Culture and Communication

Rajiv Sethi[*] Muhamet Yildiz[†]

September 20, 2016

Abstract

A defining feature of culture is similarity in the manner in which information about the world is interpreted. This makes it easier to extract information from the beliefs of those within one's own group. But this information itself may be of low quality if better informed sources lie elsewhere. Furthermore, observing individuals outside one's culture deepens our understanding not only of those individuals, but also of their group. We model this process, using unobservable, heterogeneous priors to represent fundamental belief differences across individuals; these priors are correlated within but not across groups. Within this framework, we obtain the following results. First, groups that are smaller and have higher levels of correlation in perspectives will be more likely to exhibit homophily to begin with. If the correlation in perspectives is sufficiently high, then this homophily persists over time, resulting in homogeneity and insularity in observational patterns. If not, then persistent behavioral heterogenity can arise both within and across groups, even if individuals in the same group are identical at the outset. Patterns of observation exhibit considerable structure. Under certain conditions—which depend on the variability across individuals in the quality of information, initial uncertainty about the perspectives through which this information is filtered, and the degree to which these perspectives are correlated—individuals in each group can be partitioned into two categories. One of these exhibits considerable homophily, rarely if ever stepping outside group boundaries, while the other is unbiased and seeks information wherever it is most precise.

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

1 Introduction

A defining feature of culture is similarity in the manner in which information about the world is interpreted. Two individuals who share a common culture—defined by ethnicity, religion, or even politics—will tend to have correlated mental models of the world, which facilitates communication. In particular, it easier to extract the informational content of a statement when both the speaker and the listener belong to the same group. This is a force for informational homophily: in seeking information people will often turn to those whose perspectives they understand.

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

This trade-off between these two forces changes over time, based on an individual's observational history. Previously observed sources become better understood and hence more likely to be observed again. But the degree to which an individual's understanding of another deepens through observation depends on how well-informed the observer herself happens to be in the period of observation. This is a force for symmetry-breaking, and divergence over time in the behavior of individuals who are initially identical and belong to the same group.

This intra-group heterogeneity is further reinforced because those who repeatedly observe individuals outside their own group learn not only about the perspectives of their individual targets, but also about the group to which the targets belong. That is, learning about a person from another culture teaches us not only about the person, but also about the culture to which they belong. As a consequence, such individuals become more likely to step outside the boundaries of their own group in the future.

We model this process, using unobservable, heterogeneous priors to represent fundamental belief differences across individuals. These priors—which we call *perspectives*—are correlated within but not across groups. That is, individuals initially have more precise beliefs about the perspectives of those within their culture than those outside it.

Within this framework, we obtain the following results. Groups that are smaller and have higher levels of correlation in perspectives will be more likely to exhibit homophily to begin with. If the correlation in perspectives is sufficiently high, then this homophily persists over time, resulting in homogeneity and insularity in observational patterns. If not, then heterogenity in behavior both within and across groups gan arise, even if individuals in the same group are identical at the outset. But patterns of observation exhibit considerable structure. Under certain conditions—which depend on the variability across individuals in the quality of information, initial uncertainty about the perspectives through which this information is filtered, and the degree to which these perspectives are correlated—individuals in each group can be partitioned into two categories. One of these exhibits considerable homophily, rarely if ever stepping outside group boundaries, while the other is unbiased and seeks information wherever it is most precise. This bimodality in observational behavior is the key testable prediction of the model.

2 Literature

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

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

This literature attempts to explain persistent and public differences across groups in beliefs, but not with the sources of information that individuals actively seek. In order to address this latter question, a theory needs to accommodate both fundamental belief differences across groups as well as idiosyncratic differences across individuals in the quality of information about the world. The framework developed in Sethi and Yildiz (2012, 2016)—where heterogeneous priors represent fundamental belief differences and signals of varying precision represent information—allows for such an exploration. Although priors are initially unobserved, they are drawn from a commonly known distribution, so individuals can reason and update their beliefs about these as time unfolds and information is received. In our earlier work we have used this framework to explore conditions under which distributed information is aggregated, and to study the endogenous formation of information networks in a population without distinct identity groups. Here we build on this by allowing for different cultures, with a particular correlation structure on the distribution from which priors are drawn.

Lazarsfeld and Merton (1954) are credited with coining the term homophily, and associating it with the proverb “birds of a feather flock together” (McPherson et al., 2001). Although homophily arises along multiple dimensions of interaction, our concern here is with the tendency to turn to individuals within one’s own identity group or culture for the purpose of information gathering. Gentzkow and Shapiro (2011) have examined this issue, in the context of ideological identity (conservative and liberal) in the United States. They find that ideological homophily in access to online news sources is greater than that in access to offline news, though considerably smaller than that in face-to-face interactions in neighborhoods, workplaces and voluntary associations. Here news sources are themselves are placed on an ideological spectrum based on the distribution of their users across political identity groups.

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

The idea that individuals can extract information more easily from those with whom they share a culture is the basis for a branch of the statistical discrimination literature descended from Phelps (1972); see especially Aigner and Cain (1977) and Cornell and Welch (1996).[1] 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 heterogneity.

3 The Model

The model in this section builds on Sethi and Yildiz (2016), by allowing for separate social groups with correlation in prior beliefs within (but not between) groups.

3.1 Groups and Perspectives

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

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

$n_k \geq 3$ denote the size of group k and $w_k = n_k/n$ its population share. For each $i \in N$, let $g(i) \in \{1, 2\}$ denote the group to which i belongs.

There is a sequence of periods $t = 0, 1, ...$ and in each period 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 regarding 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:

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

We refer to prior mean μ_i as the *perspective* of i, and note that it is stable over time. 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 directly observable by others, but it is commonly believed that the perspectives $\mu = (\mu_1, ..., \mu_n)$ are jointly distributed distributed according to

$$\mu \sim N(0, \Sigma),$$

where Σ is a variance-covariance matrix with typical element σ_{ij}.[2]

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

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

where the submatrices Σ_k have diagonal elements 1 and off-diagonal elements $\rho_k \in (0, 1]$. That is, for any $i \neq j$, $\sigma_{ij} = \sigma_0^2 \rho_k$ if $g(i) = g(j) = k$, and $\sigma_{ij} = 0$ otherwise. We can think of a group with high correlation ρ_k as being relatively homogeneous or tightly-knit.

The parameter σ_0^2 reflects the degree to which one individual's beliefs about the perspectives of others are imprecise, before one's own prior has been observed. Having observed one's own prior, however, beliefs about the perspectives of others within one's own group are updated (while beliefs about the perspectives of those in the other group remain unchanged). Specifically, if i and j both belong to group k, beliefs are updated as follows. Observing μ_i, i believes that μ_j is distributed normally with mean

$$E_i[\mu_j | \mu_i] = \rho_k \mu_i$$

and variance

$$Var_i (\mu_j | \mu_i) = \sigma_0^2 \left(1 - \rho_k^2\right).$$

[2]The analysis to follow does not depend on the means all being equal to zero, and would remain intact for any choice of means that are fixed and commonly known.

If $\rho_k = 1$, there is complete within-group homogeneity of perspectives. If $\rho_k < 1$, there is some heterogeneity of perspectives even within the group, and subjective uncertainty about these perspectives, but less uncertainty than there is about the perspectives of those outside the group.

Put differently, individuals within one's own group are *better understood* (beliefs about their perspectives is more precise), when compared with those outside ones group. This need not persist over time, however, and will depend on patterns of mutual observation as information is received and perspectives are learned.

Beliefs about the perspectives of others (both within and outside one's own group) will change over time through the observation of posterior beliefs, in a manner to be described below. We write $\Sigma(i,t)$ for the history-dependent variance-covariance matrix of μ at time t for player i; $\Sigma(i,t)$ 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. Note that $\sigma_j^2(i,0) = \sigma_0^2$ if i and j belong to different groups, and $\sigma_j^2(i,0) = \sigma_0^2(1-\rho_k^2)$ if they both belong to group k. Moreover, for all i and t, $\sigma_{jj'}(i,t) = 0$ whenever j and j' belong to distinct groups.[3]

3.2 Information and Expertise

Next we allow for informative signals. Suppose that in each period t, each individual i privately 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* about the period t state. We assume that these variances are public information, and are independently and identically distributed across individuals and over time in accordance with an absolutely continuous distribution function F having support $[\underline{\tau}^2, \overline{\tau}^2]$, where $0 < \underline{\tau}^2 < \overline{\tau}^2 < \infty$. That is, no individual is ever perfectly informed of the state, but all signals carry at least some information.

Having observed the signal x_{it}, individual i updates her belief about the period t state according to Bayes' rule. This results in the following posterior belief for i:

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

where y_{it} is the expected value of θ_t according to i, which we refer to as individual i's *opinion* in period t. This is computed as

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

These opinions are potentially observable by others, although the priors and signals are not separately observable.

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

3.3 Choosing Targets

Suppose that individuals can observe exactly one other opinion in each period, and would like to choose a target with the goal of having the most precise beliefs about the current state. In making this choice, an individual will generally face a trade-off between those who are well-informed (in the sense of having precise signals) and those who are well-understood (in the sense of having perspectives that are known with high precision to the observer in question).

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

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

The signal is noisy for two reasons: j is neither perfectly informed, nor perfectly understood by i in the period of observation. Taken together, the variance of the noise in this signal is

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

Here, the first term is due to the noise in the information of j, and is simply the variance of ε_j. It decreases as j becomes better informed. The second term comes from the uncertainty i faces regarding the perspective μ_j of j, and is the variance of $\tau_{jt}^2 \mu_j$ (where τ_{jt}^2 is public information and hence has zero variance). This term decreases as i develops more precise beliefs about the perspective μ_j, that is, as j becomes better understood by i.

The expression for the variance $\gamma_{ij}(t)$ reveals that in choosing a target j, an individual i has to trade-off the noise in the information of j against the noise in i's understanding of j's perspective, normalized by the level of j's expertise. This is the trade-off between targets who are well-informed and those who are well-understood.

The assignment of individuals to targets in period t may be represented by a function $\lambda_t : N \to N$, where $\lambda_t(i)$ is an element of $N \setminus \{i\}$. We shall assume that each individual in each period chooses the target that provides the most informative opinion, with ties being broken in favor of the one with the smallest label. That is

$$\lambda_i(t) = \min \left\{ \arg\min_j \gamma_{ij}(t) \right\}.$$

This function may be interpreted as a directed graph, with each node i associated with an outgoing edge to the node $\lambda_t(i)$. The graph can (and in general will) change from one period to the next, as new levels of expertise are realized. But these changes will be history dependent, since the understanding of others in the population will depend on past linkages and the conditions under which they form.

3.4 Learning Perspectives

Each time a target is observed, the observer learns something not only about the current state, but also about the target's perspective, and indeed (since perspectives are correlated within groups) about the perspectives of others in the target's group.

Given some individual i and period t, let $l = \lambda_t(i)$ denote the target chosen by i in t. That is, i observes the opinion y_{lt} in period t. This opinion has been formed in accordance with (2-3), and hence provides the following signal for μ_l:

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

The signal contains an additive noise term with variance

$$\alpha(\tau_{it}^2, \tau_{lt}^2) = \frac{1}{\tau_{lt}^4}\left(\frac{\tau_{it}^2}{1+\tau_{it}^2} + \tau_{lt}^2\right). \tag{5}$$

Note that this variance depends on the expertise of the *observer* as well as that of the target, through the observer's uncertainty about θ_t. In particular, the variance α is increasing in τ_{it}^2 and decreasing in τ_{lt}^2. Hence i obtains a less noisy signal of her target's perspective if the target is poorly informed, or if i herself is well-informed. This is intuitive: a poorly informed target will have an opinion close to her prior, while a better-informed observer will make a sharper inference from the target's opinion. Note that there are lower and upper bounds within which α must lie, given by $\underline{\alpha} = \alpha(\underline{\tau}^2, \overline{\tau}^2)$ and $\overline{\alpha} = \alpha(\overline{\tau}^2, \underline{\tau}^2)$ respectively.

The opinion y_{lt} of i's target also provides information about the perspective μ_j of each $j \in g(l)$ because μ_j and μ_l are correlated. Individual i updates her beliefs about μ using this signal, yielding a new variance-covariance matrix $\Sigma(i, t+1)$ with entries

$$\sigma_{jj'}(i, t+1) = \sigma_{jj'}(i, t) - \frac{\sigma_{jl}(i,t)\sigma_{j'l}(i,t)}{\alpha(\tau_{it}^2, \tau_{lt}^2) + \sigma_l^2(i,t)} \tag{6}$$

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 l belongs, even though these are unobserved by i in t. That is, $\sigma_{jj}(i, t+1) < \sigma_{jj}(i, t)$ for $j \in g(l)$. In the special case of perfect correlation, we have $\sigma_{jj}(i, t+1) = \sigma_{ll}(i, t+1)$ for all $j \in g(l)$. Note that i does not update her beliefs about the members of the other group: $\sigma_{jj'}(i, t+1) = \sigma_{jj'}(i, t)$ for each pair $j, j' \notin g(l)$.

After observing her target's opinion, the precision of i's belief about l's perspective increases. Specifically, replacing both j and j' with l in (6), we obtain

$$\sigma_l^2(i, t+1) = \sigma_l^2(i, t) - \frac{\sigma_l^4(i,t)}{\alpha(\tau_{it}^2, \tau_{lt}^2) + \sigma_l^2(i,t)}. \tag{7}$$

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

To summarize, given the correlation of perspectives, changes in i's belief about j's prior will cause i to update her beliefs about the priors of all those in j's group. That is, observing a target is informative about the current state, the target's perspective, and the perspectives of all others in the target's group. In the extreme case of perfect correlation, i will know the perspectives of all those in her own group at the outset, and learn something about the perspectives of those in the other group each time a target from the group is selected for observation.

Given the distribution governing expertise realizations, the dynamics of belief updating define a Markov process where the period t state consists of the variance-covariance matrices $\Sigma(i,t)$ for $i \in N$. These matrices, together with the expertise realizations in t, fully determine the pattern of observation that will arise in each period.

With imperfect correlation, heterogeneity in beliefs about perspectives will emerge and persist, both within and across groups, and these beliefs will be sensitive to historical realizations of expertise.

4 Observational Patterns

In this section, we investigate the communication patterns that can arise with positive probability. Define the threshold variance

$$\overline{\sigma}^2 = \frac{\overline{\tau}^2 - \underline{\tau}^2}{\underline{\tau}^4}. \tag{8}$$

Here, $\overline{\sigma}^2$ is defined by equality $\underline{\tau}^2 + \underline{\tau}^4 \overline{\sigma}^2 = \overline{\tau}^2$, so that by (4) an individual i is indifferent between a target j with maximally precise signal and $\sigma_j^2(i,t) = \overline{\sigma}^2$ and a target j' with minimally precise signal and $\sigma_{j'}^2(i,t) = 0$. Hence, if $\sigma_j^2(i,t) < \overline{\sigma}^2$ at some period t, then i links to j with positive probability at period t. He links to j when j has very high and all individuals $j' \neq j$ have very low expertise. Furthermore, since $\sigma_j^2(i,t)$ is non-increasing over time, the inequality $\sigma_j^2(i,t) < \overline{\sigma}^2$ will continue to be satisfied thereafter, and i will observe j infinitely often almost surely.

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

$$\beta\left(s^2\right) = \frac{\underline{\tau}^4}{\overline{\tau}^4}(s^2 - \overline{\sigma}^2). \tag{9}$$

Here, $\beta\left(s^2\right)$ 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 (4) 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 $\sigma^2_{j'}(i,t) = \beta\left(s^2\right)$. Hence, if $\sigma^2_{j'}(i,t) < \beta(\sigma^2_j(i,t))$, then i never links to j at period t, regardless of expertise realizations. This is because i prefers j' to j even when j' has minimally precise signal and j has minimally precise signal. Clearly the same preference must hold for all other expertise realizations. In general, this does not mean that the link ij is broken forever because i may learn about the perspective of j from observing the opinions of the other members of the group j belongs to, and he may link to j later when is more familiar with j and j has higher expertise than others. On the other hand, if there is j' from a group k' with

$$\sigma^2_{j'}(i,t) < \max_{j \in N_k} \beta(\sigma^2_j(i,t)), \tag{10}$$

then all the links ij to group k is permanently broken because i does not update his beliefs about the perspectives in group k thereafter.

If $\sigma^2_0 < \overline{\sigma}^2$, an individual would prefer to observe someone outside her group if the latter had sufficiently high expertise, provided that each member of her own group has sufficiently low expertise. This is true even if no out-group members have been previously observed, and all in-group members are perfectly well understood. As we show below, a consequence of this assumption is that in the long run, targets will be chosen on the basis of expertise alone, regardless of group membership.

For $\sigma^2_0 < \overline{\sigma}^2$ to hold, at least one of two things must be true: informational differences across individuals in the population cannot be too small, and uncertainty about the perspectives of those outside one's group cannot be too great. Under these conditions informational differences will matter enough to eventually overpower the tendency to homophily the comes from differences in understanding.

If, instead, we have $\sigma^2_0 > \overline{\sigma}^2$, then observational patterns even in the long run will depend on the degree to which perspectives are correlated within each group, as we show below.

4.1 Initial Observation

While this process can give rise to complicated observational patterns over time, we begin with the first period, in which only two classes of outcome are possible. Consider the following mutually exclusive events:

(E_1) there exists $j \in N_1$ such that $\lambda_0(i) = j$ for all $i \neq j$

(E_2) there exists $j \in N_2$ such that $\lambda_0(i) = j$ for all $i \neq j$

(E_3) there exist $j_1 \in N_1$ and $j_2 \in N_2$ such that $\lambda_0(i) = j_1$ for all $i \in N_1 \setminus \{j_1\}$ and $\lambda_0(i) = j_2$ for all $i \in N_2 \setminus \{j_2\}$.

The events E_1 and E_2 correspond to homophily in one group, and extreme heterophily in the other. All individuals in the population observe the globally best informed individual j, who may or may not observe the second best informed.[4] Event E_3 involves homophily in both groups, with all individuals (except possibly the best-informed in each group) observing an in-group member.

Note that for any i and j from a group k, i updates her beliefs about the perspective of the fellow group member j from her own perspective, so that she has a more informed belief about μ_j:

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

The precision of her beliefs about the perspectives of the other group remains at v_0. For $\sigma_0^2 > \overline{\sigma}^2$, if ρ_k is sufficiently high, then the likelihood that i will link to someone outside her group in the initial period—and hence also in every other period—is zero. This is because $\sigma_j^2(i,0) < \beta(\sigma_0^2)$ whenever $\rho_k > \overline{\rho}$, where $\overline{\rho}$ satisfies $\sigma_0^2(1-\overline{\rho}^2) = \beta(\sigma_0^2)$ and is equal to

$$\overline{\rho} = \sqrt{1 - \left(\underline{\tau}^4/\overline{\tau}^4\right)\left(1 - \overline{\sigma}^2/\sigma_0^2\right)} \in (0,1)\,. \tag{11}$$

What can and cannot happen in the initial period of observation depends on whether $\sigma_0^2 < \overline{\sigma}^2$ and, if not, whether the degree of correlation in perspectives exceeds the threshold $\overline{\rho}$.

Proposition 1. *Only E_1, E_2 and E_3 can arise with positive probability:* $\Pr(E_1) + \Pr(E_2) + \Pr(E_3) = 1$*, and E_3 arises with positive probability for all parameter values. If $\sigma_0^2 < \overline{\sigma}^2$, then all three events have a positive probability of occurrence. If $\sigma_0^2 > \overline{\sigma}^2$, then* $\Pr(E_k) > 0$ *if and only if $\rho_{k'} < \overline{\rho}$ for $k' \neq k$.*

This result states that homophily arises with positive probability for all parameter values, and that at least one group must exhibit homophily in the initial period. This is intuitive. If the expertise levels of the two best-informed individuals in each group are sufficiently close, there will be no crossing of group boundaries. If there are large differences in expertise between the best informed individuals in the two groups, and the correlation in perspectives is not too great in either group, then the best-informed individual in the population as a whole will attract all observers. But if $\sigma_0^2 > \overline{\sigma}^2$, then any group with highly correlated perspectives will exhibit homophily. The high degree to which members of such groups understand each other will overwhelm any informational disadvantage that might arise.

Proposition 1 tells us which observational patterns can and cannot arise in the initial period of observation, but is silent on the likelihood of those events that can occur. The following result addresses this.

Proposition 2. *For distinct groups k and k',* $\Pr(E_k)$ *is increasing in n_k and decreasing in $n_{k'}$ and $\rho_{k'}$.*

[4] If there is little difference in expertise between the second best informed and the best informed in j's own group, and these are distinct individuals, then j may observe the latter.

This states that heterophily in a group is more likely if the group is small and the other group is large. This is intuitive: heterophily arises when the best informed in one's own group has substantially lower expertise than the best informed in the other group, and this in turn is more likely when one's own group is small and the other group large. The result also states that heterophily is less likely in tight-knit groups, which again is intuitive. In such groups individuals understand each other well and a greater difference in expertise across the best informed in the two groups is needed to generate heterophily.

All this applies to the initial period of observation, in which all individuals within a group are symmetrically placed with respect to each other. In subsequent periods this symmetry is broken and more complex patterns can arise.

4.2 Cross-Cutural Communication

At the end of the initial period, each individual has a better understanding of their target in accordance with (7). They also have a better understanding of the perspectives of those in the target's group, since perspectives are correlated. But the size of these effects vary across observers, even if they have a common target, since they depend on the expertise of the observer. Specifically, from (5), an observer with higher expertise in the initial period learns more about her target (and her target's group) than one with lower expertise. As a result, many more complex patterns of observation can arise over time.

To illustrate, consider the following example.

Example 1. *Suppose* $n = 6$ *with* $N_1 = \{1, 2, 3\}$ *and* $N_2 = \{4, 5, 6\}$*,* $\sigma_0^2 = 10$*, and* $\rho_1 = \rho_2 = 0.1$*. Expertise levels are* $(1, 0.1, 0.1, 0.9, 0.8, 0.8)$ *in the first period and* $(0.1, 0.1, 0.1, 0.25, 0.2, 0.2)$ *in the second. In this case the first period targets are* $\lambda_0(i) = 1$ *for all* $i \neq 1$ *and* $\lambda_0(1) = 4$*. The second period targets are* $\lambda_1(i) = 4$ *for all* $i \in N_1$ *and* $\lambda_1(i) = 1$ *for all* $i \in N_2$*.*

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

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

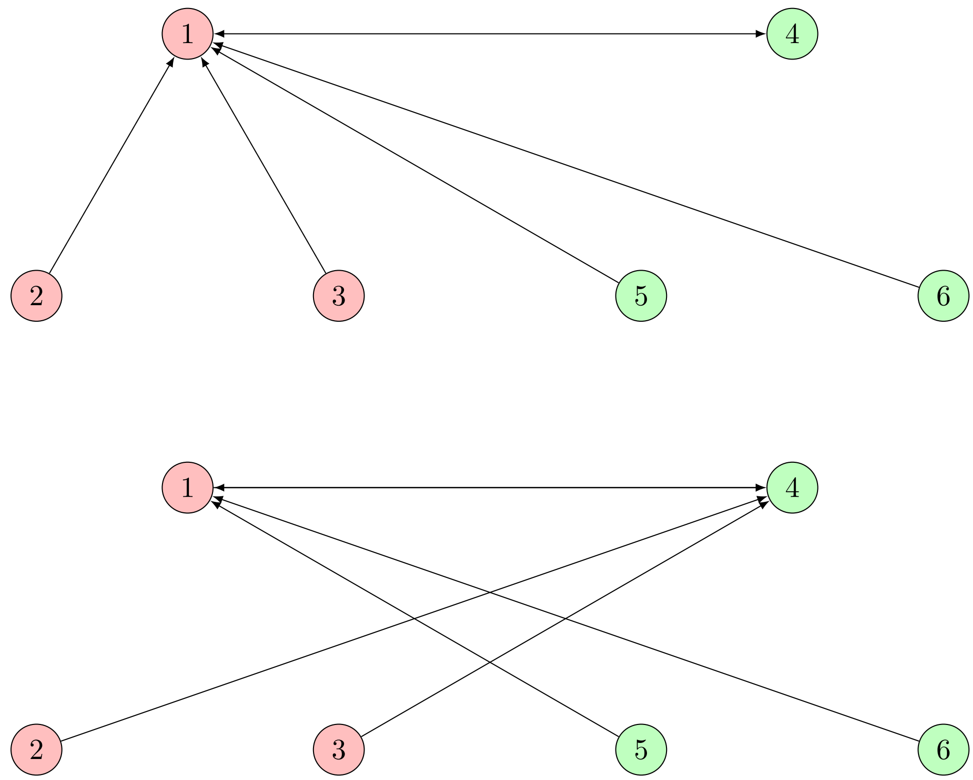

Figure 1: Homophily and heterophily followed by heterophily in both groups

also the case for 4, who has learned even more about the perspective of 1 in the initial period, and whose best within-group option in the second period is worse.

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

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

Example 1 exhibits maximal cross-cultural communication, with extreme heterophily in both groups. But this relies on extreme heterophily in one group in the first period, and raises the question of whether heterophily in both groups can arise after initial homiphily in both. The

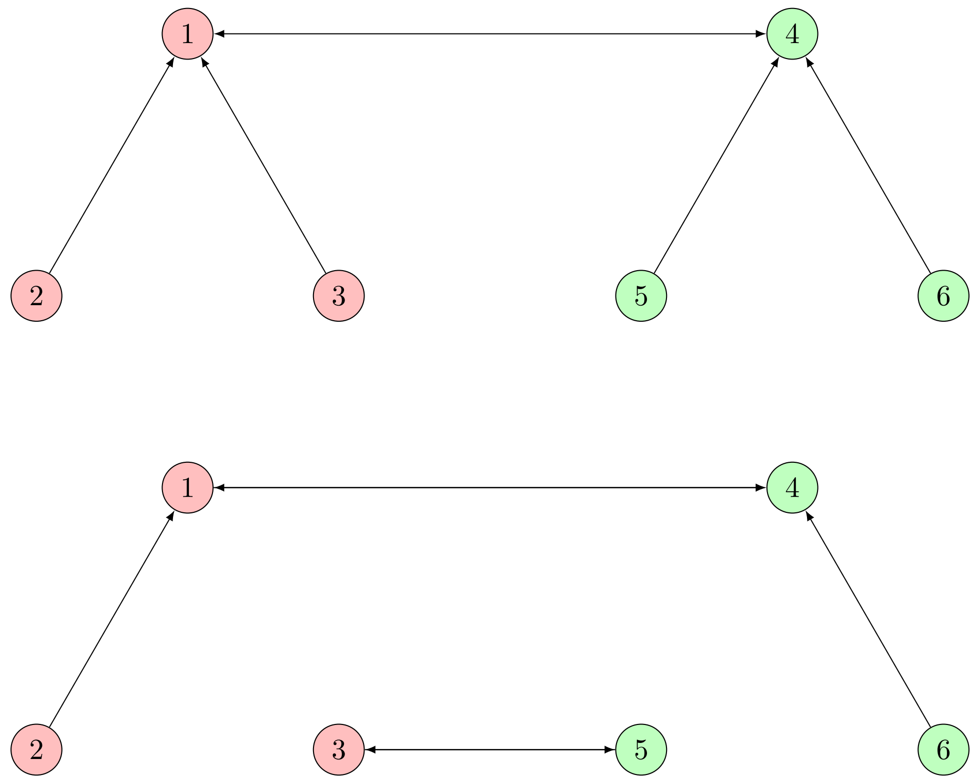

Figure 2: Homophily in both groups followed by heterophily in both

following example shows that it can.[5]

Example 2. *All parameters are as in Example 1 except for expertise levels* $(2, 1, 0.1, 2, 0.1, 1)$ *in the first period and* $(0.85, 0.1, 1.4, 0.85, 1.4, 0.1)$ *in the second. The first period targets are* $\lambda_0(i) = 1$ *for all* $i \in N_1 \setminus \{1\}$*,* $\lambda_0(i) = 4$ *for all* $i \in N_2 \setminus \{4\}$*,* $\lambda_0(1) = 4$ *and* $\lambda_0(4) = 1$*. The second period targets are the same as the first except for* $\lambda_1(3) = 5$ *and* $\lambda_1(5) = 3$*.*

In this example each individual observes the best informed in her own group initially, and the two best informed individuals in the population (1 and 4) observe each other, as shown in the top panel of Figure 2. In the second period 3 and 5 are globally best informed and observe each other, while all others remain with their initial targets, given the knowledge about perspectives obtained in the first period. The reason why these two switch away from their initial targets is that they were both very poorly informed in the initial period, and thus learned little about the perspectives of their respective targets. This second period observational structure is shown in the bottom panel of Figure 2. We get homophily in both groups followed by heterophily in both.

These examples show that there is very little structure that can be placed on realized observational networks as time elapses, at least in the short run. However, we can rule out extreme

[5] As in the case of Example 1, this example is robust in the sense that there exists a neighborhood of the set of specified expertise levels within which the same observational patters arise.

heterophily in both groups in the second period conditional on homophily in both groups in the first.

Proposition 3. *The probability that both groups exhibit extreme heterophily in the second period is positive only if one group exhibits extreme heterophily in the first period.*

This is very intuitive. Given homophily in the initial period, there is at least one member of each group who does not observe anyone outside her group, and is not the best informed in her own group in either of the first two periods. At least one of these two individuals must belong to a group with a globally best informed person in the second period. They will observe this person, thus precluding extreme heterophily in both groups.

We turn now to the observational patterns that arise in the long run.

5 Long-Run Structures

Over time, each individual i sharpens her understanding of the targets she observes and, to a lesser extent, also her understanding of those who share a culture with these targets. As in Sethi and Yildiz (2016), after a finite number of periods, each potential link j either becomes free—if $\sigma_j^2(i,t)$ has fallen below $\overline{\sigma}^2$—or breaks—if there exists some j' such that $\sigma_{j'}^2(i,t) < \beta(\sigma_j^2(i,t)$. 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 within the set J_i in each period. We next consider what form these long-run expert sets must take, with a focus on the degree of homophily in observational patterns.

For an individual i in group k, we define a (history-dependent) index of homophily as follows:

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

This is the proportion of i's long run experts who belong to i's own group. The index lies in the unit interval and equals zero if i's long run experts all lie outside her group, and equals 1 if they all lie within her group. We say that an individual i is *unbiased* in the long run if $\eta_i = \eta_i^*$

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

An individual who eventually chooses targets based only on their expertise levels will be unbiased.[6] If $\eta_i > \eta_i^*$ then i exhibits *homophily*, and if $\eta_i < \eta_i^*$ then she exhibits *heterophily*. If the index equals 1 then i exhibits extreme homophily, and she exhibits extreme heterophily if the index is 0.

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

Proposition 4. *If $\sigma_0^2 < \overline{\sigma}^2$, then $J_i = N \setminus \{i\}$ for all i almost surely; all individuals are unbiased in the long run.*

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

If, instead, $\sigma_0^2 > \overline{\sigma}^2$, then the long run levels of homophily depend on the correlation in perspectives.

Proposition 5. *Suppose $\sigma_0^2 > \overline{\sigma}^2$. Then there is a positive probability that each individual $i \in N$ exhibits extreme homophily. If, in addition, $\rho_k > \overline{\rho}$ for group k, then all individuals in k exhibit extreme homophily almost surely.*

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

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

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

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

$$\phi\left(m, \rho\right) = \frac{1 - \rho/\left(m\rho + 1\right)}{1 - \rho} \geq 1. \tag{12}$$

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. Note that this factor is increasing in both m_k and ρ, as one might expect.

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 $m_{ik} = n_k$. This happens when

$$\sigma_0^2/\overline{\sigma}^2 < \phi(m_{ik}, \rho_k) = \frac{(m_{ik} - 1)\rho_k + 1}{(m_{ik}\rho_k + 1)(1 - \rho_k)}. \tag{13}$$

This leads us to:

Proposition 6. *Consider any $i \in N$, any group k, and any $m \geq 1$. If $\sigma_0^2/\overline{\sigma}^2 < \phi(m, \rho_k)$, then, almost surely,*

$$m_{ik} = |(J_i \cup \{i\}) \cap N_k| < m \text{ or } m_{ik} = |(J_i \cup \{i\}) \cap N_k| = n_k.$$

This result states that the set of individuals in group k whom i consults infinitely often must either fall below some threshold (that depends on the correlation in group k perspectives), or must constitute the entire group.

Note that $1/\phi(1, \rho_k) = (1 - \rho_k^2)$, and this gives us a very simple and intuitive condition under which i either exhibits extreme homophily or completely unbiased behavior. Suppose $g(i) = k$. If $\sigma_0^2 < \overline{\sigma}^2/(1 - \rho_k^2)$ then i must link to all in-group members. This follows from Proposition 6 and the fact that $J_i \cup \{i\} \cap N_k$ contains at least one member, i herself. Now suppose that $g(i) \notin k$, and $\sigma_0^2 < \overline{\sigma}^2/(1 - \rho_k^2)$. Then i must link to all or none of those in group k. Taken together, this means that i exhibits either extreme homophily in the long run, or complete unbiasedness, in the sense that she simply observes the individual who is globally best informed.

Corollary 1. *For any $i \in N$ and any group k, if $\sigma_0^2(1 - \rho_k^2) < \overline{\sigma}^2$, then*

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

In particular, if $\sigma_0^2(1 - \min\{\rho_1^2, \rho_2^2\}) < \overline{\sigma}^2$, then, in the long run, each individual either exhibits extreme homophily or is unbiased.

For $\sigma_0^2 > \overline{\sigma}^2$, define

$$\overline{m}\left(\sigma_0^2, \rho\right) = \frac{1}{1 - (1 - \rho)\sigma_0^2/\overline{\sigma}^2} - 1/\rho = \frac{1 - \rho}{\rho} \frac{\sigma_0^2 - \overline{\sigma}^2}{\overline{\sigma}^2 - (1 - \rho)\sigma_0^2}, \tag{14}$$

which solves the equation $\sigma_0^2/\phi(m, \rho) = \overline{\sigma}^2$. Since $\phi(m, \rho) < 1/(1 - \rho)$ by (13), $\overline{m}\left(\sigma_0^2, \rho\right)$ is infinite when $\overline{\sigma}^2/\sigma_0^2 \leq 1 - \rho$; it is finite otherwise. When $m_{ik} > \overline{m}\left(\sigma_0^2, \rho_k\right)$, we have $\sigma_0^2/\phi(m_{ik}, \rho_k) < \overline{\sigma}^2$.

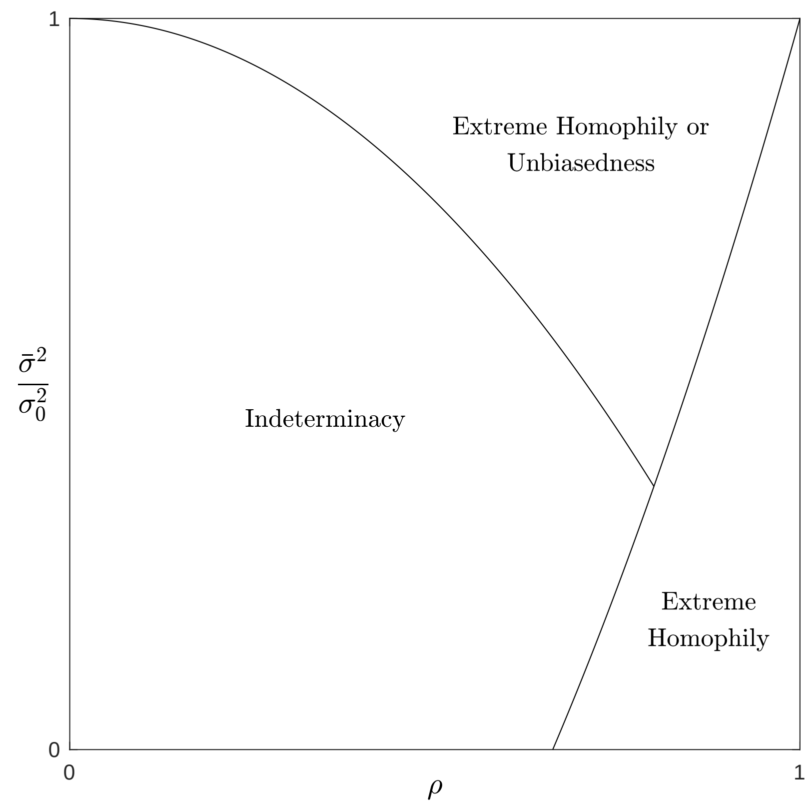

Figure 3: Regions of parameter space giving rise to homophily, unbiasedness, and indeterminacy.

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.

Corollary 2. *For any individual i from group k, in the long run, i either links to everyone in her group ($J_i = N_k \backslash \{i\}$) or links to at most $\overline{m}\left(\sigma_0^2, \rho_k\right) - 1$ of them ($|J_i| \leq \overline{m}\left(\sigma_0^2, \rho_k\right) - 1$); she either links to everyone in the other group k', or links to at most $\overline{m}\left(\sigma_0^2, \rho_{k'}\right)$ of them.*

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

Figure 3 shows regions of the parameter space under which various long run structures arise. When $\rho_k > \overline{\rho}$, we have extreme homophily and all individuals in k observe only in-group members. In the region defined by $\overline{\sigma}^2/\sigma_0^2 > 1 - \rho_k^2$, individuals belonging to k observe all others in their own group, and either all or none in the other group. If $\overline{\sigma}^2/\sigma_0^2 < 1-\rho_k^2$ and $\rho < \overline{\rho}$, more complex patterns of observation can arise, and there is considerable indeterminacy in outcomes. Here individuals may link to some members of a group while ignoring others, and there may be considerable within-group heterogeneity in behavior.

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 3. *Suppose* $n_1 = 3, n_2 = 6, \rho_1 = 2/3, \rho_2 = 1/4, \underline{\tau}^2 = 1/2, \overline{\tau}^2 = 1,$ *and* $\sigma_0^2 = 3$. *Then* $\overline{\sigma}^2 = 2$, *so* $\overline{\sigma}^2/\sigma_0^2 = 2/3$. *In this case* $1-\rho_1^2 = 5/9 < 2/3$. *Hence 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, it can be verified that* $\sigma_0^2/\overline{\sigma}^2 > \phi(m, \rho_2)$ *for all* $m = 1, ...6$. *Hence, for all* $i \in N$, *the number of long-run links to members of group 2 is unconstrained.*

Figure 4 shows the long-run structures that arise in this example for a particular realization of expertise levels. All links are resolved (either broken or free) 38 periods have elapsed. The figure shows the long-run expert sets for each of the nine agents in a separate cell, with colors indicating group membership and a black boundary identifying the subject or observer in each cell. Consistent with the results, each of the three members of group 1 link to the other two infinitely often, as shown in the top row. They differ only with respect to their links to the other group, which range from none to three. Under the parameter specifications in the example, group 2 individuals must link to all or none of those in group 1, and this is also seen in the figure: three link to all and three to none.

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

6 Large Groups

Our analysis to this point concerns groups of arbitrary size. Somewhat more can be said for groups that are sufficiently large.

One implication of the results obtained above is that when $\overline{m}$ is finite and a group is sufficiently large, repeatedly observing the opinions of a small fraction of group members is enough to learn the perspectives of all others in the group to a high degree, even if they have not been directly observed. As a result, these individuals will eventually come to be observed, whenever they happen to be substantially better informed than others in the population. This leads to the following bang-bang result, which holds under a wider range of parameter values than in the small group case:

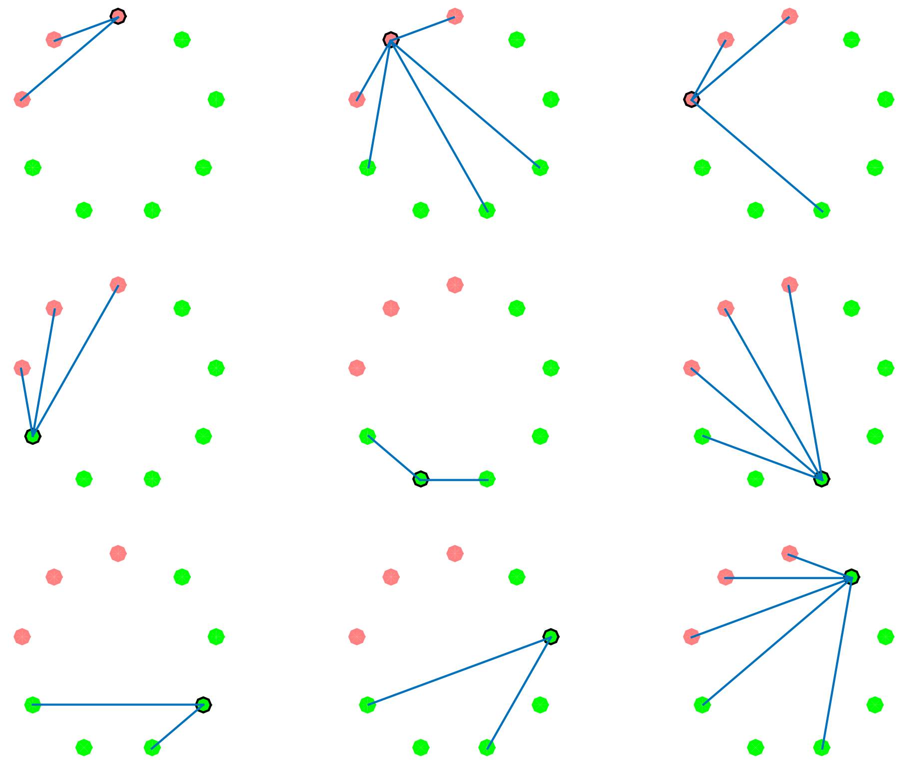

Figure 4: Free links in the long run for nine agents in two groups. Each cells is the observational network for one agent, depicted with a black boundary, and links are to the agent's long-run targets.

Corollary 3. *For any $i \in N$, any group k with $\sigma_0^2(1-\rho_k) < \overline{\sigma}^2$, and every $\varepsilon > 0$, there exists $\overline{n}$ such that, if $n_k > \overline{n}$, then either $m_{ik} < \varepsilon n_k$ 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.

We next present a lower bound on the probability that all individuals in a given group will exhibit homophily in the long run. As the group size gets large, this probability approaches 1 under any $\rho_k > 0$. There are many such bounds. We start with one that can be easily derived using the techniques in Sethi and Yildiz (2016). For simplicity, we also use a binary expertise distribution:

$$\tau_{it}^2 = \begin{cases} \underline{\tau}^2 & \text{with probability } q, \\ \overline{\tau}^2 & \text{with probability } 1-q. \end{cases} \tag{15}$$

Consider any individual i and period t. Let l denote i's target $\lambda_t(i)$ in this period, and note from (7) that

$$1/\sigma_l^2(i,t+1) = 1/\sigma_l^2(i,t) + 1/\alpha\left(\tau_{it}^2, \tau_{lt}^2\right) \geq 1/\sigma_l^2(i,t) + 1/\alpha\left(\overline{\tau}^2, \underline{\tau}^2\right). \tag{16}$$

Moreover, $\sigma_j^2(i,0) = \sigma_0^2\left(1-\rho_k^2\right)$ for any j from group $k = g(i)$. Then, as i observes an in-group member 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 (16), one can easily show that the number of repetitions for this to occur is at most $\max\{\lceil\kappa\rceil, 0\}$ where

$$\kappa = \frac{\alpha\left(\overline{\tau}^2, \underline{\tau}^2\right)}{\sigma_0^2}\left[\frac{1}{1-\overline{\rho}^2} - \frac{1}{1-\rho_k^2}\right] = \frac{\alpha\left(\overline{\tau}^2, \underline{\tau}^2\right)}{\sigma_0^2}\frac{\overline{\rho}^2-\rho_k^2}{\left(1-\overline{\rho}^2\right)\left(1-\rho_k^2\right)}. \tag{17}$$

Here, $\overline{\rho}$ is the cutoff for the correlation above which i never links to an out-group member, as defined in (11). If $\rho_k > \overline{\rho}$, we have $\kappa \leq 0$, and extreme homophily with probability 1 (see Proposition 5). 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 $\overline{\rho}$. Note that

$$\kappa < \frac{\alpha\left(\overline{\tau}^2, \underline{\tau}^2\right)}{\sigma_0^2}\frac{\overline{\rho}^2}{1-\overline{\rho}^2} \equiv \overline{\kappa}. \tag{18}$$

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 $(1-q)^{n_k-1} - (1-q)^{n-1}$.

If i has linked to an in-group member j initially, what is the probability that she will link to j again in the next round? Note that she is now most familiar with j, then with other in-group members, and least familiar with out-group members. Hence she will link to j if j has high expertise, which happens with probability q, or nobody has high expertise, which occurs with probability $(1-q)^{n-1}$. 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 she continues to observe only in-group members, the probability that she links to the most familiar in-group member remains $q + (1-q)^{n-1}$, and the probability that she observes an out-group member remains $(1-q)^{n_k-1} - (1-q)^{n-1}$, until the latter probability drops to zero. At this point all links to out-group members break permanently and we have extreme homophily. The following lower bound on the likelihood of extreme homophily is based on the probability of this event.

Proposition 7. *Under (15), for any group k and any $i \in N_k$,*

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

As $n_k \to \infty$, $p^* \to 1$.

As the group size gets large, extreme homophily becomes virtually certain. This is intuitive. Recall from Proposition 2 that the likelihood of extreme homophily in the initial period is increasing in group size; the larger the group the more likely it will be that the group contains one of the globally best-informed individuals. If this happens repeatedly for some initial set of periods, all links to out-group members break and expertise realizations in subsequent periods become irrelevant. The number of needed repetitions may be large, but is finite for any $\rho_k > 0$. The probability of this event can be made arbitrarily close to 1 by increasing group size.

7 Conclusions

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

A natural process of symmetry-breaking, arising from differences across observers in their own quality of information, can give rise to heterogeneity within groups in observation patterns. The extent of this heterogeneity if 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.

Appendix

Proof of Proposition 1. Let j and j' respectively denote individuals with the highest expertise in groups 1 and 2 respectively. If there are multiple such individuals in any group, consider the one with the lowest label. If j and j' have the same level of expertise, then each individual in N_1 will strictly prefer to observe j while each individual in N_2 will strictly prefer j'. The same is true by continuity if the expertise levels of j and j' are sufficiently close. This is a positive probability event, and corresponds to E_3.

Next suppose that $\sigma_0^2 < \overline{\sigma}$. Then

$$\underline{\tau}^2 + \underline{\tau}^4 \sigma_0^2 < \overline{\tau}^2 < \overline{\tau}^2 + \overline{\tau}^4 \sigma_0^2 (1 - \rho^2)$$

for any $\rho \in (0, 1)$. If $\tau_{j0} = \underline{\tau}^2$ and $\tau_{j'0} = \overline{\tau}^2$, then all individuals other than j observe j. The same is true by continuity if τ_{j0} is sufficiently close to $\underline{\tau}^2$, while $\tau_{j'0}$ is sufficiently close to $\overline{\tau}^2$, which is a positive probability event corresponding to E_2. That E_1 can arise with positive probability can be shown by switching the values of τ_{j0} and $\tau_{j'0}$ in this argument.

To see that no other event can occur with positive probability, note that each individual must observe either j or j', since we are considering the first period. Furthermore, all individuals in $N_1 \setminus j$ have the same preference over j and j'. If all choose j' then so does j, since this individual faces a choice between j and the member of N_1 with the second highest expertise. Similarly, if any individual in $N_2 \setminus j'$ chooses j, then so does every member of N_2. This implies that one of the three cases in the proposition must occur.

Now suppose that $\sigma_0^2 < \overline{\sigma}$. If $\rho_2 \geq \overline{\rho}$, then $\sigma_0^2(1 - \rho_2) < \beta(\sigma_0^2)$. Hence if $i \in N_2$, $\Pr(\lambda_0(i)) \in N_1) = 0$ so $\Pr(E_1) = 0$. The same reasoning may be used to prove the claim about E_2. □

Proof of Proposition 2. Fix any k and k', where $k \neq k'$. For $\tau^2 \in [\underline{\tau}^2, \overline{\tau}^2]$ and $\rho_{k'}$, σ_0^2 given, let $z(\tau^2) \in \mathbb{R}$ denote the unique solution to

$$z^2 + z^4 \sigma_0^2 (1 - \rho_{k'}) = \tau^2 + \tau^4 \sigma_0^2. \tag{19}$$

Note that $z(\tau^2)$ is strictly increasing and satisfies $z(\tau^2) > \tau^2$. This function can be interpreted as follows: if the best-informed individual in group N_k has signal variance at most τ^2, and all individuals in $N_{k'}$ have signal variance above $z(\tau^2)$, then all those in $N_{k'}$ will observe the best-informed in N_k. So will all those in N_k other than this best-informed individual, so E_k will occur. The probability of this is

$$\Pr(E_k) = \int_{\underline{\tau}^2}^{\overline{\tau}^2} \left(1 - (1 - F(x))^{n_k}\right) \left(1 - F(z(x))\right)^{n_{k'}} dx. \tag{20}$$

The claims regarding the effect of n_k and $n_{k'}$ follow immediately. Note that an increase in $\rho_{k'}$ raises $z(\cdot)$ from (19) and hence lowers the integrand in (20), completing the proof. □

Proof of Proposition 3. Suppose that both groups exhibit homophily in the first period (so E_3 occurs). At least one of the two groups must contain an individual with the globally lowest signal variance in period 2. Suppose that k is such a group, and let $j \in N_k$ be such an individual. Given E_3 and $n_k \geq 3$, there must be at least one individual in $i \in N_k$ such that $i \neq j$ and $\lambda_0(i) \in N_k$. Let $l \in N_k$ denote $\lambda_0(i)$. We then have

$$\sigma_l^2(i,1) \geq \sigma_j^2(i,1) \geq \sigma_{j'}^2(i,1)$$

for all $j' \notin N_k$. Since j is globally best informed in the second period, we must have either $\lambda_1(i) = j \in N_k$ or $\lambda_1(i) = l \in N_k$. In either case extreme heterophily in N_k in the second period is precluded. $\square$

Proof of Proposition 4. Suppose $\sigma_0^2 < \overline{\sigma}^2$ and consider any period t and any pair of distinct individuals i and j. We show that there is a positive probability that i links to j in period t regardless of the history of expertise realizations prior to t. 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 expertise realizations sufficiently close to these, so i links to j with positive probability in each period, regardless of history. Hence, for each pair of individuals i and j, $j \in J_i$ almost surely, so $J_i = N \setminus \{i\}$ as claimed. $\square$

Proof of Proposition 5. Suppose $\sigma_0^2 > \overline{\sigma}^2$, in which case $\beta(\sigma_0^2)$ is well-defined and finite. It is easily verified that in the initial period, there is a positive probability that for each $i \in N$, $\lambda_0(i) \in g(i)$. 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.) Suppose that the first period network does indeed satisfy $\lambda_0(i) \in g(i)$ for all $i \in N$. 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 $\lambda_0(i) \in g(i)$ for each $i \in N$) 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 to all members outside $g(i)$ 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 $j \in g(i) = k$ 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 $j' \notin g(i)$ is zero in the initial period. It is clearly also zero in all subsequent periods. $\square$

Proof of Proposition 6. 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. $\square$

References

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

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

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

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

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

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

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

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

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

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

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

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, in press, 2016.