Culture and Communication*

Rajiv Sethi† Muhamet Yildiz‡

March 8, 2026

Abstract

A defining feature of culture is similarity in the manner in which information is interpreted. This facilitates extraction of information from the beliefs of those in one's own group, a force for homophily. However, observation of individuals outside one's group deepens our understanding of their culture, counteracting this effect. We model this process, using unobservable, heterogeneous priors (correlated within but not across groups) to represent initial beliefs, and characterize long-run communication patterns. At moderate levels of correlation in priors, individuals can be partitioned into two categories: some exhibit extreme informational homophily, while others seek information wherever it is most precise.

*For helpful comments we thank the editor Marco Ottaviani, an anonymous associate editor, two anonymous referees, David Budescu, Ben Golub, Navin Kartik, Rich McLean, Jean-Phillippe Platteau, Debraj Ray, Thomas Sjöström, and seminar participants at Bocconi, Cambridge, Fordham, Harvard, HSE, ISI Delhi, Johns Hopkins, MIT, Namur, NUS, Oxford, Queen Mary, Rutgers, University College London, and Wisconsin. Sethi thanks the Center for Advanced Study in the Behavioral Sciences at Stanford for fellowship support.

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

‡Department of Economics, MIT.

1 Introduction

Effective communication requires us to reason about the worldviews of others, and to recognize that they may differ from our own. Such differences are not idiosyncratic. An emerging body of literature on *cultural cognition* reveals that on a range of contentious issues—such as the risks associated with climate change, private gun ownership, and the storage of nuclear waste—people disagree in ways that are correlated within social groups.[1] That is, individuals who share a worldview have correlated mental models, and interpret information in similar ways. Knowing this, they can extract information from each other's expressed beliefs more easily. This produces a static bias towards homophily in information acquisition, but also has some interesting dynamic implications.[2]

When one person observes the expressed beliefs of another, the observer learns three things—something about the world, something about the person being observed, and something about the observed person's culture. This knowledge casts a shadow over the future. Those who repeatedly observe individuals outside their own social group learn not only about people, but also about the groups to which those people belong. Over time, they develop an understanding of other worldviews that can lead to a breakdown of homophily, even as some individuals with whom they share a social identity continue to behave in insular ways. Such within-group heterogeneity arises because the degree to which one learns about another person or group depends on the quality of one's own information. As a consequence, considerable asymmetry can emerge endogenously in an initially symmetric population.

This is the process that we attempt to formalize and explore. We do so by building on a framework introduced in Sethi and Yildiz (2016), where unobservable, heterogeneous priors represent worldviews or perspectives. In that paper, we distinguished between the property of being *well-informed* and that of being *well-understood*. The former refers to precision about one's signals about the state of the world, while the latter is a relational property, describing the precision of one person's beliefs about another person's prior. That is, it is possible for one person to be well-understood by some observers and poorly understood by others. While our earlier work assumed uncorrelated priors and an *ex ante* identical population, we here introduce distinct social groups or cultures, with perspectives 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. As we discuss below, the introduction of correlation gives rise to observational patterns that cannot arise in the uncorrelated case.

[1] See Kahan et al. (2011) and the additional references in Section 2.

[2] Homophily—the propensity to interact with people similar to oneself—is widely observed along a number of dimensions (Lazarsfeld and Merton, 1954; McPherson et al., 2001). Our concern here is with homophily in the gathering of information.

The model has the following structure. There is a population of individuals partitioned into two social groups, and a sequence of periods. In each period, there is an unobserved state about which individuals would like to have precise beliefs, and these beliefs are formed based on two sources of information. First, each individual receives a private signal that is informative about the current state, and updates beliefs accordingly. Next, each individual selects a target whose posterior is then observed. The precision of each private signal is public information, but this does not mean that the person with the most precise signal will be observed by all. This is because prior beliefs about the distribution from which states are drawn are unobserved. These priors or perspectives are known to be correlated within groups, so people who share a culture are initially better able to extract information from each other's beliefs.

Once a target is observed, the observer updates beliefs about the current state, the perspective of the target, and the perspectives of all those who share a culture with the target. The target becomes better understood, as does the social group to which the target belongs. 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. Second, there is ongoing and indirect learning, based on inferences from observed opinions, about each chosen target and the target's culture. This indirect learning plays an important role in shaping long-run communication structures.

We show that when initial uncertainty about the perspectives of outsiders is sufficiently small, cultures merge in the long run: all individuals eventually seek information from wherever it is most precise and group boundaries essentially dissolve. When initial uncertainty about perspectives is large, however, an extreme form of homophily—where all individuals seek information only from those in their own group—emerges with positive probability. This is especially likely to happen when groups are large, or when the correlation in perspectives within groups is large.

At lower levels of correlation, heterogeneity in behavior both within and across groups can arise, even if individuals in each group are initially identical. 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 homophily, rarely if ever stepping outside group boundaries, while the other seeks information wherever it is most precise. This heterogeneity in the behavior of ex ante identical individuals arises because of a symmetry-breaking effect: the degree to which an individual learns about the perspectives of a target (and those who share the target's culture) depends on the precision of the observer's own private signal in the period of observation.

The 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 sufficiently well informed. And this means that their perspectives will also be eventually learned, giving 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.

2 Related Literature

The idea that culture affects cognition has been explored extensively in anthropology, psychology, political philosophy, and law. Kahan and Braman (2006) 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. The authors argue that this is due to *cultural cognition*, which they explain as follows: "Culture is prior to facts... 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, Sowell (1987) and Pinker (2002) argue that beliefs about specific issues derive from different *visions*, with a high degree of correlation in beliefs among those who share a vision, but substantial disagreement across groups with different visions. Sowell's constrained and unconstrained visions map roughly to Pinker's tragic and utopian visions. And in anthropology, *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 (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.

While this literature establishes the existence of different belief systems, it leaves several interesting questions unaddressed. For example, when do people seek information from those with different perspectives or visions? How does this process of information seeking modify their understanding of other groups? And what does this imply about patterns of information exchange in the long run?

We examine these questions using a framework introduced in Sethi and Yildiz (2012), where heterogeneous priors represent fundamental belief differences and signals of vary-

ing precision represent information. Although priors are 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. This framework was extended in Sethi and Yildiz (2016), where we studied the endogenous formation of information networks in a population with independent priors, private posteriors, and a sequence of states. Here we build on this by allowing for distinct identity groups and correlated priors, allowing us to obtain new insights into homophily and within-group heterogeneity in information gathering.

In related work, Liang and Mu (2020) 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.

An extensive literature on belief revision on networks dates back to DeGroot (1974), who considered a (non-Bayesian) averaging process that brings each individual's belief closer to those of her neighbors. This process leads to consensus regardless of the extent of homophily in initial beliefs. Allowing for biased assimilation—where individuals are more receptive to evidence that supports their existing beliefs—can give rise to polarization (Dandekar et al., 2013). Similar effects arise if ambiguous information is interpreted in accordance with existing beliefs and treated as unambiguous thereafter, never to be reinterpreted in light of new information (Fryer et al., 2019).

There are several models of social learning on networks that build on the canonical DeGroot framework in interesting ways. Dasaratha et al. (2023) consider a Bayesian model of learning from neighbors about a changing state. Individuals receive independent informative signals and also observe the beliefs (but not the private signals) of their neighbors. Their interest is in equilibrium updating rules, and in the conditions under which good information aggregation is achieved. Banerjee and Compte (2024) also consider equilibrium updating rules but allow for the use of heuristics. Specifically, individuals choose updating rules that depend on their current beliefs, and informative signal, and the beliefs of their neighbors. When the beliefs of others are observed with noise, individuals place positive (but inefficiently low) weight on their own private signals.

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.

Baccara and Yariv (2013) have examined the endogenous formation of peer groups prior to strategic decisions regarding the provision of local public goods. Their concern is with preference homophily—the formation of groups in which members agree on which of two possible public goods is the more desirable. They show that stable groups exhibit homophily in this sense. Extending this framework to allow for costs of contribution and connection, Baccara and Yariv (2016) identify conditions under which the opposite effect can arise. When costs of contribution are large and costs of connection small, the free-rider problem is more acute. Within-group preference heterogeneity can then result in greater public good provision, since those with extreme preferences are willing to contribute larger amounts to fund their preferred public good.

Kets and Sandroni (2019) 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 information extraction.

Gentzkow and Shapiro (2011) have examined informational homophily empirically, in the context of ideological identity (conservative and liberal) in the United States.[3] 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), Aigner and Cain (1977), and Cornell and Welch (1996). 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]See also Bakshy et al. (2015) and Flaxman et al. (2016) on ideological segregation in online social networks.

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 social group. We refer to these as group 1 and group 2 respectively. Let $n_k \geq 2$ 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 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. 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 $i \neq j$, the covariance between μ_i and μ_j is $\sigma_{ij} = \sigma_0^2 \rho_k$ 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.

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. Expertise levels are distributed independently from perspectives.

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 own 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: $\underline{\tau}^2 \leq \tau_{it}^2 \leq \overline{\tau}^2$ 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 for any given t (regardless of the history to that point). 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 in all periods, although it may be vanishingly small as $\epsilon \to 0$.

In each period, each individual can observe one other person's opinion, but cannot separately observe the constituent priors and signals on which the opinion is based. That is, after the expertise levels $(\tau_{1t}^2, \ldots, \tau_{nt}^2)$ have been realized and publicly observed, each i chooses a target $\lambda_t(i) \in N \backslash \{i\}$ and observes the opinion $y_{\lambda_t(i)t}$. In choosing targets, individuals seek the most informative opinion. That is, they 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. Third, no signal is perfect or completely uninformative. This assumption plays a crucial role in our argument. If an individual happened to be perfectly well-informed in any period, their opinion would perfectly reveal the current state and no other target would be observed. Similarly, if a person were completely uninformed, their opinion would reveal their perspective, and they would immediately become perfectly well-understood.

Finally, we have also assumed that expertise levels are publicly observable.[4] If expertise were not public information, the precision of one person's belief about another person's perspective and culture would depend on the realization of target opinions, inducing individuals to make inferences about the past opinions of unobserved individuals from the target choices of others. The public observability of expertise makes the precision of one individual's belief about the perspectives and culture of others independent of opinion realizations and thus common knowledge throughout. Under these conditions, our analysis is robust to many informational assumptions, such as whether individuals observe each other's past target choices.

4 Preliminaries

Beliefs about the perspectives of others change through the observation of their opinions. 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.[5]

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. There is less initial uncertainty about the perspectives of those in one's own group. Having observed μ_i, an individual i from a group k updates her belief about the perspective μ_j of any individual j from group k, coming to believe that μ_j is distributed normally with variance

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

If $\rho_k < 1$, 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.

4.1 Choosing Targets

In each period, each individual selects a single target for observation, with the goal of reducing their subjective uncertainty about the current state as much as possible. But the most informative opinion need not come from the best-informed source.

[4] See Gentzkow et al. (2025) for a recent model with uncertain signal accuracies.

[5] 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 $j \in N$.

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

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

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, as measured by τ_{jt}^2, against uncertainty in i's understanding of j's perspective, as measured by $\sigma_j^2(i,t)$. 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 expertise is realized, observational choices made, 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

$$\lambda_t(i) \in \underset{j\in N\setminus\{i\}}{\arg\min}\, \gamma_{ij}(t). \tag{6}$$

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 that i chooses the target with the smallest label. This tie-breaking rule does not play a role in our results.

4.2 The Initial Period

Consider the observational patterns that can arise in the initial period. 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 that the best-informed person in the population as a whole will be observed by all others within her group. In fact, there are only two possible types of observational pattern in the initial period: either everyone observes the best informed in the whole population, or everyone observes the best informed in their own group (the two best-informed people in each group may observe each other, or the second-best in their respective groups).

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 homophily in both groups:[6]

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

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. In this case we have homophily in one group and heterophily in the other.

As the correlation in perspectives gets larger, homophily becomes more likely. We show below that there exist thresholds $\overline{\sigma}^2$ and $\overline{\rho}$ such that homophily in group k is ensured if $\sigma_0^2 > \overline{\sigma}^2$ and $\rho_k > \overline{\rho}$. When these conditions hold, then all individuals in group k initially observe the best-informed target within their own group, regardless of how poorly informed this target may be. Outsiders are too poorly understood relative to insiders, and their opinions are too hard to interpret, even when these opinions are based on very precise information.

To summarize, 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 person in the population will be observed by all others. But if $\sigma_0^2 > \overline{\sigma}^2$, then any group with highly correlated perspectives will exhibit homophily. The degree to which members of such groups understand each other will overwhelm any informational disadvantage that arises.

This applies to the initial period, in which all individuals within a group are symmetrically placed with respect to each other. This symmetry is broken over time as perspectives are learned, and more complex patterns can arise.

4.3 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 the perspective 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

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

[6]By homophily (heterophily) we mean a tendency to observe more (fewer) in-group members than would arise by random chance.

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:[7]

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

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 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 period t. Unless correlation is perfect, we therefore have $\sigma_{jj}(i, t+1) < \sigma_{jj}(i, t)$ for each j from the target's group.

In summary, 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 stochastic 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.

4.4 Free and Broken Links

Consider an individual i and two potential targets j and j' at time t. Target j is maximally well-informed but minimally well-understood: her signal has variance $\underline{\tau}^2$, while the vari-

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

ance of i's belief about j's perspective is σ_0^2. Meanwhile j' is minimally well-informed but perfectly well-understood: her signal has variance $\overline{\tau}^2$ but the variance of i's belief about her perspective is zero. Whom should i observe? By (5), i is indifferent between these two potential targets if and only if $\sigma_0 = \overline{\sigma}$ where $\overline{\sigma}$ is as defined as

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

Hence j would be observed if $\sigma_0 < \overline{\sigma}$, while j' would be observed if $\sigma_0 > \overline{\sigma}$.

More generally, 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 j's perspective in the long run.

On the other hand, if $\sigma_0 > \overline{\sigma}$, j' would be observed in preference to j even though j' is minimally well-informed and j is maximally well-informed. This is because j is so poorly understood that very little information can be extracted from her opinion. The possibility then arises that j may *never* be observed, unless there exists some way for her to become better understood *without being observed*. There does indeed exist such a mechanism, through the correlation in perspectives: j may become better understood by i over time through the observation of targets who share a culture with j.

Even so, a situation may arise in which some other individuals have become so well-understood over time that there is no circumstance in which j's opinion will be the most informative at t, or in any subsequent period. In this case we say that the link from i to j is *broken*. Specifically, if there exists j' such that $\sigma_{j'}^2(i,t) < \beta(\sigma_j^2(i,t))$, then i will not observe j at period t regardless of expertise realizations, where the function β is defined by[8]

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

In particular, if the above inequality holds for all members j of a group k, then the links from i to group k will be broken at t.

Observe that a free link can never be broken, a broken link can never become free, and all links must eventually be either broken or free.[9] 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

[8]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 $\sigma_{j'}^2(i,t) = \beta(s^2)$.

[9]The first two claims here are easily verified, while the third follows from the fact that the precision of one's belief about a target's perspective rises by an amount bounded away from zero after each observation, from (7)–(8).

alone, regardless of group membership, in the long run. If, instead, we have $\sigma_0^2 > \overline{\sigma}^2$, then not all links will be free at the outset. Some may become free over time, while others break. Observational patterns can be complex in this case, and will depend on group sizes and correlations in perspectives.

4.5 Long-Run Experts

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. 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 $J_i \subseteq N \backslash \{i\}$ 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.

5 Main Results

Before presenting our results formally, we provide a brief overview. There are two key parameters of interest, the degree of within-group correlation in perspectives, ρ_k, and the variance of initial beliefs about the perspectives of others, σ_0^2. For our purposes, it is more convenient to work with the *relative precision* of initial beliefs about the perspectives of others, defined as

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

Figure 1 provides a schematic overview of our results, showing regions of the parameter space under which various long-run structures arise.[10] When the relative precision exceeds 1, all links are free at the outset. Individuals may favor within-group targets in the short run, but eventually seek information wherever it is most precise. Group boundaries dissolve, at least from an information gathering perspective; this case is labelled "merging of cultures" in the figure.

When the relative precision lies below 1, a range of observational patterns can emerge. If correlation in perspectives is very high while relative precision is low, we get extreme

[10] The figure is based on equal within-group correlation in perspectives ($\rho_1 = \rho_2 = \rho$).

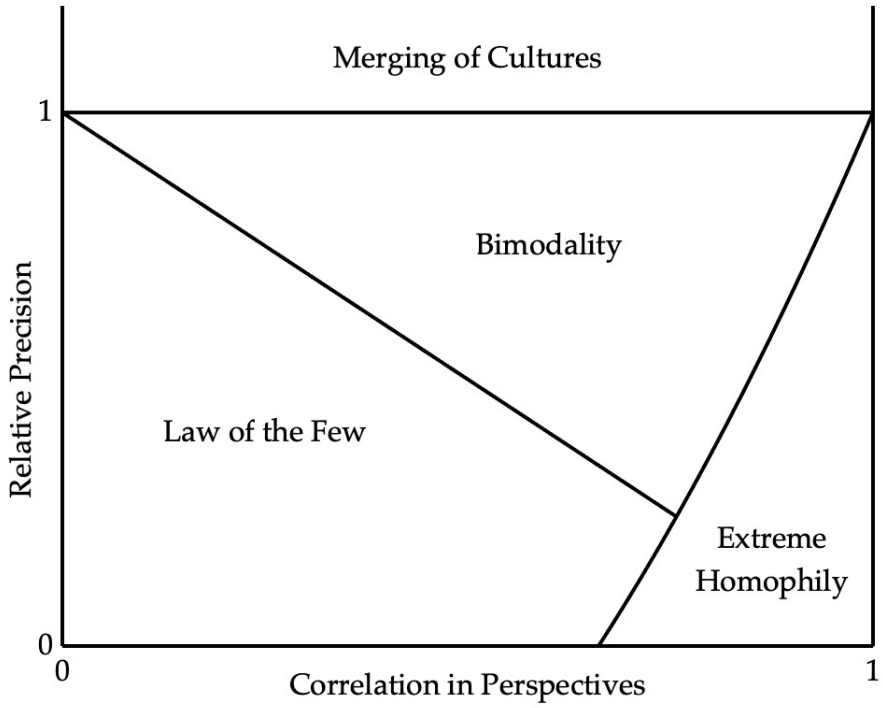

Figure 1: Long run outcomes in relation to the relative precision of initial beliefs about perspectives and within-group correlation in perspectives.

homophily: all individuals seek information only within their own group. When both correlation and relative precision are low, we get a law of the few: experts sets are small in expectation, and with high probability, each individual links to just a few sources from each group in the long run.

When correlation is in an intermediate range, and relative precision is not too low, each group can be partitioned into two subgroups, which we may call insular and cosmopolitan. The former is composed of individuals who get information largely (though not necessarily exclusively) from those within their own group. The latter is composed of individuals who seek information from the best-informed sources in the population as a whole, regardless of group boundaries. Insular individuals may have a few members of the other group in their long run expert sets, but this number is not large enough to tip them into a level of cultural understanding that would induce them to seek information wherever it is most precise. The result is observational bimodality (as indicated in Figure 1) in the sense that people either fail to develop the capacity to make inferences from out-group opinions, or develop this capacity to a very high degree.[11]

[11]The analysis in Sethi and Yildiz (2016) takes place on the vertical axis of Figure 1, with zero correlation. Allowing for correlation thus yields two new observational patterns, homophily and bimodality.

5.1 Merging of Cultures

When initial uncertainty about perspectives is low (relative precision is high) all links are free to begin with, individuals learn all perspectives, and eventually target the best informed people in the population as a whole.

Proposition 1. *If $\sigma_0^2 < \overline{\sigma}^2$, then $J_i = N \setminus \{i\}$ for all i almost surely.*

Homophily could well arise in early periods, for instance if the best informed individuals in each group have comparable levels of expertise, but these effects are overwhelmed in the long run. This happens because for each pair of individuals i and j belonging to different groups, and each period t, there is a positive lower bound for the probability that i will target j in t. Each such link will form an infinite number of times, almost surely, in the long run. Moreover, when a link forms, there is a positive lower bound for the resulting reduction in uncertainty about perspectives. Hence every individual will eventually learn every perspective. Cultures will merge in the long run, at least from an informational perspective, as each individual chooses the most informed target in the population at large; see the region marked "merging of cultures" in Figure 1.

Note the importance here of the assumption that each individual has some (possibly small) probability of having very high expertise in some periods and very low expertise in others. If this were not the case—for example if the upper and lower bounds for expertise were different for some individuals, the threshold $\overline{\sigma}^2$ would not be common to all. In this case Proposition 1 would require that σ_0^2 be below the lowest $\overline{\sigma}^2$ across all agents, otherwise some perspectives in the population may never be learned.

We next consider cases where the initial uncertainty about perspectives is high.

5.2 Extreme Homophily

If $\sigma_0^2 > \overline{\sigma}^2$, then long-run observational patterns depend on the correlation in perspectives as well as the history of expertise realizations.

We say that an individual i exhibits *extreme homophily* at each t if she observes only own-group members throughout:

$$\lambda_t(i) \in N_k \qquad (\forall t)$$

where $i \in N_k$. The following result states that unless all links are free to begin with, there is a positive probability that all individuals exhibit extreme homophily in all periods. And in groups with sufficiently high correlation in perspectives, extreme homophily is ensured.

Proposition 2. *When $\sigma_0^2 > \overline{\sigma}^2$, there is a positive probability that each individual exhibits extreme homophily at each t. In addition, if $\rho_k > \overline{\rho}$, where*

$$\overline{\rho} = \sqrt{1 - (\underline{\tau}^4/\overline{\tau}^4)(1 - \overline{\sigma}^2/\sigma_0^2)}, \tag{12}$$

then all individuals in group k exhibit extreme homophily at each t almost surely.

With low correlation in perspectives, 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 they do not observe any member from the other group subsequently, regardless of the expertise realizations that may arise. In this case, while extreme homophily is not ensured, it arises with positive probability.

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. This corresponds to the region marked "extreme homophily" in Figure 1.

Proposition 2 makes a qualitative claim, identifying conditions under which there is a positive probability that each individual exhibits extreme homophily throughout. We show in the appendix that it is possible to obtain a lower bound for this probability under the assumption that expertise realizations are binomially distributed, and to show that this bound approaches 1 as group size gets large, regardless of the extent of correlation in perspectives.

5.3 Bimodality

More complex and interesting observational patterns can emerge if $\sigma_0^2 > \overline{\sigma}^2$, and $\rho_k < \overline{\rho}$ for at least one group k. In this section, for moderate values of ρ_k, we show that individuals will exhibit extreme behavior in the long run—they will either observe only a few members of group k, or will observe them all. This corresponds to the region marked bimodality in Figure 1.

For each individual i and group k, let

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

denote 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 $\sigma_0^2/\phi(m_{ik}, \rho_k)$ where[12]

$$\phi(m, \rho) = \frac{1-\rho/(m\rho+1)}{1-\rho} \geq 1. \tag{13}$$

That is, i's uncertainty about the perspectives of all individuals in group k shrinks by a factor of (at least) $\phi(m_{ik}, \rho_k)$ if she observes m_k members of the group infinitely often. This is the *indirect learning* channel through which culture affects the dynamics of communication patterns.

Note that the factor $\phi(m, \rho)$ 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, there is a limit to indirect learning: as $m \to \infty$, $\phi(m, \rho)$ approaches to $1/(1-\rho)$, and indirect learning can at most reduce the variance by a factor of $1-\rho$. In contrast, the very first observation reduces the uncertainty by a factor of

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

That is, the effect of initial sampling is a fraction $1/(1+\rho)$ of the effect 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

$$\underline{\rho} = 1 - \overline{\sigma}^2/\sigma_0^2, \tag{14}$$

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 play important role in our analysis below, as long-run communication patterns will depend dramatically on where the correlations ρ_k lie with respect to these cutoffs.[13]

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

$$\overline{m}(\sigma_0^2, \rho) = \left(\frac{1-\rho}{\rho}\right) \frac{\sigma_0^2 - \overline{\sigma}^2}{\overline{\sigma}^2 - (1-\rho)\sigma_0^2}. \tag{15}$$

[12]Expression (13) can be obtained using standard formulas for Bayesian updating with multivariate normal distributions.

[13]Note that the cutoff $\underline{\rho}$ (like the threshold $\overline{\rho}$ introduced earlier) depends on the relative precision $\overline{\sigma}^2/\sigma_0^2$.

This solves the equation

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

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 $\sigma_0^2/\phi\left(m_{ik}, \rho_k\right) < \overline{\sigma}^2$. 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 3. *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. Recall that $\overline{m}\left(\sigma_0^2, \rho\right)$ 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 indirect learning occurs through the first sampling of perspectives, making homophily more likely.

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 3 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, or simply observes the individual who is globally best informed in the long run.

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 either observes only her own group members, or observes the most informed person at large.

One implication of these results 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 be observed, whenever they happen to be substantially better informed than all others. 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 2 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) when 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

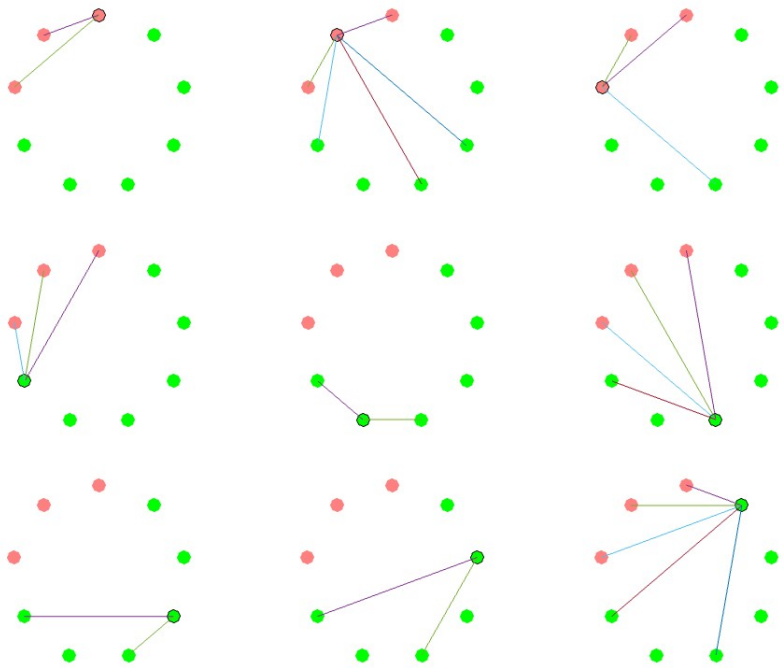

Figure 2: Long-run targets in Example 1. Each cell is the observational network for one agent, shown with a solid boundary, and links are to the agent's long-run targets.

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 observational patterns are constrained by the correlation in perspectives within groups.

5.4 Law of the Few

We now consider the case in which both the correlation and the relative precision of initial beliefs are low, as in the region marked "Law of the Few" in Figure 1. We find an upper bound for the size of the long-run expert sets in sense of first-order stochastic dominance, where the bound has geometric tails, leading to small expert sets with high probability.

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. We show that this leads to small experts sets by

obtaining a probabilistic upper bound on their size.

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, she also learns indirectly by observing others from N_k. However, there is a limit to such learning:[14]

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

That is, the variance of i's belief about j's perspective 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. This can at most be what one could get by learning the perspectives of all others in 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 independent priors. In particular, we next show that individuals are likely to focus on a small set of long-run experts in the long run, under the assumption that expertise levels are binomially distributed:

$$\tau_{it}^2 = \begin{cases} \underline{\tau}^2 & \text{with probability } q, \\ \overline{\tau}^2 & \text{with probability } 1-q. \end{cases} \tag{16}$$

[14] One can compute $\sigma_j^2(i,t)$ by first incorporating indirect learning and then direct learning. Since indirect learning can at most reduce σ_0^2 to $\sigma_0^2/\phi(n_k-1,\rho_k)$, we have

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

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

The following generalizes one of the main results in Sethi and Yildiz (2016) to accommodate correlated perspectives.[15]

Proposition 4. *Assume that expertise levels are independently and identically distributed across individuals and over time according to the binomial distribution in (16). 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*

$$\Pr\left(|J_i \cap N_k| \leq m\right) \geq \left(\frac{q}{q + (1-q)^m}\right)^{\kappa_s} \equiv p_k\left(m\right) \qquad (\forall m \geq 1) \tag{17}$$

where

$$\kappa_s = \left\lfloor \frac{\alpha(\overline{\tau}^2, \underline{\tau}^2)}{\beta\left(\left(1-\rho_k\right)\sigma_0^2\right)} - \frac{\alpha(\overline{\tau}^2, \underline{\tau}^2)}{\sigma_0^2} \right\rfloor. \tag{18}$$

Proposition 4 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 tails, the probability of having more than m long-run experts 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 goes to ∞ as ρ_k approaches $\underline{\rho}$. In that case $p_k\left(m\right)$ goes to zero, allowing a large number of experts from the group. This is intuitive because when $\rho_k > \underline{\rho}$, by Proposition 3, everybody in group k becomes a long-run expert when sufficiently many of them become one. However, $p_k\left(m\right)$ is only a lower bound on the probability $\Pr\left(|J_i \cap N_k| \leq m\right)$, and this probability can be much higher if i is also familiar with members of the other group $k' \neq k$.

The proof of Proposition 4 is based on the following logic. Under the binomial distribution, an individual i has a simple rule: if the individual j_1 with the lowest variance $\sigma_j^2\left(i,t\right)$ 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 $\sigma_j^2\left(i,t\right)$ 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. Second, she also learns indirectly about the perspectives of the others in her target's group, but

[15]The operator $\lfloor \cdot \rfloor$ in the result rounds down to the nearest integer.

there is an upper bound for such indirect learning, and the variance of her belief about a unobserved individual's perspective remains above $\sigma_0^2(1-\rho_k)$. 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 $\beta\left(\sigma_0^2(1-\rho_k)\right)$, at which point all her links to previously-unobserved individuals from group k break. We show in the proof of Proposition 4 that the number of repetitions for this to occur is at most κ_s.

Proposition 4 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 consulting new ones is no longer worthwhile. When groups are large and exhibit homophily, the expert sets of two individuals from the same group will 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.

6 Discussion

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. 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 positive feedback effects, as one learns not just about an individual from a different culture but about the culture itself.

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 has been based on a number of simplifying assumptions: myopic choices, a single observational target in each period, two groups, and no cross-group correlation in perspectives. These are simplifying assumptions, neither of which are essential to the

argument. In the appendix, we consider extensions that allow for forward-looking behavior, multiple targets per period and multiple groups with cross-cultural correlation. This serves two purposes—demonstrating the robustness of the results reported here, and obtaining new insights. For instance, when multiple observations can be made in a period, it is not generally the case that one would choose those with the most informative opinions; a potential target whose opinion would be relatively uninformative in isolation may be very useful in combination with another target. And with cross-cultural correlation, one can learn about a group by observing those in a proximate culture, which affects the nature and extent to which group boundaries dissolve in the gathering of information.

Many other questions related to group identity and informational networks could potentially be explored using the framework developed here. For instance, people may wish to conceal their opinions to avoid adverse inferences about their perspectives, as in the literature on self-censorship (Loury, 1994). Or they may engage in insincere expression that is not aligned with the beliefs they hold, thus engaging in preference falsification (Kuran, 1998). Such effects could arise even if their perspectives are closely aligned with others in their group, because neither perspectives nor signals are independently observed. These are particularly promising directions for future research.

A Proofs

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; see (11). 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. 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 4. If i observes j at period t, then (8) implies

$$1/\sigma_j^2(i, t+1) \geq 1/\sigma_j^2(i, t) + 1/\alpha(\overline{\tau}^2, \underline{\tau}^2). \tag{19}$$

To prove the result, it suffices to show that the bound in the proposition applies to the probability conditional on i having observed exactly m individuals from group k at least once. 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 $(1-q)^m$. Thus, the probability of targeting the most familiar (time and history-dependent) individuals $j_t \in J_t$ for T periods before targeting any $j' \in N_k \setminus N_k^*$ is at least

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

By the inequality (22), after T 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

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

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

B Extensions

We have made several restrictive assumptions in the main text: individuals are myopic, they can observe just one opinion at a time, and there are only two groups. In this section

we consider forward-looking behavior, multiple observations in each period and multiple groups. In addition, we obtain a bound for the probability that each individual exhibits extreme homophily throughout.

B.1 Probability Bounds

The following result establishes a a lower for the probability that an individual exhibits extreme homophily, under the assumption that expertise realizations are binomially distributed.

Proposition 5. *Suppose that expertise levels are independently and identically distributed across individuals and over time according to the binomial distribution in (16). Then, for any $\sigma_0^2 > \overline{\sigma}^2$, any group k, and any $i \in N_k$, individual i exhibits extreme homophily in the long run with probability at least $p^*(n_k, \rho_k, q, \sigma_0)$, where this lower bound is (i) increasing in n_k and approaches 1 as $n_k \to \infty$, (ii) is increasing in ρ_k and approaches 1 as $\rho_k \to \overline{\rho}$, (iii) increasing in q and approaches 1 as $q \to 1$ and (iv) is increasing in σ_0^2 and approaches 0 as $\sigma_0^2 \to \overline{\sigma}^2$.*

Proof of Proposition 5. We prove the result by showing that

$$\Pr\left(J_i \subset N_k\right) \geq \left(\frac{q + (1-q)^{n-1}}{q + (1-q)^{n_k - 1}}\right)^{\max\{\lceil \kappa \rceil, 0\}} \equiv p^* \tag{20}$$

where

$$\kappa = \frac{\alpha(\overline{\tau}^2, \underline{\tau}^2)}{\sigma_0^2}\left[\frac{1}{1-\overline{\rho}^2} - \frac{1}{1-\rho_k^2}\right]. \tag{21}$$

The lower bound p^* contains a term in parentheses and an exponent. The term in parentheses measures the relative likelihood of targeting the most informed in-group member, rather than someone from the other group. The exponent $\max\{\lceil \kappa \rceil, 0\}$ measures the number of repeated observations of an in-group member that is sufficient to break all links to the other group.[16] 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. All comparative statics claims in the result may be deduced directly from the inequality (20), which we now derive.

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

$$1/\sigma_j^2(i, t+1) \geq 1/\sigma_j^2(i,t) + 1/\alpha(\overline{\tau}^2, \underline{\tau}^2), \tag{22}$$

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 (22),

[16] The operator $\lceil \cdot \rceil$ rounds up to the nearest integer.

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. □

Consider each of these claims in turn. The lower bound p^* for the probability that any given member of group k will exhibit extreme homophily is increasing in group size, because the larger the group the more likely it will be that the group contains an individual who is globally best-informed in any given period. 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.[17] For similar reasons, the lower bound is also increasing in the probability q of high expertise. As group size gets arbitrarily large, or high expertise becomes virtually certain, extreme homophily becomes almost inevitable.

The bound p^* is also increasing in ρ_k, the degree of correlation in perspectives—the more tight-knit the group, the stronger is the force for homophily. As might be expected, the bound approaches 1 as ρ_k approaches the cutoff $\overline{\rho}$, after which extreme homophily arises with certainty (by Proposition 2). Finally, the bound is increasing in initial uncertainty σ_0^2 about the perspectives of the other group, and approaches 0 as σ_0^2 goes down to $\overline{\sigma}^2$ (recall that all links are free and so extreme homophily is ruled out in the long run when $\sigma_0^2 < \overline{\sigma}^2$ by Proposition 1).

B.2 Forward-Looking Behavior

In this section, we extend our main results to optimal forward-looking behavior for a given discount factor δ. Fix a player $i \in N_1$ from group 1 and assume that

$$\rho_1 = 1, \tag{23}$$

[17] 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.

so i knows the perspectives of all members of her own group at the outset. This allows us to focus on learning about the other group, and also provides the best case scenario for homophily. We also assume that at the end of each period, i learns the perspective of her latest target (if it is not already known). This simplifies the analysis by assuming away issues related to partial learning, where less informed targets reveal more about themselves.

At any given period t, the population can be divided into two groups: N_0, the set of those individuals $j \in N \setminus \{i\}$ whose perspectives i knows (these are fellow group members and others whom she has observed), and $N' = N_2 \setminus N_0$ (these are members of the other group whom she has not previously observed). Her payoff from observing the opinion of a target j at period t is

$$u_i = \begin{cases} \hat{u}(\tau_{it}, \tau_{jt}) & \text{if } j \in N_0 \\ u(\tau_{it}, \tau_{jt}, \phi(m, \rho_2)) & \text{otherwise} \end{cases} \tag{24}$$

where m is the number of individuals from group 2 that i has observed and ϕ is defined in (13). We assume that u is decreasing in τ_{it} and τ_{jt}, increasing and concave in ϕ, approaching $\hat{u}(\tau_{it}, \tau_{jt})$ as $\phi \to \infty$, and $\hat{u}(\tau_{it}, \overline{\tau}) = u(\tau_{it}, \underline{\tau}, \overline{\sigma}^2/\sigma_0^2)$.[18] The objective of i is to maximize the expected value of

$$U = (1-\delta)\sum_t \delta^t u_i \tag{25}$$

where $\delta \in (0,1)$ is a given discount factor. We assume that expertise levels are i.i.d. serially and across players. For any $N_0 \subseteq N \setminus \{i\}$ with $|N_0| = n_0$, we write

$$V_0(n_0) = E\left[\hat{u}(\tau_i, \min_{j \in N_0} \tau_j)\right] \tag{26}$$

for the expected payoff from observing the best informed member of N_0 when their perspectives are known. Finally, we assume that $V_0(n_0+1) - V_0(n_0)$ is decreasing in n_0 and vanishes as $n_0 \to \infty$.

Let $V(m)$ be the continuation value of i when she knows the perspectives of m individuals from group 2. Then:

$$V(m) \geq V_0(n_1 + m - 1).$$

Observe that V depends on m, n_1, ρ_2, δ and other parameters of the model while V_0 depends only on n_0 and the distribution of expertise levels. Let $\tau_1 = \min_{j \in N_0} \tau_j$ and $\tau_2 = \min_{j \in N'} \tau_j$ be the noise variances for the best informed members in groups N_0 of known sources and N' of unknown sources, respectively.

Player i faces the following trade off: (i) she can observe the opinion of the most informed known source, in which case she obtains the payoff of $\hat{u}(\tau_{it}, \tau_1)$ today and her continuation value remains as $V(m)$, or (ii) she can observe the opinion of the best informed

[18] These conditions can be microfounded by assuming that i takes an action a and obtains the payoff of $1 - (\theta - a)^2$.

unknown source, obtaining the payoff $u(\tau_{it}, \tau_2, \phi(m, \rho_2))$ today and increasing the continuation value to $V(m+1)$ at the next period. Hence, experimentation with an unknown source comes with a benefit

$$V(m+1) - V(m), \tag{27}$$

which is realized from the next period on, and a cost

$$C(\tau_{it}, \tau_1, \tau_2, m) = \hat{u}(\tau_{it}, \tau_1) - u(\tau_{it}, \tau_2, \phi(m, \rho_2)), \tag{28}$$

which is incurred only at the current period. Hence, i chooses to experiment with an unknown source if and only if

$$\delta(V(m+1) - V(m)) \geq (1-\delta)C(\tau_{it}, \tau_1, \tau_2, m). \tag{29}$$

Since the continuation values do not depend on current expertise levels, she experiments if and only if the cost is sufficiently low. For any given m, she experiments at *some* realization of expertise levels if and only if

$$\delta(V(m+1) - V(m)) \geq (1-\delta)C_{\min}(m) \tag{30}$$

where

$$C_{\min}(m) = \min_{\tau_i, \tau_1, \tau_2} C(\tau_i, \tau_1, \tau_2, m) \tag{31}$$

is the minimum cost of experimentation. Note that the minimum cost of experimentation can be negative.

Since N_2 is finite, the process of experimentation ends at some finite $m^\infty \leq n_2$. Moreover, one can easily check that m^∞ is deterministic (almost surely) because the minimal cost will be realized infinitely often almost surely. Moreover, when experimentation stops, the continuation value is $V(m^\infty) = V_0(m^\infty + n_1 - 1)$. The following generalizes our main results to allow for experimentation:

Proposition 6. *The following statements are true in the forward-looking model:*

1. *(Merging of Cultures) $m^\infty = n_2$ if $\sigma_0 < \overline{\sigma}$ or if δ is sufficiently large.*
2. *(Extreme Homophily) $m^\infty = 0$ if $\sigma_0 > \overline{\sigma}$ and δ is sufficiently small.*
3. *(Bimodality) For any $\sigma_0 > \overline{\sigma}$ and $\rho_2 > 1 - \overline{\sigma}^2/\sigma_0^2$, either $m^\infty \leq \overline{m}(\sigma_0^2, \rho_2)$ or $m^\infty = n_2$.*
4. *(Law of the Few) For any $\sigma_0 > \overline{\sigma}$ and $\rho_2 < 1 - \overline{\sigma}^2/\sigma_0^2$, we have $m^\infty \leq \hat{m}$ where $\hat{m}$ solves*

$$\frac{V_0(n_1+m) - V_0(n_1+m-1)}{\min_{\tau_i, \tau_1, \tau_2} \hat{u}(\tau_i, \tau_1) - u(\tau_i, \tau_2, 1/(1-\rho_2))} = \frac{1-\delta}{\delta}. \tag{32}$$

Proof of Proposition 6. (Part 1) Assume first that $\sigma_0 < \overline{\sigma}$. Then, $C_{\min}(m) < 0$ for each m. Hence, the cost of experimentation becomes negative infinitely often, when it is optimal to experiment and target the best informed player from the set N' until it becomes empty. Therefore, $m^\infty = n_2$. To prove the claim for large δ, observe that

$$\min_{m<n_2} \frac{V_0(n_1+m) - V_0(n_1+m-1)}{C_{\min}(0)} > \frac{1-\delta}{\delta};$$

when δ is sufficiently large; pick any such δ. We will show that she experiments at any $m < n_2$ whenever the cost is sufficiently small, and thus $m^\infty = n_2$. Suppose this is not the case: she chooses a known target with probability 1 at some $m < n_2$. Then, $V(m) = V_0(n_1+m-1)$. Hence,

$$\delta(V(m+1) - V(m)) \geq \delta(V_0(n_1+m) - V_0(n_1+m-1)) > (1-\delta)C_{\min}(0) \geq (1-\delta)C_{\min}(m),$$

showing that she must experiment when the cost is sufficiently low, a contradiction. (Here, the first inequality is by $V(m+1) \geq V_0(n_1+m)$, as the latter payoff can be obtained by stopping at $m+1$; the strict inequality is by the assumption on δ, and the last equality is by the fact that $C_{\min}$ is weakly decreasing.)

(Part 2) Assume $\sigma_0 > \overline{\sigma}$. Then, $C_{\min}(0) > 0$. Then,

$$\frac{V_0(n_1+n_2-1) - V_0(n_1-1)}{C_{\min}(0)} < \frac{1-\delta}{\delta}$$

when δ is sufficiently small. Take any such δ. Observe that $V(1) \leq V_0(n_1+n_2-1)$ and $V(0) \geq V_0(n_1-1)$. Hence,

$$\frac{V(1) - V(0)}{C_{\min}(0)} \leq \frac{V_0(n_1+n_2-1) - V_0(n_1-1)}{C_{\min}(0)} < \frac{1-\delta}{\delta},$$

showing that she will choose a target in set $N_1 \setminus \{i\}$ at the beginning with probability one, and therefore $m^\infty = 0$.

(Part 3) By definition, $C_{\min}(m) < 0$ whenever $m > \overline{m}\left(\sigma_0^2, \rho_2\right)$. Hence, at any $m > \overline{m}\left(\sigma_0^2, \rho_2\right)$, almost surely, the cost becomes negative infinitely often, and she chooses a target from the set N', experimenting with an unknown target. Thus, if she observes $m > \overline{m}\left(\sigma_0^2, \rho_2\right)$ targets from group 2, then she will almost surely observe everybody from group 2 eventually. That is, if $m^\infty > \overline{m}\left(\sigma_0^2, \rho_2\right)$, then $m^\infty = n_2$.

(Part 4) Take any $\sigma_0 > \overline{\sigma}$ and $\rho_2 < 1 - \overline{\sigma}^2/\sigma_0^2$. Observe that, for each m, $\phi(m, \rho_2) < 1/(1-\rho_2)$, and hence

$$C_{\min}(m) > \min_{\tau_i, \tau_1, \tau_2} \hat{u}(\tau_i, \tau_1) - u(\tau_i, \tau_2, 1/(1-\rho_2)) > 0. \tag{33}$$

Using mathematical induction backwards on m, we will show that she chooses a known target with probability 1 at each $m > \hat{m}$, showing that $m^\infty \leq \hat{m}$. To do this, we first observe

that the statement is vacuously true when $m \geq n_2$. For mathematical induction, assume that she chooses a known target with probability one at some $m+1$ with $m > \hat{m}$. Then, $V(m+1) = V_0(n_1+m)$, and hence $V(m+1) - V(m) \leq V_0(n_1+m) - V_0(n_1+m-1)$. Thus,

$$\frac{V(m+1)-V(m)}{C_{\min}(m)} < \frac{V_0(n_1+m)-V_0(n_1+m-1)}{\min_{\tau_i,\tau_1,\tau_2} \hat{u}(\tau_i,\tau_1) - u(\tau_i,\tau_2,1/(1-\rho_2))} < \frac{1-\delta}{\delta}, \tag{34}$$

where the first equality is by (33) and the second inequality is by $m > \hat{m}$. Therefore, she chooses a known target with probability 1 at m. □

Proposition 6 extends our main results to allow for forward-looking behavior as follows. The first part states that, as in Proposition 1, when $\sigma_0 < \overline{\sigma}$, cultures merge eventually, and individual i observes the best informed individual in the population as a whole. This follows from the fact that, when $\sigma_0 < \overline{\sigma}$, the minimum cost $C_{\min}$ is negative, and hence the cost becomes negative infinitely often almost surely, when she experiments by observing an unknown source (if there is one). She thus learns the perspectives of all individuals eventually. This is also the outcome when i is sufficiently patient, even when the the minimum cost $C_{\min}$ is positive, since the cost is outweighed by the perpetual benefit.

Likewise, as in Proposition 3, the third part states that when $\sigma_0 > \overline{\sigma}$ and $\rho_2 > 1 - \overline{\sigma}^2/\sigma_0^2$, in the long run, she will either observe only $m \leq \overline{m}(\sigma_0^2, \rho_2)$ individuals from the other group or she will observe everyone. This follows from the fact that, when $m > \overline{m}(\sigma_0^2, \rho_2)$, the minimum cost $C_{\min}(m)$ is negative, and she chooses to observe everyone eventually if she observes more than $\overline{m}(\sigma_0^2, \rho_2)$ individuals from group 2. These parts of the result extend our earlier findings on the merging of cultures and bimodality in a straightforward fashion.

The remaining parts extend our results on homophily and the law of the few for low and moderate discount factors. The second part states that there will be extreme homophily with probability 1 when players are sufficiently impatient, extending Proposition 2 from $\delta = 0$ to small δ. Together with Part 3, this also illustrates that when individuals have varying degrees of impatience, the distribution will be bimodal: some impatient individuals will exhibit homophily by observing at most a few sources from the other group, while more patient players will observe everyone.

To illustrate the rationale for homophily, take $\rho_2 = 1$. This is the worst case for homophily when $\rho_1 = 1$. Now, if i observes the opinion of a target from group 2, she will learn the perspectives of everybody, and her continuation value will be $V(1) = V_0(n_1+n_2-1)$. If she does not observe anybody from group 2, her continuation value will be $V(0) = V_0(n_1-1)$. Hence, the maximal benefit to cost ratio from observing a single individual from group 2 is

$$\frac{V_0(n_1+n_2-1) - V_0(n_1-1)}{C_{\min}(0)},$$

which is positive. When i is sufficiently impatient so that the above ratio is below $(1-\delta)/\delta$, she will never observe anybody from group 2, exhibiting extreme homophily. Otherwise she will observe the best informed player from group 2 when the cost is near the minimal cost, which will happen almost surely eventually.

Finally, for $\rho_2 < 1 - \overline{\sigma}^2/\sigma_0^2$, Part 4 provides a deterministic upper bound $\hat{m}$ for the number of targets that i observes from the other group in the long run, extending Proposition 4 to forward looking behavior. The upper bound $\hat{m}$ is increasing in the discount factor δ, approaching 0 as $\delta \to 0$ and ∞ as $\delta \to 1$. More interestingly, it is increasing in the correlation ρ_2 of perspectives in group 2. This is because the higher the correlation the cheaper it is to experiment with additional members from group 2. Finally, the upper bound is decreasing in the size n_1 of i's own group, approaching 0 as $n_1 \to \infty$. As in Proposition 5, the incentive to seek opinions from outsiders vanishes as one's own group grows large.

B.3 Multiple Observations

We have assumed to this point that individuals observe just one opinion at a time. If perspectives were independently distributed, this would not affect our qualitative results. But with correlated perspectives, allowing individuals to observe multiple opinions in a given period can have subtle and interesting effects. When one learns from biased sources, the correlation between those biases plays an important role. On the one hand, it helps by allowing players identify the bias. On the other hand, it hurts because there is less independent information in each signal. Either effect can dominate, and the overall effect is non-monotonic in correlation.

To see this, suppose an individual i observes the opinions of two individuals j and j'. This observation increases the precision of her belief about θ by

$$\Delta = \frac{\gamma_{ij}(t) + \gamma_{ij'}(t) - 2c}{\gamma_{ij}(t)\gamma_{ij'}(t) - c^2} \tag{35}$$

where $c = \tau_{jt}^2 \tau_{j't}^2 \sigma_{jj'}(i,t)$ is the covariance of the noise terms in the opinions of j and j' according to i. When μ_j and $\mu_{j'}$ are uncorrelated, Δ is simply the sum of the precisions $1/\gamma_{ij}(t)$ and $1/\gamma_{ij'}(t)$ of the opinions. In this case i would simply target the two sources with the most informative opinions, and multiple observations do not change the analysis much. When perspectives are correlated, the correlation affects Δ non-monotonically: it is decreasing in c if and only if $(\gamma_{ij}(t) - c)(\gamma_{ij'}(t) - c) > 0$. In particular, correlation hurts when the sources have similar precision or the correlation is low, but it may help when the sources have substantially different precisions. Figure 3 plots Δ as a function of the correlation in perspectives. Δ is decreasing in $\sigma_{jj'}(i,t)$ until it reaches $1/2$ and increases there after. Observe that Δ is higher than the uncorrelated case when the correlation is very high.

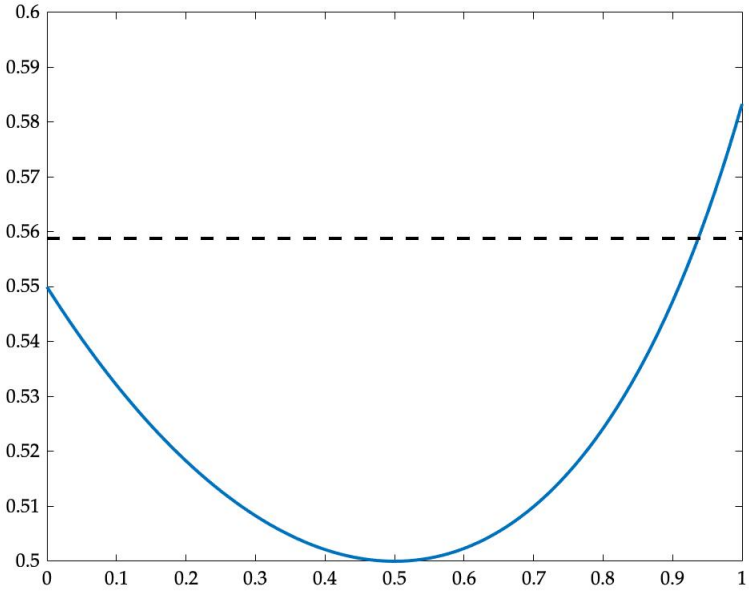

Figure 3: The precision gain Δ from two opinions as a function of covariance $\sigma_{jj'}(i,t)$ between the underlying perspectives; $\tau_{jt} = 1; \tau_{j't} = 2; \sigma_j^2(i,t) = \sigma_{j'}^2(i,t) = 1$.

To see how this affects our analysis, consider four potential sources: $\{1,2\}$ from i's own group, and $\{3,4\}$ from the other group. Sources 1 and 3 have high expertise with $\tau_{1t} = \tau_{3t} = 1$ while 2 and 4 have low expertise with $\tau_{2t} = \tau_{4t} = 2$. Player i knows the perspectives of sources from her own group better: $\sigma_1^2(i,t) = \sigma_2^2(i,t) = 1$ while $\sigma_3^2(i,t) = \sigma_4^2(i,t) = 16$. Viewed in isolation, the informativeness of the four opinions satisfies:

$$\gamma_{i1}(t) = 2 < \gamma_{i3}(t) = 17 < \gamma_{i2}(t) = 20 < \gamma_{i4}(t) = 260.$$

If perspectives were uncorrelated, i would simply target the two best-informed sources, 1 and 3, who also happen to have the most informative opinions in this example. However, perspectives are correlated within each group. If i chooses to observe 1 and 2, her precision gain is

$$\Delta = \frac{22 - 8\sigma_{12}(i,t)}{40 - 16\left(\sigma_{12}(i,t)\right)^2},$$

which is plotted in Figure 3. If i chooses to observe 1 and 3, her precision gain would be $\Delta = 1/2 + 1/17$, corresponding to the dashed horizontal line in the figure. When the correlation is sufficiently high, i prefers to target fellow group members 1 and 2, as this results in a larger precision gain, using information on correlation.

In this example, correlation increases the incentive for homophily. But it can also have the opposite effect. For example, suppose that $\sigma_3^2(i,t) = 24$, so that $\gamma_{i3}(t) = 25$, and i would prefer 1 and 2 when viewed individually. In this case the precision gain from 1 and 3 is 0.54. If the correlation $\sigma_{12}(i,t)$ is near $1/2$, the precision gain from 1 and 2 is nearly

0.5, and she would prefer to observe the opinions of 1 and 3. Thus correlation can also decrease incentives for homophily.

B.4 Multiple Cultures

For simplicity, we have considered only two groups, with zero cross-group correlation in perspectives. In this section, we extend our results to the case of multiple cultural groups with the possibility of cross-group correlation. In this case one can learn about one group by observing those in a proximate culture, and this gives rise to even richer patterns of behavior.

Suppose that there are K groups, $N_1, \ldots, N_K$, such that the correlation between any pair of perspectives μ_i and μ_j is ρ_k if both i and j belong to the same group k, and the correlation is ρ_{kl} if $i \in N_k$ and $j \in N_l$ for some distinct groups k and l. We assume that $\rho_k > \rho_{kl} \geq 0$ for all distinct groups k and l. Our main results extend to this case as follows.

Proposition 7. *In the model with multiple correlated cultures, the following statements are true.*

1. *If $\sigma_0^2 < \overline{\sigma}^2$, then $J_i = N \setminus \{i\}$ for all i almost surely.*
2. *Assume $\rho_{kl} = 0$ for all distinct groups k, l and $\sigma_0^2 > \overline{\sigma}^2$. Then, there is a positive probability that each individual exhibits extreme homophily at each t. In addition, if $\rho_k > \overline{\rho}$, then all individuals in group k exhibit extreme homophily at each t almost surely.*
3. *Assume that $\sigma_0^2(1 - \rho_{kl}^2) > \overline{\sigma}^2$ for all distinct k, l. Then, there is a positive probability that each individual exhibits extreme homophily at each t. There exists $\tilde{\rho} < 1$ such that, whenever $\rho_k > \tilde{\rho}$, all individuals in group k exhibit extreme homophily at each t almost surely.*
4. *Consider any individual i and any group k with $\rho_k > \underline{\rho} \equiv 1 - \overline{\sigma}^2/\sigma_0^2$. Then:*

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

The first part extends Proposition 1 to the case of multiple groups, without any assumption on cross-cultural correlations. Such correlations only facilitate and expedite the merging of cultures. The second and third parts extend Proposition 2 on extreme homophily. When there is no cross-cultural correlation, Proposition 2 extends to the multicultural model verbatim. When there is cross-cultural correlation, then each individual's initial uncertainty about the perspectives in other groups is also lowered. In particular, the variance of the initial belief of an individual $i \in N_k$ about the perspective of $j \in N_l$ is

$$\sigma_j^2(i, 0) = (1 - \rho_{kl}^2)\sigma_0^2. \tag{36}$$

Part 3 extends our homophily result to the case of multiple correlated cultures by assuming that all of these initial variances are above the cutoff $\overline{\sigma}^2$. The cutoff on within-group correlation that ensures extreme homophily is now higher.

Finally, Part 4 extends our result on bimodality to the case of multiple groups without making any restriction on cross-cultural correlations. Such correlations would only enlarge the scope of bimodality, by lowering the cutoff $\overline{m}(\sigma_0^2, \rho_k)$ for tipping. When there is no cross-cultural correlation, our probability bound in Proposition 4 for the size of expert sets also extends to the multiple group case verbatim. (The proofs also extend to the case with cross-cultural correlation *mutatis mutandis*, and we omit them.)

The correlation ρ_{kl} can be viewed as a measure of similarity between the cultures of groups k and l. When this is positive, observers can learn about members of one culture by observing those of the other. This has an impact on information flows, and can give rise to richer patterns of behavior. We illustrate this here by stating two variants of our earlier results. First, we observe that two similar cultures can merge even when initial uncertainty is so high that two dissimilar cultures would not:

Proposition 8. *For any distinct groups k and l, if $(1-\rho_{kl}^2)\sigma_0^2 < \overline{\sigma}^2$, then $N_l \subseteq J_i$ for all $i \in N_k$ almost surely.*

This result points to an interesting pattern of behavior. For example, consider four groups, $N_1, \ldots, N_4$, where within-group correlations and the across-group correlations ρ_{12} and ρ_{34} are higher than $1 - \overline{\sigma}^2/\sigma_0^2$ while all the other cross-group correlations are zero. Then, the cultures $\{1,2\}$ and $\{3,4\}$ will merge in the long-run while there can be no communication across $N_1 \cup N_2$ and $N_3 \cup N_4$.

Moreover, since individuals can learn indirectly about one group by observing members of another, there will also be tipping behavior across groups:

Proposition 9. *For any $i \in N$ and any two distinct groups k and l with*

$$\rho_{kl}^2/\rho_k > 1 - \overline{\sigma}^2/\sigma_0^2, \tag{37}$$

if $|J_i \cap N_k| > \overline{m}_{kl}$, then $N_l \subset J_i$ almost surely where

$$\overline{m}_{kl} = \frac{1-\rho_k}{\rho_{kl}^2/(1-\overline{\sigma}^2/\sigma_0^2) - \rho_k}$$

Proof of Proposition 9. If i observes the perspectives of m individuals from group k, then the variance of her belief about any individual $j \in N_l$ reduces to σ_0^2/ϕ_{kl} where

$$\phi_{kl} = \frac{1+(m-1)\rho_k}{1+(m-1)\rho_k - m\rho_{kl}^2}.$$

The cutoff $\overline{m}_{kl}$ in the proposition is the solution to the equation $1/\phi_{kl} = \overline{\sigma}^2/\sigma_0^2$, so that $\sigma_0^2/\phi_{kl} < \overline{\sigma}^2$ whenever $m > \overline{m}_{kl}$. Hence, if $|J_i \cap N_k| > \overline{m}_{kl}$, then $\sigma_j^2(i,t) < \overline{\sigma}^2$ in the long run. Thus, the link from i to j must become free at some t, showing that $j \in J_i$ almost surely. □

As an individual i learns the perspectives in group k, she also indirectly learns about the perspectives in group l. The condition (37) ensures that such indirect learning renders links to group l free if i observes infinitely many such perspectives. Thus, those links will become free if i learns more than some finite number $\overline{m}_{kl}$ of perspectives of individuals in group k, as computed in the proposition. In particular, if i observes more than $\overline{m}_{kl}$ individuals from group k in the long run, she will come to learn enough about those in group l to observe them whenever their expertise is sufficiently higher relative to those outside the group.

Example 2. *Consider three large groups, N_1, N_2, N_3. The groups are located on a line, and their perspectives are correlated with those of their neighbors, i.e., ρ_{12} and ρ_{23} are positive, but $\rho_{13} = 0$. Specifically, for each player $i \in N_k$, $\mu_i = \overline{\mu}_k + \nu_i$ and*

$$\overline{\mu}_k = X_k + X_{k-1,k} + X_{k,k+1}$$

for some independent standard normal random variables $X_{01}, X_1, X_{12}, X_2, X_{23}, X_3, X_{34}$ and ν_i. Observe that $\sigma_0^2 = 4$, $\rho_k = 3/4$, and $\rho_{12} = \rho_{23} = 1/4$ while $\rho_{13} = 0$. Take $4 > \overline{\sigma}^2 > 3.75$. First, imagine that group 2 is absent. In this case, by Proposition 2, with some positive probability, there would be extreme homophily at each period t, and nobody from group 1 would learn from anyone from group 3. Now, introduce group 2. Observe that $(1 - \rho_{12}^2)\sigma_0^2 = 3.75$, which is less than $\overline{\sigma}^2$. Hence, by Proposition 8, groups 1 and 2 will merge in the long run. In particular, J_i contains N_2 for each $i \in N_1$. But observe also that $\rho_{23}^2/\rho_2 = 1/12 > 1 - \overline{\sigma}^2/\sigma_0^2$. Hence, by Proposition 9, J_i must also contain N_3. That is, in the long run, each individual i in group 1 must also target each individual j in group 3 whenever j is more informed than others.

This example illustrates the kinds of interesting observational phenomena that could arise with multiple correlated groups. It also makes clear that changes in cross-group correlation—for instance through migration—can have an impact on third parties, and set off cascading effects that spread across groups over time.

References

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

Mariagiovanna Baccara and Leeat Yariv. Homophily in peer groups. American Economic Journal: Microeconomics, 5(3):69–96, 2013.

Mariagiovanna Baccara and Leeat Yariv. Choosing peers: Homophily and polarization in groups. Journal of Economic Theory, 165:152–178, 2016.

Eytan Bakshy, Solomon Messing, and Lada A Adamic. Exposure to ideologically diverse news and opinion on facebook. Science, 348(6239):1130–1132, 2015.

Abhijit Banerjee and Olivier Compte. Consensus and disagreement: Information aggregation under (not so) naive learning. Journal of Political Economy, 132(8):2790–2829, 2024.

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

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.

Pranav Dandekar, Ashish Goel, and David T Lee. Biased assimilation, homophily, and the dynamics of polarization. Proceedings of the National Academy of Sciences, 110(15): 5791–5796, 2013.

Krishna Dasaratha, Benjamin Golub, and Nir Hak. Learning from neighbours about a changing state. Review of Economic Studies, 90(5):2326–2369, 2023.

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

Seth Flaxman, Sharad Goel, and Justin M Rao. Filter bubbles, echo chambers, and online news consumption. Public Opinion Quarterly, 80(S1):298–320, 2016.

Roland G Fryer, Philipp Harms, and Matthew O Jackson. Updating beliefs when evidence is open to interpretation: Implications for bias and polarization. Journal of the European Economic Association, 17(5):1470–1501, 2019.

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

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

Matthew Gentzkow, Michael B Wong, and Allen T Zhang. Ideological bias and trust in information sources. American Economic Journal: Microeconomics, 17(2):162–213, 2025.

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

Dan M Kahan, Hank Jenkins-Smith, and Donald Braman. Cultural cognition of scientific consensus. Journal of Risk Research, 14(2):147–174, 2011.

Willemien Kets and Alvaro Sandroni. A belief-based theory of homophily. Games and Economic Behavior, 115:410–435, 2019.

Timur Kuran. Private Truths, Public Lies: The Social Consequences of Preference Falsification. Harvard University Press, 1998.

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. Complementary information and learning traps. Quarterly Journal of Economics, 135(1):389–448, 2020.

Glenn C Loury. Self-censorship in public discourse: A theory of "political correctness" and related phenomena. Rationality and Society, 6(4):428–461, 1994.

Miller McPherson, Lynn Smith-Lovin, and James M Cook. Birds of a feather: Homophily in social networks. Annual Review of Sociology, 27(1):415–444, 2001.

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

Steven Pinker. The Blank Slate: The Modern Denial of Human Nature. Penguin, 2002.

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.

Thomas Sowell. A Conflict of Visions: Ideological Origins of Political Struggles. William Morrow & Co, 1987.