$

This solves the equation

$$\frac{\sigma_0^2}{\overline{\sigma}^2} = \phi(m,\rho).$$

Note that $\overline{m}\left(\sigma_0^2,\rho\right)$ is positive and finite if and only if $\overline{\sigma}^2/\sigma_0^2 > 1-\rho$, i.e., $\rho > \underline{\rho}$.

When $m_{ik} > \overline{m}\left(\sigma_0^2,\rho_k\right)$, we have $\sigma_0^2/\phi(m_{ik},\rho_k) < \overline{\sigma}^2$. This implies that if i learns the perspectives of $\overline{m}\left(\sigma_0^2,\rho_k\right)$ or more members of group k, she must link to all members of group k in the long run. Bearing in mind that each individual knows her own perspective to begin with, this reasoning leads to the following result.

Proposition 4. *Consider any individual i and any group k with $\rho_k > \underline{\rho} \equiv 1 - \overline{\sigma}^2/\sigma_0^2$. Then:*

$$(J_i \cup \{i\}) \cap N_k = N_k \text{ or } |(J_i \cup \{i\}) \cap N_k| < \overline{m}(\sigma_0^2,\rho_k).$$

This result states that each individual i either links to every other member of her own group with positive probability in the long run, or links to at most $\overline{m}\left(\sigma_0^2,\rho_k\right) - 1$ of them. In addition, either i links to every member of the other group k', or links to at most $\overline{m}\left(\sigma_0^2,\rho_{k'}\right)$ of them. In either case, the set of individuals whom i consults infinitely often must either fall below some threshold (that depends on the correlation in group perspectives), or must constitute the entire group.

Observing a sufficiently large number of individuals in a group many times leads to sharp beliefs about the perspectives of others in the group, which makes these also desirable candidates for observation when they are well-informed. Since one's own perspective is known at the outset, the critical number of own-group targets that need to be observed for this effect is lower. Note that the threshold $\overline{m}\left(\sigma_0^2, \rho\right)$ is increasing in σ_0^2 and decreasing in ρ. Hence high correlation and low initial uncertainty about perspectives lead to lower thresholds for this tipping process to arise, and make it more likely that all members of the group will eventually be observed.

While the two critical thresholds (own group and other group) may seem virtually the same, the probability of homophily can be significantly greater than that of heterophily, since early observations are more likely to involve own group members. This is especially the case for large and tight-knit groups, as we have seen. More importantly, as we discussed above, most of indirect learning happens through the first sampling of perspectives, leading to homophily.

The cutoff $\sqrt{\underline{\rho}}$ yields a very simple and intuitive condition under which i either exhibits extreme homophily or completely unbiased behavior. Suppose $i \in N_k$. If $\rho_k > \sqrt{\underline{\rho}}$, then i must link to all in-group members. This follows from Proposition 4 and the fact that $\overline{m}\left(\sigma_0^2, \rho_k\right) < 1$ in this case. Now suppose that $i \notin N_k$, and $\rho_k > \sqrt{\underline{\rho}}$. Then, i must link to all or none of those in group k. Taken together, this means that i exhibits either extreme homophily in the long run, or complete unbiasedness, in the sense that she simply observes the individual who is globally best informed.

Corollary 1. *For any $i \in N$ and any group k with $\rho_k > \sqrt{\underline{\rho}}$, we have*

$$\begin{aligned} J_i \cap N_k &= N_k \setminus \{i\} \text{ if } i \in N_k \text{ and} \\ J_i \cap N_k &\in \{\varnothing, N_k\} \text{ if } i \notin N_k. \end{aligned}$$

In particular, if $\rho_k > \sqrt{\underline{\rho}}$ for each k, then, in the long run, each individual exhibits either maximal or baseline homophily.

One implication of the results obtained above is that when $\overline{m}$ is finite and a group is sufficiently large, repeatedly observing the opinions of a small fraction of group members is enough to learn the perspectives of all others in the group to a high degree, even if they have not been directly observed. As a result, these individuals will eventually come to

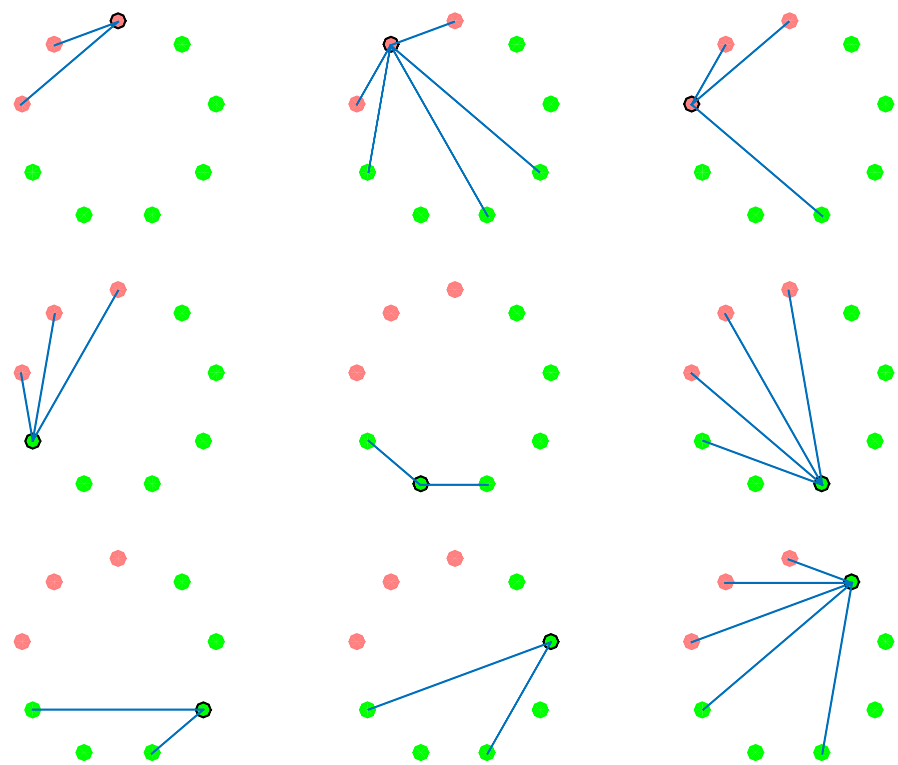

Figure 3: Free links in the long run for nine agents in two groups. Each cells is the observational network for one agent, depicted with a black boundary, and links are to the agent's long-run targets.

be observed, whenever they happen to be substantially better informed than others in the population. This leads to the following bang-bang result, which holds under a wider range of parameter values than in the small group case:

Corollary 2. *For any $i \in N$, any group k with $\rho_k > \sqrt{\underline{\rho}}$, and every $\epsilon > 0$, there exists $\overline{n}$ such that, if $n_k > \overline{n}$, then either $m_{ik} < \epsilon n_k$ or $m_{ik} = n_k$.*

That is, if group k is sufficiently large, all individuals in the population either observe only a small fraction of those in this group, or observe all members of the group in the long run.

The following example illustrates a case in which one group exhibits varying levels of homophily across individuals, while the other is characterized by considerable heterogeneity, with some individuals exhibiting homophily while the others exhibit heterophily.

Example 1. *Take $n_1 = 3, n_2 = 6, \rho_1 = 2/3, \rho_2 = 1/4, \underline{\tau}^2 = 1/2, \overline{\tau}^2 = 1$, and $\sigma_0^2 = 3$. Then $\overline{\sigma}^2 = 2$, so $\overline{\sigma}^2/\sigma_0^2 = 2/3$, yielding $\underline{\rho} = 1/3$. Since $\rho_1 > \sqrt{\underline{\rho}}$, all individuals in group 1 link to all* *others in their group, while all those in group 2 either link to all or none in group 1. Furthermore, since $\rho_2 < \underline{\rho}$, for all $i \in N$, the number of long-run links to members of group 2 is unconstrained.*

Figure 3 shows the long-run structures that arise in this example for a particular realization of expertise levels. All links are resolved (either broken or free) 38 periods have elapsed. The figure shows the long-run expert sets for each of the nine agents in a separate cell, with colors indicating group membership and a black boundary identifying the subject or observer in each cell. Consistent with the results, each of the three members of group 1 link to the other two infinitely often, as shown in the top row. They differ only with respect to their links to the other group, which range from none to three. Under the parameter specifications in the example, group 2 individuals must link to all or none of those in group 1, and this is also seen in the figure: three link to all and three to none.

While the figure shows only one realization of the process, and one set of possible long-run structures, it illustrates the manner in which long-run structures are constrained by the correlation in perspectives within groups.

5.6 The Size of Expert Sets

There is a limit to how much one can learn about an individual indirectly by observing other individuals from his group. Indirect learning can at most scale down the variance by $1 - \rho_k$. Hence, when $\rho_k < \underline{\rho}$, the effect of indirect learning is not sufficient for links to unobserved individuals to become free. In that case, our model inherits the long-run communication patterns under independent perspectives in Section 5.2.

For any individuals i and j and any history, define

$$\hat{\sigma}_j^2(i, \hat{t}) = \left(1/\sigma_0^2 + \sum_{\left\{t<\hat{t} | \gamma_t(i)=j\right\}} 1/\alpha(\tau_{it}^2, \tau_{jt}^2) \right)^{-1}.$$

This would be the variance of the belief of i about the perspective of $j \in N_k$ if she ignored the implications of the information she gleaned from observing the others in group N_k—including her own perspective if she also belongs to N_k. This is the variance from direct learning. Of course, as we have seen above, she also learns indirectly by observing others from N_k. However, there is a limit to such learning:[10]

$$\sigma_j^2(i, t) \geq \hat{\sigma}_j^2(i, t) / \phi(n_k - 1, \rho_k) > (1 - \rho_k) \hat{\sigma}_j^2(i, t).$$

[10]One can compute $\sigma_j^2(i, t)$ by first incorporating indirect learning and then direct learning. Since indirect

That is, the variance of the belief of i about the perspective of j is at least $(1-\rho_k)\,\hat{\sigma}_j^2(i,t)$. Intuitively, one can compute $\sigma_j^2(i,t)$ from $\hat{\sigma}_j^2(i,t)$ by conditioning on the information obtained indirectly from others in group N_k, and this can at most be what one could get if she learned the perspectives of all the others in group N_k, which would scale down the variance by $\phi(n_k-1,\rho_k)$.

This limitation of indirect learning causes the links to previously unobserved sources to break as one gets increasingly familiar with one source. In particular, when $\rho_k < \underline{\rho}$, we have $(1-\rho_k)\,\sigma_0^2 > \overline{\sigma}^2$, and hence indirect learning is not sufficient for an individual to observe previously unobserved members of group k even when they are uniquely well-informed about the current state. Hence, if $\sigma_j^2(i,t) < \beta\left((1-\rho_k)\,\sigma_0^2\right)$ for some j, individual i will never target those $j' \in N_k$ she has not observed yet, and her links to them will break permanently. This is because, although she may learn indirectly about these sources, the variance of her beliefs about them will remain above $(1-\rho_k)\,\sigma_0^2$.

The resulting long-run communication patterns will be similar to those under the independent priors discussed in Section 5.2. In particular, we next show that individuals are likely to focus on a small set of long-run experts in the long run, generalizing one of the main results in Sethi and Yildiz (2016) to the case of correlated perspectives. We use the binomial expertise distribution in (14) for simplicity.

Proposition 5. *Assume that expertise levels are independently and identically distributed across individuals and over time according to the binomial distribution in (14). Then, for any $i \in N$ and any group k with $\rho_k < \underline{\rho} \equiv 1-\overline{\sigma}^2/\sigma_0^2$, we have*

$$\Pr\left(|J_i \cap N_k| \le m\right) \ge \left(\frac{q}{q+(1-q)^m}\right)^{\kappa_s} \equiv p_k(m) \qquad (\forall m \ge 1) \tag{19}$$

where

$$\kappa_s = \left\lfloor \frac{\alpha(\overline{\tau}^2,\underline{\tau}^2)}{\beta\left((1-\rho_k)\,\sigma_0^2\right)} - \frac{\alpha(\overline{\tau}^2,\underline{\tau}^2)}{\sigma_0^2} \right\rfloor. \tag{20}$$

Proposition 5 establishes that the probability distribution p_k first-order stochastically dominates the number of the long-run experts from group k. Since p_k has exponential

learning can at most reduce σ_0^2 to $\sigma_0^2/\phi(n_k-1,\rho_k)$, we have

$$\sigma_j^2(i,t) \ge \left(\frac{\phi(n_k-1,\rho_k)}{\sigma_0^2} + \sum_{\{t<\hat{t}\,|\,\gamma_t(i)=j\}} \frac{1}{\alpha(\tau_{it}^2,\tau_{jt}^2)}\right)^{-1}.$$

Using the definition of $\hat{\sigma}_j^2(i,t)$ and simple algebra, one can show that the right-hand side is at least $\hat{\sigma}_j^2(i,t)/\phi(n_k-1,\rho_k)$.

tails, the probability of having more than m long-run expert from group k is exponentially decreasing in k, resulting in a small expected number of long-run experts when κ_s is low. Note that κ_s is increasing in ρ_k, and it goes to ∞ as ρ_k approaches $\underline{\rho}$. In that case $p_k(m)$ goes to zero, allowing large number of experts from group. This is intuitive because when $\rho_k > \underline{\rho}$, by Proposition 4, everybody in group k becomes a long run expert when sufficiently many of them become one. However, $p_k(m)$ is only a lower bound on the probability $\Pr(|J_i \cap N_k| \leq m)$, and this probability can be much higher if i is also familiar with members of the other group $k' \neq k$.

The logic of the proof is similar to the proof of Proposition 3. Under the binomial distribution, an individual i has a simple rule: if the individual j_1 with the lowest variance $\sigma_j^2(i,t)$ has high expertise (i.e., $\tau_{j_1 t}^2 = \underline{\tau}^2$) she targets j_1; if j_1 has low expertise, then she checks the individual j_2 with the second lowest variance $\sigma_j^2(i,t)$ and targets her if she has high expertise; if she also has low expertise, then she goes on to the next familiar source, and so on. Thus, she targets j_1 with at least probability q. On the other hand, if she has observed m individuals from group k some times in the past, she will not target any previously unobserved individual from group k unless all those m previously-observed targets have low expertise, which has probability $(1-q)^m$. Thus, the probability of observing the most familiar individual T times before observing any previously unobserved individual from group k is at least

$$\left(\frac{q}{q+(1-q)^m}\right)^T.$$

As she is making these observations, she learns about two things. First, she learns directly about her target's perspective, and there is a lower bound for that information, as in (21). Second, she also learns indirectly about the perspectives of the others in her target's group, but there is an upper bound for such indirect learning, and the variance of her belief about a unobserved individual's perspective remains above $\sigma_0^2(1-\rho_k)$, as we have discussed above. Then, as i observes a most familiar individual repeatedly, the lowest variance $\min_{j\in N\setminus\{i\}} \sigma_j^2(i,t)$ eventually drops below $\beta\left(\sigma_0^2(1-\rho_k)\right)$, when all her links to previously-unobserved individuals from group k breaks for good. Using the inequality in (21), one can easily show that the number of repetitions for this to occur is at most κ_s.

Proposition 5 provides a theoretical explanation for the empirical regularity that individuals seek information from just a few sources, even when many better informed sources exist. The reason is because previously observed sources have become better understood, and experimenting with new ones is no longer worthwhile. When groups are large and exhibit homophily, the expert sets of two individuals from the same group will

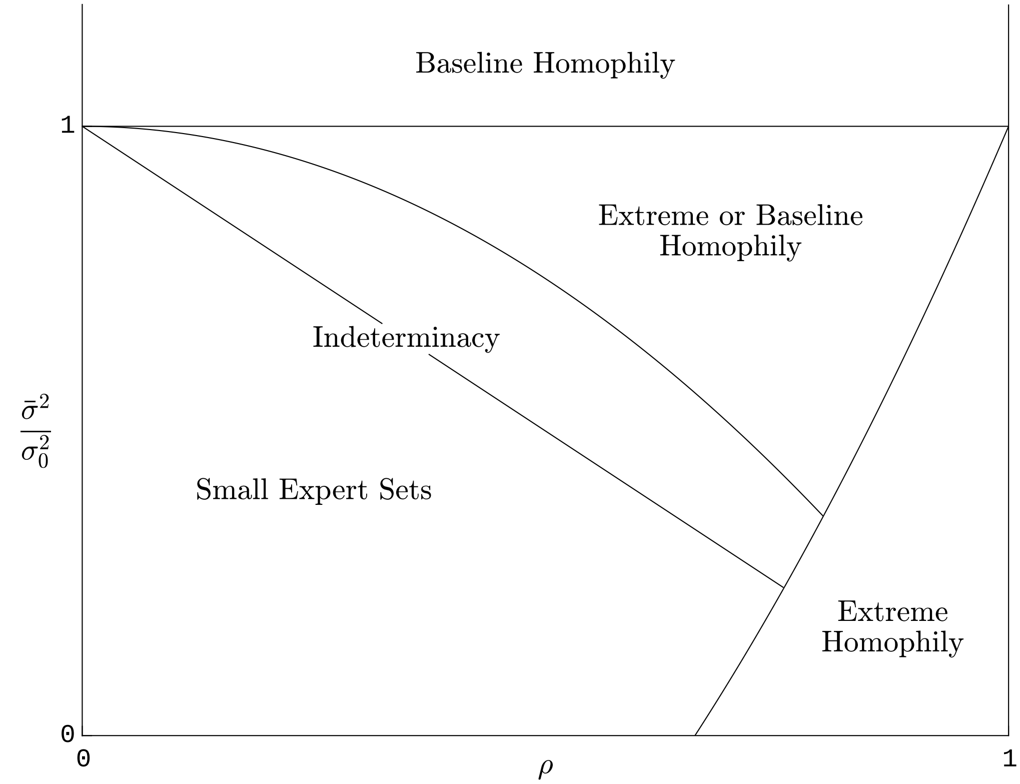

Figure 4: Regions of parameter space giving rise to homophily, unbiasedness, and indeterminacy.

largely coincide, and each group will have a small set of opinion leaders who are the main source of information for others in their group. This may be viewed as a multi-group version of what Galeotti and Goyal (2010) call the law of the few, though their model provides a very different theory of the phenomenon.

Our findings in this section are summarized in Figure 4 for $\rho_1 = \rho_2 = \rho$. This figure shows regions of the parameter space under which various long run structures arise. When $\rho > \overline{\rho}$, we have extreme homophily and all individuals observe only in-group members. When $\rho > \sqrt{\underline{\rho}}$, individuals observe all others in their own group, and either all or none in the other group. This is the region defined by $\overline{\sigma}^2/\sigma_0^2 > 1 - \rho^2$. When $\rho < \min\{\sqrt{\underline{\rho}}, \overline{\rho}\}$, more complex patterns of observation can arise, and there is considerable indeterminacy in outcomes. Here individuals may link to some members of a group while ignoring others, and there may be considerable within-group heterogeneity in behavior. Despite such indeterminacy, when $\rho < \underline{\rho} \equiv 1 - \overline{\sigma}^2/\sigma_0^2$, experts sets are small in expectation: each individual links to a few sources from each group in the long run with high probability.

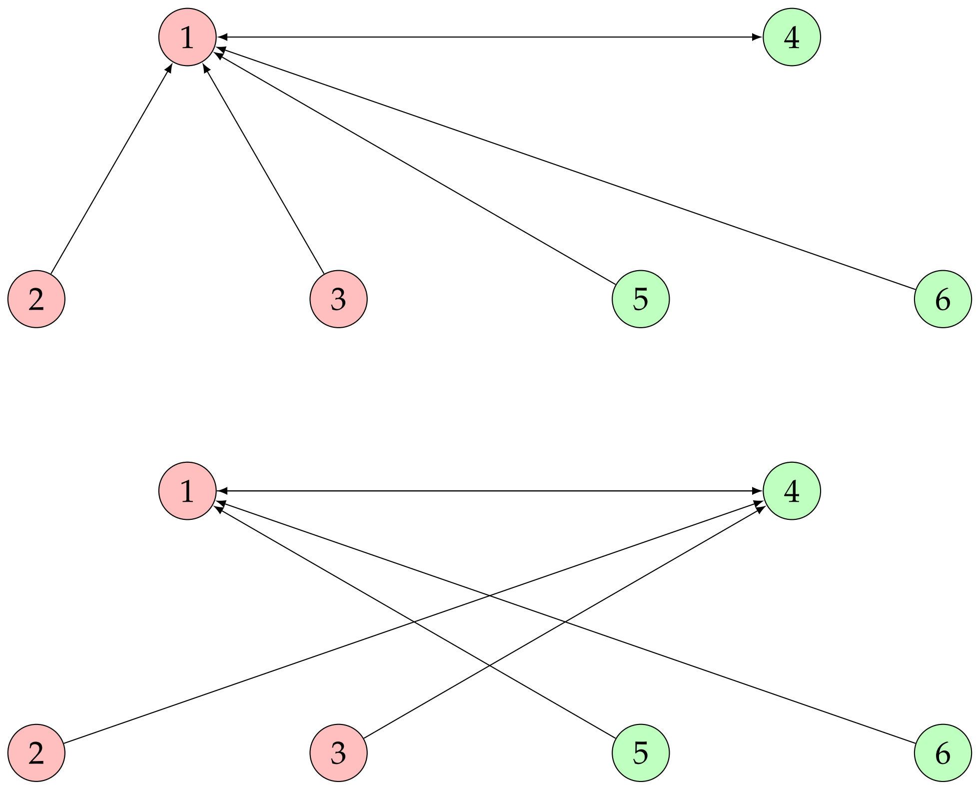

Figure 5: Homophily and heterophily followed by heterophily in both groups

5.7 Heterophily

When $\sigma_0^2 > \overline{\sigma}^2$ and $\rho_k < \overline{\rho}$, although homophily remains as a distinct possibility, other patterns of communication may emerge in the long run with positive probability. In particular, when ρ_k is low, learning effects may overwhelm cultural affinity, and individuals may exhibit heterophily in the long run. The next example illustrates this.

Example 2. *Suppose that $N_1 = \{1, 2, 3\}$, $N_2 = \{4, 5, 6\}$, $\sigma_0^2 = 10$, and $\rho_1 = \rho_2 = 0.1$. Consider the following sequence of expertise realizations and associated links:*

t	τ_{1t}^2	τ_{2t}^2	τ_{3t}^2	τ_{4t}^2	τ_{5t}^2	τ_{6t}^2	$\lambda_t(1)$	$\lambda_t(2)$	$\lambda_t(3)$	$\lambda_t(4)$	$\lambda_t(5)$	$\lambda_t(6)$
0	1	10	10	1.1	1.25	1.25	4	1	1	1	1	1
1	10	10	10	4	5	5	4	4	4	1	1	1

Then, at $t = 0$, individual 1 emerges as an opinion leader: $\lambda_0(i) = 1$ for all $i \neq 1$ and $\lambda_0(1) = 4$. At $t = 1$, we have extreme heterophily: $\lambda_1(i) = 4$ for all $i \in N_1$ and $\lambda_1(i) = 1$ for all $i \in N_2$.

In this example the two groups are each of size 3. In the first period the two individuals with the greatest expertise are 1 and 4, with 1 being the best informed globally. Clearly 2 and 3 observe 1, since they all belong to the same group. Since perspectives are not highly correlated, all those in N_2 also observe 1, and 1 observes 4. This pattern is shown in the top panel of Figure 5.

Now consider the second period. Since all those in N_2 were better informed than 2 and 3 in the initial period, they learn more about the perspective of 1 after the first observation. This follows directly from (7). As a result, they are more inclined to observe 1 again in the second period, even if there is a better informed individual in the population. Given the small correlation in perspectives, this effect is strong enough to overcome the fact that 4, a member of their own group, is globally the best informed in the second period. Hence 5 and 6 observe 1 in the second period. This is also the case for 4, who has learned even more about the perspective of 1 in the initial period, and whose best within-group option in the second period is worse.

Finally, consider the members of N_1 in the second period. Since 2 and 3 were poorly informed in the first period and learned little about the perspective of 1, they observe 4, who is the globally best informed individual in the second period. So does 1, who has already observed 4 in the initial period. As a result all members of the population observe someone outside their own group. This outcome is shown in the bottom panel of Figure 5. We get extreme heterophily in both groups.

It is easily verified that this example is robust, in that there exists an open set of expertise realizations that generates the same observational patterns. That is, each individual in each period strictly prefers her chosen target to any target not chosen. This is the case even for those in group N_2 in period 2, whose chosen target has the same expertise as others in group a; they have all previously observed only individual 1 and hence have a strictly better understanding of her perspective.

Moreover, if the expertise levels in the subsequent periods were similar to the expertise levels in the second period, everybody repeats their behavior at the second period in those periods, learning their targets increasingly well so much so that their links to others break, in that each i targets $\lambda_1(i)$ for ever even when $\lambda_1(i)$ is minimally informed and a member of her own group is maximally employed. Therefore, under full support assumption for expertise levels, extreme heterophily will emerge with positive probability in the long run.

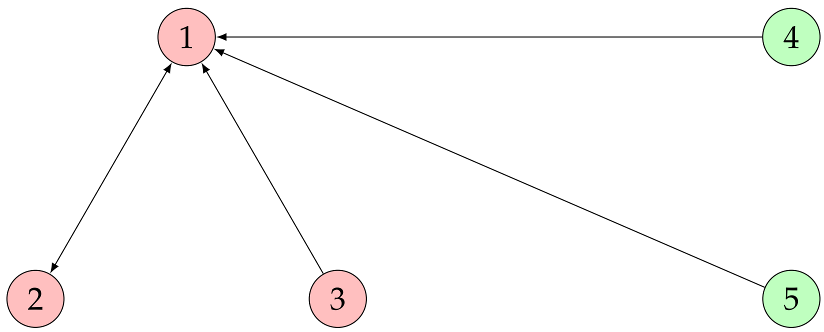

Figure 6: Informational network after three periods with independent priors (Example 3).

5.8 Divided Attention

In the initial period more tight-knit communities (with more strongly correlated perspectives) have a stronger tendency to homophily. Furthermore, larger groups are more likely to exhibit homophily than smaller groups, for the simple reason that the globally best-informed individual is more likely to be found in a larger group. These properties do not extend to the multi-period setting, because of a divided attention problem, as illustrated in the following sequence of examples.

Example 3. *Suppose* $N_1 = \{1, 2, 3\}$, $N_2 = \{4, 5\}$, $\overline{\tau}^2 = 4$, $\underline{\tau}^2 = 2$, *and* $\sigma_0^2 = 3/2$, *in which case* $\overline{\sigma}^2 = 1/2$. *There is also an intermediate level of expertise:* $\tau_m^2 = 3$. *Priors are independent in both groups:* $\rho_1 = \rho_2 = 0$. *In this case, it takes one observation to make a link free. Moreover, observing an individual's opinion once when she has high expertise* $\underline{\tau}^2$ *and once when she has medium expertise* τ_m^2 *breaks to the links to those who have never been observed. Consider the following sequence of signal variance realizations and associated links:*

t	τ_{1t}^2	τ_{2t}^2	τ_{3t}^2	τ_{4t}^2	τ_{5t}^2	$\lambda_t(1)$	$\lambda_t(2)$	$\lambda_t(3)$	$\lambda_t(4)$	$\lambda_t(5)$
0	2	3	4	4	4	2	1	1	1	1
1	3	2	4	4	4	2	1	1	1	1
2	4	4	4	2	2	2	1	1	1	1

At the end of period $t = 1$, *each individual has observed somebody twice—as described above—and all links that have not been used are broken. Therefore* $\lambda_t(i) = \lambda_1(i)$ *for each* i *and each* t *thereafter.*

The observational network for this example, which is the same in all periods, is shown in Figure 6. It is characterized by inbreeding homophily in the larger group and heterophily in the smaller group.

Now consider the same example but with greater correlation in perspectives in the larger group.

Example 4. *Suppose that all specifications and signal variance realizations are as in Example 3 except that* $\rho_1' = 1/2$. *Then the links formed are as follows:*

t	$\lambda_t(1)$	$\lambda_t(2)$	$\lambda_t(3)$	$\lambda_t(4)$	$\lambda_t(5)$
0	2	1	1	1	1
1	2	1	2	2	2
2	2	1	4	5	4

Now, at $t = 1$, *individuals* 3, 4, *and* 5 *switch to* 2, *because the correlation in priors allowed them to learn about all perspectives in the larger group during the initial period. Since one observation is not enough to break links in this example, individuals* 3, 4, *and* 5 *all switch to highly informed targets from the smaller group in period* $t = 2$. *Now since the links to all previously observed individuals become free, individual 3 targets out-group member 4 whenever the latter is more informed than others in the population. Likewise, members of the smaller group also target each other in the long run whenever they are highly informed.*

Figure 7 shows the third period network in this example. Since this network can arise with positive probability in all subsequent periods, we see lower levels of long run homophily in the larger group despite a higher level of correlation in perspectives.

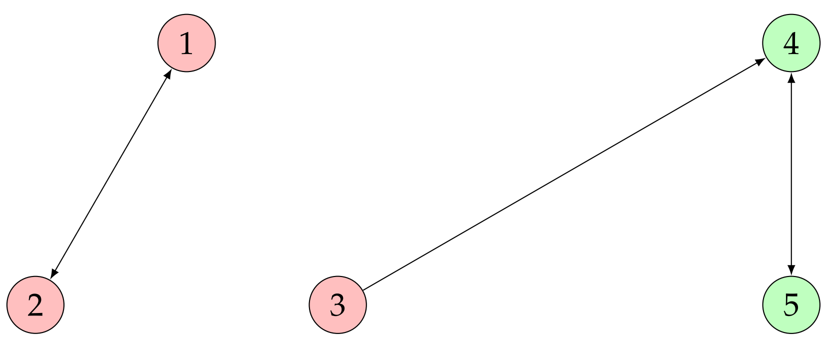

Figure 7: Informational network after three periods when $\rho_1 = 1/2$ (Example 4).

Examples 3–4 illustrate a more general problem of divided attention that makes it difficult to break other links. When $\rho_1 = 0$, in the first two periods, all individuals focus on individual 1, and she focuses on 2. In particular, individuals 3, 4, and 5 do not switch to the better informed individual 2 at $t = 1$—because the variance of the perspective of 1 goes below 0.477 at the end of $t = 0$ and leads to a lower noise for 1: $\gamma_{i1} < 7.3 < 8 = \gamma_{i2}$ for $i > 2$. This results in breaking all the links to group 2. Indeed, since $\sigma_1^2(3,1) \cong 0.213$, at $t = 2$, for individual 3, the noise in individual 1's opinion is lower than the noise in the opinions of the members of group 2: $\gamma_{31} \cong 7.1 < 8 = \gamma_{34} = \gamma_{35}$. Since 1 has the lowest possible expertise while the members of group 2 have the highest possible expertise levels in this case, she will never target them after that. Likewise, they will not link to each other either.

In contrast, when the correlation in group 1 is higher, individuals learn more *indirectly* about the perspectives of the other members of the group in the initial period of observation. This should make members of the larger group more attractive targets in later rounds, for any given history of past observation. However, the observations are endogenous in our model. When there are many attractive targets, individuals lose focus and divide their attention between these targets. As a result, they may not learn about any of them sufficiently well to focus on this group in the long run. Specifically, in the above example, when the correlation is 0.5, individual 3 learns a lot about 2 from observing 1: the variance of her belief about 2 goes below 1.05. Now in the second round, as 2 becomes better informed than 1, she switches to 2. She now does not know either of them well enough to focus on them in later periods. Indeed, at the end of period 1, the variance of her belief about 1 and 2 is approximately 0.42. As a result, in period 2, the noise variance of the opinions of 1 and 2 is approximately 10.7, higher than the noise variance of 8 in the opinions of the members of other group. She switches to the latter individuals as a result, resulting in lower levels of homophily thereafter.

Examples 3–4 show that intuitions about the effects of greater correlation can fail as one moves to a multi-period setting. The same is true for intuitions about group size. In the initial period, a larger group is more likely to exhibit homophily, but the following example shows that this too does not hold generally, again because of divided attention.

Example 5. *Suppose that all specifications and signal variance realizations are as in Example 4 except that individual 2 is removed from the population. Then the links formed are as follows:*

t	τ^2_{1t}	τ^2_{3t}	τ^2_{4t}	τ^2_{5t}	$\lambda_t(1)$	$\lambda_t(3)$	$\lambda_t(4)$	$\lambda_t(5)$
0	2	4	4	4	3	1	1	1
1	3	4	4	4	3	1	1	1
2	4	4	2	2	3	1	1	1

Now, without individual 2, individuals 3, 4, and 5 all focus on 1 in the first two periods; their attention is not divided. Consequently, in the long run, individual 1 targets 3 and everybody else targets 1, leading to extreme homophily and informational leadership by

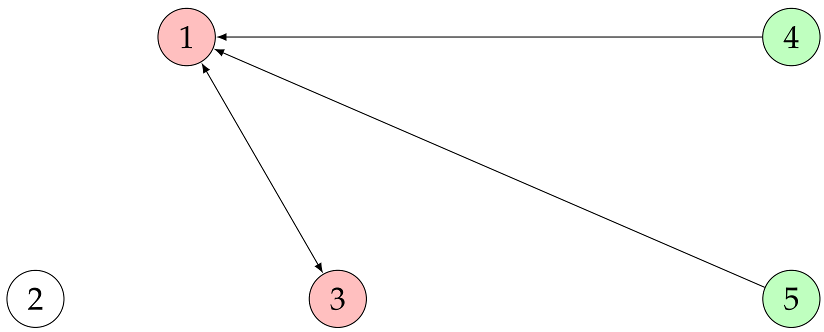

Figure 8: Informational network after three periods when $\rho_1 = 1/2$ and individual 2 is removed (Example 5).

group 1. This is despite the fact that it is now smaller, and hence less likely to contain the globally best informed individual in any given period.

6 Simulation Results

To identify some additional properties of long run observational structures, we simulated the model 1000 times for a given individual i under binomial distribution for various values of q and common correlation parameter $\rho = \rho_1 = \rho_2$. We start with the case of (roughly) equal sized groups, and then consider the consequences of asymmetries.

6.1 Non-Monotonic Homophily

In our first set of simulations, we consider an individual with 10 other members in her own group and 10 members in the other group, so baseline homophily is 0.5. We took $\overline{\tau}^2 = 1$ and $\underline{\tau}^2 = 1/2$, so that the signal of a low-expertise individual is as informative as his prior while the signal of a high-expertise individual is twice as precise. The resulting critical variance is $\overline{\sigma}^2 = 2$, and we took $\sigma_0^2 = 4$. We have $\alpha\left(\overline{\tau}^2, \underline{\tau}^2\right) = 4$, so it takes two observations for $\sigma_j^2(i,t)$ to drop to the critical level $\overline{\sigma}^2$. The number of observations needed to break a link depends on ρ and will typically be much higher; with minimum steps, it takes 14 repeated observations to break the links to unobserved individuals when $\rho = 0$.

In Figure 9, we present the average number of i's long-run experts from each group, and the corresponding levels of inbreeding homophily. On the left panel, we take $q =$

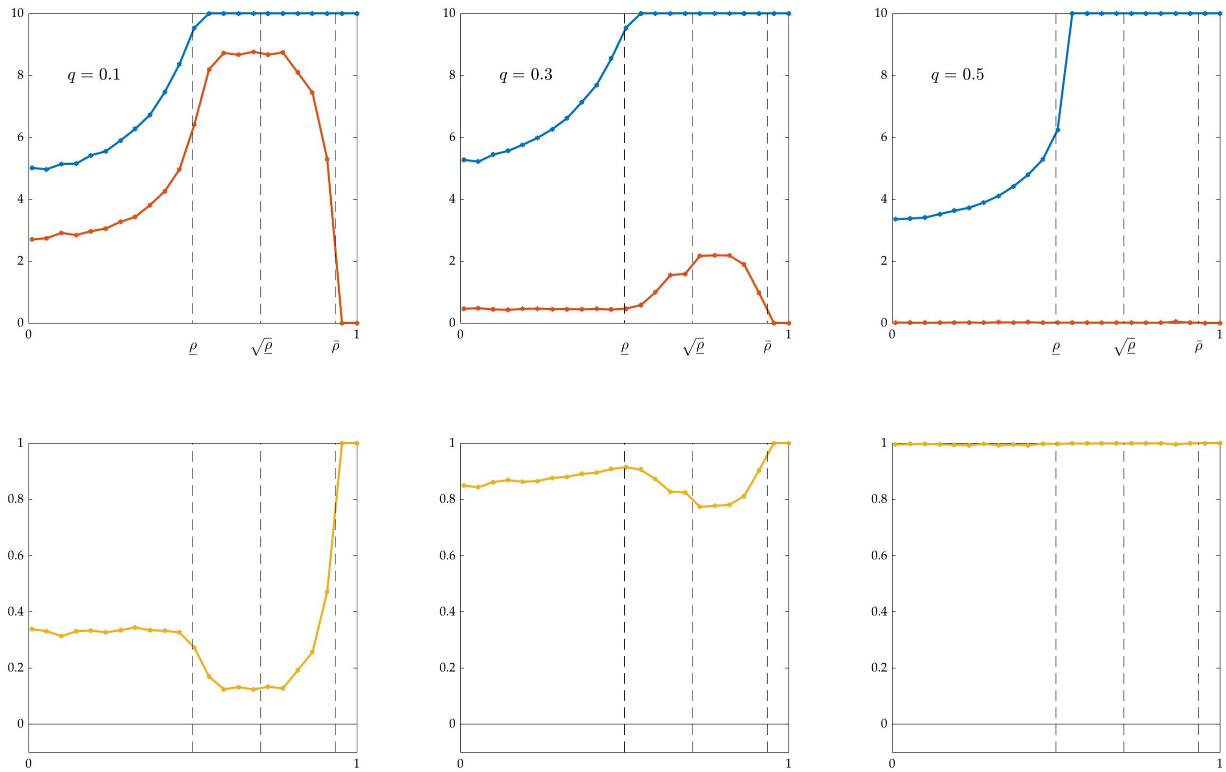

Figure 9: Top: Number of long run experts from own group (blue) and other group (red) as a function of ρ for three values of q. Bottom: corresponding levels of inbreeding homophily.

0.1 so that it is not common to have a high expertise. In this case the number of long-run experts is relatively high from each group, for most values of ρ. For example, the individual has about 5-6 experts from her own group and 3-4 experts from the other group when ρ is less than 0.4. As ρ rises, the number of experts from each group sharply rises. At around the cutoff $\underline{\rho}$, she consults almost everybody from her own group and 8 individuals from the other. This is because since q is low, the probability of the same individuals having high expertise repeatedly is low, and this allows her to experiment with multiple sources before any links break. The number from the other group sharply drops to 0 at the cutoff $\overline{\rho}$, after which she exhibits extreme homophily.

It is remarkable that the number of long-run experts from the other group is also very high when ρ is in between $\underline{\rho}$ and $\overline{\rho}$. At this level, there is good chance that nobody in her group has high expertise but there is somebody with expertise in the other group—with approximate probability 0.23. In those instances she gets to observe someone from the other group. When ρ is high, she learns not only about her targets but also about the others from that group, leading her to observe many of them in the long run.

We plot the results for $q = 0.3$ on the middle panel. While the number of long-run experts from her own group remains similar to the case of $q = 0.1$, the number of experts from the other group declines substantially, remaining below 2 for all values of ρ. The probability of the same individuals having high expertise repeatedly is still low, and she still gets to experiment various members of her own group before her links to them break. However, the probability of nobody in her group having high expertise is now very small and hence she gets to observe the other group's members with small probability of approximately 0.03. Her links to the other group breaks quickly as a result.

We plot the results for $q = 0.5$ on the right panel. Now, she exhibits virtually extreme homophily at all correlation levels, and the number of experts from her own group is also low when $\rho < \underline{\rho}$.

The indices of inbreeding homophily are shown in the bottom panel of the figure. The value of the index is mainly determined by probability q of high expertise level, and it increases with q. When $q = 0.5$, the homophily index remains near 1 at all correlation levels, exhibiting near extreme homophily. When q drops to 0.3, homophily remains high. The homophily index starts at 0.92 for $\rho = 0$ and mildly increases with ρ until $\rho = \underline{\rho}$, then it takes a dip in between $\underline{\rho}$ and $\overline{\rho}$ and sharply increases to 1 at $\overline{\rho}$. When q drops further down to 0.1, the homophily index exhibits a similar non-monotone pattern albeit at lower levels. In fact, for an intermediate range of ρ values, we have approximately baseline homophily.

This non-monotonicity is due to the effect of indirect learning about the other culture when $\rho > \underline{\rho}$. By Proposition 4, when $\rho > \underline{\rho}$, indirect learning leads her to link to all members of a group if she observes sufficiently many of them. In particular, when $\rho > \sqrt{\underline{\rho}}$, she will link to everybody in her own group, and she would also link to everybody from the other group if she learns even one of them well. When q is low, this increases the number of long-run experts from the other group, without affecting the already high number of experts from own group, resulting in a sharp drop in homophily towards baseline levels.

The binomial expertise distribution is somewhat special in that it puts all the probabilities to extremes. When q is moderately high and ρ is small, this biases the results in favor of homophily because small initial familiarity due to shared culture gives a large advantage to own group members, as it only matters whether an individual has high or low expertise. Although analytical results here extend to arbitrary distributions, quantitative results could be less extreme for other distributions.

6.2 Group Size Effects

The numerical results above assumed that the groups were of roughly equal size, with the level of baseline homophily being 0.5 for those in the (slightly) larger group. When there are sharp differences across groups in size, asymmetries in the level of inbreeding homophily arise.

Figure 10 shows indexes of inbreeding homophily when the total population is 21 (as in the previous section) but groups are of size 16 and 5 respectively. Again we consider three different values of q. In all cases individuals in the larger group have much higher levels of inbreeding homophily, and these levels are increasing in q. When q is small, the smaller group can exhibit heterophily in expectation for values of ρ close to $\underline{\rho}$. This is because high expertise is rare, and it is likely that the only targets with high expertise in the initial period are found in the larger group. Given the positive feedback effects that arise through the learning of perspectives, members of the smaller group can end up relying disproportionately on the larger group for sources of information.

That is, those in a small minority reach a better understanding of majority group members than the reverse. This is a common feature of many aspects of culture, but the informational channel explored here offers a novel perspective on the phenomenon.

7 Conclusions

The basic premise underlying our analysis here is that members of an identity group share a common worldview, and filter information about the world in a similar manner. We have modeled these worldviews using heterogeneous prior beliefs, assumed to be correlated within but not across groups. When seeking information about the world, this leads individuals to exhibit an initial preference for observing the opinions of in-group members, since these opinions are easier to interpret. But this bias need not overwhelm differences in the quality of information: outsiders may be observed if they have significantly more precise signals than insiders. And observing outsiders gives rise to additional

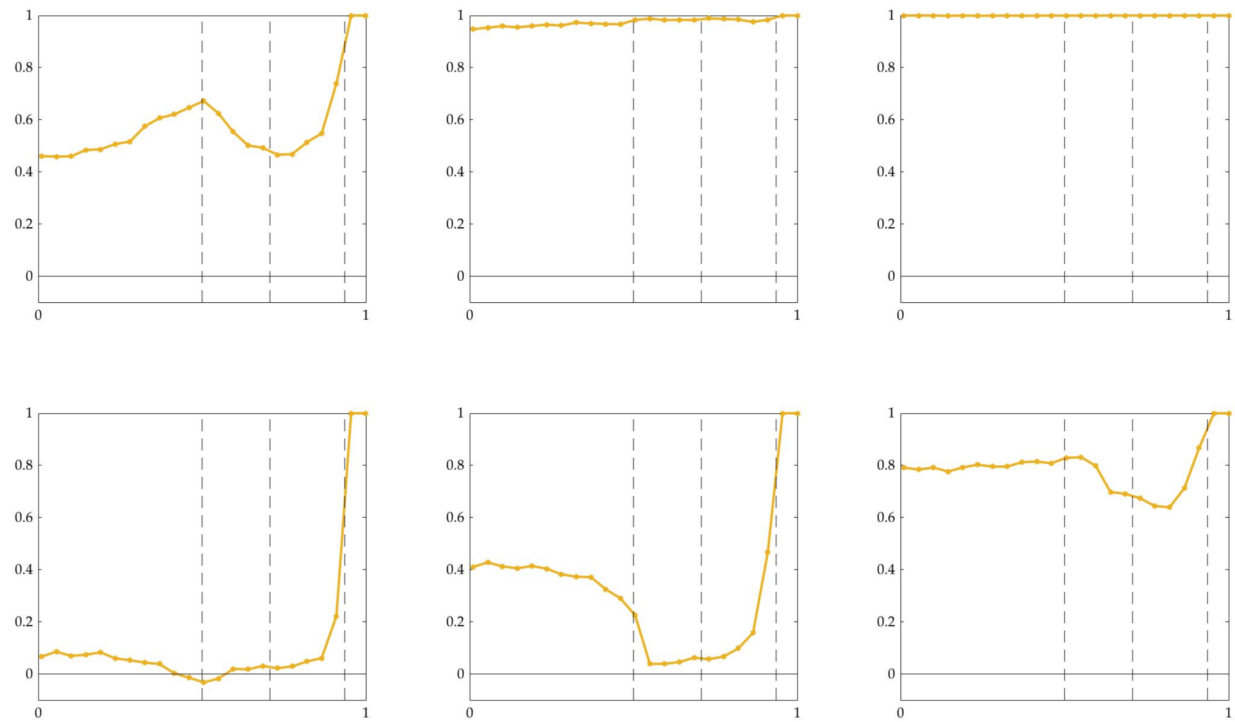

Figure 10: Inbreeding homophily with groups of size $n_1 = 16$ (top row) and $n_2 = 5$ (bottom), for $q = 0.1$ (left panel), 0.3 (center), and 0.5 (right).

positive feedback effects, as one learns not just about a different individual but also about a different culture.

A natural process of symmetry-breaking, arising from differences across observers in their own quality of information, can give rise to heterogeneity within groups in observation patterns. The extent of this heterogeneity is constrained, however, and under certain conditions results in a sharp separation of individuals into two categories: those who exhibit extreme homophily, and those who shed all initial biases towards in-group members. This bimodality of observation patterns is potentially testable using data on communication networks.

Our analysis also makes clear that the extent of homophily depends in systematic ways on the degree of correlation in perspectives, though not always in the manner one might expect. Homophily is lowest at intermediate levels of correlation, and small groups can exhibit heterophily in expectation.

Many other questions related to group identity and informational networks could potentially be explored using the framework developed here. For instance, one might explore the possibility of an informational gradient—with members of larger groups seeking information from a larger set of sources—analogous to the friendship gradient found in social networks. This seems to be a particularly promising direction for future research.

Appendix

Proof of Proposition 1. Suppose $\sigma_0^2 < \overline{\sigma}^2$ and consider any period t and any pair of distinct individuals i and j. We will obtain a strictly positive uniform lower bound on the probability that i links to j in period t. Then, i links to j infinitely often almost surely, showing that $j \in J_i$ and that $J_i = N \setminus \{i\}$ as claimed. To obtain the lower bound, note that $\sigma_j^2(i,t) \leq \sigma_0^2$ while $\sigma_{j'}^2(i,t) > 0$ for all $j' \notin \{i,j\}$ and all histories. Suppose that $\tau_{jt} = \underline{\tau}$ and $\tau_{j't} = \overline{\tau}$ for all $j' \notin \{i,j\}$. Then

$$\gamma_{ij}(t) = \underline{\tau}^2 + \underline{\tau}^4 \sigma_j^2(i,t) < \underline{\tau}^2 + \underline{\tau}^4 \overline{\sigma}^2 = \overline{\tau}^2 < \overline{\tau}^2 + \overline{\tau}^4 \sigma_{j'}^2(i,t) = \gamma_{ij'}(t).$$

Hence i links to j under these expertise realizations. The same is true for an open set of expertise realizations sufficiently close to these. There is a positive lower bound on the probability of the open set, providing a lower bound on the probability that i links to j at any given period. □

Proof of Proposition 2. Suppose $\sigma_0^2 > \overline{\sigma}^2$, in which case $\beta(\sigma_0^2)$ is well-defined and finite. It is easily verified that in the initial period, there is a positive probability that for each group k and each $i \in N_k$ we have $\lambda_0(i) \in N_k$. That is, there is a positive probability that in the initial period, each individual links to a member of their own group. (This will happen if the four best-informed individuals in the population all have sufficiently similar levels of expertise, and two of these are in the first group while the other two are in the second. There is a strictly positive lower bound on the probability of this event.) Suppose that the first period network does indeed satisfy $\lambda_0(i) \in N_k$ for each $i \in N_k$. Note that the likelihood that this same network will form again in the second period is also positive, as is the probability that it will form in each of the first s periods for any given, finite s. Suppose that the same network (with $\lambda_0(i) \in N_k$ for each $i \in N_k$) forms in each of the first s periods. If s is sufficiently large, then for each $i \in N$, we reach

$$\sigma_{\lambda_0(i)}^2(i,s) < \beta(\sigma_0^2).$$

At this point all links from $i \in N_k$ to all individuals outside N_k break, and i subsequently exhibits extreme homophily. We have shown that this is a positive probability event.

To prove the second claim, consider any group k with $\rho_k > \overline{\rho}$. Then, for each pair i and j such that $i, j \in N_k$ we have

$$\sigma_j^2(i,0) = \sigma_0^2(1 - \rho_k^2) < \beta(\sigma_0^2).$$

Hence the probability that i links to any $j' \notin N_k$ is zero in the initial period. It is clearly also zero in all subsequent periods. □

Proof of Proposition 3. Suppose i observes an in-group member j at period t. By (8),

$$1/\sigma_j^2(i,t+1) \geq 1/\sigma_j^2(i,t) + 1/\alpha(\overline{\tau}^2, \underline{\tau}^2), \tag{21}$$

where $\sigma_j^2(i,0) = \sigma_0^2\left(1-\rho_k^2\right)$. If i observes j repeatedly, the variance $\sigma_j^2(i,t)$ eventually drops below $\beta\left(\sigma_0^2\right)$. If i has not observed any out-group member in the meantime, her links to all out-group members break permanently at that point. Using the inequality in (21), one can easily show that the number of repetitions for this to occur is at most $\max\{\lceil\kappa\rceil, 0\}$ where κ is as defined in the proposition. If $\rho_k \geq \overline{\rho}$, we have $\kappa \leq 0$, and extreme homophily with probability 1 (see Proposition 2). If not, then $\kappa > 0$, and some repetition may be needed in order for all out-group links to break. The extent of this repetition depends on the distance of ρ_k from the cutoff $\overline{\rho}$.

To obtain our lower bound, consider the case of binary expertise. In the initial period, since $\sigma_j^2(i,0) = \sigma_0^2\left(1-\rho_k^2\right) > \sigma_0^2$, individual i links to an in-group member whenever there is any such individual with high expertise, which occurs with probability $1-(1-q)^{n_k-1} \geq q$, or there is no individual with high expertise in the population, which occurs with probability $(1-q)^{n-1}$. She links to an out-group member otherwise, which happens with probability $(1-q)^{n_k-1} - (1-q)^{n-1}$.

Suppose that i has linked to an in-group member j initially, and consider the probability that she will link to j again in the next round. Since she is now most familiar with j, she will link to j if j has high expertise, which happens with probability q, or nobody has high expertise, which occurs with probability $(1-q)^{n-1}$. Since she is least familiar with out-group members, she links to an out-group member with probability $(1-q)^{n_k-1} - (1-q)^{n-1}$ as in the previous case. With the remaining probability she links to a different in-group member, and learns more about that individual, as well as about j and other in-group members.

As i continues to observe only in-group members, the probability that she links to the most familiar in-group member remains $p_{in} = q + (1-q)^{n-1}$, and the probability that she observes an out-group member remains $p_{out} = (1-q)^{n_k-1} - (1-q)^{n-1}$, until the latter probability drops to zero. But the latter probability must drop to zero after $\max\{\lceil\kappa\rceil, 0\}$ observations of the most familiar in-group member. To see this, suppose that i has not targeted any out-group member so far and

$$\min_{j \in N_k \setminus \{i\}} \sigma_j^2(i,t) \geq \beta(\sigma_0^2).$$

The probability that i will target some $j_t \in N_k \setminus \{i\}$ with $\sigma_{j_t}^2(i,t) = \min_{j \in N_k \setminus \{i\}} \sigma_j^2(i,t)$ is at least $p_{in} = q + (1-q)^{n-1}$ while the probability that she targets an out-group member is $p_{out} = (1-q)^{n_k-1} - (1-q)^{n-1}$. Thus, probability that she targets a most-familiar in-group member T times without targeting any out-group member is at least

$$\left(\frac{p_{in}}{p_{in} + p_{out}}\right)^T,$$

where we allow i to target some other in-group members in between. But after $T = \max\{\lceil\kappa\rceil, 0\}$ observations, we have

$$\min_{j \in N_k \setminus \{i\}} \sigma_j^2(i,t+1) < \beta(\sigma_0^2),$$

and the links to out-group members are all broken. Specifically, if i targets $j_{t'}$ at period t', we have

$$\frac{1}{\min_{j \in N_k \setminus \{i\}} \sigma_j^2(i,t'+1)} \geq \frac{1}{\min_{j \in N_k \setminus \{i\}} \sigma_j^2(i,t')} + \frac{1}{\alpha(\overline{\tau}^2, \underline{\tau}^2)}.$$

Thus, if i has targeted the most-familiar in-group member $j_{t'}$ for $\max\{\lceil\kappa\rceil, 0\}$ times, we have

$$\frac{1}{\min_{j \in N_k \setminus \{i\}} \sigma_j^2(i,t+1)} \geq \frac{1}{\sigma_0^2\left(1-\rho_k^2\right)} + \frac{\max\{\lceil\kappa\rceil, 0\}}{\alpha(\overline{\tau}^2, \underline{\tau}^2)} > \frac{1}{\beta\left(\sigma_0^2\right)}.$$

Hence the probability of extreme homophily is at least

$$p^* = \left(\frac{p_{in}}{p_{in} + p_{out}}\right)^{\max\{\lceil\kappa\rceil, 0\}}$$

as claimed. □

Proof of Proposition 4. The claim follows from the analysis in the text. Indeed, take any history in which $m_{ik} \geq m$ where $\sigma_0^2/\overline{\sigma}^2 < \phi(m, \rho_k)$. Then, along that history, for any $j \in N_k \setminus \{i\}$, $\sigma_j^2(i,\infty) \equiv \lim_{t\to\infty} \sigma_j^2(i,t) \leq \sigma_0^2/\phi(m_{ik}, \rho_k) \leq \sigma_0^2/\phi(m, \rho_k) < \overline{\sigma}^2$. There then exists t^* such that $\sigma_j^2(i,t^*) < \overline{\sigma}^2$. Since the probability of histories in which i does not link to j infinitely often is zero, this completes the proof. □

Proof of Proposition 5. It suffices to show that the bound in the proposition applies to the probability conditional on that i has observed exactly m individuals from group k one or more periods. She may have also observed some other individuals from the other group k'. Denote the set of previously observed m members by N_k^*. For each t, let $J_t = \arg\min_{j \in N \setminus \{i\}} \sigma_j^2(i,t)$ be the set of individuals with whom i is most familiar at time t. Recall that the probability of i targeting some $j_t \in J_t$ at t is at least q. Moreover, since $\sigma_j^2(i,t) < \sigma_{j'}^2(i,t)$ for any $j \in N_k^*$ and any $j' \in N_k \setminus N_k^*$, the probability of i targeting any $j' \in N_k \setminus N_k^*$ is at most $(1-q)^m$. (This probability is often lower because there could be members j'' from the other group with $\sigma_{j''}^2(i,t) < \sigma_{j'}^2(i,t)$). Thus, the probability of targeting the most familiar (time and history-dependent) individuals $j_t \in J_t$ for T times before targeting any $j' \in N_k \setminus N_k^*$ is at least $\left(q/\left(q + m\left(1-q\right)^m\right)\right)^T$. Moreover, by inequality in (21), after such observations, we have

$$\min_{j \in N \setminus \{i\}} \sigma_j^2(i,t) \leq \frac{\sigma_0^2 \alpha(\overline{\tau}^2, \underline{\tau}^2)}{(T+1)\,\sigma_0^2 + \alpha(\overline{\tau}^2, \underline{\tau}^2)}$$

where we use the fact that each $j_t \in J_t$ has been observed at least once when we conditioned. For $T = \kappa$, the expression on right-hand side is less than

$$\beta\left(\left(1-\rho_k\right)\sigma_0^2\right),$$

and thus all the links to $j' \in N_k \setminus N_k^*$ are broken. □

References

Dennis J. Aigner and Glen G. Cain. Statistical theories of discrimination in labor markets. Industrial and Labor Relations Review, pages 175–187, 1977.

Kenneth J. Arrow. The theory of discrimination. In Orley Ashenfelter and Albert Rees, editors, Discrimination in Labor Markets. Princeton, NJ: Princeton University Press, 1973.

William H. Batchelder and Royce Anders. Cultural consensus theory: Comparing different concepts of cultural truth. Journal of Mathematical Psychology, 56(5):316–332, 2012.

Stephen Coate and Glenn C. Loury. Will affirmative-action policies eliminate negative stereotypes? The American Economic Review, 83(5):1220–1240, 1993.

Bradford Cornell and Ivo Welch. Culture, information, and screening discrimination. Journal of Political Economy, pages 542–571, 1996.

Sergio Currarini, Matthew O. Jackson, and Paolo Pin. An economic model of friendship: Homophily, minorities, and segregation. Econometrica, 77(4):1003–1045, 2009.

Mary Douglas and Aaron Wildavsky. Risk and culture: An essay on the selection of technological and environmental dangers. University of California Press, 1982.

Andrea Galeotti and Sanjeev Goyal. The law of the few. The American Economic Review, 100(4):1468–1492, 2010.

John D. Geanakoplos and Heraklis M. Polemarchakis. We can't disagree forever. Journal of Economic Theory, 28(1):192–200, 1982.

Matthew Gentzkow and Jesse M. Shapiro. Ideological segregation online and offline. The Quarterly Journal of Economics, 126(4):1799–1839, 2011.

Dan M. Kahan and Donald Braman. Cultural cognition and public policy. Yale Law & Policy Review, 24:147, 2006.

Willemien Kets and Alvaro Sandroni. A belief-based theory of homophily. Unpublished Manuscript, Northwestern University, 2015.

Paul F. Lazarsfeld and Robert K. Merton. Friendship as a social process: A substantive and methodological analysis. In Morroe Berger, editor, Freedom and Control in Modern Society, pages 18–66. New York: Van Nostrand, 1954.

Annie Liang and Xiaosheng Mu. Overabundant information and learning traps. PIER Working Paper No. 18-008, 2018.

Miller McPherson, Lynn Smith-Lovin, and James M. Cook. Birds of a feather: Homophily in social networks. Annual review of sociology (2001): 415-444., pages 415–444, 2001.

Edmund S. Phelps. The statistical theory of racism and sexism. American Economic Review, 62(4):659–661, 1972.

A. Kimball Romney, Susan C. Weller, and William H. Batchelder. Culture as consensus: A theory of culture and informant accuracy. American Anthropologist, 88(2):313–338, 1986.

Rajiv Sethi and Muhamet Yildiz. Public disagreement. American Economic Journal: Microeconomics, 4(3):57–95, 2012.

Rajiv Sethi and Muhamet Yildiz. Communication with unknown perspectives. Econometrica, 84(6):2029–2069, 2016.