ECONOMETRICA

JOURNAL OF THE ECONOMETRIC SOCIETY

An International Society for the Advancement of Economic Theory in its Relation to Statistics and Mathematics

http://www.econometricsociety.org/

Econometrica, Vol. 84, No. 6 (November, 2016), 2029–2069

COMMUNICATION WITH UNKNOWN PERSPECTIVES

RAJIV SETHI
Barnard College, Columbia University, New York, NY 10027, U.S.A.

MUHAMET YILDIZ
MIT, Cambridge, MA 02142, U.S.A.

Econometrica, Vol. 84, No. 6 (November, 2016), 2029–2069

COMMUNICATION WITH UNKNOWN PERSPECTIVES

BY RAJIV SETHI AND MUHAMET YILDIZ[1]

Consider a group of individuals with unobservable *perspectives* (subjective prior beliefs) about a sequence of states. In each period, each individual receives private information about the current state and forms an *opinion* (a posterior belief). She also chooses a target individual and observes the target's opinion. This choice involves a trade-off between *well-informed* targets, whose signals are precise, and *well-understood* targets, whose perspectives are well known. Opinions are informative about the target's perspective, so observed individuals become better understood over time. We identify a simple condition under which long-run behavior is history independent. When this fails, each individual restricts attention to a small set of experts and observes the most informed among these. A broad range of observational patterns can arise with positive probability, including opinion leadership and information segregation. In an application to areas of expertise, we show how these mechanisms generate own field bias and large field dominance.

KEYWORDS: Communication, heterogeneous priors, networks, signal extraction, opinion leadership.

1. INTRODUCTION

THE SOLICITATION AND INTERPRETATION OF OPINIONS PLAYS A CENTRAL ROLE in information gathering. In academic professions, for instance, reviews and recommendation letters are important inputs in graduate admissions, junior hiring, publications in scientific journals, and internal promotions. However, opinions convey not just objective information but also subjective judgments that are not necessarily shared or even fully known by an observer. For example, a reviewer's recommendation might depend on her subjective views and the reference group she has in mind, and the most crucial assessments are often conveyed using ambiguous terms such as excellent or interesting. Hence, as informative signals, opinions are contaminated with two distinct sources of noise, one stemming from the imprecision of the opinion holder's information, and the other from the observer's uncertainty about the subjective perspective of the opinion holder.

In choosing which opinions to observe, one therefore faces a trade-off between *well-informed* sources—with more precise information—and *well-*

[1]For their helpful suggestions, we thank Daron Acemoglu, Sanjeev Goyal, Navin Kartik, Bart Lipman, the editor, four anonymous referees, and participants at the 2014 Econometric Society Latin American Meetings, 2014 Workshop on Economic Theory (FGV), 2014 Time, Uncertainties and Strategies workshop (PSE), and seminars at ASU, University of Arizona, Bocconi, BU, University of Chicago, University of Chile, Collegio Carlo Alberto, Columbia, Harvard, MIT, Microsoft Research Center—New England, Northwestern, and UCL. We also thank the Becker Friedman Institute at the University of Chicago, the Institute for Advanced Study at Princeton, University of Pennsylvania, Princeton, Microsoft Research Center—New England, and Harvard for hospitality and support.

 DOI: 10.3982/ECTA13320

understood sources—with better known perspectives. Here, a person is well-understood by another if the opinion of the former reveals her information to the latter with a high degree of precision. The better one knows a source's perspective, the easier it becomes to extract information from the source's opinion. One may therefore be able to extract more information from the opinion of a less well-informed source if this source is sufficiently well-understood. For example, in choosing reviewers for a promotion case, one may prefer a senior generalist with a long track record of reviews to a young specialist with deep expertise in the specific area, but possibly strong subjective judgments that are unknown to observers. Similarly, in forecasting elections, one might learn more from pollsters whose methodological biases or house effects are well known than from those with larger samples but unknown biases.

This trade-off between being well-informed and being well-understood has some interesting dynamic implications, since the observation of an opinion not only provides a signal about the information that gave rise to it, but also reveals something about the observed individual's perspective. In other words, the process of being observed makes one better understood. This can give rise to complex patterns of linkages over time, even if all individuals are identical to begin with. It is these effects with which the present paper is concerned.

Specifically, we model a finite set of individuals facing a sequence of periods. Corresponding to each period is a distinct, unobserved state. Individuals all believe that the states are independently and identically distributed, but differ with respect to their prior beliefs about the distribution from which these states are drawn. These beliefs, which we call *perspectives*, are themselves unobservable, although each individual holds beliefs about the perspectives of others. In each period, each individual receives a signal that is informative about the current state; the precision of this signal is the individual's *expertise* in that period. The expertise levels are stochastic and their realized values are public information. Individuals update their beliefs on the basis of their signals, resulting in posterior beliefs that we call *opinions*. Each person then chooses a target individual and observes the target's opinion. This choice is made by selecting the target whose opinion reveals the most precise information about the current state.

The observation of an opinion has two effects. First, it makes the observer's belief about the current period state more precise. Second, the observer's belief about the target's perspective itself becomes more precise. Because of the latter effect, the observer develops an attachment to the target, in that the target becomes more likely to be selected again in subsequent periods. Importantly, the level of attachment to previously observed targets depends on the expertise realizations of *both* observer and observed in the period in which the observation occurs. Better informed observers learn more about the perspectives of their targets since they have more precise beliefs about the signal that the target is likely to have received. This gives rise to *symmetry breaking* over time: two observers who select the same target initially will develop different levels of attachment to that individual. Hence they might make different observational choices in subsequent periods, despite the fact that all expertise realizations are public information.

In the long run, an individual may develop so great an attachment to some set of experts that she stops observing all others. Over time, she learns the perspectives of these *long-run experts* to an arbitrarily high level of precision, and eventually chooses among them on the basis of their expertise alone. Due to the symmetry breaking effects, ex ante identical individuals may end up with very different—or even non-overlapping—sets of long-run experts. However, we show that when the precision of initial beliefs about the perspectives of others is above a certain threshold, all individuals become long-run experts, and everyone links to the most informed individual in each period. All effects of path-dependence eventually disappear, and we have *long-run efficiency*.

When the precision of initial beliefs about the perspectives of others is below this threshold, we show that each individual's set of long-run experts is likely to be small, containing only a negligible fraction of all individuals in large populations. The mechanism giving rise to this is the following. In any period, each individual i links to a more familiar expert unless there is a less familiar expert who is substantially better informed. That is, there is a pecking order for potential experts based on i's familiarity with them: the most familiar expert is observed with greatest likelihood, and so on. Hence, if there are already m experts who are more familiar than a potential expert j, then i will link to j only if j is substantially better informed than *each* of these m experts. This is an exponentially long shot event. Therefore, before i chooses to observe any such j, she links to more familiar individuals many times, learning something about them on each occasion, and eventually develops enough attachment to these that she stops observing j permanently.

Under certain conditions, the long-run expert sets of various individuals are not only small but also overlapping. That is, a few individuals emerge as *opinion leaders*, and are observed even when some individuals outside this set are better informed. But, as a consequence of symmetry breaking, a variety of other complex and interesting observational patterns can also arise. For intermediate levels of the precision of initial beliefs about the perspectives of others, we show that *any* given network emerges as the long-run network with positive probability. In this case, the limiting outcome is a static network, with each individual observing the same target in each period, regardless of expertise realizations. Another interesting linkage pattern is *information segregation*: the population is partitioned into subgroups, and individuals observe only those within their own subgroup. In fact, for any given partition of individuals to groups with at least two members, we show that information segregation according to the given partition emerges in the long run with positive probability as long as initial uncertainty about the perspectives of others is neither too high nor too low.

As an application of the model, we consider the case of a principal with a given area of expertise, dealing with a sequence of cases that may lie within or outside this area. We show that principals will tend to consult experts within their own area of expertise even when the case in question lies outside it, a phenomenon we call *own field bias*. We also show that those with expertise in larger fields—in which individual cases are more likely to lie—will be consulted on cases outside their area of expertise with disproportionately high frequency. We call this *large field dominance*.

Our approach to social communication may be contrasted with the literature descended from DeGroot (1974), which deals with the spread of a given amount of information across an exogenously fixed network, and focuses on the possibility of double counting and related inference problems. We believe that in many applications information is relatively short-lived, while the manner in which it is subjectively processed by individuals is enduring. By observing a given person's opinion, one learns about both the short-lived information and the more enduring subjective perspective through which it is filtered. This makes one more inclined to observe the opinions of the person on other issues. This is the environment we explore here, with particular attention to the endogenous formation of social communication networks.

The remainder of the paper is structured as follows. We develop the baseline model in Section 2, and examine the evolution of beliefs and networks in Section 3. The set of networks that can arise in the long run are characterized in Section 4. Section 5 identifies conditions under which various network structures, such as opinion leadership and information segregation, can emerge with positive probability. Bounds on the size of long-run expert sets are obtained in Section 6, and the application to areas of expertise is developed in Section 7. Various extensions and variations of the model are discussed in Section 8. Section 9 reviews related literature, and Section 10 is a conclusion. The Appendix contains omitted proofs.

2. THE MODEL

Consider a population $N = \{1, \dots, n\}$, and a sequence of periods $T = \{0, 1, 2, \dots\}$. In each period $t \in T$, there is an unobservable state $\theta_t \in \mathbb{R}$. All individuals agree that the sequence of states $\theta_1, \theta_2, \dots$ are independently and identically distributed, but they disagree about the distribution from which they are drawn. According to the prior belief of individual i, the states are normally distributed with mean μ_i and variance 1:

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

We refer to the prior mean μ_i as the *perspective* of individual i. This is not directly observable by others, but it is commonly known that the perspectives $\mu_1, \dots, \mu_n$ are independently distributed according to

$$\mu_i \sim N(\overline{\mu}_i, 1/v_0)$$

for some real numbers $\overline{\mu}_1, \ldots, \overline{\mu}_n$ and $v_0 > 0$. This describes the beliefs held by individuals about each others' perspectives prior to the receipt of any information. Note that the precision in beliefs about perspectives is symmetric in the initial period, since v_0 is common to all. This symmetry is broken as individuals learn about perspectives over time, and the revision of these beliefs plays a central role in the analysis to follow.

In each period t, each individual i privately observes an informative signal

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

where $\varepsilon_{it} \sim N(0, 1/\pi_{it})$. The signal precisions π_{it} capture the degree to which any given individual i is well-informed about the state in period t. We shall refer to π_{it} as the *expertise* of individual i regarding the period t state.

We allow expertise levels π_{it} to be random and vary over time. We assume that these are uniformly bounded:

$$a \leq \pi_{it} \leq b$$

everywhere for some positive constants a and b with $a < b$. That is, no individual is ever perfectly informed of the state in any period, but all signals carry at least some information. In addition, we make the following assumption.

ASSUMPTION 1—Full Support: *For every non-empty open subset Π of $[a, b]^n$, there exists $\lambda(\Pi) > 0$ such that the conditional probability that $(\pi_{1t}, \ldots, \pi_{nt}) \in \Pi$ given any history of expertise levels is at least $\lambda(\Pi)$.*

That is, the support of the expertise levels $(\pi_{1t}, \ldots, \pi_{nt})$ remains $[a, b]^n$ at all histories, and the probability of a given open subset is uniformly bounded away from zero. This is more demanding than required for our results; for the most part, it suffices that we have positive probabilities at all corners. Finally, we assume that the expertise levels π_{it} are publicly observable at t.

REMARK 1: Since priors are heterogeneous, each individual has her own subjective beliefs. We use the subscript i to denote the individual whose belief is being considered. For example, we write $\theta_t \sim_i N(\mu_i, 1)$ to indicate that θ_t is normally distributed with mean μ_i according to i. When all individuals share a belief, we drop the subscript. For example, $\varepsilon_{it} \sim N(0, 1/\pi_{it})$ means that all individuals agree that the noise in x_{it} is normally distributed with mean 0 and variance $1/\pi_{it}$. While an individual j does not infer anything about θ_t from the value μ_i, j does update her belief about θ_t upon receiving information about x_{it}.

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

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

where y_{it} is the expected value of θ_t according to i and $1+\pi_{it}$ is the precision of the posterior belief. We refer to y_{it} as individual i's *opinion* at time t. The opinion is computed as

$$y_{it} = \frac{1}{1+\pi_{it}}\mu_i + \frac{\pi_{it}}{1+\pi_{it}}x_{it}. \tag{2}$$

A key concern in this paper is the process by means of which individuals choose targets whose opinions are then observed. We model this choice as follows. In each period t, each individual i chooses one other individual, denoted by $j_{it} \in N$, and observes her opinion $y_{j_{it}t}$ about the current state θ_t. This information is useful because i then chooses an action $\hat{\theta}_{it} \in \mathbb{R}$ in order to minimize

$$E\big[(\hat{\theta}_{it} - \theta_t)^2\big]. \tag{3}$$

This implies that individuals always prefer to observe a more informative signal to a less informative one. We specify the actions and the payoffs only for the sake of concreteness; our analysis is valid so long as the desire to seek out the most informative signal is assumed. (In many applications, this desire may be present even if no action is to be taken.) The timeline of events at each period t is as follows:

1. The levels of expertise $(\pi_{1t}, \dots, \pi_{nt})$ are realized and publicly observed.
2. Each i observes her signal x_{it}, forms her opinion y_{it}, and chooses a target $j_{it} \in N \setminus \{i\}$.
3. Each i observes the opinion $y_{j_{it}t}$ of her target and takes an action $\hat{\theta}_{it}$.

It is convenient to introduce the variable l_{ij}^t which takes the value 1 if $j_{it} = j$ and zero otherwise. That is, l_{ij}^t indicates whether or not i observes or *links to* j in period t. The set of all such links defines a directed graph that describes who listens to whom in any given period. We represent such directed graphs by functions $g : N \to N$ with $g(i) \neq i$ for each $i \in N$ and write G for the set of all such functions.

REMARK 2: We assume that individuals are myopic, do not observe the actions or past targets of others, and do not observe the realization of the state. As discussed in Section 8, our results extend for the most part to the case of forward-looking behavior, as well as delayed observability of states, actions, and past targets of others.

REMARK 3: The inference problems at any two dates t and t' are related because each individual's ex ante expectation of θ_t and $\theta_{t'}$ is the same; this expectation is what we call the individual's perspective. As we show below, any information about the perspective μ_j of an individual j is useful in interpreting j's opinion y_{jt}, and this opinion in turn is informative about j's perspective. Consequently, the choice of target at date t affects the choice of target at any later date t'. In particular, the initial symmetry is broken after individuals choose their first targets, potentially leading to highly asymmetric outcomes.

3. EVOLUTION OF BELIEFS AND NETWORKS

We now describe how a given individual i selects a target j, and what i learns about the state θ_t and j's perspective μ_j from observing the opinion y_{jt}. This determines the network of information flows and the evolution of beliefs over time.

Under our assumptions, the posterior beliefs held by any individual about the perspectives of any other individual will continue to be normally distributed throughout the process of belief revision. Write v_{ij}^t for the precision of the distribution of μ_j according to i at beginning of t. Initially, these precisions are identical: for all $i \neq j$,

$$v_{ij}^0 = v_0. \tag{4}$$

In subsequent periods, the precisions v_{ij}^t depend on the history of realized expertise levels and observational networks. These precisions of beliefs about the perspectives of others are central to our analysis; the expected value of an individual's perspective is irrelevant as far as the target choice decision is concerned. What matters is how well a potential target is understood, not how far their perspective deviates from that of the observer.

3.1. *Interpretation of Opinions and Selection of Targets*

Suppose that i has chosen to observe the opinion y_{jt} of j, knowing that y_{jt} is formed in accordance with (2). Since $x_{jt} = \theta_t + \varepsilon_{jt}$, this observation provides the following signal regarding θ_t:

$$\frac{1+\pi_{jt}}{\pi_{jt}} y_{jt} = \theta_t + \varepsilon_{jt} + \frac{1}{\pi_{jt}}\mu_j.$$

The signal is noisy in two respects. First, the information x_{jt} of j is noisy, with signal variance ε_{jt}. Second, since the opinion y_{jt} depends on j's unobservable perspective μ_j, the signal observed by i has an additional source of noise, reflected in the term μ_j/π_{jt}.

Taken together, the variance of the noise in the signal observed by i is

$$\gamma(\pi_{jt}, v_{ij}^t) \equiv \frac{1}{\pi_{jt}} + \frac{1}{\pi_{jt}^2}\frac{1}{v_{ij}^t}. \tag{5}$$

Here, the first component $1/\pi_{jt}$ comes directly from the noise in the information of j, and is simply the variance of ε_{jt}. It decreases as j becomes better informed. The second component, $1/(\pi_{jt}^2 v_{ij}^t)$, comes from the uncertainty i faces regarding the perspective μ_j of j, and corresponds to the variance of μ_j/π_{jt} (where π_{jt} is public information and hence has zero variance). This component decreases as i becomes better acquainted with the perspective μ_j, that is, as j becomes better understood by i.

The variance γ reveals that in choosing a target j, an individual i has to trade off the noise $1/\pi_{jt}$ in the information of j against the noise $1/(\pi_{jt}^2 v_{ij}^t)$ in i's understanding of j's perspective, normalized by the level of j's expertise. The trade-off is between targets who are *well-informed* and those who are *well-understood*.

Since i seeks to observe the most informative opinion, she chooses a target for whom the variance γ is lowest. For completeness, we assume that ties are broken in favor of the individual with the smallest label:

$$j_{it} = \min\Big\{\arg\min_{j\neq i} \gamma\big(\pi_{jt}, v_{ij}^t\big)\Big\}. \tag{6}$$

Note that j_{it} has two determinants: the current expertise levels π_{jt} and the precision v_{ij}^t of beliefs regarding the perspectives of others. While π_{jt} is randomly drawn from an exogenously given distribution, v_{ij}^t is endogenous and depends on the sequence of prior target choices, which in turn depends on previously realized levels of expertise.

3.2. *Evolution of Beliefs*

We now describe how the beliefs v_{ij}^t are revised over time. In particular, we show that the belief of an observer about the perspective of her target becomes more precise once the opinion of the latter has been observed, and that the strength of this effect depends systematically on the realized expertise levels of both observer and observed.

Suppose that $j_{it} = j$, so i observes y_{jt}. Recall that j has previously observed x_{jt} and updated her belief about the period t state in accordance with (1)–(2). Hence, observation of y_{jt} by i provides the following signal about μ_j:

$$(1+\pi_{jt})y_{jt} = \mu_j + \pi_{jt}\theta_t + \pi_{jt}\varepsilon_{jt}.$$

The signal contains an additive noise term $\pi_{jt}\theta_t + \pi_{jt}\varepsilon_{jt}$, the variance of which is

$$\pi_{jt}^2\left(\frac{1}{1+\pi_{it}} + \frac{1}{\pi_{jt}}\right).$$

This variance depends on the expertise of the *observer* as well as that of the target, through the observer's uncertainty about θ_t. Accordingly, the precision of the signal is $\Delta(\pi_{it}, \pi_{jt})$, defined as

$$\Delta(\pi_{it}, \pi_{jt}) = \frac{1 + \pi_{it}}{\pi_{jt}(1 + \pi_{it} + \pi_{jt})}. \tag{7}$$

Hence we obtain

$$v_{ij}^{t+1} = \begin{cases} v_{ij}^t + \Delta(\pi_{it}, \pi_{jt}) & \text{if } j_{it} = j, \\ v_{ij}^t & \text{if } j_{it} \neq j, \end{cases} \tag{8}$$

where we are using the fact that if $j_{it} \neq j$, then i receives no signal of j's perspective, and so her belief about μ_j remains unchanged. This leads to the following closed-form solution:

$$v_{ij}^{t+1} = v_0 + \sum_{s=0}^{t} \Delta(\pi_{is}, \pi_{js}) l_{ij}^s. \tag{9}$$

Each time i observes j, her beliefs about j's perspective become more precise. But, by (7), the increase $\Delta(\pi_{it}, \pi_{jt})$ in precision depends on the specific realizations of π_{it} and π_{jt} in the period of observation, in accordance with the following.

LEMMA 1: *$\Delta(\pi_{it}, \pi_{jt})$ is strictly increasing in π_{it} and strictly decreasing in π_{jt}. Hence*,

$$\underline{\Delta} \leq \Delta(\pi_{it}, \pi_{jt}) \leq \overline{\Delta},$$

where $\underline{\Delta} \equiv \Delta(a, b) > 0$ and $\overline{\Delta} \equiv \Delta(b, a)$.

In particular, 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.

The fact that individuals with different expertise levels learn about the perspective of a common target to different degrees can result in *symmetry breaking*, as the following example illustrates. Suppose that $n = 4$, and $\pi_{1t} > \pi_{2t} > \pi_{4t} > \pi_{3t}$ at $t = 0$. Then individual 1 links to 2 and all the others link to 1. The resulting graph is shown in the left panel of Figure 1. In the initial period, individuals 2, 3, and 4 all learn something about the perspective of individual 1, but those who are better informed about the state learn more: $v_{21}^2 > v_{41}^2 > v_{31}^2$. Now

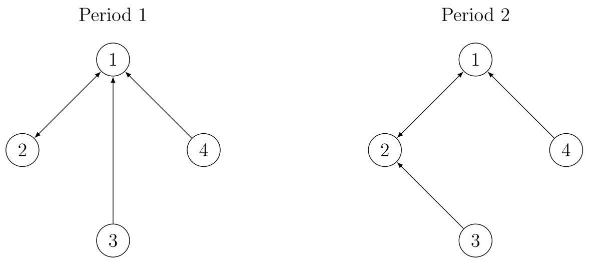

FIGURE 1.—Asymmetric effects of first period observations on second period links.

consider period $t = 1$, and suppose that this time $\pi_{2t} > \pi_{1t} > \pi_{4t} > \pi_{3t}$. There is clearly no change in the links chosen by individuals 1 and 2, who remain the two best informed individuals. But there is an open set of expertise realizations for which individuals 3 and 4 choose different targets: 3 switches to the best informed individual while 4 links to her previous target. This outcome is shown in the right panel of Figure 1.

In this example, the difference between the expertise levels of 1 and 2 in the second period is large enough to overcome the attachment of 3 to 1, but not large enough to overcome the stronger attachment of individual 4, who was more precisely informed of the state in the initial period, and hence learned more about the perspective of her initial target. Hence two individuals with a common observational history can start to make different choices over time.

3.3. *Network Dynamics*

Given the precisions v_{ij}^t at the start of period t, and the realizations of the levels of expertise π_{it}, the links chosen by each individual in period t are given by (6). This then determines the precisions v_{ij}^{t+1} at the start of the subsequent period in accordance with (8), with initial precision $v_{ij}^0 = v_0$.

For any period t, let $h_t := (v_{ij}^{t'})_{t'<t}$ denote the history of precisions of beliefs (regarding perspectives) up to the start of period t; h_0 denotes the initial empty history. Observe that for $t \geq 1$, h_t also implicitly contains information about all past links. The target choice $j_{it}(h_t, \pi_t)$ in period t is a function of h_t and the realized values of expertise levels π_{jt}. Hence, h_t induces a probability distribution on all subsequent links.

We say that the link ij is *active* in period t if $j_{it} = j$. Given any history h_t, we say that the link ij is *broken* in period t if, conditional on h_t, the probability of $j_{it} = j$ is zero. It is easily verified that under Assumption 1, if a link is broken in period t, then it is broken in all subsequent periods. This follows from the fact that the precisions v_{ij}^t are non-decreasing over time, and v_{ij} increases in period t if and only if $j_{it} = j$. Finally, we say that a link ij is *free* in period t conditional on history h_t if the probability that it will be broken in this or any subsequent period is zero conditional on h_t. If a link ij is free at time t, there is a positive probability that $j_{is} = j$ for all $s \geq t$.

We next identify conditions under which a link breaks or becomes free. Define a threshold

$$\overline{v} = \frac{a}{b(b-a)}$$

for the precision v_{ij} of an individual's belief about another individual's perspective. Note that $\overline{v}$ satisfies the indifference condition

$$\gamma(a, \infty) = \gamma(b, \overline{v})$$

between a minimally informed individual whose perspective is known and a maximally informed individual whose perspective is uncertain with precision $\overline{v}$. Define also the function $\beta : (0, \overline{v}) \to \mathbb{R}_+$, by setting

$$\beta(v) = \frac{b^2}{a^2}\left(\frac{1}{v} - \frac{1}{\overline{v}}\right)^{-1}.$$

This satisfies the indifference condition

$$\gamma\big(a, \beta(v)\big) = \gamma(b, v)$$

between a maximally informed individual whose perspective is uncertain with precision v and a minimally informed individual whose perspective is uncertain with precision $\beta(v)$. When $v^t_{ik} > \beta(v^t_{ij})$ for some k, individual i never links to j because the variance $\gamma(\pi_{kt}, v^t_{ik})$ of the information from k is always lower than the variance $\gamma(\pi_{jt}, v^t_{ij})$ of the information from j. Since v^t_{ij} remains constant and v^t_{ik} cannot decrease, i never links to j thereafter, that is, the link ij is broken. Conversely, if $v^t_{ik} < \beta(v^t_{ij})$ for all k, i links to j when j is sufficiently well-informed and all others are sufficiently poorly informed.

When $v^t_{ij}(h_t) > \beta(v^t_{ik}(h_t))$ for all $k \in N \setminus \{i, j\}$, all links ik are broken, so i links to j in all subsequent periods and ij is free. Moreover, assuming that the support of π_t remains $[a, b]^n$ throughout, when $v_{ij} > \overline{v}$, i links to j with positive probability in each period, and each such link causes v_{ij} to increase further. Hence the probability that i links to j remains positive perpetually, so ij is free. Conversely, in all remaining cases, there is a positive probability that i will link to some other node k repeatedly until v_{ik} exceeds $\beta(v^t_{ij}(h_t))$, resulting in the link ij being broken. (This happens when i links to k at least $(\beta(v^t_{ij}(h_t)) - v^t_{ik}(h_t))/\underline{\Delta}$ times in a row.) Note that along every infinite history, every link eventually either breaks or becomes free.

Define the cutoff $\tilde{v} \in (0, \overline{v})$ as the unique solution to the equation

$$\beta(\tilde{v}) - \tilde{v} = \underline{\Delta}. \tag{10}$$

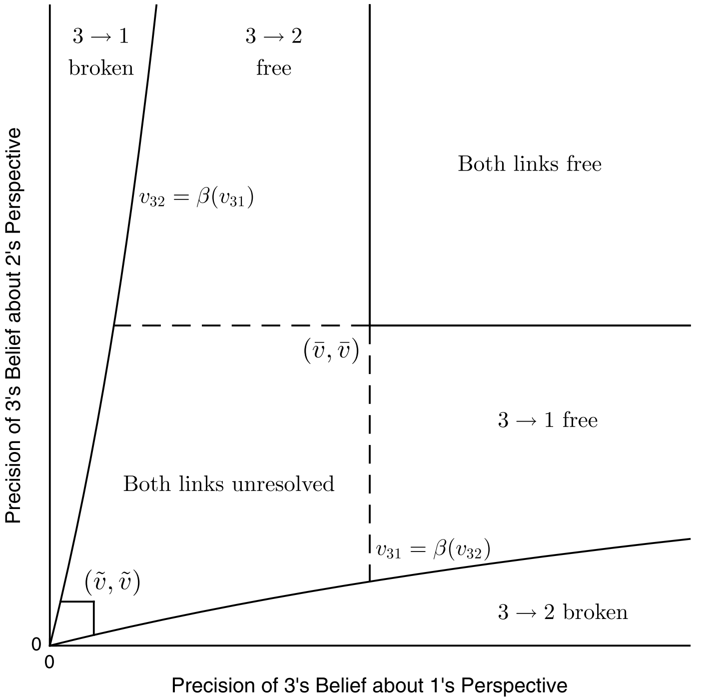

FIGURE 2.—Regions of state space with broken and free links.

Note that $\beta(v) - v$ is increasing, so if the initial precision v_0 (of beliefs about the perspectives of others) is below $\tilde{v}$, then each individual will link in all periods to their first target. This is because if $v_0 < \tilde{v}$ and i observes k initially, then $v_{ik}^1 \geq v_0 + \underline{\Delta}$, and hence $v_{ik}^1 > \beta(v_0) = \beta(v_{ij})$ for all $j \neq k$. All links except those that form initially break, and the initial observational network is persistent.[2]

To illustrate these ideas, consider a simple example with $N = \{1, 2, 3\}$. Figure 2 plots regions of the state space in which the links $3 \to 1$ and $3 \to 2$ are broken, free, or unresolved for various values of v_{31} and v_{32} (the precisions of individual 3's beliefs about the perspectives of 1 and 2, respectively). The figure is based on parameter values $a = 1$ and $b = 2$, which imply $\overline{v} = 0.5$. In the orthant above $(\overline{v}, \overline{v})$, links to both nodes are free. Individual 3 links to each of

[2]Note that the thresholds $\overline{v}$ and $\tilde{v}$ both depend on the support $[a, b]$ from which expertise realizations are drawn, though we suppress this dependence for notational simplicity. We are assuming $a < b$ throughout, but it is useful to briefly consider the limiting case of constant expertise ($a = b$). In this case $\overline{v} = \infty$, so no link is free to begin with, no matter how great the initial precision in beliefs about perspectives happens to be. Moreover, $\beta(v) = v$, so (10) has no solution, and all links break except those that form in the initial period. The resulting outcome is efficient. This clarifies the importance for our analysis of the assumption that issues vary across periods and expertise accordingly varies across individuals.

the other two individuals with positive probability thereafter, eventually learning both their perspectives with arbitrarily high precision. Hence, in the long run, she links with likelihood approaching 1 to whichever of the two is better informed in any given period. In this case, long-run behavior is independent of past realizations.

When $v_{32} > \beta(v_{31})$, the region above the steeper curve in the figure, the link $3 \to 1$ breaks. Individual 3 links only to 2 thereafter, learning her perspective and therefore fully incorporating her information in the long run. But this comes at the expense of failing to link to individual 1 even when the latter is better informed. Along similar lines, in the region below the flatter curve, 3 links only to 1 in the long run.

Now consider the region between the two curves but outside the orthant with vertex at $(\overline{v}, \overline{v})$. Here one or both of the two links remains to be resolved. If $\overline{v} < v_{32} < \beta(v_{31})$, then although the link $3 \to 2$ is free, the link $3 \to 1$ has not been resolved. Depending on subsequent expertise realizations, either both links will become free or $3 \to 1$ will break. Symmetrically, when $\overline{v} < v_{31} < \beta(v_{32})$, the link $3 \to 1$ is free while $3 \to 2$ will either break or become free in some future period. Finally, in the region between the two curves but below the point $(\overline{v}, \overline{v})$, individual 3 may attach to either one of the two nodes (with the other link being broken) or enter the orthant in which both links are free. But when $v_0 < \tilde{v} \cong 0.07$, then any link not formed in the initial period will break right away, so there is no possibility of both links becoming free. Hence, other things equal, the likelihood that all links will become free is increasing in the initial precision in beliefs about perspectives.

4. LONG-RUN EXPERTS

In this section, we show that in the long run, each individual has a history-dependent set of experts, and links with high probability to the most informed among them.

For each infinite history h, define the mapping $J_h : N \to 2^N$ as

$$J_h(i) = \{j \mid j_{it}(h) = j \text{ infinitely often}\} \quad (\forall i \in N). \tag{11}$$

Here $J_h(i)$ is the (nonempty) set of individuals to whom i links infinitely many times along the history h; these are i's *long-run experts*. On this path, eventually, the links ij with $j \in J_h(i)$ become free, and all other links break. Individual i then links exclusively to individuals $j \in J_h(i)$. But each time i links to j, v_{ij}^t increases by at least $\underline{\Delta}$. Hence, given the history h, i knows the perspective of j with arbitrarily high precision after a finite number of periods. This, of course, applies to all individuals $j \in J_h(i)$, so i comes to know all perspectives within $J_h(i)$ very well, and chooses targets from within this set largely on the basis of their expertise levels. This leads to the following characterization.

PROPOSITION 1: *For every $\varepsilon > 0$ and history h, there exists a period $\tau(h)$ such that*

$$j_{it}(h_t, \pi_t) \in \{j \in J_h(i) | \pi_{jt} \geq \pi_{j't} - \varepsilon \, \forall j' \in J_h(i)\} \quad (\forall i \in N, \forall t \geq \tau(h)).$$

PROOF: Observe that there exists $\overline{v}_\varepsilon < \infty$ such that if $v^t_{ij} > \overline{v}_\varepsilon$ and $\pi_{jt} > \pi_{j't} + \varepsilon$, then $j_{it} \neq j'$. By Lemma 1, for every $i, j \in N$ with $j \in J_h(i)$, we have $v^t_{ij}(h_t) \to \infty$. (This follows from the fact that i observes each $j \in J_h(i)$ infinitely often along h.) Hence, there exists $\tau_{ij\varepsilon}(h)$ such that $v^t_{ij}(h_t) > \overline{v}_\varepsilon$ whenever $t \geq \tau_{ij\varepsilon}(h)$. Since N is finite, we can set $\tau(h) = \max\{\overline{t}(h), \max_{i \in N, j \in J_h(i)} \tau_{ij\varepsilon}(h)\}$, where $\overline{t}(h) = \max\{t | j_{it} \notin J_h(i) \text{ for some } i\}$ is the last time a transient link occurs along h. Then, for any $t > \tau(h)$, we have $j_{it}(h_t, \pi_t) \in J_h(i)$ because $t > \overline{t}(h)$, and $\pi_{j_{it}(h_t,\pi_t)t} \geq \pi_{jt} - \varepsilon$ for all $j \in J_h(i)$ because $v^t_{ij} > \overline{v}_\varepsilon$—as claimed. *Q.E.D.*

This result establishes that for any given history of expertise realizations and any $\varepsilon > 0$, there exists some period τ after which each individual i's target has expertise within ε of the best informed individual *among* her long-run experts $J_h(i)$. There may, of course, be better informed individuals outside $J_h(i)$ to which i does not link. The requirement that all individuals simultaneously link to the best informed among their long-run experts sharply restricts the set of possible graphs. For example, in the long run, if two individuals i and i' each link to both j and j', then i cannot link to j in a period in which i' links to j'.

In the Supplemental Material (Sethi and Yildiz (2016b)), we show that when expertise levels $(\pi_{1t}, \dots, \pi_{nt})$ are serially i.i.d., the long-run graphs are also serially i.i.d. with a history-dependent long-run distribution. The long-run distribution is revealed at a finite, history-dependent time τ, in that $J_h = J_{h'}$ for continuations h and h' of h_τ with probability 1. Furthermore, if it has been revealed at a history h_t that the set of long-run experts is J, then for all $\varepsilon > 0$, there exists $t^* > t$ such that

$$P\Big(j_{it'} \in \arg\max_{j \in J(i)} \pi_{jt'} \big| h_t\Big) > 1 - \varepsilon \tag{12}$$

for all $t' > t^*$ and $i \in N$. That is, given ε arbitrarily small, after a known finite time t^*, everyone links to her best informed long-run expert with arbitrarily high probability $1 - \varepsilon$.

Since we have abstracted away from strategic concerns, the outcome in our model is necessarily ex ante optimal, maximizing the payoff of the myopic self at $t = 0$. This, of course, does not mean that future selves do not regret the choices made by earlier ones. Indeed, if expertise levels $(\pi_{1t}, \dots, \pi_{nt})$ are serially i.i.d., at any history h, the expected payoff at the start of each period t converges to

$$u_{\infty,i,h} = -E\left[\frac{1}{1 + \pi_i + \max_{j \in J_h(i)} \pi_j}\right].$$

We call $u_{\infty,i,h}$ the *long-run payoff* of i at history h. This payoff is an increasing function of the cardinality of the set $J_h(i)$ of long-run experts, and is maximized at $J_h(i) = N \setminus \{i\}$. The outcome here is constrained efficient when the social planner has the same myopic time preference as the agents in the model. However, a social planner who places weight on the payoffs of future selves would choose differently, and may even choose a sequence of linkages that maximizes the long-run payoff.[3]

5. LONG-RUN NETWORKS

We have established that, in the long run, individuals restrict attention to a history-dependent and individualized set J_h of long-run experts and seek the opinion of the best-informed among these. We next describe several observational patterns that might result, each corresponding to a specific mapping J_h, and characterize the parameter values (a, b, v_0) under which each of these arises with probability 1 or with positive probability. We start by describing these patterns.

Long-Run Efficiency. We say that long-run efficiency obtains at history h if

$$J_h(i) = N \setminus \{i\} \quad (i \in N).$$

This outcome maximizes the long-run payoff.

Static Networks. We say that the static network $g \in G$ emerges at history h if

$$J_h(i) = \{g(i)\} \quad (i \in N).$$

That is, independent of expertise levels, each individual i links to $g(i)$, the target that graph g assigns to her.

Extreme Opinion Leadership. We say that extreme opinion leadership emerges at history h if there exist individuals i_1 and i_2 such that

$$J_h(i_1) = \{i_2\} \quad \text{and} \quad J_h(i) = \{i_1\} \quad (\forall i \neq i_1).$$

That is, all individuals link to a specific individual i_1, who links to i_2, regardless of expertise realizations. When players are ex ante identical, extreme opinion leadership minimizes the long-run payoff, although it may not be the worst possible situation in asymmetric environments—for example, if i_1 is better informed than others in expectation.

[3]For a version of our model with forward-looking agents, see our working paper (Sethi and Yildiz (2016a)), and the discussion in Section 8.1 below. In this case, individuals may sacrifice current payoffs to learn about the perspectives of unfamiliar targets. The resulting outcome can deviate substantially from long-run efficiency but remains constrained efficient as long as the planner discounts the welfare of future selves at the same rate as the agents in the model.

Information Segregation. We say that *segregation over a partition* $\{S_1, S_2, \dots, S_m\}$ of N emerges at history h if

$$J_h(i) \subset S_k \quad (i \in S_k, \forall k).$$

Under information segregation, clusters emerge in which individuals within a cluster link only to others within the same cluster in the long run. In this case, there may even be a limited form of long-run efficiency within clusters, so that individuals tend to link to the best informed in their own group, but avoid linkages that cross group boundaries.

These patterns are clearly not exhaustive. For example, a weaker form of opinion leadership can arise in which some subset of individuals are observed with high frequency even when their levels of expertise are known to be low, while others are never observed.

The following result identifies conditions under which various long-run outcomes can arise:

PROPOSITION 2: *Under Assumption 1, for any $v_0 \notin \{\tilde{v}, \overline{v}\}$, the following are true*:

(a) *Long-run efficiency obtains with probability 1 if and only if $v_0 > \overline{v}$.*

(b) *Extreme opinion leadership emerges with positive probability if and only if $v_0 < \overline{v}$, and with probability 1 if and only if $v_0 < \tilde{v}$.*

(c) *For any partition $\{S_1, S_2, \dots, S_m\}$ of N such that each S_k has at least two elements, there is segregation over $\{S_1, S_2, \dots, S_m\}$ with positive probability if and only if $v_0 \in (\tilde{v}, \overline{v} - \underline{\Delta})$.*

(d) *Assume that $v_0 < \overline{v} - \Delta(b, b)$ and suppose that there exists $\pi \in (a, b)$ such that $\gamma(\pi, v_0) < \gamma(a, v_0 + \Delta(\pi, b))$ and $\gamma(b, v_0) < \gamma(\pi, v_0 + \Delta(\pi, b))$. Then every $g \in G$ emerges as a static network with positive probability.*

Parts (a) and (b) of this result are highly intuitive. If $v_0 > \overline{v}$, then all links are free to begin with so long-run efficiency is ensured. If this inequality is reversed, then no link is initially free. This implies that extreme opinion leadership can arise with positive probability, for the following reason. Any network that is realized in period t has a positive probability of being realized again in period $t + 1$ because the only links that can possibly break at t are those that are inactive in this period. Hence there is a positive probability that the network that forms initially will also be formed in each of the first s periods for any finite s. For large enough s, all links must eventually break except those that are initially active, resulting in extreme opinion leadership. This proves both that extreme opinion leadership arises with positive probability when $v_0 < \overline{v}$, and that long-run efficiency is not ensured.

Moreover, when $v_0 < \tilde{v}$, we have $v_0 + \underline{\Delta} > \beta(v_0)$ and each individual adheres to their very first target regardless of subsequent expertise realizations. The most informed individual in the first period emerges as the unique information leader and herself links perpetually to the individual who was initially the second best informed. Hence we get extreme opinion leadership with certainty.

Next consider segregation. When $v_0 > \overline{v} - \underline{\Delta}$, segregation cannot arise because all links to the best informed individual in the first period become free. Neither can segregation arise when $v_0 < \tilde{v}$, since we have extreme opinion leadership with certainty. The strength of Proposition 2(c) lies in showing that not only is $v_0 \in (\tilde{v}, \overline{v} - \underline{\Delta})$ necessary for segregation, it is also *sufficient* for segregation over *any* partition to arise with positive probability.

Proposition 2(d) goes a step further. Under an additional condition, it establishes that *any* static network can arise with positive probability. That is, each individual may be locked into a single, arbitrarily given target in the long run. The additional condition may be understood as follows. There exists some feasible expertise level π such that: (i) a previously unobserved target with expertise π is strictly preferred to a once-observed target with minimal expertise, provided that the latter had maximal expertise while the observer had expertise π in the period of prior observation, and (ii) a previously unobserved target with maximal expertise is strictly preferred to a once-observed target with expertise π, provided that the latter had expertise π while the observer had maximal expertise in the period of prior observation. This allows us to construct a positive probability event that results in convergence to an arbitrarily given static network.[4]

While the emergence of opinion leadership is intuitive, convergence to a segregated network or an arbitrary static network is less so. Insights from the literature on multi-armed bandits provide some intuition for the case of extreme opinion leadership. In that literature, individuals choose among targets with uncertain value, and learn something about the value of each target they observe. They eventually settle on a single target because the opportunity cost of learning makes further exploration undesirable. There is a positive probability of any given target being selected in this manner, as in the case of extreme opinion leadership here. But note that this very same intuition implies strong restrictions *across* individuals in the kinds of observational networks that can arise. This is because all observers face the same distribution of expertise in the population, and all but one link to the same target in the initial period. The intuition from bandit problems therefore suggests that all but one will settle on the same target in the long run, which rules out segregation and most static networks. Along similar lines, the social learning literature (which we discuss in some detail below) can generate inefficiencies and opinion leadership but not the kinds of asymmetries we find here.

[4]Note that the assumption holds whenever $v_0 > v^*$ where v^* is defined by $\beta(v^*) - v^* = 2\Delta(b, b)$. A sufficient condition for such convergence to occur is accordingly $v_0 \in (v^*, \overline{v} - \Delta(b, b))$, and it is easily verified that this set is nonempty. For instance, if $(a, b) = (1, 2)$, then $(v^*, \overline{v} - \Delta(b, b)) = (0.13, 0.20)$.

6. THE SIZE OF EXPERT SETS

We have focused to this point on long-run outcomes that can or will emerge for various parameter values. In particular, when $v_0 < \overline{v}$, individuals may limit themselves to a small set of potential experts even when individuals outside this set are better informed. This begs the question of how likely such outcomes actually are. Indeed, in proving Proposition 2, we use specific scenarios that arise with positive but possibly very low probability.

The following result identifies bounds on the probability distribution over long-run expert sets, and shows that these are very likely to be small in absolute size. In large populations, therefore, expert sets constitute a negligible fraction of all potential targets.

PROPOSITION 3: *Assume that π_{it} are independently and identically distributed with distribution function F, such that $0 < F(\pi) < 1$ for all $\pi \in (a, b)$. Then, for any $v_0 < \overline{v}$,*

$$\Pr\big(|J_h(i)| \leq m\big) \geq \left(\frac{1 - F(\hat{\pi})}{1 - F(\hat{\pi}) + mF(\hat{\pi})^m}\right)^{\frac{\beta(v_0)-v_0}{\underline{\Delta}}} \equiv p^*(m) \quad (\forall i \in N, \forall m),$$

where $\hat{\pi} = \min\{\pi | \gamma(\pi, v_0 + \underline{\Delta}) \leq \gamma(b, v_0)\} < b$. In particular, for every $\varepsilon > 0$, there exists $\overline{n} < \infty$ such that

$$\Pr\left(\frac{|J_h(i)|}{n-1} \leq \varepsilon\right) > 1 - \varepsilon \quad (\forall i \in N, \forall n > \overline{n}).$$

PROOF: The second part immediately follows from the first because $mF(\hat{\pi})^m \to 0$ as $m \to \infty$ and the lower bound does not depend on n. To prove the first part, we obtain a lower bound on the conditional probability that $|J_h(i)| \leq m$ given that i has linked to exactly m distinct individuals so far; we call these m individuals *insiders* and the rest *outsiders*. This is also a lower bound on the unconditional probability of $|J_h(i)| \leq m$.[5] Now, at any such history h_t, a lower bound for the probability that i links to the j with the highest v_{ij}^t is

$$\big(1 - F(\hat{\pi})\big)/m.$$

To see this, observe that $v_{ij}^t \geq v_{ij'}^t \geq v_0 + \underline{\Delta} > v_{ij''}^t = v_0$ for all insiders j' and outsiders j''. Hence, if $\pi_{jt} > \hat{\pi}$, individual i prefers j to all outsiders j''. Moreover,

[5]For each history h, define $\tau(h)$ as the first time i has linked to m distinct individuals, where $\tau(h)$ may be ∞. Observe that $\Pr(|J_h(i)| \leq m | h_{\tau(h)}) = 1$ if $\tau(h) = \infty$ and $\Pr(|J_h(i)| \leq m | h_{\tau(h)}) \geq p^*(m)$ otherwise.

since $v_{ij}^t \geq v_{ij'}^t$ for all insiders j', the probability that i prefers j to all other insiders is at least $1/m$, and this is also true when we condition on $\pi_{jt} > \hat{\pi}$. Thus, the probability that i prefers j to all other individuals is at least $(1 - F(\hat{\pi}))/m$. Likewise, the probability that i links to an outsider cannot exceed

$$F(\hat{\pi})^m$$

because i links to an insider whenever there is an insider with expertise exceeding $\hat{\pi}$. Therefore, the probability that i links to the best-known insider at the time (i.e., the j' with highest $v_{ij'}^{t'}$ at date t') for k times before ever linking to an outsider is

$$\left(\frac{\big(1 - F(\hat{\pi})\big)/m}{\big(1 - F(\hat{\pi})\big)/m + F(\hat{\pi})^m}\right)^k.$$

Note that the best-known insider may be changing over time since this event allows paths in which i links to lesser-known insiders until we observe k occurrences of linking to the initially best-known insider. Now, at every period t' in which i links to the best-known individual, her familiarity $v_{t'}^* \equiv \max_j v_{ij}^{t'}$ with the latter increases by at least $\underline{\Delta}$. Hence, after k occurrences, we have $v_{t'}^* \geq v_0 + \underline{\Delta} + k\underline{\Delta}$. Therefore, for any integer $k > (\beta(v_0) - v_0)/\underline{\Delta} - 1$, after k occurrences, we have $v_{t'}^* > \beta(v_0)$. Links to all outsiders are accordingly broken, since $v_{ij''}^{t'}$ remains equal to v_0 for all outsiders j'' throughout. *Q.E.D.*

The first part of this result provides a lower bound $p^*(m)$ on the probability $\Pr(|J_h(i)| \leq m)$ that the size of the set of long-run experts does not exceed m, uniformly for all population sizes n. Here, $p^*(m)$ depends on the distribution F of expertise levels and the parameter v_0, and is decreasing in v_0 and $F(\hat{\pi})$. Since $p^*(m)$ approaches 1 as m gets large, and is independent of n, the fraction of individuals in the set of long-run experts becomes arbitrarily small, with arbitrarily high probability, as the population grows large.

The mechanism giving rise to an absolute bound on the expected size of expert sets is the following. Given a history of expertise realizations and observational networks, each individual i faces a ranking of potential experts based on their familiarity to her. If there are m experts who are already more familiar than some potential expert j in this ranking, i will link to j only if the latter is substantially better informed than each of the m individuals who are more familiar. This is an exponentially long shot event. Before i elects to observe any such j, she will link with high probability to more familiar individuals many times, learning more about them on each occasion, until her link to j breaks permanently.

Tighter Bounds. Proposition 3 contains a simple but loose lower bound $p^*(m)$ on the probability that once the set of insiders reaches size m, no outsider is ever observed. Using this, one can obtain a tighter but more complicated lower bound:

$$\Pr\left(\left|J_h(i)\right| \leq m\right) \geq \sum_{k \leq m}\left(p^*(k) \prod_{k' < k}\left(1 - p^*\left(k'\right)\right)\right) \equiv p^{**}(m).$$

To get yet another bound, note that p^* is not monotonic: it decreases up to $1/\log(1/F(\hat{\pi}))$ and increases after that. One can therefore obtain a tighter bound $\overline{p}^*$ by ironing p^*, where $\overline{p}^*(m) = \max_{m' \leq m} p^*(m)$. Then the bounds $\overline{p}^*$ and p^{**} are both (weakly increasing) cumulative distribution functions, and they all first-order stochastically dominate the distribution over $|J_h(i)|$ generated by the model.

The Binomial Case. Consider a binomial distribution of expertise, with $\pi_{it} = b$ with probability q and $\pi_{it} = a$ with probability $1 - q$. In this case, i links to one of the most familiar insiders whenever any such individual has expertise b. The probability of this is at least q, from which we obtain

$$\Pr\left(\left|J_h(i)\right| \leq m\right) \geq \left(\frac{q}{q + (1-q)^m}\right)^{\frac{\beta(v_0) - v_0}{\Delta}} \equiv p_b^*(m).$$

This is a tighter bound than p^*. One can obtain even a tighter bound p_b^{**} by substituting p_b^* for p^* in the definition of p^{**}.

To illustrate, take $b = 2$ and $a = 1$. From (2), an individual with low expertise puts equal weight on her prior and her information, and one with high expertise puts weight $2/3$ on her information and $1/3$ on her prior. Note that $\overline{v} = 0.5$ and $\tilde{v} \cong 0.07$ in this case, and suppose that $v_0 = 0.3$. In Figure 3, we plot simulated values of $|J_h(i)|$, averaged across all individuals i and across 1000 trials, for various values of n as a function of q. We also plot the theoretical upper bound for the expected value of $|J_h(i)|$ obtained from p_b^{**}, which is uniform for all n. As the figure shows, the set $J_h(i)$ is small in absolute terms: when $q \geq 1/2$, the average expert set has at most 4 members in all simulations, and our theoretical bounds imply that the expected value of the number of members cannot exceed 5 no matter how large the population happens to be.

Law of the Few. While Proposition 3 tells us that the size of expert sets is small for each individual, it does not tell us the extent to which these sets overlap. It has been observed that, in practice, most individuals get their information from a small set of experts; Galeotti and Goyal (2010) referred to this as the *law of the few*. In their model, individuals can obtain information either directly from a primary source or indirectly through the observation of others. In an equilibrium model of network formation, they showed that a small group of experts will be the source of information for everyone else. The equilibrium experts in their model are either identical ex ante to those who observe them, or have a cost advantage in the acquisition of primary information.

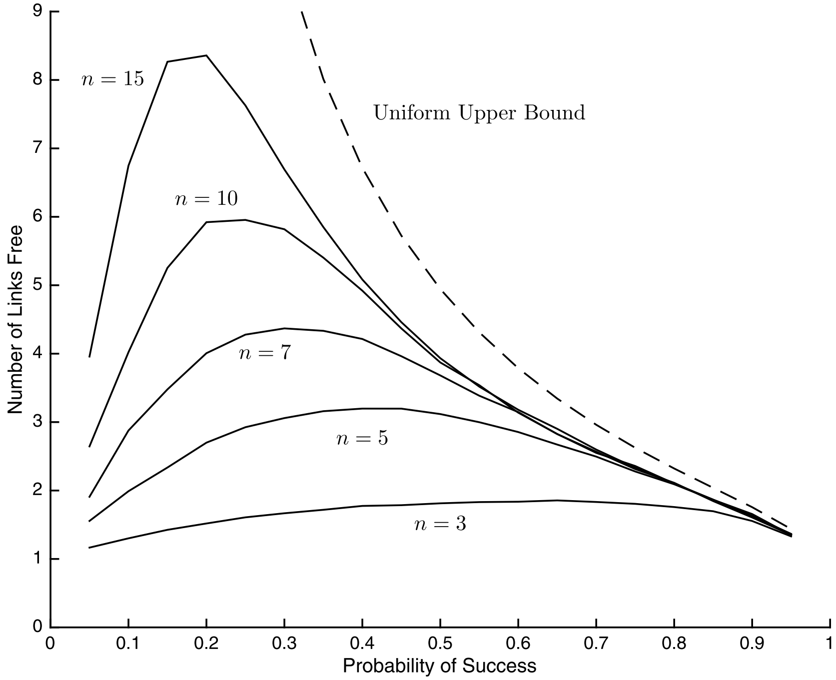

FIGURE 3.—The average number of links per person as a function of q in the binomial example. Solid lines are simulation results; the dashed line is the uniform upper bound p_b^{**}.

Building on Proposition 3, we can show that the law of the few is also predicted by a variant of our model, but through a very different mechanism and with the potential for experts to be consulted even when better information is available elsewhere. We establish this using the following variation.

Two-Sided Model With Observable States. The population N is partitioned into a set of decision makers N_d and a set of potential experts N_e, where the only links allowed are from decision makers to experts. Moreover, the state θ_t becomes publicly observable at the end of period t for each t. (See Section 8.4 for details.)

COROLLARY 1: *Consider the two-sided model with observable states. Assume that π_{it} are independently and identically distributed with distribution function F, such that $0 < F(\pi) < 1$ for all $\pi \in (a, b)$. For every history h, there exists a set $J_h^* \subset N_e$ of experts, such that $J_h(i) = J_h^*$ for every $i \in N_d$. Moreover, for every $\varepsilon > 0$, there exists $\overline{n} < \infty$ such that, whenever $|N_e| > \overline{n}$, we have*

$$\Pr\left(\frac{|J_h^*|}{|N_e|} \leq \varepsilon\right) > 1 - \varepsilon.$$

PROOF: The first part follows from the observability of the state. If $v^t_{ij} = v^t_{i'j}$ for all $j \in N_e$, then $v^{t+1}_{ij} = v^{t+1}_{i'j} = v^t_{ij} + l^t_{ij}/\pi_{jt}$. Since $v^0_{ij} = v^0_{i'j} = v_0$, this shows that $v^t_{ij} = v^t_{i'j}$ throughout, yielding $j_{it} = j_{i't}$ everywhere. Therefore, $J_h(i) = J_h(i')$ for all $i, i' \in N_d$. The second part follows from the first part and Proposition 3, which also holds for the two-sided model with observable states. *Q.E.D.*

Hence there is a history-dependent set J^*_h of core experts who become opinion leaders. Every decision-maker seeks the opinion of the best informed core expert in the long run. Moreover, as the set of potential experts becomes large, the fraction who are actually consulted becomes negligible, and we obtain a law of the few. In contrast with Galeotti and Goyal (2010), however, the group of observed individuals may have poorer information than some who remain unobserved.

Finite-Population Bounds. Proposition 3 provides a uniform bound for all n. For a fixed n, one can find a tighter bound. In the binomial example above, we have

$$\Pr\left(\left|J_h(i)\right| \leq m\right) \geq \left(\frac{q + (1-q)^{n-1}}{q + (1-q)^m}\right)^{\frac{\beta(v_0)-v_0}{\Delta}} \equiv p^*_b(m|n, q).$$

This is because with probability $(1-q)^{n-1}$, all $j \in N \setminus \{i\}$ have low expertise, and i links to her most familiar expert, rather than someone previously unobserved. Note that for $q \in \{0, 1\}$, $p^*_b(m|n, q) = 1$. Hence $|J_h(i)| = 1$ with probability close to 1 when q is close to 0 or 1. For intermediate values of q, the expected size of $J_h(i)$ is greater, leading to a non-monotone relation—as in the simulation results in Figure 3. When n is small and q is not close to 0 and 1, the relative size $|J_h(i)|/(n-1)$ of long-run experts can be large, and we may have long-run efficiency with high probability. Nonetheless, the size of $J_h(i)$ does not grow much as n increases, leading to small sets of experts. This is illustrated in Table I, which reports the estimated probability of $|J_h(i)|$ for $q = 1/2$ in our simulations. The distribution of $|J_h(i)|$ is increasing with n in the sense of first-order stochastic dominance, but the impact of n diminishes quickly at small values of n—the probability of $|J_h(i)| \in \{3, 4, 5\}$ is approximately 0.9 for

TABLE I

FREQUENCY OF A GIVEN SIZE $|J_h|$ AS A FUNCTION OF n (NUMBER OF TRIALS = 10,000)

$n \setminus \|J_h\|$	1	2	3	4	5	6	7	8	$\|J_h\| > 8$
3	0.1842	0.8158							
5	0.0131	0.1714	0.4973	0.3183					
7	0.0035	0.0701	0.3313	0.4182	0.1604	0.0164			
10	0.0017	0.0514	0.2811	0.4137	0.2105	0.0392	0.0024	0.0001	0
15	0.0013	0.0512	0.2730	0.4112	0.2140	0.0453	0.0039	0.0001	0

$n = 7$, 10, and 15. Although we have long-run efficiency ($|J_h(i)| = 2$) with a high probability of 0.82 for $n = 3$, the probability of long-run efficiency drops sharply to 0.32 for $n = 5$ and to less than 0.02 for $n = 7$.

Speed of Convergence. In our model, the set of free links grows slowly at an exponentially decreasing rate while the links are broken quite quickly, resulting in small sets $J_h(i)$ of long-run experts. In the binomial model, it takes

$$E[t_m] = 1 + 1/(1-q) + \cdots + 1/(1-q)^{m-1} = \frac{1-q}{q}\left(\frac{1}{(1-q)^m} - 1\right)$$

rounds for i to select m distinct targets in expectation; it may take substantially longer to get m links free. Note that $E[t_m]$ grows exponentially: $E[t_{10}] = 2^{10} - 1 = 1023$ and $E[t_{20}] = 2^{20} - 1 > 10^6$ for $q = 1/2$. In contrast, it takes at most

$$E\big[t_b^*(v)\big] = \frac{\big(\beta(v) - v_0\big)/\underline{\Delta}}{q}$$

rounds to break a link ij with $v_{ij} = v$ in expectation; we have $\max_{j'} v_{ij'}^t > v$ at $t = t_b^*(v)$. For the parameter values used in our simulations, a link ij becomes free as soon as i first links to j. For $q = 1/2$, $E[t_b^*(v_0)] = 21.6$, and all non-free links are broken at round 22 in expectation.

7. AREAS OF EXPERTISE

The framework developed here can be applied to a number of settings, and we now explore one such application in detail.

Consider a principal who faces a sequence of decisions that we call *cases*, each of which lies in one of two areas of expertise or *fields* 1 and 2. For concreteness, one may think of a journal editor facing a sequence of submissions, or a university administrator facing a sequence of promotion cases; in either scenario, the principal must make a decision based on an assessment of quality. For each case, the principal can consult an outside expert or referee drawn from a pool of potential experts.

All individuals (principals and experts) themselves have expertise in one of the two fields. In addition, all individuals have prior beliefs regarding the quality of each case, and these are drawn independently and identically from a normal distribution with precision v_0 as before. The field to which any given case belongs is observable.

We adopt the convention of using i_1 and i_2 to denote principals with expertise in fields 1 and 2, respectively, and j_1 and j_2 for experts in these respective fields. If the period t case lies in field 1, then a principal i_1 has expertise $\pi_{i_1 t} = b$, while a principal i_2 has expertise $\pi_{i_2 t} = a < b$. If the case lies instead in field 2, these expertise levels are reversed, and we have $\pi_{i_1 t} = a$ and $\pi_{i_2 t} = b$. The same applies to experts: those asked to evaluate a case in their field have expertise b while those whose field is mismatched with that of the case have lower expertise a.[6]

As before, we assume that in any given period, the principal consults the expert whose opinion on the current case is most informative. It is clear that in the initial period, since no expert is better understood than any other, the principal will choose an expert to match the field of the case (regardless of the field to which the principal herself belongs). This follows directly from the fact that $\gamma(b, v_0) < \gamma(a, v_0)$. Furthermore, if there exists a period t in which the fields of the case and the chosen expert differ, then the same expert will also be selected in all subsequent periods.

Hence, along any history of cases h, a principal i selects an expert who is matched to the field of the case until some period $t_i \leq \infty$, and subsequently chooses the same expert regardless of field match. If $t_i(h) = \infty$, then we have long-run efficiency: experts and cases are always matched by field. Otherwise, the principal i attaches to an expert in a specific field, which may or may not match the field in which the principal herself has expertise.

Note that when faced with the *same* history h, the principal i_1 may behave differently from the principal i_2: they may attach to experts in different fields, and may do so at different times, or one may attach while the other does not. But not all events can arise with positive probability. The principals exhibit the following own field bias.

PROPOSITION 4: *Given any history h, if principal i_1 attaches to expert j_2 in period $t_{i_1}(h)$, then principal i_2 must attach to j_2 in some period $t_{i_2}(h) \leq t_{i_1}(h)$.*

PROOF: The result follows from the following claim: given any history h, if principal i_1 consults expert j_2 in any period t, then principal i_2 also consults j_2 in t. We prove this claim by induction. It is clearly true in the first period, since j_2 is consulted by each type of principal if and only if the first case is in field 2. Suppose the claim is true for the first $t-1 \geq 1$ periods, and let η denote the proportion of these periods in which i_1 consults j_2. Then, since $\Delta(a, b) < \Delta(b, b)$, and i_2 consults j_2 at least $\eta(t-1)$ times in the first $t-1$ periods by hypothesis, we obtain

$$v^t_{i_1 j_2} = v_0 + \eta(t-1)\Delta(a, b) < v_0 + \eta(t-1)\Delta(b, b) \leq v^t_{i_2 j_2}. \tag{13}$$

Similarly,

$$v^t_{i_1 j_1} = v_0 + (1-\eta)(t-1)\Delta(b, b) > v_0 + (1-\eta)(t-1)\Delta(a, b) \geq v^t_{i_2 j_1}. \tag{14}$$

[6] Experts in any given field are ex ante identical, though they may have different realized priors over the quality of the cases. Given that a previously consulted expert becomes better understood by a principal, and thus certain to be selected over previously unobserved experts in the same field, nothing essential is lost by assuming that there are just two experts in the pool, one in each field.

If i_1 consults j_2 in period t along h, then it must be because

$$\gamma\left(\pi_{j_2 t}, v_{i_1 j_2}^t\right) \le \gamma\left(\pi_{j_1 t}, v_{i_1 j_1}^t\right).$$

But given (13)–(14), this implies

$$\gamma\left(\pi_{j_2 t}, v_{i_2 j_2}^t\right) < \gamma\left(\pi_{j_1 t}, v_{i_2 j_1}^t\right),$$

so i_2 also consults j_2 in period t. *Q.E.D.*

This result rules out many possibilities. When facing a common history of cases h, if i_1 attaches to j_2, then so must i_2, ruling out the possibility that i_2 attaches to j_1 or attaches to no expert at all (thus matching the field of the expert to that of the case in all periods). This leaves four qualitatively different possibilities: (i) dominance by a field (both principals attach to the same expert), (ii) partial dominance by a field (one principal attaches to an expert in her own field while the other does not attach at all), (iii) segregation (each principal attaches to an expert in her own field), and (iv) long-run efficiency (neither principal attaches to any expert).

Since total or partial dominance can involve either one of the two fields, we have six possible outcomes in all. Each of these can be represented by a function $J : \{i_1, i_2\} \to 2^{\{j_1, j_2\}} \setminus \{\emptyset\}$, where $J(i)$ denotes the set of experts consulted infinitely often by principal i. Each function J corresponds to a distinct event. Table II identifies the six events that can arise with positive probability (and the three that are ruled out by Proposition 4).

Note that segregation can arise despite priors about case quality being independently and identically distributed across principals and experts. This happens because a principal is able to learn faster about the prior beliefs of an expert when both belong to the same field and are evaluating a case within that common field.[7]

TABLE II

POSITIVE PROBABILITY EVENTS

	$J(i_2) = \{j_1\}$	$J(i_2) = \{j_1, j_2\}$	$J(i_2) = \{j_2\}$
$J(i_1) = \{j_1\}$	Dominance by 1	Partial Dominance by 1	Segregation
$J(i_1) = \{j_1, j_2\}$	—	Long-Run Efficiency	Partial Dominance by 2
$J(i_1) = \{j_2\}$	—	—	Dominance by 2

[7]This mechanism is quite different from that driving other models of information homophily. For instance, Baccara and Yariv (2013) considered peer group formation for the purpose of information sharing. In their model, individuals have heterogeneous preferences with respect to the issues they care about and, in equilibrium, groups are characterized by preference homophily. This then implies information homophily, since individuals collect and disseminate information on the issues of greatest concern to them.

Next, we show that experts in larger fields are more likely to rise to dominance, in the sense that their opinions are solicited even for cases on which they lack expertise. In order to express this more precisely, let p_1 denote the (time-invariant) probability that the period t case is in field 1, with $p_2 = 1 - p_1$ being the probability that it is in field 2. Assume without loss of generality that $p_1 \geq p_2$. For each principal i and event J, define

$$q_i(J) = \frac{\left|J(i) \cap \{j_1\}\right|}{\left|J(i)\right|}.$$

This is the frequency with which expert j_1 is consulted by principal i in the long run. For instance, dominance by field 1 corresponds to $q_1 = q_2 = 1$, efficiency to $q_1 = q_2 = 0.5$, segregation to $(q_1, q_2) = (1, 0)$, and partial dominance by field 1 to $(q_1, q_2) = (1, 0.5)$. We can use this to define a partial order on the set of six positive probability events identified above:

$$J \succeq J' \quad \iff \quad \left[\forall i, q_i(J) \geq q_i\left(J'\right)\right].$$

That is, $J \succeq J'$ if and only if *both* principals consult expert j_1 with weakly greater long-run frequency. Note that segregation and long-run efficiency are not comparable, and moving upwards and/or to the left in Table II leads to events that ordered higher:

$$D_1 \succeq PD_1 \succeq S, LRE \succeq PD_2 \succeq D_2. \tag{15}$$

We can also partially order individual histories according to the occurrence of field 1 as follows:

$$h \succeq h' \quad \iff \quad \left[\forall t, \pi_{i_1 t}(h) < \pi_{i_2 t}(h) \Rightarrow \pi_{i_1 t}\left(h'\right) < \pi_{i_2 t}\left(h'\right)\right].$$

That is, $h \succeq h'$ if and only if the case in each period is in field 1 under h whenever it is in field 1 under h'.

Next, we define the probability distribution $P(\cdot|p_1)$ on mappings J, by assigning the probability of $\{h|J_h = J\}$ under p_1 to J for each J. We are interested in the manner in which this distribution varies with p_1. Accordingly, we rank probability distributions on mappings J according to first-order stochastic dominance with respect to the order $\succeq$:

$$P \succeq_{\text{FOSD}} Q \quad \iff \quad \left[\forall J, P\left(\left\{J'|J \succeq J'\right\}\right) \leq Q\left(\left\{J'|J \succeq J'\right\}\right)\right],$$

with strict inequality for some J.

Segregation and efficiency are not ranked in the partial order (15), and for some purposes it is important to distinguish between these. To do this, we define the events $D_1^* = D_1 \cup PD_1 \cup S$ and $D_2^* = D_2 \cup PD_2 \cup S$; the purpose of these definitions will become apparent below. The following proposition formalizes the idea that experts in larger fields are consulted disproportionately often on cases outside their area of expertise:

PROPOSITION 5: *The following are true for all* h, h', p_1, *and* p'_1:

(a) *if* $h \succeq h'$, *then* $J_h \succeq J_{h'}$.

(b) *if* $p_1 > p'_1$, *then* $P(\cdot|p_1) \succeq_{\text{FOSD}} P(\cdot|p'_1)$. *Furthermore*, $P(D_1|p_1) > P(D_1|p'_1)$, $P(D_1^*|p_1) > P(D_1^*|p'_1)$, $P(D_2|p_1) < P(D_2|p'_1)$ *and* $P(D_2^*|p_1) < P(D_2^*|p'_1)$.

(c) $\lim_{p_1 \to 1} P(D_1|p_1) = 1$.

This result establishes that experts in larger fields are more likely to be consulted on cases outside their area of expertise than experts in smaller fields. Here, a larger field is interpreted as one in which an arbitrary case is more likely to lie. The first part establishes this by comparing realized histories, and the second by comparing ex ante probabilities. Specifically, when the probability p_1 rises, cases in field 2 are assigned to field 1 experts with greater likelihood, and field 1 cases are assigned to field 2 experts with smaller likelihood.

This is true regardless of the principal's own field of expertise. To see why, consider first the case of a principal in field 1. She will assign a field 2 case to a field 1 expert in the long run if D_1^* occurs, and a field 1 case to a field 2 expert only if D_2 occurs. Proposition 5(b) states that the first of these events becomes more likely and the second less likely as p_1 rises. Similarly, if the principal belongs to field 2, she will assign a field 2 case to a field 1 expert only if D_1 occurs, and a field 1 case to a field 2 expert if D_2^* occurs. Again, the result tells us that the former event becomes more likely and the latter less likely as p_1 rises.

To summarize, when decision makers, experts, and cases are all associated with specific areas of specialization, the heterogeneity and unobservability of perspectives gives rise to two sharp predictions: *own field bias* (other things equal, principals are more likely to consult experts in their own fields) and *large field dominance* (experts in larger fields are more likely to be consulted on cases outside their area of expertise).

8. EXTENSIONS AND VARIATIONS

In this section, we briefly discuss various extensions and variants of the model. These are explored in detail in the Supplemental Material, and in our working paper (Sethi and Yildiz (2016a)).

8.1. *Forward-Looking Behavior*

In order to explore the trade-off between well-informed and well-understood targets, we have assumed throughout that individuals seek the most informative opinion in each period. But one might expect that forward-looking individuals may sometimes choose to observe an opinion that is relatively uninformative about the current period state, in order to *build* familiarity with a target, in the hope that this might be useful in future periods. The benefit of doing so, relative to the cost, is exponentially decreasing in the number of already familiar targets. It turns out that when the initial precision of beliefs about the perspectives of others is below the threshold $\overline{v}$ for long-run efficiency, the cost of building familiarity with new potential experts quickly exceeds the benefit, resulting in a small set of long-run experts. As in the myopic case, this set contains only a negligible fraction of all individuals in large populations. We establish this formally in our working paper (Sethi and Yildiz (2016a)), using a simplified version of our baseline model in which individuals learn the perspectives of their targets completely after a single observation, and maximize the expected sum of discounted payoffs over an infinite horizon. Like their myopic counterparts, forward-looking individuals restrict attention to a relatively small set of long-run experts, and link to the most informed among them, even when there are better informed individuals outside this set.

8.2. *Shifting Perspectives*

We have assumed throughout that perspectives are fixed: each individual believes that θ_t is i.i.d. with a specific distribution, and does not update her beliefs about this distribution as she observes realizations of θ_t or signals about θ_t—even when the emerging data are highly unlikely under the presumed distribution. This is motivated by our interpretation of a perspective as a stable characteristic of individual cognition that governs the manner in which information about a variety of issues is processed; see Section 9 for more on the existence of such frames of reference. Perspectives in this sense can sometimes be subject to sudden and drastic change, as in the case of ideological conversions, but will not generally be subject to incremental adjustment in the face of evidence.

Nevertheless, it is worth considering the theoretical implications of continuously and gradually changing perspectives. If one views a perspective as a model of the world, perspectives can adapt to incoming data if there is model uncertainty. In the Supplemental Material, we present an extension in which θ_t is an exchangeable process. Individuals update their perspectives as they observe θ_t, while recognizing that others are also doing so. As might be expected, all perspectives converge to the publicly observed empirical frequency in the long run, and all individuals eventually link to the most informed person in the population at large, resulting in long-run efficiency.

While this result is of theoretical interest, it is subject to a couple of caveats. First, when initial beliefs about the distribution of θ_t are firm, perspectives are slow to change and the medium-run behavior of the learning model resembles long-run behavior with fixed perspectives. In particular, our qualitative results apply for the medium run with shifting perspectives when the initial beliefs are sufficiently firm. Second, we show in the Supplemental Material that learning actually strengthens path dependence in early periods: relative to the baseline model, it induces individuals to discount expertise vis-à-vis familiarity with the target. Intuitively, model uncertainty implies a more diffuse prior over the state, which makes expertise realizations that differ from the prior mean less surprising, and makes posteriors less sensitive to private signals. The *effective* expertise level of targets is accordingly lower, and the trade-off between being well-informed and being well-understood is shifted as a result. This effect is eventually overwhelmed by knowledge of the empirical frequency of states, and the fact that all perspectives converge to this empirical distribution.[8]

8.3. *Observability of Targets and Actions*

We have assumed for the most part that an individual's actions and target choices are not observable by others. If one could observe the actions of one's target, as well as their own choice of target, one could infer something about the perspective of the latter. This would then affect subsequent observational choices. However, observing the targets of others (without observing their actions) would be irrelevant for our analysis. This is because individuals do not learn anything from the target choices of others: each individual can compute j_{it} using publicly available data even before j_{it} has been selected.[9] This simplifies the analysis considerably, due to the linear formula for normal variables; see (16) in the Appendix. In a more general model, one may be able to obtain useful information by observing j_{it}. For example, without linearity, $v_{ij}^{t+1} - v_{ij}^{t}$ could depend on y_{jt} for some i with $j_{it} = j$. Since y_{jt} provides information about μ_j, and v_{ij}^{t+1} affects $j_{it'}$ for $t' \geq t + 1$, one could then infer useful information about μ_j from $j_{it'}$ for such t'. The formula (8) would not be true for t' in that case, possibly allowing for other forms of inference at later dates.

8.4. *A Two-Sided Model and Observable States*

All individuals are symmetrically placed in our baseline model, in the sense that they are both observers and potential experts. For some applications, it is more useful to consider a population that is partitioned into two groups: a set

[8]The convergence of all perspectives to the empirical distribution requires that individuals know the exact relation between the distribution of θ_t and the signals they observe. This is quite demanding when individuals observe only signals rather than the state itself. Acemoglu, Chernozhukov, and Yildiz (2016) showed that when individuals learn about the relation between signals and states, the intuition provided by the learning model is fragile. Although individuals manage to learn the frequency of future signals, their asymptotic beliefs about the underlying parameters are highly sensitive to their initial beliefs about the relation between signals and states.

[9]One can prove this inductively as follows. At $t = 1$, one can compute j_{it} from (6) using $(\pi_{1t}, \dots. \pi_{nt})$ and v_0 without observing j_{it}. Suppose now that this is indeed the case for all $t' < t$ for some t, that is, j_{it} does not provide any additional information about μ_i. Then all beliefs about perspectives are given by (8) up to date t. One can see from this formula that each v_{ij}^{t} is a known function of past expertise levels $(\pi_{1t'}, \dots, \pi_{nt'})_{t'<t}$, all of which are publicly observable. That is, one knows v_{ij}^{t} for all distinct $i, j \in N$. Using $(\pi_{1t}, \dots, \pi_{nt})$ and these values, one can then compute j_{it} from (6).

of observers or decision markers who are never themselves observed, and a set of potential experts whose opinions are solicited but who do not themselves seek opinions. We examine this case in the Supplemental Material, obtaining a crisper version of Proposition 2. When $v_0 > \overline{v}$ and $v_0 < \tilde{v}$, we have long-run efficiency and extreme opinion leadership, respectively, as in the baseline model. For intermediate values $v_0 \in (\tilde{v}, \overline{v} - \underline{\Delta})$, each pattern of long-run behavior identified in Proposition 2—including information segregation and convergence to an arbitrary static network—emerges with positive probability.

Another variant of the model allows for states to be publicly observable with some delay. If the delay is zero, the period t state is observed at the end of the period itself; an infinite delay corresponds to our baseline model. This case is also examined in the Supplemental Material. Observability of past states retroactively improves the precision of beliefs about the perspectives of those targets who have been observed at earlier dates, without affecting the precision of beliefs about other individuals, along a given history. Such an improvement only enhances the attachment to previously observed individuals. This does not affect our results concerning any single individual's behavior, such as the characterization of long-run outcomes in Proposition 1. Nor does it affect patterns of behavior that are symmetric on the observer side, such as long-run efficiency and opinion leadership in parts (a) and (b) of Proposition 2. However, observability of past states has a second effect: two individuals with identical observational histories have identical beliefs about the perspectives of all targets observed sufficiently far in the past. This makes asymmetric linkage patterns—such as non-star-shaped static networks and information segregation—less likely to emerge. Nevertheless, with positive delay, private signals do affect target choices, and symmetry breaking remains possible. Our results on information segregation and static networks extend to the case of delayed observability for a sufficiently long delay.

9. RELATED LITERATURE

A key idea underlying our work is that there is some aspect of cognition that is variable across individuals and stable over time, and that affects the manner in which information pertaining to a broad range of issues is filtered. This aspect of cognition is what we have called a *perspective*. In our model, knowledge of others' perspectives changes endogenously through the observation of their opinions. Differences in political ideology, cultural orientation, and even personality attributes can give rise to such stable variability in the manner in which information is interpreted. This is a feature of the cultural theory of perception (Douglas and Wildavsky (1982)) and the related notion of cultural cognition (Kahan and Braman (2006)).

Evidence on persistent and public belief differences that cannot realistically be attributed to informational differences is plentiful. For instance, political ideology correlates quite strongly with beliefs about the religion and birthplace of Barack Obama, the accuracy of election polling data, the reliability of official unemployment statistics, and even perceived changes in local temperatures (Thrush (2009), Pew Research Center for the People & the Press (2008), Plambeck (2012), Voorhees (2012), Goebbert, Jenkins-Smith, Klockow, Nowlin, and Silva (2012)). Since much of the hard evidence pertaining to these issues is in the public domain, it is unlikely that such stark belief differences arise from informational differences alone. In some cases, observable characteristics of individuals (such as racial markers) can be used to infer biases, but this is less easily done with biases arising from different personality types or worldviews.

Our analysis is connected to several stands of literature on observational learning, network formation, and heterogeneous priors. DeGroot (1974) was among the first to examine the spread of information across an exogenous network through heuristic belief updating across multiple rounds. Golub and Jackson (2010) revisited this model, and identified conditions under which such simple learning rules result in full aggregation of distributed information as the network grows large. DeMarzo, Vayanos, and Zweibel (2003) also considered the spread of information across a given network or "listening structure" when individuals are subject to persuasion bias—a failure to allow completely for the fact that two different sources of information may not be independent. Agents with particular network positions in this case can have disproportionate influence. Network architecture also affects the optimality of learning in Bala and Goyal (1998), Gale and Kariv (2003), and Acemoglu, Dahleh, Lobel, and Ozdaglar (2011), where actions rather than beliefs are observed, in the tradition of the early observational learning literature (Banerjee (1992), Bikhchandani, Hirschleifer, and Welch (1992), Smith and Sorensen (2000)).

Key early contributions to the network formation literature include Jackson and Wolinsky (1996) and Bala and Goyal (2000); see Bloch and Dutta (2011) for a survey. Two papers especially relevant for our work are Galeotti and Goyal (2010) and Acemoglu, Bimpikis, and Ozdaglar (2014). Galeotti and Goyal (2010) developed a model to account for the law of the few, which refers to the empirical finding that the population share of individuals who invest in the direct acquisition of information is small relative to the share of those who acquire it indirectly via observation of others, despite minor differences in attributes across the two groups. All individuals are ex ante identical in their model and can choose to acquire information directly, or can choose to form costly links in order to obtain information that others have paid to acquire. All strict Nash equilibria in their baseline model have a core-periphery structure, with all individuals observing those in the core and none linking to those in the periphery. Hence all equilibria are characterized by opinion leadership: those in the core acquire information directly and this is then accessed by all others in the population. Since there are no problems with the interpretation of opinions in their framework, and hence no variation in the extent to which different individuals are well-understood, information segregation cannot arise.

Acemoglu, Bimpikis, and Ozdaglar (2014) also considered communication in an endogenous network. Individuals can observe the information of anyone to whom they are linked either directly or indirectly via a path, but observing more distant individuals requires waiting longer before an action is taken. Holding constant the network, the key trade-off in their model is between reduced delay and a more informed decision. They showed that dispersed information is most effectively aggregated if the network has a hub and spoke structure with some individuals gathering information from numerous others and transmitting it either directly or via neighbors to large groups. This structure is then shown to emerge endogenously when costly links are chosen prior to communication, provided that certain conditions are satisfied. One of these conditions is that friendship cliques, defined as sets of individuals who can observe each other at zero cost, not be too large. Members of large cliques are well-informed, have a low marginal value of information, and will not form costly links to those outside the clique. Hence both opinion leadership and information segregation are possible equilibrium outcomes in their model, though the mechanisms giving rise to these are clearly distinct from those explored here.

The literature on communication in organizations also explicitly considers the precision of messages sent and received, in an environment in which adaptation to local information and coordination of actions across individuals both matter; see Dessein, Galeotti, and Santos (2015) for a recent contribution. Message precision is an object of choice, subject to costs and determined endogenously. This literature is concerned with questions related to organizational form and focus, somewhat orthogonal to those considered here. Most closely related is work by Calvó-Armengol, de Martí, and Prat (2015), who explored the extent of influence exerted by individuals at different points in an exogenously given network.

A trade-off between being well-informed and well-understood appears in Dewan and Myatt (2008), who considered communication by leaders of political parties. As in the literature on communication in organizations, both adaptation to information and coordination of actions matter, but instead of local states there is a global state and only leaders receive signals regarding its value. Leaders vary in the degree to which they are well-informed (their sense of direction, in the language of the authors) and also vary in the clarity with which they can communicate their information. Influential leaders have the right mix of attributes, which the authors showed is tilted towards clarity of communication. The potential for clear communication is a parameter in their static model, rather than a consequence of prior observation as in ours.

Finally, strategic communication with observable heterogeneous priors has previously been considered by Banerjee and Somanathan (2001), Che and Kartik (2009), and Van den Steen (2010), building on Crawford and Sobel (1982). In our own previous work, we have considered truthful communication with unobservable priors, but with a single state and public belief announcements (Sethi and Yildiz (2012)). Communication across an endogenous network with unobserved heterogeneity in prior beliefs and a sequence of states has not previously been explored as far as we are aware. Furthermore, the theory we offer to account for the size and structure of expert sets, own field bias, and large field dominance is novel, and this constitutes our main contribution to the literature.

10. CONCLUSION

Interpreting the opinions of others is challenging because such opinions are based in part on private information and in part on prior beliefs that are not directly observable. Individuals seeking informative opinions may therefore choose to observe those whose priors are well-understood, even if their private information is noisy. This problem is compounded by the fact that observing opinions is informative not only about private signals but also about perspectives, so preferential attachment to particular persons can develop endogenously over time. As a result, when there is sufficient initial uncertainty about the perspectives of others, individuals limit attention to a small set of experts who have become familiar through past observation, and neglect others who may be better-informed on particular issues. These sets are of negligible relative size in large populations, even when individuals are forward-looking. Moreover, the extent of attachment that develops in any period depends on how well-informed the observer happens to be. This gives rise to symmetry breaking and allows for a broad range of networks to emerge over time, including opinion leadership and information segregation.

Our basic premise is that it is costly to extract information from less familiar sources. These costs arise from the difficulty of making inferences when opinions are contaminated by unobserved prior beliefs. The degree of such difficulty changes endogenously in response to historical patterns of observation. We have explored one application of this idea in detail, showing that it gives rise to own field bias when a principal is tasked with evaluating a sequence of cases, and leads to experts in larger fields being consulted with disproportionately high frequency on cases outside their area of expertise. We believe that the framework developed here can be usefully applied to a variety of other settings, including but not limited to informational segregation across identity groups.

APPENDIX

The following formula is used repeatedly in the text and stated here for convenience. Given a prior $\theta \sim N(\mu, 1/v)$ and signal $s = \theta + \varepsilon$ with $\varepsilon \sim N(0, 1/r)$, the posterior is $\theta \sim N(y, 1/w)$ where

$$y = E[\theta|s] = \frac{v}{v+r}\mu + \frac{r}{v+r}s \tag{16}$$

and $w = v + r$.

As a step towards proving Proposition 2, the following lemma identifies sufficient conditions for a link to be broken or free; see the Supplemental Material for a full characterization.

LEMMA 2: *Under Assumption 1, at a history h_t, a link ij is free if $v_{ij}^t(h_t) > \overline{v}$ and broken if there exists $k \in N$ with $v_{ik}^t(h_t) > \beta(v_{ij}^t(h_t))$.*

PROOF: To prove the first part, take any i, j with $v_{ij}^t(h_t) > \overline{v}$. Then, by definition of $\overline{v}$, for any $k \notin \{i, j\}$,

$$\gamma\big(b, v_{ij}^t(h_t)\big) < \gamma(b, \overline{v}) \leq \gamma\big(a, v_{ik}^t(h_t)\big),$$

where the first inequality holds because γ is decreasing in v and the second is by definition of $\overline{v}$. Hence, by continuity of γ, there exists $\eta > 0$ such that for all $k \notin \{i, j\}$,

$$\gamma\big(b - \eta, v_{ij}^t(h_t)\big) < \gamma\big(a + \eta, v_{ik}^t(h_t)\big).$$

Consider the event Π in which $\pi_{jt} \in [b - \eta, b]$ and $\pi_{kt} \in [a, a + \eta]$ for all $k \neq j$. This has positive probability under Assumption 1, and on this event $j_{it} = j$. For any $s \geq t$, since $v_{ij}^s \geq v_{ij}^t \geq \overline{v}$, we have $\Pr(j_{is} = j) > 0$, showing that the link ij is free.

To prove the second part, take $v_{ik}^t(h_t) > \beta(v_{ij}^t(h_t))$. By definition of β,

$$\gamma\big(a, v_{ik}^t(h_t)\big) < \gamma\big(a, \beta\big(v_{ij}^t(h_t)\big)\big) = \gamma\big(b, v_{ij}^t(h_t)\big),$$

where the inequality is by monotonicity of γ and the equality is by definition of β. Hence, $\Pr(l_{ij}^t = 1|h_t) = 0$. Moreover, by (9), at any h_{t+1} that follows h_t, $v_{ij}^{t+1}(h_{t+1}) = v_{ij}^t(h_t)$ and $v_{ik}^{t+1}(h_{t+1}) \geq v_{ik}^t(h_t)$, and hence the previous argument yields $\Pr(l_{ij}^{t+1} = 1|h_t) = 0$. Inductive application of the same argument shows that $\Pr(l_{ij}^s = 1|h_t) = 0$ for every $s \geq 0$, showing that the link ij is broken at h_t. *Q.E.D.*

PROOF OF PROPOSITION 2: Part (a). Assume $v_0 > \overline{v}$ for all distinct $i, j \in N$. Then, for each h_t, the probability of $j_{it}(h_t) = j$ is bounded from below by $\lambda(\Pi) > 0$ for the event Π defined in the proof of Lemma 2. Hence, with probability 1, i links to j infinitely often, showing that $j \in J_h(i)$.

Part (b). Clearly, when $v_0 > \overline{v}$, the long-run outcome is history independent by Part (a), and hence opinion leadership is not possible. Accordingly, suppose that $v_0 < \overline{v}$. Consider the positive probability event A that for every $t \leq t^*$, $\pi_{1t} > \pi_{2t} > \max_{k>2} \pi_{kt}$ for some $t^* \geq (\beta(v_0) - v_0)/\underline{\Delta}$. Clearly, on event A, for any $t \leq t^*$ and $k > 1$, $j_{kt} = 1$ and $j_{1t} = 2$, since these are the best-informed and best-known targets for all those who link to them. Then, on event A, for $ij \in S \equiv \{12, 21, 31, \dots, n1\}$,

$$v_{ij}^{t^*+1} = v_0 + \sum_{t=0}^{t^*} \Delta(\pi_{is}, \pi_{js}) \geq v_0 + \left(t^* + 1\right)\underline{\Delta} > \beta(v_0),$$

while $v_{ik}^{t^*+1} = v_0$ for any $ik \notin S$. (Here, the equalities are by (9); the weak inequality is by Lemma 1, and the strict inequality is by definition of t^*.) Therefore, by Lemma 2, all the links $ik \notin S$ are broken by t^*, resulting in extreme opinion leadership as claimed.

To prove the second part of the statement, note that for any $v_0 \leq \tilde{v}$ and $i \in N$,

$$v_{ij_{i1}}^{1} = v_0 + \Delta(\pi_{i1}, \pi_{ij_1}) \geq v_0 + \underline{\Delta} \geq \beta(v_0),$$

while $v_{ik}^1 = v_0$ for all $k \neq j_{i1}$, showing by Lemma 2 that all such links ik are broken after the first period. Since $j_{i1} = \min \arg\max_i \pi_{i1}$ for every $i \neq \min \arg\max_i \pi_{i1}$, this shows that extreme leadership emerges at the end of the first period with probability 1. The claim that extreme opinion leadership arises with probability less than 1 if $v_0 > \tilde{v}$ follows from Part (c) below.

Part (c). For simplicity, we take $t = 1, 2, 3, \dots$. Take any $v_0 \in (\tilde{v}, \overline{v} - \underline{\Delta})$ and any partition $\{S_1, \dots, S_m\}$ where each cluster S_k has at least two elements i_k and j_k. We will now construct a positive probability event on which the process exhibits segregation over partition $\{S_1, \dots, S_m\}$. Since $v_0 \in (\tilde{v}, \overline{v} - \underline{\Delta})$, there exists $\varepsilon > 0$ such that

$$v_0 + \Delta(a + \varepsilon, b - \varepsilon) < \min\left\{\beta(v_0), \overline{v}\right\} \tag{17}$$

and

$$\Delta(b - \varepsilon, b) > \Delta(a + \varepsilon, b - \varepsilon). \tag{18}$$

By (18) and by the continuity and monotonicity properties of γ, there also exist $\pi^* \in (a, b)$ and $\varepsilon' > 0$ such that

$$\begin{aligned} &\gamma\left(\pi^* - \varepsilon', v_0 + \Delta(b - \varepsilon, b)\right) < \gamma(b, v_0), \\ &\gamma\left(\pi^* + \varepsilon', v_0 + \Delta(a + \varepsilon, b - \varepsilon)\right) > \gamma(b - \varepsilon, v_0). \end{aligned} \tag{19}$$

For every $t \in \{1, \dots, m\}$, the realized expertise levels are as follows:

$$\begin{aligned} &\pi_{i_t t} > \pi_{j_t t} > \pi_{it} > b - \varepsilon \quad (\forall i \in S_t), \\ &\pi^* + \varepsilon' > \pi_{i_k t} > \pi_{j_k t} > \pi_{it} > \pi^* - \varepsilon' \quad (\forall i \in S_k, k < t), \\ &\pi_{it} < a + \varepsilon \quad (\forall i \in S_k, k > t). \end{aligned}$$

Fixing

$$t^* > \big(\beta\big(v_0 + \Delta(a+\varepsilon, b-\varepsilon)\big) - v_0\big)/\underline{\Delta},$$

the realized expertise levels for $t \in \{m+1, \dots, m+t^*\}$ are as follows:

$$\pi^* + \varepsilon' > \pi_{i_k t} > \pi_{j_k t} > \pi_{it} > \pi^* - \varepsilon' \quad (\forall i \in S_k, \forall k).$$

The above event clearly has positive probability. We next show that the links ij from distinct clusters are all broken by $m + t^* + 1$.

Note that at $t = 1$, $j_{i_1 1} = j_1$ and $j_{i1} = i_1$ for all $i \neq i_1$. Hence,

$$v^2_{ii_1} \geq v_0 + \Delta(b-\varepsilon, b) > v_0 + \Delta(a+\varepsilon, b-\varepsilon) \geq v^2_{ji_1} \quad (\forall i \in S_1, \forall j \notin S_1),$$

where the strict inequality is by (18). Therefore, by (19), at $t = 2$, each $i \in S_1$ sticks to her previous link

$$j_{i_1 1} = j_1 \quad \text{and} \quad j_{i1} = i_1 \quad \forall i \in S_1 \setminus \{i_1\},$$

while each $i \notin S_1$ switches to a new link

$$j_{i_2 2} = j_2 \quad \text{and} \quad j_{i2} = i_2 \quad \forall i \in N \setminus \big(S_1 \cup \{i_2\}\big).$$

Using the same argument inductively, observe that for any $t \in \{2, \dots, m\}$, for any $i \in S_k$ and $i' \in S_l$ with $k < t \leq l$, and for any $s < t$,

$$v^t_{ij_{i(t-1)}} \geq v_0 + \Delta(b-\varepsilon, b) > v_0 + \Delta(a+\varepsilon, b-\varepsilon) \geq v^2_{i' j_{i's}}.$$

Hence, by (19),

$$j_{it} = \begin{cases} j_{i(t-1)} & \text{if } i \in S_k \text{ for some } k < t, \\ j_t & \text{if } i = i_t, \\ i_t & \text{otherwise.} \end{cases}$$

In particular, at $t = m$, for any $i \in S_k$, $j_{im} = i_k$ if $i \neq i_k$ and $j_{i_k m} = j_k$. Once again,

$$v^t_{ij_{im}} \geq v_0 + \Delta(b-\varepsilon, b).$$

Moreover, i could have observed any other j at most once, when $\pi_{it} < a^* + \varepsilon$ and $\pi_{jt} > b - \varepsilon$, yielding

$$v^t_{ij} \leq v_0 + \Delta(a+\varepsilon, b-\varepsilon).$$

Hence, by (19), i sticks to j_{im} by date $m + t^*$, yielding

$$v^{m+t^*+1}_{ij_{im}} \geq v_0 + \Delta(b-\varepsilon, b) + t^*\underline{\Delta} > \beta\big(v_0 + \Delta(a+\varepsilon, b-\varepsilon)\big) \geq \beta\big(v^{m+t^*+1}_{ij}\big)$$

for each $j \neq j_{im}$. By Lemma 2, this shows that the link ij is broken. Since $j_{im} \in S_k$, this proves the result.

Part (d). Take v_0 as in the hypothesis, and take any $g: N \to N$. We will construct some t^* and a positive probability event on which

$$j_{it} = g(i) \quad \forall i \in N, t > n + t^*.$$

Now, let π be as in the hypothesis. By continuity of Δ and γ, there exists a small but positive ε such that

$$\gamma(\pi, v_0) < \gamma\big(a, v_0 + \Delta(\pi + \varepsilon, b - \varepsilon)\big), \tag{20}$$

$$\gamma(b - \varepsilon, v_0) < \gamma\big(\pi + \varepsilon, v_0 + \Delta(\pi + \varepsilon, b - \varepsilon)\big), \tag{21}$$

$$\Delta(b - \varepsilon, \pi + \varepsilon) > \Delta(\pi + \varepsilon, b - \varepsilon). \tag{22}$$

Fix some

$$t^* > \big(\beta\big(v_0 + \Delta(\pi + \varepsilon, b - \varepsilon)\big) - v_0\big)/\underline{\Delta},$$

and consider the following positive probability event:

$$\begin{aligned} &\pi_{t,t-1} \geq b - \varepsilon > \pi + \varepsilon \geq \pi_{g(t),t-1} \geq \pi > a + \varepsilon \geq \pi_{j,t-1} \\ &\quad \big(\forall j \in N \setminus \{t, g(t)\}, \forall t \in N\big), \\ &(\pi_{1t}, \dots, \pi_{nt}) \in A \quad \big(\forall t \in \{n, \dots, n + t^* - 1\}\big), \end{aligned}$$

where

$$\begin{aligned} A \equiv \big\{(\pi_1, \dots, \pi_n) | \gamma\big(\pi_i, v_0 + \Delta(\pi + \varepsilon, b - \varepsilon)\big) \\ > \gamma\big(\pi_j, v_0 + \Delta(b - \varepsilon, \pi + \varepsilon)\big) \; \forall i, j \in N\big\}. \end{aligned}$$

Note that A is open and nonempty (as it contains the diagonal set). Note that for every $t \in N$, at date $t - 1$, the individual t becomes an ultimate expert (with precision nearly b), and her target $g(t)$ is the second best expert.

We will next show that the links ij with $j \neq g(i)$ are all broken by $n + t^*$. Towards this goal, we will first make the following observation:

> For every $t \in N$, at date $t - 1$, t observes $g(t)$; every $i < t$ observes either t or $g(i)$, and every $i > t$ observes t.

At $t = 0$, the above observation is clearly true: 1 observes $g(1)$, while everybody else observes 1. Suppose that the above observation is true up to $t - 1$ for some t. Then, by date $t - 1$, for any $i \geq t$, i has observed each $j \in \{1, \dots, t - 1\}$ once, when her own precision was in $[a, \pi + \varepsilon]$ and the precision of j was in $[b - \varepsilon, b]$. Hence, by Lemma 1, $v_{ij}^{t-1} \leq v_0 + \Delta(\pi + \varepsilon, b - \varepsilon)$. She has not observed any other individual, and hence $v_{ij}^{t-1} = v_0$ for all $j \geq t$. Thus, by (21), for any $i > t$, $\gamma(\pi_{t,t-1}, v_{it}^{t-1}) < \gamma(\pi_{j,t-1}, v_{ij}^{t-1})$ for every $j \in N \setminus \{i, t\}$, showing that i observes t, that is, $j_{i,t-1} = t$. Likewise, by (20), for $i = t$, $\gamma(\pi_{g(t),t-1}, v_{tg(t)}^{t-1}) < \gamma(\pi_{j,t-1}, v_{tj}^{t-1})$ for every $j \in N \setminus \{t, g(t)\}$, showing that t observes $g(t)$, that is, $j_{t,t-1} = g(t)$. Finally, for any $i < t$, by the inductive hypothesis, i has observed any $j \neq g(i)$ at most once, yielding $v_{ij}^{t-1} \leq v_0 + \Delta(\pi + \varepsilon, b - \varepsilon)$. Hence, as above, for any $j \in N \setminus \{i, t, g(i)\}$, $\gamma(\pi_{t,t-1}, v_{it}^{t-1}) < \gamma(\pi_{j,t-1}, v_{ij}^{t-1})$, showing that i does not observe j, that is, $j_{i,t-1} \in \{g(i), t\}$.

By the above observations, after the first n periods, each i has observed any other $j \neq g(i)$ at most once, so that

$$v_{ij}^{n} \leq v_0 + \Delta(\pi + \varepsilon, b - \varepsilon) \quad \bigl(\forall j \neq g(i)\bigr). \tag{23}$$

She has observed $g(i)$ at least once, and in one of these occasions (i.e., at date i), her own precision was in $[b - \varepsilon, b]$ and the precision of $g(i)$ was in $[\pi, \pi + \varepsilon]$, yielding

$$v_{ig(i)}^{n} \geq v_0 + \Delta(b - \varepsilon, \pi + \varepsilon). \tag{24}$$

By definition of A, inequalities (23) and (24) imply that each i observes $g(i)$ at n. Consequently, the inequalities (23) and (24) also hold at date $n + 1$, leading each i again to observe $g(i)$ at $n + 1$, and so on. Hence, at dates $t \in \{n, \dots, t^* + n - 1\}$, each i observes $g(i)$, yielding

$$\begin{aligned} v_{ig(i)}^{n+t^*} &\geq v_{ig(i)}^{n} + t^* \underline{\Delta} \\ &> v_0 + \Delta(b - \varepsilon, \pi + \varepsilon) + \beta\bigl(v_0 + \Delta(\pi + \varepsilon, b - \varepsilon)\bigr) - v_0 \\ &> \beta\bigl(v_0 + \Delta(\pi + \varepsilon, b - \varepsilon)\bigr). \end{aligned}$$

For any $j \neq g(i)$, since $v_{ij}^{n+t^*} = v_{ij}^{n+1}$, together with (23), this implies that

$$v_{ig(i)}^{n+t^*} > \beta\bigl(v_{ij}^{n+t^*}\bigr).$$

Therefore, by Lemma 2, the link ij is broken at date $t^* + n$. *Q.E.D.*

PROOF OF PROPOSITION 5: Part (a). Take any h and h' with $h \succeq h'$ and any principal i. Consider first the case $t_i(h) \geq t_i(h')$, that is, i attaches to an expert under h' at an earlier date $t_i(h')$. If $t_i(h') = \infty$, we have $J_h(i) = J_{h'}(i)$, and the claim clearly holds. If $t_i(h')$ is finite, then we must have

$$v_{ij_1}^{t_i(h')}(h) \geq v_{ij_1}^{t_i(h')}\bigl(h'\bigr) \quad \text{and} \quad v_{ij_2}^{t_i(h')}(h) \leq v_{ij_2}^{t_i(h')}\bigl(h'\bigr).$$

This follows from the facts that (i) at any $t \leq t_i(h')$ at which behavior under h and h' are different, i observes j_1 instead of j_2, so v_{ij_1} increases and v_{ij_2} remains constant under h, while v_{ij_2} increases and v_{ij_1} remains constant under h', and (ii) at any $t \leq t_i(h')$ at which behavior under h and h' are the same, v_{ij_1} and v_{ij_2} change by the same amount. Hence, if i attaches to j_1 at $t_i(h')$ under h', she must also attach to j_1 under h at this period or earlier. That is,

$$v_{ij_1}^{t_i(h')}(h) \geq v_{ij_1}^{t_i(h')}(h') > \beta\big(v_{ij_2}^{t_i(h')}(h')\big) \geq \beta\big(v_{ij_2}^{t_i(h')}(h)\big),$$

where the strict inequality is because i attaches to j_1 at $t_i(h')$ under h', and the last inequality is because β is increasing. Hence $q_i(J_h) \geq q_i(J_{h'})$. Similarly, if i attaches to j_2 under h at some finite $t_i(h)$, then she must also attach to j_2 under h' at $t_i(h)$ or earlier, so $q_i(J_h) \geq q_i(J_{h'})$. Since this is true for both principals, we have $J_h \succeq J_{h'}$.

Parts (b)–(c). By Part (a), it suffices to show that we can obtain the distribution on the set of all histories under p_1 from the distribution under p_1' by the following transformation. For each history h' of realized expertise levels under p_1', change every π_t with $\pi_{j_1t}(h') = a < b = \pi_{j_2t}(h')$ to $\pi_{j_2t}(h') = a < b = \pi_{j_1t}(h')$ with probability $\hat{p} = (p_1 - p_1')/(1 - p_1') \in [0, 1]$ independently. That is, flip the case to field 1 with probability $\hat{p}$ if it happens to be in field 2. This leads to a probability distribution on histories h with $h \succeq h'$, whence $J_h \succeq_1 J_{h'}$ by Part (a). Observe that the resulting probability distribution is also i.i.d., with probability that a case is in field 1 being $p_1' + \hat{p}(1 - p_1') = p_1$. To complete the proof, we show that $p_1 > p_1'$ implies

$$P(D_1|p_1) > P\big(D_1|p_1'\big).$$

Let n denote the largest integer such that

$$v_0 + n\Delta(b, b) < \overline{v},$$

and for each $m = 0, 1, \dots, n$, define k_m as the smallest integer such that

$$v_0 + k_m\Delta(a, b) > \beta\big(v_0 + m\Delta(b, b)\big).$$

Note that principal i_2 attaches to expert j_1 if and only if at least k_m of the first $k_m + m$ cases lie in field 1 for some $m \leq n$. From Proposition 4, i_1 must also attach to j_1 in this case, so D_1 occurs. It is easily verified that the probability of this event is strictly increasing in p_1 and approaches 1 as $p_1 \to 1$.

We have already shown that $P(D_1|p_1) > P(D_1|p_1')$. Now let n denote the largest integer such that

$$v_0 + n\Delta(a, b) < \overline{v},$$

and for each $m = 0, 1, \dots, n$, define k_m as the smallest integer such that

$$v_0 + k_m\Delta(b, b) > \beta\big(v_0 + m\Delta(a, b)\big).$$

Note that D_1^* occurs if and only if principal i_1 attaches to expert j_1, and this occurs if and only if at least k_m of the first $k_m + m$ cases lie in field 1 for some $m \leq n$. It is easily verified that the probability of this event is strictly increasing in p_1. The claims regarding D_2 and D_2^* may be proved analogously. *Q.E.D.*

REFERENCES

ACEMOGLU, D., K. BIMPIKIS, AND A. OZDAGLAR (2014): "Dynamics of Information Exchange in Endogenous Social Networks," *Theoretical Economics*, 9, 41–97. [2059,2060]

ACEMOGLU, D., V. CHERNOZHUKOV, AND M. YILDIZ (2016): "Fragility of Asymptotic Agreement Under Bayesian Learning," *Theoretical Economics*, 11, 187–225. [2057]

ACEMOGLU, D., M. A. DAHLEH, I. LOBEL, AND A. OZDAGLAR (2011): "Bayesian Learning in Social Networks," *Review of Economic Studies*, 78, 1201–1236. [2059]

BACCARA, M., AND L. YARIV (2013): "Homophily in Peer Groups," *American Economic Journal: Microeconomics*, 5, 69–96. [2053]

BALA, V., AND S. GOYAL (1998): "Learning From Neighbours," *Review of Economic Studies*, 65, 595–621. [2059]

——— (2000): "A Noncooperative Model of Network Formation," *Econometrica*, 68, 1181–1229. [2059]

BANERJEE, A. V. (1992): "A Simple Model of Herd Behavior," *Quarterly Journal of Economics*, 107, 797–817. [2059]

BANERJEE, A. V., AND R. SOMANATHAN (2001): "A Simple Model of Voice," *Quarterly Journal of Economics*, 116, 189–227. [2060]

BIKHCHANDANI, S., D. HIRSCHLEIFER, AND I. WELCH (1992): "A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades," *Journal of Political Economy*, 100, 992–1023. [2059]

BLOCH, F., AND B. DUTTA (2011): "Formation of Networks and Coalitions," in *Handbook of Social Economics*, ed. by J. Benhabib, A. Bisin, and M. O. Jackson. Amsterdam: Elsevier. [2059]

CALVÓ-ARMENGOL, A., J. DE MARTÍ, AND A. PRAT (2015): "Communication and Influence," *Theoretical Economics*, 10, 649–690. [2060]

CHE, Y.-K., AND N. KARTIK (2009): "Opinions as Incentives," *Journal of Political Economy*, 117, 815–860. [2060]

CRAWFORD, V. P., AND J. SOBEL (1982): "Strategic Information Transmission," *Econometrica*, 50, 1431–1451. [2060]

DEGROOT, M. (1974): "Reaching a Consensus," *Journal of the American Statistical Association*, 69, 118–121. [2032,2059]

DEMARZO, P., D. VAYANOS, AND J. ZWEIBEL (2003): "Persuasion Bias, Social Influence, and Unidimensional Opinions," *Quarterly Journal of Economics*, 118, 909–968. [2059]

DESSEIN, W., A. GALEOTTI, AND T. SANTOS (2015): "Rational Inattention and Organizational Focus," Unpublished manuscript, Columbia Business School. [2060]

DEWAN, T., AND D. P. MYATT (2008): "The Qualities of Leadership: Direction, Communication, and Obfuscation," *American Political Science Review*, 102, 351–368. [2060]

DOUGLAS, M., AND A. WILDAVSKY (1982): *Risk and Culture: An Essay on the Selection of Technical and Environmental Dangers*. Berkeley: University of California Press. [2058]

GALE, D., AND S. KARIV (2003): "Bayesian Learning in Social Networks," *Games and Economic Behavior*, 45, 329–346. [2059]

GALEOTTI, A., AND S. GOYAL (2010): "The Law of the Few," *American Economic Review*, 100, 1468–1492. [2048,2050,2059]

GOEBBERT, K., H. C. JENKINS-SMITH, K. KLOCKOW, M. C. NOWLIN, AND C. L. SILVA (2012): "Weather, Climate, and Worldviews: The Sources and Consequences of Public Perceptions of Changes in Local Weather Patterns," *Weather, Climate and Society*, 4, 132–144. [2059]

GOLUB, B., AND M. O. JACKSON (2010): "Naive Learning in Social Networks and the Wisdom of Crowds," *American Economic Journal: Microeconomics*, 2, 112–149. [2059]

JACKSON, M. O., AND A. WOLINSKY (1996): "A Strategic Model of Social and Economic Networks," *Journal of Economic Theory*, 71, 44–74. [2059]

KAHAN, D. M., AND D. BRAMAN (2006): "Cultural Cognition and Public Policy," *Yale Law & Policy Review*, 24, 147–170. [2058]

PEW RESEARCH CENTER FOR THE PEOPLE & THE PRESS (2008): "12% Still Think Obama Is Muslim," Accessed at http://pewresearch.org/databank/dailynumber/?NumberID=509 on October 27. [2059]

PLAMBECK, J. (2012): "From Jack Welch, a Conspiracy Theory," *New York Times* (October 5). [2059]

SETHI, R., AND M. YILDIZ (2012): "Public Disagreement," *American Economic Journal: Microeconomics*, 4, 57–95. [2061]

——— (2016a): "Communication With Unknown Perspectives," Available at SSRN http://ssrn.com/abstract=2324154. [2043,2055,2056]

——— (2016b): "Supplement to 'Communication With Unknown Perspectives'," *Econometrica Supplemental Material*, 84, http://dx.doi.org/10.3982/ECTA13320. [2042]

SMITH, L., AND P. SORENSEN (2000): "Pathological Outcomes of Observational Learning," *Econometrica*, 68, 371–398. [2059]

THRUSH, G. (2009): "58 Percent of GOP Not Sure/Doubt Obama Born in US," *Politico.com* (July 31). [2059]

VAN DEN STEEN, E. (2010): "Culture Clash: The Costs and Benefits of Homogeneity," *Management Science*, 56, 1718–1738. [2060]

VOORHEES, J. (2012): "The Skewed-Polling Conspiracy Theory Goes Mainstream," *Slate* (October 2). [2059]

Dept. of Economics, Barnard College, Columbia University, 3009 Broadway, New York, NY 10027, U.S.A.; rs328@columbia.edu

and

Dept. of Economics, MIT, 50 Memorial Drive, Cambridge, MA 02142, U.S.A.; myildiz@mit.edu.

Co-editor Joel Sobel handled this manuscript.

Manuscript received March, 2015; final revision received May, 2016.