Communication with Unknown Perspectives*

Rajiv Sethi[†] Muhamet Yildiz[‡]

January 14, 2016

Abstract

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.

*For their helpful suggestions, we thank Daron Acemoglu, Sanjeev Goyal, Navin Kartik, Joel Sobel, four anonymous referees, and participants at the 2014 Econometric Society Latin American Meetings, 2014 Workshop on Economic Theory (FGV), 2014 Time, Uncertainities and Strategies workshop (PSE), and seminars at ASU, University of Arizona, Bocconi, University of Chicago, University of Chile, Columbia, Harvard, MIT, and Northwestern. We also thank the Becker Friedman Institute at University of Chicago, the Institute for Advanced Study at Princeton, University of Pennsylvania, Princeton University, and Harvard University for hospitality and support.

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

[‡]Department of Economics, MIT.

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 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-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 graduate admissions, one may rely on recommenders with long track records whose opinions have become easier to interpret over time. And 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 unusual and interesting 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, whose opinion is observed. 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 target. 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 i 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, when the precision of initial beliefs about the perspectives of others is above a certain threshold, we show that 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, individual i will link to j only if j is substantially more informed than *each* of these m experts. This is an exponentially long shot event. Therefore, before i chooses to observe any such individual j, she links to more familiar individuals many times, learning more about them on each occasion, and develops so much attachment to these that she stops observing j permanently.

A similar result holds even if individuals are forward-looking. In this case individuals may choose to observe a less informative opinion in the current period in order to *build* familiarity with someone who could be useful to them in future periods. The benefit of doing so, relative to the cost, is exponentially decreasing in the number of already familiar targets. We show that when the initial precision of beliefs about the perspectives of others is below the threshold 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 before, this set contains only a negligible fraction of all individuals in large populations. The welfare loss, relative to the case of observable perspectives, can be substantial.

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 long run 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. One consequence of this is 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.

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 as individuals make observational choices 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. The application to areas of expertise and own-field bias is developed in Section 7. Section 8 extends our results to allow for forward-looking agents, and other extensions and variations of the model are discussed in Section 9. A review of related literature is contained in Section 10, and Section 11 concludes. The appendix contains omitted proofs, and a supplementary appendix (available online) contains a number of additional results.

2 The Model

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

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

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

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

for some real numbers $\overline{\mu}_1, ..., \overline{\mu}_n$ 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 key 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. For our general analysis, we only 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. Finally, we assume that the expertise levels π_{it} are publicly observable at t.

Remark 1. *Since priors are heterogenous, 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}. For a more extensive discussion of belief revision with incomplete information and unobservable, heterogenous priors, see Sethi and Yildiz (2012), where we study repeated communication about a single state in a group of individuals with equal levels of expertise.*

Having observed the signal x_{it} in period t, individual i updates her belief about the state according to Bayes' 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

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

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 $(\pi_{1t}, \ldots, \pi_{nt})$ 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 j in period t, and the matrix $[l_{ij}^t]$ defines a directed graph that describes who listens to whom. Consistent with this interpretation, we shall say that i *links to* j in period t if $j = j_{it}$. It is more convenient to represent such directed graphs by functions $g : N \to N$ with $g(i) \neq i$ for each $i \in N$. We write G for the set of all such functions.

Remark 2. *We assume to begin with that individuals are myopic, do not observe the actions or past targets of others, and do not observe the realization of the state. As shown in Sections 8-9 and the supplementary appendix, our results extend for the most part to the case of forward-looking behavior, as well as delayed observability of states, actions, and the 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 $\theta_{t'}$ are 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 the criterion on the basis of which a given individual i selects a target j whose opinion y_{jt} is to be observed, and what i learns about the state θ_t and j's perspective μ_j as a result of this observation. This determines the process for the evolution of beliefs and the network of information flows.

Under our assumptions, the posterior beliefs held by any individual about the perspectives of any other 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 $i \neq j$,

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

The precisions v_{ij}^t in subsequent periods 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 an individual i has chosen to observe the opinion y_{jt} of individual 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 additive 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 to observe an individual for whom the variance γ is lowest. For completeness we assume that ties are broken in favor of the individual with the smallest label:

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

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 the manner in which 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, where a solid line

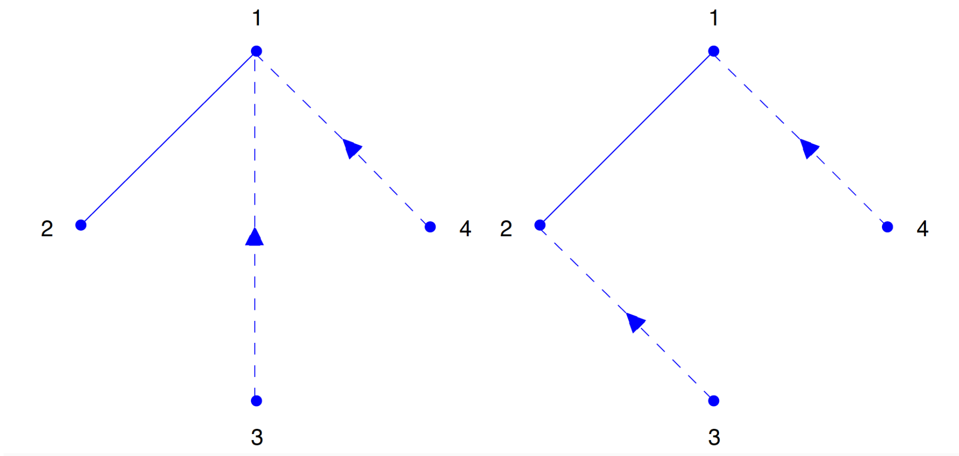

Figure 1: Asymmetric effects of first period observations on second period links.

indicates links in both directions. 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: $v^2_{21} > v^2_{41} > v^2_{31}$. Now 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^t_{ij} 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^{t+1}_{ij} at the start of the subsequent period in accordance with (8), with initial precision $v^0_{ij} = v_0$.

For any period t, let $h_t := (v^{t'}_{ij})_{t'<t}$ 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 $j_{it}\left(h_t, \pi_t\right)$ 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 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

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

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

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

This satisfies the indifference condition

$$\gamma\left(a, \beta\left(v\right)\right) = \gamma\left(b, v\right)$$

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

When $v_{ij}^t(h_t) > \beta(v_{ik}^t(h_t))$ for all $k \in N \backslash \{i, j\}$, 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 $\beta(v_{ij}^t(h_t))$, resulting in the link ij being broken. (This happens when i links to k at least $(\beta(v_{ij}^t(h_t)) - v_{ik}^t(h_t))/\underline{\Delta}$ 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

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

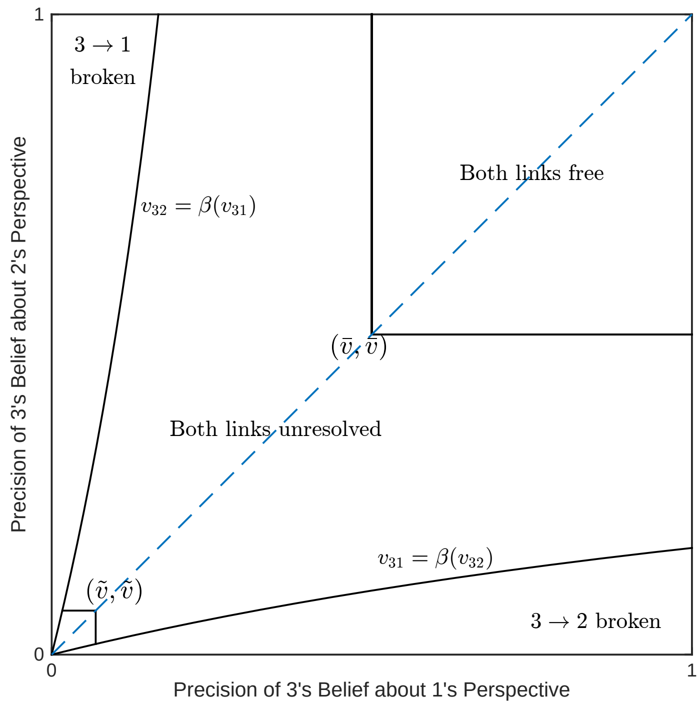

Figure 2: Regions of state space with broken and free links

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 period target. This is because if $v_0 < \tilde{v}$ and i observes k initially, then $v_{ik}^2 \geq v_0 + \underline{\Delta}$, and hence $v_{ik}^2 > \beta(v_0) = \beta(v_{ij})$ for all $j \neq k$. All links except those that form initially break by the second period, and the initial observational network is persistent.[1]

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 or free, 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 these nodes 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 individual is better

[1]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 because of symmetry in initial beliefs about perspectives, but in asymmetric models with constant expertise levels, individuals may attach to targets who are initially more familiar but have less expertise throughout. This clarifies the importance for our analysis of the assumption that issues vary across periods and expertise accordingly varies across individuals.

informed in any given period. This long-run behavior is therefore independent of past realizations when the initial precision in beliefs about perspectives is sufficiently high.

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 flatter curve, 3 links only to 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

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

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*

$$j_{it}(h_t, \pi_t) \in \left\{ j \in J_h(i) \mid \pi_{jt} \geq \pi_{j't} - \varepsilon \; \forall j' \in J_h(i) \right\} \qquad (\forall i \in N, \forall t \geq \tau(h)).$$

Proof. Observe that there exists $\overline{v}_\varepsilon < \infty$ such that if $v_{ij}^t > \overline{v}_\varepsilon$ 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 $v_{ij}^t(h_t) \to \infty$. (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 $v_{ij}^t(h_t) > \overline{v}_\varepsilon$ whenever $t \geq \tau_{ij\varepsilon}(h)$. Since N is finite, we can set $\tau(h) = \max\left\{\overline{t}(h), \max_{i\in N, j\in J_h(i)} \tau_{ij\varepsilon}(h)\right\}$, 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 $v_{ij}^t > \overline{v}_\varepsilon$)—as claimed. □

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 links to both j and j', then i cannot link to j in a period in which i' links to j'.

In the supplementary appendix we show that when expertise levels $(\pi_{1t}, \ldots, \pi_{nt})$ are serially i.i.d., the long run graphs are also serially i.i.d. with a history-dependent long-run distribution. Moreover, 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

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

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 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 $(\pi_{1t}, \ldots, \pi_{nt})$ 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 maximized at $J_h(i) = N\backslash\{i\}$.

5 Long-Run Observational 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 most 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

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

This outcome maximizes the log-run payoff.

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

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

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

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

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 payoffs of all individuals, although it may not be the worst possible situation in asymmetric environments—for example if i_1 is expected to be better informed than others.

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

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

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.

To identify conditions under which each of the above patterns arise, 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.

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 $\{S_1, S_2, \ldots, S_m\}$ of N such that each S_k has at least two elements, there is segregation over $\{S_1, S_2, \ldots, S_m\}$ 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.

Proposition 2 identifies conditions under which a variety of long-run outcomes can arise. Parts (a) and (b) 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 active in all periods, 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 if $v < \tilde{v}$ we get extreme opinion leadership with certainty.

While the emergence of opinion leadership is intuitive, convergence to a segregated network or an arbitrary static network is much less so. Since 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 possibility that they may all choose different targets in the long run, or may be partitioned into segregated clusters, is counter-intuitive. Nevertheless, there exist sequences of expertise realizations that result in such strong asymmetries.

Segregation can arise only if $v_0 < \overline{v}$ (otherwise we obtain long-run efficiency). Furthermore, if $v_0 > \overline{v} - \underline{\Delta}$, all links to the best informed individual in the first period become free. This is because all such links are active in the first period, and the precision of all beliefs about this particular target's perspective rise above $v_0 + \underline{\Delta} > \overline{v}$. This clearly rules out segregation. So v_0 cannot be too large if segregation is to arise. And it cannot be too small either: extreme opinion leadership is inconsistent with segregation and arises with certainty when $v_0 < \tilde{v}$.

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. Even more surprisingly, there exists an open set of parameter values for which any arbitrarily given network can emerge as a static network with positive probability. That is, each individual may be locked into a single, arbitrarily given target in the long run.

Proposition 2(d) goes a step further, establishing that *any* static network can arise with positive probability if a further condition is satisfied. This 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.[2]

We have assumed to this point that all individuals are symmetrically placed, 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 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 supplementary appendix, 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 supplementary appendix. 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

[2]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 $(v^*, \overline{v} - \Delta(b, b)) = (0.13, 0.20)$.

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.

6 The Size of Expert Sets and a Law of the Few

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 $v_0 < \overline{v}$, we have*

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

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

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

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 exatly 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 $|J_h(i)| \leq m$.[3] Now, at any such history h_t, a lower bound for the

[3] 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 $\Pr\left(|J_h(i)| \leq m \mid h_{\tau(h)}\right) = 1$ if $\tau(h) = \infty$ and $\Pr\left(|J_h(i)| \leq m \mid h_{\tau(h)}\right) \geq p^*(m)$ otherwise.

probability that i links to the j with the highest v_{ij}^t is

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

To see this, observe that $v_{ij}^t \geq v_{ij'}^t \geq v_0 + \Delta\left(a, b\right) > v_{ij''}^t = v_0$ for all insiders j' and outsiders j''. Hence, if $\pi_{jt} > \hat{\pi}$, individual i prefers j to all outsiders j''. Moreover, 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 $\left(1 - F\left(\hat{\pi}\right)\right)/m$. Likewise, the probability that i links to an outsider cannot exceed

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

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

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

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 $\Delta\left(a, b\right)$. Hence, after k occurrences, we have $v_{t'}^* \geq v_0 + \Delta\left(a, b\right) + k\Delta\left(a, b\right)$. Therefore, for any integer $k > \left(\beta\left(v_0\right) - v_0\right)/\Delta\left(a, b\right) - 1$, after k occurrences, we have $v_{t'}^* > \beta\left(v_0\right)$. Links to all outsiders are accordingly broken, since $v_{ij''}^{t'}$ remains equal to v_0 for all outsiders j'' throughout. □

The first part of this result provides a lower bound $p^*\left(m\right)$ on the probability $\Pr\left(\left|J_h\left(i\right)\right| \leq m\right)$ that the size of the set of long run experts does not exceed m, uniformly for all population sizes n. Here, $p^*\left(m\right)$ depends on the distribution F of expertise levels and the parameter v_0, and is decreasing in v_0 and $F\left(\hat{\pi}\right)$. Since $p^*\left(m\right)$ 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.

As an illustration of Proposition 3, consider a binomial distribution of expertise, with $\pi_{it} = b = 2$ with probability q and $\pi_{it} = a = 1$ with probability $1-q$. 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^*, as well as two tighter bounds that are discussed below. As the figure demonstrates, 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

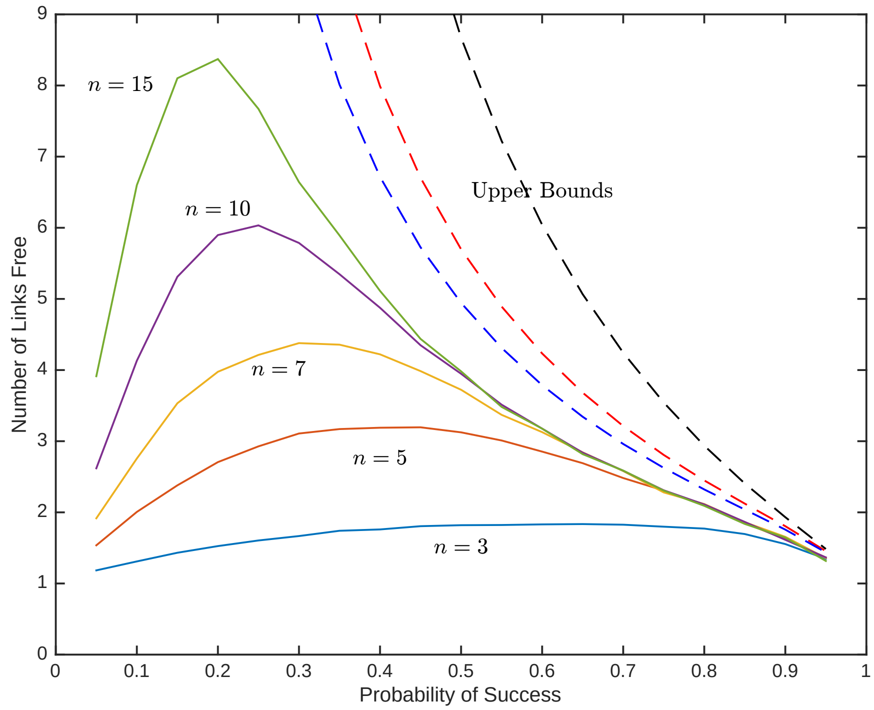

Figure 3: The average number of links per person as a function of q in the binomial example. Solid lines are simulation results, and dashed lines are theoretical upper bounds, which apply to all n uniformly.

bounds imply that the expected value of the number of members cannot exceed 5 no matter how large the population.

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.

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:

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

Moreover, the probability that i links to the most familiar insider is usually higher than the lower bound $(1 - F(\hat{\pi}))/m$ used in the proof of the result. For instance, in the binomial 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 a tighter lower bound:

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

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

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 $\overline{p}^*(m) = \max_{m' \leq m} p^*(m)$. In the binomial example, $\overline{p}^* = p^*$ when $q \geq 1/2$. Now, the bounds $\overline{p}^*$, p^{**}, p_b^*, and p_b^{**} are all (weakly increasing) cumulative distribution functions, and they all first-order stochastically dominate the distribution over $|J_h(i)|$ generated by the model. In Figure 3, which is based on the binomial example, the three upper bounds are the expected values of expert set sizes under $\overline{p}^*$, p_b^*, and p_b^{**}. These expected values are upper bounds for the expected value of $|J_h(i)|$ uniformly for all n, where $\overline{p}^*$ and p_b^{**} yield the loosest and the tightest bounds respectively.

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 core of experts; Galeotti and Goyal (2010) refer 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 show 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.

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:

Corollary 1. *Consider the two-sided model, in which the population N is partitioned into decision-makers N_d and potential experts N_e, and suppose that the state θ_t becomes publicly observable at the end of period t for each t. Then, 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 in two-sided model. If $v_{ij}^t = v_{i'j}^t$ for all $j \in N_e$, then $v_{ij}^{t+1} = v_{i'j}^{t+1} = v_{ij}^t + l_{ij}^t/\pi_{jt}$. Since $v_{ij}^0 = v_{i'j}^0 = v_0$, this shows that $v_{ij}^t = v_{i'j}^t$ 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 holds for the two-sided model with observable states. □

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.

7 Areas of Expertise and Own-Field Bias

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 lie 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.[4]

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

[4]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.

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:

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 any 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,

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

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

$$\gamma(\pi_{j_2 t}, v^t_{i_1 j_2}) \leq \gamma(\pi_{j_1 t}, v^t_{i_1 j_1}).$$

But given (13–14), this implies

$$\gamma(\pi_{j_2 t}, v^t_{i_2 j_2}) < \gamma(\pi_{j_1 t}, v^t_{i_2 j_1}),$$

so i_2 also consults j_2 in period t. □

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 $J : \{i_1, i_2\} \to 2^{\{j_1,j_2\}} \backslash \{\varnothing\}$, where $J(i)$ denotes the set of experts consulted infinitely often by principal i. Each function J corresponds to a distinct event, and following table identifies the six events that can arise with positive probability (and the three that are ruled out by Proposition 4):

	$J(i_2) = \{j_1\}$	$J(i_2) = \{j_1, j_2\}$	$J(i_2) = \{j_2\}$
$J(i_1) = \{j_1\}$	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

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.[5]

A key observation in this section is 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

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

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:

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

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 the above table leads to events that ordered higher:

$$D_1 \succeq PD_1 \succeq S, LRE \succeq PD_2 \succeq D_2.$$

[5]This mechanism is quite different from that driving other models of information homophily. For instance, Baccara and Yariv (2013) consider 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.

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

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

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 $P(\cdot \mid p_1)$ on mappings J, by assigning the probability of $\{h \mid J_h = J\}$ 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 $\succeq$:

$$P \succeq_{FOSD} Q \iff \left[\forall J, \quad P\left(\left\{J' \mid J \succeq J'\right\}\right) \geq Q\left(\left\{J' \mid J \succeq J'\right\}\right)\right].$$

with strict inequality for some J.

The following proposition formalizes the idea that experts in larger fields are consulted disproportionately often, in the sense that they are more likely to be consulted on cases outside their area of expertise:

Proposition 5. *The following are true for all h, h', p_1, and p_1':*

1. *if $h \succeq h'$, then $J_h \succeq J_{h'}$*
2. *if $p_1 > p_1'$, then $P(\cdot| p_1) \succeq_{FOSD} P(\cdot| p_1')$*
3. $\lim_{p_1 \to 1} P(D_1| p_1) = 1$

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.[6]

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).

[6]Observe that when $h \succeq h'$, the expert in field 1 is, if anything, better-informed under h than h' at all periods, i.e., $\pi_{j_1 t}(h) \geq \pi_{j_1 t}(h')$ and $\pi_{j_2 t}(h) \leq \pi_{j_2 t}(h')$. It is tempting to conclude from this that experts in larger fields become more prominent simply because they are better informed, but this is not the case. To see this, suppose that the first case lies in field 1, and change the expertise level of j_1 at $t = 1$ to $b' > b$ keeping the rest—including the realization of cases and expertise levels in all other periods—as before. This change will only lower $v_{ij_1}^t$ for all i and t and will make this expert *less* likely to be selected in the future. It is easy to construct examples in which field 1 will dominate without the above change while field 2 dominates with the change.

8 Forward-Looking Behavior

In order to explore the trade-off between well-informed and well-understood targets, we have assumed to this point 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. We now examine this case, in a simplified version of the baseline model, and show that it leads to qualitatively similar results. 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.

Simplified Forward-Looking Model Each individual i learns the perspective of j_{it} at the end of period t, and maximizes the expected value of the sum of discounted future payoffs, with discount factor δ for some $\delta \in (0,1)$:

$$(1-\delta)\sum_t \delta^t u_i\left(\pi_{it}, \pi_{jt}, v_{ij}^t\right)$$

where

$$u_i\left(\pi_i, \pi_j, v_{ij}\right) \equiv 1 - \frac{1}{1+\pi_i+1/\gamma\left(\pi_j, v_{ij}\right)};$$

we set $u_i\left(\pi_i, \pi_j, \infty\right) \equiv 1 - 1/(1+\pi_i+\pi_j)$. Expertise levels π_{it} are independently and identically distributed with distribution function F such that $0 < F(\pi) < 1$ for all $\pi \in (a,b)$.

The key simplifying assumption here is that an individual observing y_j completely learns μ_j at the end of the period. This holds approximately when b is small, since signals carry little information and opinions are close to priors. Our preferred interpretation is that a period corresponds to a series of interactions during which expertise levels remain constant and after which i faces little uncertainty about μ_j.

For any history h and individual i, let

$$J_h(i) = \{j_{it}(h) \mid t \in T\}$$

be the set of targets to whom i links along the history h, learning their perspectives after the first observation in each case. Since N is finite, there is some finite time $\tau(h)$, at which individual i knows the perspectives of all individuals $j \in J_h(i)$. For all $t \geq \tau(h)$ and $\pi_t \in [a,b]^n$, we have $j_{it}(h_t, \pi_t) \in J_h(i)$. Optimality requires that

$$j_{it}(h_t, \pi_t) \in \arg\max_{j \in J_h(i)} \pi_{jt} \qquad (\forall t \geq \tau(h), \forall \pi_t). \tag{15}$$

Just like their myopic counterparts (in Proposition 1), forward-looking individuals limit themselves to a set $J_h(i)$ of long-run experts and eventually link to the best informed within this set.

The set $J_h(i)$ can be viewed as a *portfolio* of experts who benefit i by providing information whenever their expertise is higher than that of others in the portfolio. By (15), the long-run value of a portfolio $J_h(i)$ with size $|J_h(i)| = m$ is

$$U(m) = \int \overline{U}\left(\pi_{J_h(i)}\right) dF^m\left(\pi_{J_h(i)}\right) \tag{16}$$

where F^m is the distribution of the maximum $\pi_{J_h(i)} = \max_{j \leq m} \pi_j$ of m expertise levels, and

$$\overline{U}(\pi_j) = \int u_i(\pi_i, \pi_j, \infty)\, dF(\pi_i) \tag{17}$$

is the expected utility of receiving a signal with precision π_j in addition to one's own signal. Note that $U(m)$ is strictly increasing and converges to $\overline{U}(b)$ as m goes to ∞. That is, the marginal benefit $U(m+1) - U(m)$ of adding another expert once an individual has already observed m distinct targets goes to zero as m gets large. Hence the value $V(m)$ of having m experts in the portfolio at the beginning of a period under the optimal strategy lies between $U(m)$ and $\overline{U}(b)$.

The optimal strategy solves the following dynamic investment problem. At any given (h_t, π_t), individual i faces one of the following two choices. On the one hand, i can invest in a previously unobserved individual $j \notin J_{h,t}(i) \equiv \{j_{it'}(h) \,|\, t' < t\}$ by linking to j and learning μ_j. This raises her future value to $V(m+1)$ and generates a current return $u_i(\pi_{it}, \pi_{jt}, v_0)$, yielding the payoff

$$\delta V(m+1) + (1-\delta)\, u_i(\pi_{it}, \pi_{jt}, v_0).$$

Alternatively, i can link to the best-informed known individual $j' \in J_{h,t}(i)$. This leaves her continuation payoff at $V(m)$ but results in a possibly higher current return of $u_i(\pi_{it}, \pi_{J_{h,t}(i)}, \infty)$, where $\pi_{J_{h,t}(i)} \equiv \max_{j' \in J_{h,t}(i)} \pi_{j't}$. This alternative has payoff

$$\delta V(m) + (1-\delta)\, u_i(\pi_i, \pi_{J_{h,t}(i)}, \infty).$$

Overall, investment in an additional expert has a return

$$R(m) = V(m+1) - V(m), \tag{18}$$

which is obtained in the next period and onwards, and has a one-time opportunity cost

$$C(\pi_i, \pi_{J_{h,t}(i)}, \pi_{jt}, v_0) = u_i(\pi_{it}, \pi_{J_{h,t}(i)}, \infty) - u_i(\pi_{it}, \pi_{jt}, v_0) \tag{19}$$

that is incurred in the current period. The return $R(m)$ is positive and depends only on m. This comes from the simplifying assumption that i fully learns μ_j after observing j. The opportunity cost $C(\pi_i, \pi_{J_{h,t}(i)}, \pi_{jt}, v_0)$ depends only on the various expertise levels and v_0. Hence, i invests in the best-informed unknown expert $j \notin J_{h,t}(i) \cup \{i\}$ if and only if

$$\delta R(m) > (1-\delta)\, C(\pi_i, \pi_{J_{h,t}(i)}, \pi_{jt}, v_0).$$

One can then write the value function as

$$V(m) = U(m) + \int_{\delta R > (1-\delta)C} (\delta R(m) - (1-\delta) C(\pi_i, \pi_J, \pi_j, v_0))\, dF(\pi_i)\, dF^m(\pi_J)\, dF^{n-m-1}(\pi_j)\,. \tag{20}$$

That is, the continuation payoff under the optimal strategy is her payoff from observing the best-informed known expert $j \in J_{h,t}(i)$ from now on, plus the expected net payoff from investment whenever investment is beneficial.

When does she invest? First consider the case $v_0 > \overline{v}$. Then, the cost is negative whenever an unknown expert is highly informed while all known experts have low expertise: $C(\pi_i, a, b, v_0) < 0$. Hence, she invests in new experts with positive probability until we reach $J_{h,t}(i) = N \backslash \{i\}$. This results in long-run efficiency, as in the case of myopic behavior.

Next consider the case $v_0 < \overline{v}$. Now, the opportunity cost C is always positive:

$$C \geq C_{\min}(v_0) \equiv C(b, a, b, v_0) > 0.$$

Hence, investment is beneficial only when $R(m)$ is sufficiently high. Since $R(m) \to 0$, investment stops at some m with probability 1, regardless of n. Intuitively, if i were to invest naively by comparing the long-run benefit $\delta(U(m+1) - U(m))$ to the opportunity cost $(1-\delta)C$, she would keep investing until $\delta(U(m+1) - U(m))$ goes below the minimum cost $(1-\delta)C_{\min}(v_0)$, stopping when we reach the smallest integer m^* with

$$\delta(U(m^*+1) - U(m^*)) \leq (1-\delta) C_{\min}(v_0)\,. \tag{21}$$

Under such a strategy, almost surely, the size of the long-run portfolio is m^*. It turns out that this is also true under the optimal strategy because $V(m) = U(m)$ in the long run. Our next result establishes this fact, thus showing that forward-looking and myopic individuals behave in similar fashion in the long run.

Proposition 6. *In the simplified forward-looking model, under the optimal strategy, almost surely,*

$$j_{it}(h) \in \arg\max_{j \in J_h(i)} \pi_{jt} \qquad (\forall h, \forall t \geq \tau(h), i \in N)\,.$$

Moreover, almost surely,

$$|J_h(i)| = \begin{cases} n-1 & \text{if } v_0 > \overline{v} \\ \min\{m^*, n-1\} & \text{if } v_0 < \overline{v}, \end{cases} \tag{22}$$

where m^* is as defined in (21). In particular, when $v_0 < \overline{v}$, for every $\varepsilon > 0$, there exists $\overline{n} < \infty$ such that, almost surely,

$$\frac{|J_h(i)|}{n-1} \leq \varepsilon \qquad (\forall i \in N, \forall n > \overline{n})\,. \tag{23}$$

Long-run behavior under the optimal strategy is history-dependent, and two individuals may have very different sets of long-run experts. Individuals invest in new experts when the opportunity cost is low, which happens when their own expertise is high. Hence distinct individuals invest at distinct times, and choose different targets in general. Despite this asymmetry, Proposition 6 establishes that the number $|J_h(i)|$ of long run experts is history independent and given by (21–22) for all individuals.

For the economically interesting case of $v_0 < \overline{v}$, one can derive intuitive properties of $|J_h(i)|$ from (21). First, since m^* does not depend on n, the fraction $|J_h(i)|/(n-1)$ of individuals that are included in the set of long run experts goes to zero as n gets large, as stated in the last part of the proposition. This extends Proposition 3 to the case of forward-looking individuals. Second, m^* is increasing in the discount factor δ, yielding $|J_h(i)| = m^* = 1$ when i is very impatient, and long-run efficiency (i.e. $|J_h(i)| = n-1$) when i is very patient.[7] Third, m^* is decreasing in the minimum cost $C_{\min}(v_0)$ and hence increasing in v_0; it takes the value $n-1$ when v_0 is close to $\overline{v}$.

To see that the welfare implications of perspective unobservability can be large even with forward-looking behavior, consider the case of binomially distributed expertise: $\pi_{it} = a$ with probability $1-q$ and $\pi_{it} = b$ with probability q for some $q \in (0,1)$. In this case the long-run payoff is given by

$$U(m) = \overline{U}(b) - (1-q)^m \left(\overline{U}(b) - \overline{U}(a)\right),$$

where $(1-q)^m \left(\overline{U}(b) - \overline{U}(a)\right)$ can be viewed as the inefficiency incurred by restricting attention to $J_h(i)$ because $U(m) \to \overline{U}(b)$ as $m \to \infty$. The ex-ante payoff is $V(0)$. Since

$$V(0) \leq V(m^*) = U(m^*), \tag{24}$$

the long-run payoff provides a loose upper bound for the ex-ante payoff. The marginal contribution of an additional expert is

$$U(m+1) - U(m) = q(1-q)^m \left(\overline{U}(b) - \overline{U}(a)\right).$$

Intuitively, the additional expert contributes $\overline{U}(b) - \overline{U}(a)$ when she is the only target with high expertise, which happens with probability $q(1-q)^m$. Then, by (21), the long-run payoff is approximately

$$U(m^*) = \overline{U}(b) - \frac{1-\delta}{\delta q} C_{\min}(v_0), \tag{25}$$

up to a correction for an integer constraint.

The long-run payoff is decreasing in the minimal cost $C_{\min}(v_0)$ and hence increasing in v_0, approaching the maximum payoff $\overline{U}(b)$ as $v_0 \to \overline{v}$. It is also increasing in δ. When δ is low, the long-run outcome is highly inefficient, yielding the lowest payoff $U(1)$. On the other hand, the

[7] The result for high impatience is a direct consequence of our assumption here that perspectives are fully learned after a single period of observation; this blocks the more complex observational patterns we uncovered in the case of myopia.

long-run payoff approaches the maximum payoff $\overline{U}(b)$ as $\delta \to 1$. Interestingly, the payoff is also increasing in the probability q of high expertise. When q is low, the inefficiency $(1-\delta)C_{\min}(v_0)/\delta q$ can be arbitrarily high, and the long-run payoff is at its lowest attainable level of $U(1)$. Intuitively, when q is low, the return from investment in a previously unobserved individual becomes very small. Hence individuals settle on a small set of experts with low levels of overall expertise $1-(1-q)^m$. Of course, in large populations, there are likely to be some individuals with high expertise, and an observer could obtain nearly the highest attainable payoff $\overline{U}(b)$ in expectation if perspectives were observable. Since they are unobservable, and forming a sufficiently large portfolio is prohibitively costly, experts sets are chosen to be small.

9 Extensions and Variations

We have already considered some extensions to our baseline model, including the important case of forward-looking behavior, and briefly discussed the two-sided case (with distinct groups of observers and experts) as well as observable states. In this section we consider a few other extensions and variants of the model. Some of these are explored in detail in the supplementary appendix, while others are largely left for future research.

9.1 Observational Learning with Unobservable Preferences

In our baseline model, we use heterogeneous priors to represent perspectives and study the signal extraction problem that arises when these priors are unknown. The same underlying forces are at work in other contexts, including observational learning with payoff uncertainty and common priors.

In the supplementary appendix, we present a model of observational learning with common priors where individuals have privately known preference biases as in Crawford and Sobel (1982). Instead of observing opinions, people observe actions, but since preferences are unobserved these actions do not reveal the private signals received by targets. The resulting signal extraction problem is similar to the one considered here, but the observational learning model leads to more path dependence and less long-run efficiency. That is, we show that the cutoff $\overline{v}$ for long-efficiency is scaled up while the cutoff $\beta(v)$ for links to be broken is scaled down. Under the new cutoffs, our results extend verbatim: we have long-run efficiency with probability 1 when $v_0 > \overline{v}$ and extreme opinion leadership with positive probability when $v_0 < \overline{v}$. Subject to the revised thresholds, the probability distribution over long-run experts is constrained by the bounds $p^*(m)$ identified in Proposition 3; hence the sets of long-run experts contain a negligible fraction of all individuals in large populations.

9.2 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 is 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 10 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 supplementary appendix, we present an extension in which 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.

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. And second, we show in the supplementary appendix 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]

9.3 Correlated Perspectives

When perspectives are correlated, knowledge of one's own perspective is informative about the perspectives of others. Furthermore, observed opinions can be informative not only about the

[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 et al. (2015) show 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.

perspective of the target individual, but also about the perspectives of third parties.

Suppose that it is commonly known that the perspectives $\mu = (\mu_1, ..., \mu_n)$ are jointly distributed according to

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

where $\overline{\mu} = (\overline{\mu}_1, ..., \overline{\mu}_n)$ is the expected value of the vector of n perspectives as before, and Σ is a variance-covariance matrix with typical element σ_{ij}. Let $\sigma_{ii} = \sigma^2$ for all i in the variance-covariance matrix Σ, but allow arbitrary correlations by setting $\sigma_{ij} = \sigma_{ji} = \rho_{ij}\sigma^2$. Now there is an initial mutual attraction among individuals with highly correlated perspectives because they can deduce something about each other's perspectives from knowledge of their own. In this case even the first period network will exhibit asymmetries and opinion leadership becomes less likely. When heterogeneity in expertise levels is limited, each individual initially links with high likelihood to the target j with highest correlation ρ_{ij}. In general, this may lead to a wide range of networks, as i links to j while j links to some other individual.

However, if perspectives are highly correlated within identity groups (defined by race, ethnicity or religion for example) but largely independent across groups, then an individual who chooses a target from a different group will develop an attachment not only to the observed target but also—to a lesser degree—to others in the target's group. We conjecture that this would lead to long-run observational patterns in which there is a high density of connections within groups along with sparse but persistent links across groups, as those who step outside group boundaries once become more inclined to do so repeatedly.[9]

9.4 Multiple Targets

We have assumed throughout that each individual can observe only one target in any period. Extending our results to the case of k targets per period, for some fixed positive $k < n - 1$, is straightforward. In this case a link ij becomes free if $v_{ij}^t > \overline{v}$ (as in the baseline model) and breaks if there are k individuals $j' \in N \setminus \{i, j\}$ with $v_{ij'}^t > \beta(v_{ij}^k)$. In the long run, each individual links to her k best informed long-run experts. As in the baseline model, we have long-run efficiency when $v_0 > \overline{v}$. Extreme opinion leadership must now be redefined to mean that a set of k individuals constitute the long-run experts for all those outside this set. Given this, we obtain extreme opinion leadership with certainty when $v_0 < \tilde{v}$, and with positive probability when $\tilde{v} < v_0 < \overline{v}$. Our result concerning arbitrary static networks—with k targets for each individual—can also be extended, but the necessary condition is more demanding. And our bound on the size of long-run expert sets also extends to case of multiple targets—albeit with necessary modifications in the formula.

[9] The property that one finds signals from one's own group more informative than those from other groups is commonly treated as a primitive in the literature on statistical discrimination descended from Phelps (1972) and Aigner and Cain (1977). Allowing for correlated perspectives can deepen our understanding of the foundations and limitations of this hypothesis.

One can also extend our results to allow for a cost of observing opinions. Suppose, for instance, that individuals pay a fixed cost $c > 0$ for each opinion observed. Now there exists a cutoff $\hat{v}(c)$ such that individuals do not link to anyone when $v_0 < \hat{v}(c)$, where $\hat{v}(c)$ is obtained by equating the cost c to the benefit for an individual with expertise level a of observing the opinion of a target with expertise level b and familiarity $v_0 = \hat{v}(c)$. If $v_0 > \hat{v}(c)$, then in any given period, individuals link to a set of targets with lowest γ, and eventually choose from among their long-run experts based only on expertise levels. But the number of opinions observed must vary across individuals and over time. This is because the marginal benefit of an additional opinion depends both on the observer's expertise level in the period of observation, and extent of her familiarity with others in the population. We obtain long-run efficiency with probability 1 if and only if $v_0 > \max\{\overline{v}, \hat{v}(c)\}$.

9.5 Indirect Observation and Social Networks

We have assumed that individuals observes the opinions of their selected targets after all private signals have been received and processed, but before the targets have learned from the opinions of their own respective targets. This allows us to focus on unobservable perspectives without dealing with the orthogonal problem of double-counting associated with untangling sources of information. However, it also rules out flows of information through a chain of connected individuals. One could generalize the model by allowing for multiple rounds of opinion observation within each period. In this case the first round observation reveals something about about the target's private signal, the second round observation additionally reveals something about the private signal of the target's own target, and so on. Although this is a very promising direction for future research, we do not pursue it here. We have also assumed that any individual in the population is a potential target for anyone else, so that the resulting communication network is fully endogenous. In principle, one could impose a prior social structure on the model that restricts target choices, so that communication occurs only between social neighbors. One could then explore the question of how changes in social networks affect the size and structure of the endogenous communication subnetwork.

9.6 Observability of Targets and Actions

We have assumed throughout that individuals' 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: i can compute j_{it} using publicly available data even before j_{it} has been observed.[10] This simplifies the analysis considerably, due to the linear formula for normal variables;

[10] One can prove this inductively as follows. At $t = 1$, one can compute j_{it} from (6) using $(\pi_{1t}, \ldots \pi_{nt})$ and v_0 without observing j_{it}. Suppose now that this is indeed the case for all $t' < t$ for some t, i.e., j_{it} does not provide

see (26) 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.

10 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 identity-protective cognition (Kahan et al., 2007).

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 2008, Plambeck 2012, Voorhees 2012, Goebbert et al., 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.[11] Two especially relevant contributions from the perspective of our work are by Galeotti and Goyal (2010) and Acemoglu et al. (2014). Galeotti and Goyal

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 $(\pi_{1t'}, \ldots, \pi_{nt'})_{t'<t}$, all of which are publicly observable. That is, one knows v_{ij}^{t} for all distinct $i, j \in N$. Using $(\pi_{1t}, \ldots \pi_{nt})$ and these values, one can then compute j_{it} from (6).

[11] For a survey of the observational learning literature, see Goyal (2010). Early and influential contributions include Banerjee (1992), Bikhchandani et al. (1992), and Smith and Sorensen (2000) in the context of sequential choice. For learning in networks see Bala and Goyal (1998), Gale and Kariv (2003), DeMarzo et al. (2003), Golub and Jackson (2010), Acemoglu et al. (2011), Chatterjee and Xu (2004), and Jadbabaie et al. (2012). For surveys of the network formation literature see Bloch and Dutta (2010) and Jackson (2010). Key early contributions include Jackson and Wolinsky (1996) and Bala and Goyal (2000); see also Watts (2001), Bramoulle and Kranton (2007), and Bloch et al. (2008).

(2010) develop 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 et al. (2014) also consider 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 show 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 et al. (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 et al. (2015), who explore 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 consider 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 show 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) amongst others. Dixit and Weibull (2007) have shown that the beliefs of individuals with heterogeneous priors can diverge further upon observation of a public signal, and Acemoglu et al. (2015) that they can fail to converge even after an infinite sequence of signals. 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.

11 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. And since the extent of such attachment depends on the degree to which the observer is well-informed, there is a natural process of symmetry breaking. This allows for a broad range of networks to emerge over time, including opinion leadership and information segregation. We have shown that 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.

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

$$y = E[\theta|s] = \frac{v}{v+r}\mu + \frac{r}{v+r}s \tag{26}$$

and $w = v + r$.

As a step towards proving Proposition 2, following lemma identifies sufficient conditions for a link to be broken or free; see the supplementary appendix 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 $v_{ik}^t(h_t) > \beta\left(v_{ij}^t(h_t)\right)$.*

Proof of Lemma 2. 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 $k \notin \{i, j\}$,

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

where the first inequality is 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 $k \notin \{i, j\}$,

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

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 $v_{ik}^t(h_t) > \beta\left(v_{ij}^t(h_t)\right)$. By definition of β,

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

where the inequality is by monotonicity of γ and the equality is by definition of β. Hence, $\Pr\left(l_{ij}^t = 1|h_t\right) = 0$. 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 $\Pr\left(l_{ij}^{t+1} = 1|h_t\right) = 0$. Inductive application of the same argument shows that $\Pr\left(l_{ij}^s = 1|h_t\right) = 0$ for every $s \geq 0$, showing that the link ij is broken at h_t. □

Proof of Proposition 2. (Part 1) 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 2) Clearly, when $v_0 > \overline{v}$, the long-run outcome is history independent by Part 1, 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 $j_{1t} = 2$, as the targets are best informed and best known individuals among others. Then, on event A, for $ij \in S \equiv \{12, 21, 31, \ldots, n1\}$,

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

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 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 3 below.

(Part 3) Given the parameter values for which it applies, Part 4 implies Part 3, as there are graphs that partition N as claimed. See the supplementary appendix for a direct proof of Part 3 that applies for all $v_0 \in (\tilde{v}, \overline{v} - \underline{\Delta})$.

(Part 4) 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

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

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

$$\gamma(\pi, v_0) < \gamma(a, v_0 + \Delta(b - \varepsilon, \pi + \varepsilon)) \tag{27}$$

$$\gamma(b - \varepsilon, v_0) < \gamma(\pi + \varepsilon, v_0 + \Delta(\pi + \varepsilon, b - \varepsilon)) \tag{28}$$

$$\Delta(b - \varepsilon, \pi + \varepsilon) > \Delta(\pi + \varepsilon, b - \varepsilon). \tag{29}$$

Fix some

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

and consider the following positive probability event:

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

where

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

Note that A is open and non-empty (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 $j \in \{1, \ldots, t-1\}$ 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 (28), for any $i > t$, $\gamma\left(\pi_{t,t-1}, v_{it}^{t-1}\right) < \gamma\left(\pi_{j,t-1}, v_{ij}^{t-1}\right)$ for every $j \in N \backslash \{i, t\}$, showing that i observes t, i.e., $j_{i,t-1} = t$. Likewise, by (27), for $i = t$, $\gamma\left(\pi_{g(t),t-1}, v_{tg(t)}^{t-1}\right) < \gamma\left(\pi_{j,t-1}, v_{tj}^{t-1}\right)$ for every $j \in N \backslash \{t, g(t)\}$, showing that t observes $g(t)$, i.e., $j_{t,t-1} = g(t)$. 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 $j \in N \backslash \{i, t, g(i)\}$, $\gamma\left(\pi_{t,t-1}, v_{it}^{t-1}\right) < \gamma\left(\pi_{j,t-1}, v_{ij}^{t-1}\right)$, showing that i does not observe j, i.e., $j_{i,t-1} \in \{g(i), t\}$.

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

$$v_{ij}^n \leq v_0 + \Delta(\pi + \varepsilon, b - \varepsilon) \qquad (\forall j \neq g(i)). \tag{30}$$

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

$$v_{ig(i)}^n \geq v_0 + \Delta(b - \varepsilon, \pi + \varepsilon). \tag{31}$$

By definition of A, inequalities (30) and (31) imply that each i observes $g(i)$ at n. Consequently, the inequalities (30) and (31) also hold at date $n+1$, leading each i again to observe $g(i)$ at $n+1$, and so on. Hence, at dates $t \in \{n, \ldots, t^* + n - 1\}$, each i observes $g(i)$, yielding

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

For any $j \neq g(i)$, since $v_{ij}^{n+t^*} = v_{ij}^{n+1}$, together with (30), this implies that

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

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

Proof of Proposition 5. (Part 1) Take any h and h' with $h \succeq h'$ and any principal i. Consider first the case that $t_i(h) \geq t_i(h')$, i.e., 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

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

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:

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

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 2-3) By Part 1, 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 $\pi_{j_1 t}(h') = a < b = \pi_{j_2 t}(h')$ to $\pi_{j_2 t}(h') = a < b = \pi_{j_1 t}(h')$ 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 1). 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

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

Let n denote the largest integer such that

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

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

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

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 (which completes the proof of Part 2), and approaches 1 as $p_1 \to 1$ (proving Part 3). □

Proof of Proposition 6. It suffices to prove (22), which implies (23). To this end, first consider the case $v_0 > \overline{v}$. Since $C_{\min}(v_0) < 0$ and C is continuous, the probability of the event $\Pi_J = \left\{\pi \mid C\left(\pi_i, \pi_J, \pi_{N \setminus (J \cup \{i\})}\right) < 0\right\}$ is uniformly bounded away from zero for all proper subsets J of $N \setminus \{i\}$. Hence, the event Π_J occurs infinitely often for each J almost surely. But, whenever $\pi_t \in \Pi_{J_h(i),i}$, we have $j_{it}(h, \pi_t) \notin J_{h,t}(i)$, yielding $|J_{h,t+1}(i)| = |J_{h,t}(i)| + 1$ until $|J_{h,t}(i)| = n - 1$. Now, consider the case, $v_0 < \overline{v}$. It suffices to show that

$$\delta R(m) > (1 - \delta) C_{\min}(v_0) \iff m < m^*;$$

as in the previous case, this implies that investment occurs until we reach $\min\{m^*, n-1\}$ almost surely, and investment occurs with zero probability thereafter. Since $R(m) \to 0$, there exists some $\overline{m}$ (independent of n) such that $\delta R(m) < (1 - \delta) C_{\min}(v_0)$ whenever $m \geq \overline{m}$. Now, towards an induction, assume that $\delta R(m+1) < (1 - \delta) C_{\min}(v_0)$ for some $m \geq m^*$. Then, by (20), we have $V(m+1) = U(m+1)$. Hence,

$$\begin{aligned} \delta R(m) &= \delta (V(m+1) - V(m)) = \delta (U(m+1) - V(m)) \leq \delta (U(m+1) - U(m)) \\ &\leq (1 - \delta) C_{\min}(v_0), \end{aligned}$$

where the first equality is by definition, the next equality is by the inductive hypothesis, the next inequality is by $V(m) \geq U(m)$ and the last inequality is by $m \geq m^*$. For the converse, suppose that $\delta R(m) \leq (1 - \delta) C_{\min}(v_0)$ at some $m < m^*$. Then, $V(m) = U(m)$ and $V(m+1) \geq U(m)$, and hence

$$\delta R(m) = \delta (V(m+1) - V(m)) \geq \delta (U(m+1) - U(m)) > (1 - \delta) C_{\min}(v_0),$$

a contradiction. □

References

[1] Acemoglu, Daron, Kostas Bimpikis, Asuman Ozdaglar (2014), "Dynamics of Information Exchange in Endogenous Social Networks." *Theoretical Economics* 9: 41-97.

[2] Acemoglu, Daron, Victor Chernozhukov, and Muhamet Yildiz. (2015). "Fragility of Asymptotic Agreement under Bayesian Learning." *Theoretical Economics*, forthcoming.

[3] Acemoglu, Daron, Munther A. Dahleh, Ilan Lobel, and Asuman Ozdaglar (2011), Bayesian Learning in Social Networks, *Review of Economic Studies*, 78: 1201-1236.

[4] Aigner, Dennis J. and Cain, Glen G. (1977) "Statistical Theories of Discrimination in Labor Markets." *Industrial and Labor Relations Review* 30:175-87.

[5] Baccara, M. and Yariv, L. (2013). Homophily in Peer Groups. American Economic Journal: Microeconomics, 5(3), 69-96.

[6] Bala, Venkatesh and Sanjeev Goyal 1998, Learning from Neighbours, Review of Economic Studies 65, 595-621.

[7] Bala Venkatesh and Sanjeev Goyal (2000). A Noncooperative Model of Network Formation. Econometrica 68: 1181-1229

[8] Banerjee, Abhijit V. (1992) A Simple Model of Herd Behavior, Quarterly Journal of Economics 107: 797–817.

[9] Bikhchandani, S., Hirschleifer, D. and Welch, I. (1992), "A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades", Journal of Political Economy, 100, 992-1023.

[10] Bloch, Francis and Bhaskar Dutta (2010). Formation of Networks and Coalitions. In Handbook of Social Economics, edited by Jess Benhabib, Alberto Bisin and Matthew O. Jackson.

[11] Bloch, Francis, Garance Genicot, and Debraj Ray (2008). Informal Insurance in Social Networks. Journal of Economic Theory 143: 36-58.

[12] Bramoullé, Yann and Rachel Kranton (2007). Public Goods in Networks. Journal of Economic Theory 135: 478-494

[13] Calvó-Armengol, Antoni, Joan de Martí and Andrea Prat (2015). Communication and Influence, Theoretical Economics, 10: 649-690.

[14] Chatterjee, Kalyan and Susan Xu (2004), Technology Diffusion by learning from neighbors, Advances in Applied Probability, 36, 2, 355-376.

[15] Che, Yeon-Koo and Navin Kartik (2009). Opinions as Incentives. Journal of Political Economy 117: 815-860

[16] Cornell, Bradford and Ivo Welch (1996), Culture, Information, and Screening Discrimination, Journal of Political Economy 104: 542-571.

[17] Crawford, Vincent P., and Joel Sobel. (1982). Strategic information transmission, Econometrica 1431-1451.

[18] Dessein, W., Galeotti, A., and Santos, T. (2015). Rational inattention and organizational focus. Unpublished manuscript, Columbia Business School.

[19] DeGroot, Morris (1974), Reaching a Consensus, Journal of American Statistical Association 69: 118-121.

[20] DeMarzo, Peter, Dimitri Vayanos, Jeffrey Zweibel (2003), Persuasion bias, social influence, and unidimensional opinions, Quarterly Journal of Economics 118: 909-968.

[21] Dewan, T. and Myatt, D. P. (2008). The qualities of leadership: Direction, communication, and obfuscation. American Political Science Review, 102(03), 351-368.

[22] Dixit, Avinash K. and Jorgen W. Weibull. 2007. Political Polarization, Proceedings of the National Academy of Sciences 104: 7351-7356.

[23] Gale, Douglas and Shachar Kariv (2003), Bayesian Learning in Social Networks, Games and Economic Behavior, 45: 329-346.

[24] Galeotti, Andrea and Sanjeev Goyal (2010). The Law of the Few, American Economic Review 100: 1468-1492.

[25] Golub, Benjamin and Matthew O. Jackson (2010), Naive Learning in social networks and the Wisdom of Crowds. American Economic Journal: Microeconomics, 2: 112- 149.

[26] Goyal, Sanjeev (2010). Learning in Networks. In Handbook of Social Economics, edited by Jess Benhabib, Alberto Bisin and Matthew O. Jackson.

[27] Jackson, Matthew O. (2010). An Overview of Social Networks and Economic Applications. In Handbook of Social Economics, edited by Jess Benhabib, Alberto Bisin and Matthew O. Jackson. Elsevier.

[28] Jackson, Matthew O. and Asher Wolinsky (1996). A strategic model of social and economic networks. Journal of Economic Theory 71: 44-74.

[29] Jadbabaie, Ali, Pooya Molavi, Alvaro Sandroni, and Alireza Tahbaz-Salehi (2012). "Non-Bayesian social learning." Games and Economic Behavior 76: 210-225.

[30] 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.

[31] Phelps, Edmund S. (1972). "The Statistical Theory of Racism and Sexism." American Economic Review 62: 659-61.

[32] Plambeck, Joseph (2012), "From Jack Welch, a Conspiracy Theory." New York Times (October 5, 2012).

[33] Sethi, Rajiv and Muhamet Yildiz (2012). "Public Disagreement." American Economic Journal: Microeconomics 4: 57–95.

[34] Smith, Lones and Peter Sorensen (2000). Pathological Outcomes of Observational Learning. Econometrica 68: 371-398

[35] Thrush, Glenn (2009). "58 percent of GOP not sure/doubt Obama born in US". Politico.com. (July 31, 2009).

[36] Van den Steen, Eric (2010). Culture Clash: The Costs and Benefits of Homogeneity, Management Science 56: 1718-1738.

[37] Voorhees, Josh (2012). "The Skewed-Polling Conspiracy Theory Goes Mainstream." Slate (October 2, 2012).

[38] Watts, Alison (2001). A Dynamic Model of Network Formation. Games and Economic Behavior 34: 331-341.

Supplementary Appendix (for online publication)

In this appendix we consider some additional properties of long-run dynamics in Section A. A direct proof of Proposition 2(c) for the full range of parameter values is contained in Section B. Section C contains three variants of our model: a two-sided case in which the sets of observers and experts are disjoint, a case with delayed observability of states, and a case with common priors but unobserved preferences. A model of shifting perspectives is explored in Section D.

A Further Properties of Long-Run Dynamics

We write $b_{ij}^t = 1$ if the link ij is broken at t.

The following corollary of Proposition 1 establishes the frequency with which each $g \in G$ is realized in the long run, where $j_t = (j_{1t}, \dots, j_{nt})$ is the history-dependent network realized at time t.

Corollary 2. *If expertise levels are serially i.i.d then, almost surely, the long-run frequency*

$$\phi_\infty (g|h) = \lim_{t\to\infty} \frac{\# \{s \le t \,|\, j_s (h) = g\}}{t} \qquad (\forall g \in G)$$

exists, and

$$\phi_\infty (g|h) = P\left(g(i) = \arg\max_{j\in J(i)} \pi_j \ \forall i \in N\right).$$

When expertise levels are serially i.i.d., then the realized networks are also i.i.d. in the long-run, where the history-dependent long-run distribution is obtained by selecting the best-informed long-run expert for each i. This generates a testable prediction regarding the joint distribution of behavior in the long run: if both j and j' are elements of $J_h(i) \cap J_h(i')$, then i cannot link to j while i' links to j'. Furthermore, each pattern of linkages identified in Proposition 2 has an associated long-run distribution: long run efficiency is characterized by an i.i.d. distribution on star networks, in which all players link to one player and that player links to another; the static network g is characterized by a point mass on g, and extreme opinion leadership is characterized by a point mass on a specific star network.

We now prove Corollary 2 and establish some additional results regarding long-run behavior. Let

$$D^\lambda = \{(\pi_1, \dots, \pi_n) \mid |\pi_i - \pi_j| \le \lambda\}$$

denote the set of expertise realizations such that each pair of expertise levels are within λ of each other. For any given J, let

$$p_{J,\lambda}(g) = \Pr\left(g(i) = \arg\max_{j\in J(i)} \pi_j \ \forall i \in N \mid \pi \notin D^\lambda\right)$$

denote the conditional probability distribution on g obtained by restricting expertise realizations to lie outside the set D^λ. Finally, for any probability distribution p on G, let

$$B_\varepsilon(p) = \{q \mid |q\,(g) - p\,(g)| < \varepsilon \ \forall g \in G\}$$

denote the set of probability distributions q on G such that $q(g)$ and $p(g)$ are within ε of each other for all $g \in G$.

We say that $\phi_t(\cdot \mid h) \in B_\varepsilon(p)$ *eventually* if there exists $\bar{t}$ such that $\phi_t(\cdot \mid h) \in B_\varepsilon(p)$ for all $t > \bar{t}$. The following basic observations will also be useful in our analysis.

Observation 1. *The following are true.*

1. *For every $\varepsilon > 0$, there exists $\lambda(\varepsilon) \in (0, \varepsilon)$ such that $\Pr\left(D^{\lambda(\varepsilon)}\right) < \varepsilon$.*
2. *For every $\lambda > 0$, there exists $\overline{v}_\lambda < \infty$ such that if $v_{ij}^t > \overline{v}_\lambda$ and $\pi_{jt} > \pi_{j't} + \lambda$, then $j_{it} \neq j'$.*

The first of these observations follows from the fact that $\Pr(D^\lambda)$ is continuous and approaches 0 as $\lambda \to 0$, and the second can be readily deduced using (5).

Next we establish that, along every history, each link is eventually either broken or free. Define

$$\tilde{J}_h(i) = \left\{ j \mid \lim_{t\to\infty} v_{ij}^t(h) > \overline{v} \right\}$$

as the set of individuals j for which the link ij becomes free eventually. For any $J : N \to 2^N$ with $i \notin J(i)$, we also define

$$\tilde{H}^J = \left\{ h_t \mid v_{ij}^t(h_t) > \overline{v} \text{ and } b_{ij'}(h_t) = 1 \;\; (\forall i \in N, \forall j \in J(i), \forall j' \notin J(i)) \right\}$$

as the set of histories in which all links ij with $j \in J(i)$ are free and all links ij' with $j' \notin J(i)$ are broken. We define $\tilde{H} = \cup_J \tilde{H}^J$ as the set of all histories at which all the links are resolved in the sense that they are either free or broken. Finally, we define the stopping time $\tilde{\tau}$ as the first time the process enters $\tilde{H}$, i.e., $h_{\tilde{\tau}} \in \tilde{H}$ but $h_t \notin \tilde{H}$ for any $t < \tilde{\tau}$.

Lemma 3. *The stopping time $\tilde{\tau}$ is finite, i.e., for every h, there exists $\tilde{\tau}(h) < \infty$ such that $h_{\tilde{\tau}(h)} \in \tilde{H}$ but $h_t \notin \tilde{H}$ for all $t < \tilde{\tau}(h)$. Moreover, conditional on $h_{\tilde{\tau}}$, almost surely,*

$$J_h = \tilde{J}_h = \tilde{J}_{h_{\tilde{\tau}}}$$

where $\tilde{J}_{h_{\tilde{\tau}}}$ is uniquely defined by $h_{\tilde{\tau}} \in \tilde{H}^{\tilde{J}_{h_{\tilde{\tau}}}}$. Finally, $J_h = \tilde{J}_h$ almost surely.

Proof. Consider any h. By definition, for every $i, j \in N$ with $j \in \tilde{J}_h(i)$, the link ij becomes free for the first time at some $\tau_{ij}(h)$. Moreover, by Lemma 1, for every $i, j \in N$ with $j \in J_h(i)$, we have $\lim_t v_{ij}^t(h_t) = \infty$. Hence, by Lemma 2, for every $j' \notin \tilde{J}$, the link ij is broken for the first time at some $\tau_{ij}(h)$.[12] Therefore, $h_{\tilde{\tau}(h)} \in \tilde{H}$ for the first time at $\tilde{\tau}(h) = \max_{i\in N, j \in J_h(i)} \tau_{ij}(h)$.

To prove the second part, observe that $\tilde{J}_h = \tilde{J}_{h_{\tilde{\tau}}}$ by definition. Moreover, $J_h \subseteq \tilde{J}_h$ because $\lim_t v_{ij}^t(h_t) = \infty$ whenever $j \in J_h(i)$. It therefore suffices to show that, conditional on $h_{\tilde{\tau}}$, each i links to each $j \in \tilde{J}_{h_{\tilde{\tau}}}(i)$ infinitely often almost surely. To establish this, take any i and j with $j \in \tilde{J}_{h_{\tilde{\tau}}}(i)$. Since $v_{ij}^{\tilde{\tau}}(h_{\tilde{\tau}}) > \overline{v}$, we have

$$\gamma\left(b, v_{ij}^{\tilde{\tau}}(h_{\tilde{\tau}})\right) < \gamma(b, \overline{v}) \leq \gamma(a, v) \qquad (\forall v),$$

[12] By Observation 1 and definition of h, $\sup_t v_{ij'}^t(h) < \overline{v}$.

where the first inequality is because γ is decreasing in $v_{ij}^{\tilde{\tau}}(h_{\tilde{\tau}})$ and the second is by definition of $\overline{v}$. Hence, by continuity of γ, there exists $\eta > 0$ such that

$$\gamma\left(b-\eta, v_{ij}^{\tilde{\tau}}(h_{\tilde{\tau}})\right) < \gamma(a+\eta, v) \qquad (\forall v).$$

Since $v_{ij}^{t}(h_t) \geq v_{ij}^{\tilde{\tau}}(h_{\tilde{\tau}}) > \overline{v}$ for all continuations h_t of $h_{\tilde{\tau}}$, this further implies that

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

for every history h_t that follows $h_{\tilde{\tau}}$, for every k distinct from i and j, and for every t. Consequently, $l_{ij}^{t+1} = 1$ whenever $\pi_{jt} > b - \eta$ and $\pi_{kt} \leq a + \eta$ for all other k. Thus,

$$\Pr(l_{ij}^{t+1} = 1) \geq F(a+\eta)^{n-2}(1 - F(b-\eta)) > 0$$

after any history that follows $h_{\tilde{\tau}}$ and any date $t \geq \tilde{\tau}$. Therefore, $l_{ij}^{t+1} = 1$ occurs infinitely often almost surely conditional on $h_{\tilde{\tau}}$. The last statement of the lemma immediately follows from the first two. □

Lemma 3 establishes that at some finite (history-dependent) time $\tilde{\tau}(h)$, all the links become either free or broken and remain so thereafter. That is when the set $\tilde{J}_h(i)$ of free links along infinite history h becomes known. Although the set $J_h(i)$ of long run experts is contained in this set, some of the free-links may not be activated after a while by chance. Lemma 3 establishes that such an event has zero probability, and all the free links are activated infinitely often. In that case, after a while, all individuals learn the perspectives of their long run experts to a high degree, and the behavior approaches the long-run behavior, each individual linking to her most informed long-run expert.

Although the set of all long-run experts is known at time $\tilde{\tau}(h)$, it may take considerably longer for behavior approach the long-run limit. Towards determining such time of convergence, for an arbitrary $\varepsilon > 0$, which will measure the level of approximation, and for any $J : N \to 2^N$ with $i \notin J(i)$, define the event

$$\hat{H}^{\lambda,J} = \left\{h_t | v_{ij}^{t}(h_t) > \overline{v}_\lambda \text{ and } b_{ij'}(h_t) = 1 \ \ (\forall i \in N, \forall j \in J(i), \forall j' \notin J(i))\right\}$$

where $\overline{v}_\lambda$ is as in Observation 1. Define the event

$$H^\varepsilon = \bigcup\nolimits_J \hat{H}^{\lambda(\varepsilon),J}$$

where $\lambda(\varepsilon)$ is as defined in Observation 1. When the process is in H^ε, we will have approximately the long-run behavior as identified in Proposition 1. Define the stopping time $\hat{\tau}$ as the first time the process enters H^ε, i.e., $h_{\hat{\tau}} \in H^\varepsilon$ but $h_t \notin H^\varepsilon$ for any $t < \hat{\tau}$. Define also $J_{h_{\hat{\tau}}}$ by $h_{\hat{\tau}} \in \hat{H}^{\lambda(\varepsilon),J_{h_{\hat{\tau}}}}$; this is well-defined because such $J_{h_{\hat{\tau}}}$ is unique. As discussed above, $\hat{\tau}$ may be infinite at some histories, but the total probability of such histories is zero by the last statement of Lemma 3. When $\hat{\tau}(h)$ is finite, we can take $\hat{\tau}(h)$ as $\tau(h)$ in Proposition 1. The next proposition summarizes our findings about the long-run behavior.

Proposition 7. *For every $\varepsilon \in (0, 1/n)$, there exist a set $\Pi \subset [a,b]^n$ with $\Pr(\pi_t \in \Pi) \geq 1 - \varepsilon$ such that for all continuations h_t of all $h_{\hat{\tau}}$*

1. $j_{it}(h_t, \pi_t) \in \{j \in J_{h_{\hat{\tau}}}(i) \mid \pi_{jt} \geq \pi_{j't} - \varepsilon \ \forall j' \in J_{h_{\hat{\tau}}}(i)\}$ *for all* $i \in N$;
2. $j_{it}(h_t, \pi_t) = \arg\max_{j \in J_{h_{\hat{\tau}}}(i)} \pi_{jt}$ *for all* $i \in N$ *whenever* $\pi_t \in \Pi$;

3. $\left|\Pr\left(j_t\left(h_t\right)=g\right)-p_{J_{h_{\hat{\tau}}}}\left(g\right)\right|\leq\varepsilon$ *for all* $g\in G$;

4. $J_h=J_{h_{\hat{\tau}}}$ *conditional on* $h_{\hat{\tau}}$ *almost surely.*

Proof. Fix an arbitrary $\varepsilon>0$, and set $\Pi=[a,b]^n\setminus D^{\lambda(\varepsilon)}$. Now, the first part of the proposition is by the definition. Indeed, for any continuation h_t of $h_{\hat{\tau}}$, $v_{ij}^t\left(h_t\right)\geq\overline{v}_{\lambda(\varepsilon)}$ whenever $j\in J_{h_{\hat{\tau}}}(i)$, and the link ij is broken whenever $j\notin J_{h_{\hat{\tau}}}(i)$. Hence, the statement follows from Observation 1 and from the fact that $\lambda(\varepsilon)<\varepsilon$. The second statement also immediately follows from the first one. Now, since j_{it} differs from $\arg\max_{j\in J_{h_{\hat{\tau}}}(i)}$ only when $\pi_t\in D^{\lambda(\varepsilon)}$, we have

$$\left|\Pr\left(j_t\left(h_t\right)=g\right)-p_{J_{h_{\hat{\tau}}}}\left(g\right)\right|\leq\Pr\left(D^{\lambda(\varepsilon)}\right)<\varepsilon,$$

proving the third part. To see the fourth part, for any $j\in J_{h_{\hat{\tau}}}(i)$, observe that $\Pr\left(j=\arg\max_{j'\in J_{h_{\hat{\tau}}}(i)}\pi_{j'}\right)=1/\left|J_{h_{\hat{\tau}}}(i)\right|>1/n$. Hence, by part 3, $\Pr\left(j_{it}=j|h_{\hat{\tau}}\right)>1/n-\varepsilon>0$ for all continuations. Therefore, by Kolmogorov's zero-one law, conditional on $h_{\hat{\tau}}$, $j_{it}=j$ infinitely often, i.e., $j\in J_h(i)$, among any continuation h almost surely. □

Ignoring the zero probability event in which the set of long run experts (determined by J_h) differs from the set of eventually free links (determined by $\tilde{J}_h$), Proposition 7 can be understood as follows. At some history-dependent time $\hat{\tau}$, all individuals learn the perspectives of all their long-run experts approximately. The first part states that they link to an approximately best-informed long-run expert thereafter. The second part states that they link to precisely the best-informed long-run expert with high probability. The third part states that, thereafter, the endogenous networks are approximately independently and identically distributed with $p_{J_{h_{\hat{\tau}}}}$, the distribution generated by selecting the most informed expert $j\in J_{h_{\hat{\tau}}}(i)$ for each i. Since $p_{J_{h_{\hat{\tau}}}}$ is history dependent, from and ex-ante perspective the long-run exogenous networks are only exchangeable (i.i.d. with unknown distribution).

In the remainder of this section, we will prove Corollary 2, establishing the long-run frequency of endogenous networks. The following lemma is a key step.

Lemma 4. *For any* $\lambda\in(0,1)$, t_0, J, *and* $h_{t_0}\in\hat{H}^{\lambda,J}$ *and for any* $\varepsilon>\Pr\left(D^\lambda\right)$,

$$\Pr\left(\phi_t\left(\cdot\,|\,\cdot\right)\in B_\varepsilon\left(p_{J,\lambda}\right)\text{ eventually}\,|\,h_{t_0}\right)=1.$$

Proof. For each $g\in G$ and each continuation history h of h_{t_0}, $\phi_t\left(g\,|\,h\right)$ can be decomposed as

$$\phi_t\left(g\,|\,h\right)=\phi_{t_0}\left(g\,|\,h_{t_0}\right)\frac{t_0}{t}+\phi_{t,1}\left(g\,|\,h\right)+\phi_{t,2}\left(g\,|\,h\right)$$

where

$$\phi_{t,1}\left(g\,|\,h\right)=\frac{\#\left\{t_0<s\leq t\,|\,j_{is}\left(h\right)=g\left(i\right)\forall i\in N\text{ and }\pi_s\in D^\lambda\right\}}{t}$$

and

$$\begin{aligned}\phi_{t,2}\left(g\,|\,h\right)&=\frac{\#\left\{t_0<s\leq t\,|\,j_{is}\left(h\right)=g\left(i\right)\forall i\in N\text{ and }\pi_s\notin D^\lambda\right\}}{t}\\&=\frac{\#\left\{t_0<s\leq t\,|\,g\left(i\right)=\arg\max_{j\in J(i)}\pi_{js}\forall i\in N\text{ and }\pi_s\notin D^\lambda\right\}}{t}.\end{aligned}$$

Here, the last equality is by the hypothesis in the lemma and by the definition of $\overline{v}_\lambda$ in Observation 1. Hence, by the strong law of large numbers, as $t \to \infty$,

$$\phi_{t,2}(g\,|\,h) \to \Pr\left(g(i) = \arg\max_{j \in J(i)} \pi_{js}\;\forall i \in N \text{ and } \pi_s \notin D^\lambda\right) = p_{J,\lambda}(g)(1 - \Pr(D^\lambda)),$$

where the last equality is by definition. Thus, almost surely,

$$\begin{aligned}\limsup_t \phi_t(g\,|\,h) &= \limsup_t \phi_{t,1}(g\,|\,h) + p_{J,\lambda}(g)(1 - \Pr(D^\lambda)) \\ &\leq p_{J,\lambda}(g) + \Pr(D^\lambda),\end{aligned}$$

where the inequality follows from the fact that $\limsup_t \phi_{t,1}(g\,|\,h) \leq \Pr(D^\lambda)$, which in turn follows from the strong law of large numbers and the definition of $\phi_{t,1}$. Likewise, almost surely,

$$\begin{aligned}\liminf_t \phi_t(g\,|\,h) &= \liminf_t \phi_{t,1}(g\,|\,h) + p_{J,\lambda}(g)(1 - \Pr(D^\lambda)) \\ &\geq p_{J,\lambda}(g) - \Pr(D^\lambda),\end{aligned}$$

where the inequality follows from $\liminf_t \phi_{t,1}(g\,|\,h) \geq 0$ and $p_{J,\lambda}(g) \leq 1$. Hence for any $\varepsilon > \Pr(D^\lambda)$, for almost all continuations h of h_{t_0}, there exists $\bar{t}$ such that $\phi_t(g\,|\,h) \in (p_{J,\lambda}(g) - \varepsilon, p_{J,\lambda}(g) + \varepsilon)$ for all g. That is, $\phi_t(\cdot|h) \in B_\varepsilon(p_{J,\lambda})$ eventually, almost surely. □

Proof of Corollary 2. Ignore the zero probability event in which $\tilde{J}_h \neq J_h$ and $\hat{\tau}$ is infinite (see Lemma 3). Then, by the third part of Proposition 7, $J_h = J_{h_{\hat{\tau}}}$ almost surely, where $h_{\hat{\tau}}$ is the truncation of h to the time the process enters H^ε (along h). Define

$$\hat{H}^\varepsilon = \{h \in H \,|\, \phi_t(\cdot\,|\,h) \in B_{2\varepsilon}(p_{J_h}) \text{ eventually}\},$$

and observe that $\phi_t(\cdot\,|\,h) \in B_{2\varepsilon}(p_{J_h})$ whenever $\phi_t(\cdot\,|\,h) \in B_\varepsilon\left(p_{J_h,\lambda(\varepsilon)}\right)$. But Lemma 4 states that, conditional on $h_{\hat{\tau}}$, $\phi_t(\cdot\,|\,h) \in B_\varepsilon\left(p_{J_h,\lambda(\varepsilon)}\right)$ eventually with probability 1. That is, $\Pr(\hat{H}^\varepsilon|h_{\hat{\tau}}) = 1$ for each $h_{\hat{\tau}}$. Therefore,

$$\Pr(\hat{H}^\varepsilon) = 1.$$

Clearly, $\hat{H}^\varepsilon$ is increasing in ε, and as $\varepsilon \to 0$,

$$\hat{H}^\varepsilon \to \hat{H}^0 = \{h \in H \,|\, \phi_t(\cdot\,|\,h) \to p_{J_h}\}.$$

Therefore,

$$\Pr(\hat{H}^0) = \lim_{\varepsilon \to 0} \Pr(\hat{H}^\varepsilon) = 1.$$

□

B Information Segregation

We next present a direct proof of Proposition 2(c), which applies to the full range of parameter values claimed in the result.

Proof of Proposition 2(c). Take any $v_0 \in (\tilde{v}, \overline{v} - \underline{\delta})$ and any partition $\{S_1, \ldots, S_m\}$ 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 $\{S_1, \ldots, S_m\}$. Since $v_0 \in (\tilde{v}, \overline{v} - \underline{\delta})$, there exists a small $\varepsilon > 0$ such that

$$v_0 + \delta(a + \varepsilon, b - \varepsilon) < \min\{\beta(v_0), \overline{v}\} \tag{32}$$

and

$$\delta(b - \varepsilon, b) > \delta(a + \varepsilon, b - \varepsilon). \tag{33}$$

By (33) and by continuity and monotonicity properties of γ, there also exist $\pi^* \in (a, b)$ and $\varepsilon' > 0$ such that

$$\begin{aligned} \gamma(\pi^* - \varepsilon', v_0 + \delta(b - \varepsilon, b)) &< \gamma(b, v_0) \\ \gamma(\pi^* + \varepsilon', v_0 + \delta(a + \varepsilon, b - \varepsilon)) &> \gamma(b - \varepsilon, v_0). \end{aligned} \tag{34}$$

For every $t \in \{2, \ldots, m\}$, the realized expertise levels are as follows:

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

Fixing

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

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

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

The above event has clearly positive probability. We will 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,

$$v_{ii_1}^2 \geq v_0 + \delta(b - \varepsilon, b) > v_0 + \delta(a + \varepsilon, b - \varepsilon) \geq v_{ji_1}^2 \qquad (\forall i \in S_1, \forall j \notin S_1),$$

where the strict inequality is by (33).Therefore, by (34), at $t = 2$, each $i \in S_1$ sticks to her previous link

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

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

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

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

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

Hence, by (34),

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

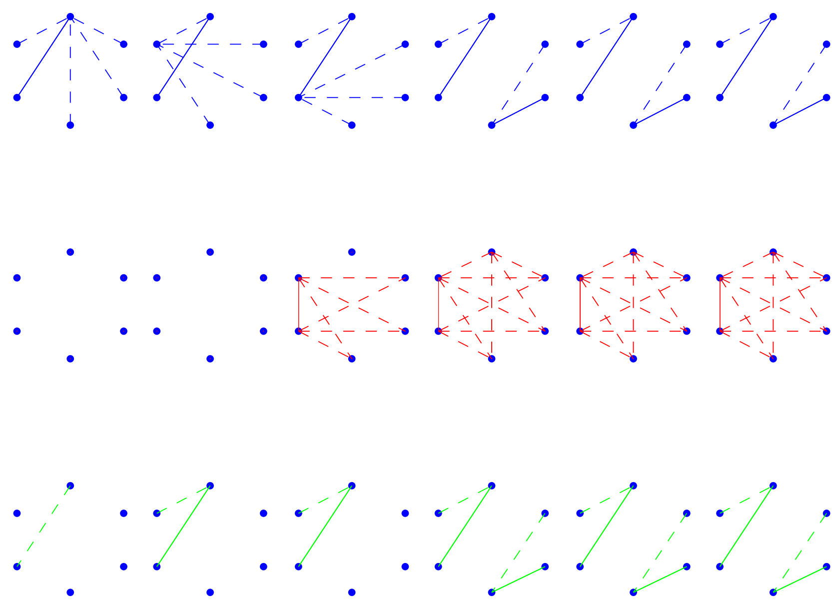

Figure 4: Emergence of segregated clusters. Each column corresponds to a period. The top row shows links formed, the middle row links broken, and the bottom row links free

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,

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

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

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

Hence, by (34), i sticks to j_{im} by date $m + t^*$, yielding

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

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. □

The forces that give rise to information segregation can be understood by considering the example depicted in Figure 4, where two segregated clusters of equal size emerge in a population of size 6. Reading anti-clockwise from the top, nodes 1, 2 and 3 are the best informed, respectively, in the first three periods. After period 4, all links from this cluster to the nodes 4–6 are broken. Following this, the nodes 4–6 are best informed and link to each other, but receive no incoming links. Although the network is not yet resolved by the end of the sixth period, it is clear that segregation can arise with positive probability because any finite repetition of the period 6 network has positive probability, and all links across the two clusters must break after a finite number of such repetitions. Here a very particular pattern of expertise realizations is required to generate segregation, and segregation over a different partition may require a very different set of expertise realizations. Nevertheless, *any* partition of the population into segregated clusters can arise with positive probability.

C Variations of the Model

The Two-Sided Case

Suppose that the set N of individuals is partitioned to two disjoint subsets: a set N_D of *decision makers*, and a set N_E of *potential experts*. Only decision makers make observational choices, and they can observe only potential experts. The domain and the range of graphs are modified accordingly; for example $j_{it} \in N_E$, and it is defined only for $i \in N_D$. The definitions of the various patterns of long-run behavior are also adjusted accordingly. For example, opinion leadership is defined by $J_h(i) = \{j^*\}$ for all $i \in N_D$, and long-run efficiency is defined by $J_h(i) = N_E$ for all $i \in N_D$. In all other respect the model is exactly as in the baseline case.

Our results concerning the behavior of a single individual clearly apply also to this variation. This includes our characterization of long run behavior in Proposition 1, our bound on the expected number long run experts in Proposition 3, and our characterization of the long-run behavior of forward looking individuals in Proposition 6. The following result presents a crisper version of Proposition 2 for the two-sided model. In this version, within $(\tilde{v}, \overline{v} - \Delta(a,b))$, every graph emerges as a stable network with positive probability. Since the networks that involve seggregation cannot arise outside of this region, this yields a sharp characterization.

Proposition 8. *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 every $v_0 \in (\tilde{v}, \overline{v} - \underline{\Delta})$, every $g : N_D \to N_E$ emerges as a static network with positive probability.

Proof. The proofs of parts (a) and (b) are as in the one-sided model. The proof of part (c) is as follows. Fix any $v_0 \in (\tilde{v}, \overline{v} - \underline{\Delta})$, where $v_0 + \Delta(a,b) < \min\{\beta(v_0), \overline{v}\}$. By continuity of Δ and γ and by definition of β, there exists $\varepsilon \in (0, (b-a)/3)$ such that $v_0 + \Delta(a+\varepsilon, b-\varepsilon) < \min\{\beta(v_0), \overline{v}\}$,

$$\Delta(b-\varepsilon, b) > \Delta(a+\varepsilon, b-\varepsilon), \tag{35}$$

and

$$\gamma(b-\varepsilon, v_0) < \gamma(a+\varepsilon, v_0 + \Delta(a+\varepsilon, b-\varepsilon)). \tag{36}$$

Fix any such ε. Finally, fix any $g : N_D \to N_E$ and denote $g(N_D) = \{j_0, \ldots, j_k\}$. Consider the following event Π: At any $t = 0, \ldots, k$, the expertise levels of j_t and all i with $g(i) = j_t$ are greater than $b - \varepsilon$, and the expertise levels of all other individuals are less than $a+\varepsilon$. For $t=k+1,\ldots,k+K$ with $v_0+K\Delta(a,b)>\beta(v_0+\Delta(a+\varepsilon,b-\varepsilon))$, all the expertise levels are in a neighborhood of the diagonal such that

$$\gamma(\pi_{jt},v_0+\Delta(b-\varepsilon,b))<\gamma(\pi_{j't},v_0+\Delta(a+\varepsilon,b-\varepsilon)) \tag{37}$$

for all $j,j'\in N_E$. There is such an open non-empty neighborhood by (35). Now, at any $t=0,\ldots,k$, if $j_{it}=j_t$, then $v^{t+1}_{ij_{it}}>v^t_{ij_{it}}+\Delta(b-\varepsilon,b)$ when $g(i)=j_t$ and $v^{t+1}_{ij_{it}}<v^t_{ij_{it}}+\Delta(a+\varepsilon,b-\varepsilon)$ when $g(i)\neq j_t$. Hence, by (36), we have $j_{it}=j_t$ for all i with $g(i)=j_{k'}$ with $k'\geq t$, and $j_{it}\in\{g(i),j_t\}$ for all other i. Thus, $v^{k+1}_{ig(i)}>v_0+\Delta(b-\varepsilon,b)$ and $v^{k+1}_{ij}<v_0+\Delta(a+\varepsilon,b-\varepsilon)$ for all i and $j\neq g(i)$. Then, by (37), $j_{it}=g(i)$ for all i and all $t=k+1,\ldots,k+K$. Therefore, all the links ij with $j\neq g(i)$ are broken at $k+K+1$, and $J_h(i)=\{g(i)\}$ for all i and all $h\in\Pi$.

□

Using the ideas in the previous proofs, the following result delineates a subset of $(\tilde{v},\overline{v})$ on which every non-empty correspondence $J:N_D\rightrightarrows N_E$ arises with positive probability. That is, one cannot say more than Proposition 1 about the behavior that arises with positive probability in the long run.

Proposition 9. *Assume that there exists an integer m such that*

$$v_0+m\Delta(a,b) < \overline{v}, \tag{38}$$
$$v_0+m\Delta(b,b) > \overline{v}, \tag{39}$$
$$v_0+\Delta(b,b) < \beta(v_0). \tag{40}$$

Then, for every non-empty $J:N_D\rightrightarrows N_E$, there exists a positive probability event on which $J_h=J$.

Proof. Given the stated assumptions, there clearly exists $\varepsilon\in(0,(b-a)/3)$ such that

$$v_0+m\Delta(a+\varepsilon,b-\varepsilon) < \overline{v}, \tag{41}$$
$$v_0+m\Delta(b-2\varepsilon,b) > \overline{v}, \tag{42}$$
$$v_0+\Delta(b-\varepsilon,b-\varepsilon) < \beta(v_0), \tag{43}$$
$$\gamma(a+\varepsilon,v_0+\Delta(b-\varepsilon,b-\varepsilon)) > \gamma(b-\varepsilon,v_0), \tag{44}$$

where (43) and (44) follow from (40). Fix any non-empty correspondence $J:N_D\rightrightarrows N_E$, and define $J(N_D)=\cup_{i\in N_D}J(i)$. For the first $m|J(N_D)|$ periods, consider the periodic sequence of experts $j^*_t\in J(N_D)$ obtained by cycling through the members of $J(N_D)$. That is, j^*_1 is the first member of $J(N_D)$, j^*_2 is the second member of $J(N_D)$, $\ldots$, $j^*_{|J(N_D)|}$ is the last member of $J(N_D)$, and $j^*_{k|J(N_D)|+t}=j^{*t}_t$. At any $t\leq m|J(N_D)|$, we have

$$\pi_{j,t}\in\begin{cases}[b-\varepsilon,b] & \text{if } j=j^*_t\\ [b-2\varepsilon,b-\varepsilon) & \text{if } j\in J^{-1}(j^*)\equiv\{i\in N_E|J(i)=j^*\}\\ [a,a+\varepsilon] & \text{otherwise.}\end{cases}$$

That is, j^*_t has the highest expertise, the decision makers who would have j^* a long-run expert according to J have the second highest level of expertise, and all the other individuals have the lowest level of expertise. Note that that, by (43), in the first iteration of the cycle (the first $|J(N_D)|$ periods), we have $j_{it}=j^*_t$ for each $i \in N_E$. Since $\beta(v) - v$ is non-decreasing, this further implies that $j_{it} = j_t^*$ for each $i \in N_E$ at every $t \leq m\,|J(N_D)|$. Therefore, by the definition of m, at the end of period $m\,|J(N_D)|$, we have the link ij free (i.e. $v_{ij} > \overline{v}$) if and only if $j \in J(i)$. □

Observable States

Next, we consider the possibility that states are publicly observable with some delay. In particular, we assume that there exists $\tau \geq 0$ such that, for all t, θ_t becomes publicly observable at the end of period $t+\tau$. Note that $\tau = 0$ corresponds to observability of θ_t at the end of period t itself, as would be the case if one's own payoffs were immediately known. At the other extreme is the case where the state is never observed (as in our baseline model), which corresponds to the limit $\tau = \infty$.

With observable states, given any history at the beginning of date t, the precision of the belief of an individual i about the perspective of individual j is

$$v_{ij\tau}^t = v_{ij}^0 + \sum_{\{t'<t-\tau : j_{it'}=j\}} 1/\pi_{jt'} + \sum_{\{t-\tau \leq t' < t : j_{it'}=j\}} \Delta\left(\pi_{it'}, \pi_{jt'}\right). \tag{45}$$

For $t' < t - \tau$, individual i retrospectively updates her belief about the perspective of her target j at t' by using the true value of $\theta_{t'}$ instead of her private signal $x_{it'}$. This causes her belief about j's perspective to become more precise, rising by $1/\pi_{jt'}$ instead of $\Delta(\pi_{it'}, \pi_{jt'})$. Note that knowledge of the state does not imply knowledge of a target's perspective, since the target's signal remains unobserved.

This is the main effect of observability of past states: it 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 one individual's behavior, such as the characterization of long-run behavior in Proposition 1 and the bound on the expected number of long run experts in Proposition 3. Nor does it affect results concerning patterns of behavior that are symmetric on the observer side, such as long-run efficiency and opinion leadership in the first two parts of Proposition 2.[13]

Observability of states has a second effect, which relates to the asymmetry of observers. For $t' < t - \tau$, since an individual i already observes the true state $\theta_{t'}$, her signal $x_{it'}$ does not affect her beliefs at any *fixed* history, as seen in (45). Consequently, two individuals with identical observational histories have identical beliefs about the perspectives of all targets observed before $t - \tau$. This makes asymmetric linkage patterns, such as non-star-shaped static networks and information segregation, less likely to emerge. Nevertheless, when $\tau > 0$, individuals *do* use their private information in selecting targets until the state is observed. Therefore, under delayed observability, individuals' private signals do impact their target choices, leading

[13] To be precise, with observable states we have long-run efficiency whenever $v_0 > \overline{v}$, opinion leadership with positive probability when $v_0 < \overline{v}$, and opinion leadership with probability 1 when $v_0 < \tilde{v}$ (as in the first two parts of Proposition 2). However the probability of opinion leadership may be 1 even when $v_0 > \tilde{v}$. Indeed, when $\tau = 0$, opinion leadership emerges with probability 1 whenever $v_0 < \tilde{v}'$, where $\tilde{v}' > \tilde{v}$ is defined by $\beta(\tilde{v}') - \tilde{v}' = 1/b$.

them to possibly different paths of observed targets. Indeed, our results about information segregation and static networks extend to the case of delayed observability for a sufficiently long delay τ.

Specifically, for a sufficiently long delay, every network emerges as the static network with positive probability. The reason for this is quite straightforward. Without observability, on a history under which g emerges as a static network, individuals become attached to their respective targets under g arbitrarily strongly over time. Hence, even if individuals start observing past states and learn more about other targets, the new information will not be sufficient to mend those broken links once enough time has elapsed. Moreover, for any partition of the population into sets of two or more individuals, there exists some $g \in G$ that maps each player i to a member of the same set in the partition. In this case we have information segregation over the given partition.

To summarize, allowing for the observability of states with some delay does not alter the main message of this paper, and in some cases gives it greater force. The trade-off between being well-informed and being well-understood has interesting dynamic implications because those whom we observe become better understood by us over time. This effect is strengthened when a state is subsequently observed, since an even sharper signal of a target's perspective is obtained.

Observational Learning with Unobservable Preferences

There is a close connection between communication with unobservable priors and observational learning with payoff uncertainty, which we explore next. Modify the information structure in our model by imposing a common-prior: the states θ_t are independently and identically distributed with $N(0,1)$, i.e., it is common knowledge that $\mu_1 = \cdots = \mu_n = 0$; the signals x_{it} and the expertise levels π_{it} are as in model. Rather than having heterogenous priors, the individuals have heterogenous preferences now:

$$u_i(a_{it}, \theta_t) = -(a_{it} - \theta_t - B_i)$$

where $a_{it} \in R$ is the action of i at t and B_i is a preference parameter (namely preferncial bias) of i, privately known by i and i.i.d. with $N(0, 1/v_0)$—as in Crawford and Sobel (1982); in our model we set $B_1 = \cdots = B_n = 0$. As in the two sided model, partition the set N of individuals to decision makers N_D and experts N_E. Each expert j takes her action a_{jt} based on her information x_{jt}. Each decision maker i can (only) observe the action $a_{j_{it}}$ of her target before taking her action a_{it}.

By substituing B_j and a_{jt} for μ_j and y_{jt}, respectively, one can apply our results to this model as follows. Now,

$$a_{jt} = B_j + \frac{\pi_{jt}}{1+\pi_{jt}} x_{jt}. \tag{46}$$

Accordingly, as a signal for θ_t, the noise variance in a_{jt} is

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

Viewed as a signal for B_j, (46) leads to $v_{ij}^{t+1} = v_{ij}^t + \Delta(\pi_{it}, \pi_{jt})$ where

$$\Delta(\pi_{it}, \pi_{jt}) = \frac{1+\pi_{it}}{\pi_{jt}(1+\pi_{it}+\pi_{jt})} (1+\pi_{jt})^2.$$

Because of slight difference in (46), the functions γ and Δ are slightly different from their counterparts in our models. This leads to a scaling of the key variables

$$\begin{aligned}\overline{v} &= \frac{a}{b\,(b-a)}\,(b+1)^2 \\ \beta\,(v) &= \left(\frac{b}{a}\right)^2 \frac{\overline{v}v}{\overline{v}-v}\left(\frac{1+a}{1+b}\right)^2.\end{aligned}$$

Here, the last multiplicative term in each formula is new, scaling up $\overline{v}$ by $(b+1)^2$ and scaling down $\beta\,(v)$ by $((1+a)/(1+b))^2$. This implies a higher threshold for links to be free, and links break with greater likelihood, resulting to an increased probability of path dependence. Our qualitative results, however, remain intact. Once again, we have long-run efficiency with probability 1 when $v_0 > \overline{v}$ and extreme opinion leadership with positive probability $v_0 < \overline{v}$ (although the cutoff $\overline{v}$ is now higher). Our Proposition 3 extends to this model verbatim

$$\Pr\left(|J_h\,(i)| \leq m\right) \geq p^*\,(m) = \left(\frac{1-F\,(\hat{\pi})}{1-F\,(\hat{\pi})+mF\,(\hat{\pi})^m}\right)^{\frac{\beta(v_0)-v_0}{\Delta(a,b)}},$$

but of course using the new functions β, γ, and Δ, leading to smaller sets of long-run experts. Likewise, our Proposition 6 extends to the new model verbatim with the new utility and cost functions.

D Shifting Perspectives

In our baseline model we assumed that all perspectives were fixed: each individual assumes that θ_t is i.i.d. with a known distribution, and does not update her beliefs about this distribution as she observes realizations of θ_t or signals about θ_t. We now consider the possibility that individuals recognize that they do not know the mean of θ_t and update their perspectives over time. We take

$$\theta_t = \mu + z_t \tag{47}$$

where the random variables $(\mu, z_1, z_2, \ldots)$ are stochastically independent and

$$\mu \sim_i N\,(\mu_{i0}, 1)\,, \tag{48}$$

$$z_t \sim N\,(0, 1/\alpha_0)\,. \tag{49}$$

Recall that $\sim_i$ indicates the belief of individual i, who believes that θ_t is i.i.d. with mean μ and variance 1, but does not know the mean μ; she believes—initially—that μ is normally distributed with mean μ_{i0} and precision α_0. We refer to the mean μ_{i0} as the *initial perspective* of i, and to the precision α_0 as the *initial firmness* of her perspective. We assume that θ_t is publicly observed at the end of the period t—as in the case of $\tau = 0$ in our discussion of observable states. This simplifies the analysis because individuals update their perspectives purely based on θ_t, rather than the signals and opinions they observe at t.

At the end of period t, the perspective of an individual i is

$$\mu_{it} \equiv E\,(\mu\,|\,\theta_1, \ldots, \theta_t) = \frac{\alpha_0}{\alpha_0+t}\mu_{i0} + \frac{t}{\alpha_0+t}\frac{\theta_1+\cdots+\theta_t}{t}, \tag{50}$$

and the firmness of her perspective (i.e. the precision of the belief about μ) is

$$\alpha_t \equiv \alpha_0 + t. \tag{51}$$

Note that the perspective μ_{it} is a convex combination of the initial perspective μ_{i0} and the empirical average $\overline{\theta}_t = (\theta_1 + \cdots + \theta_t)/t$ of the realized states with deterministic weights α_0/α_t and t/α_t, respectively.[14] As time progresses, the weight α_0/α_t of the initial perspective decreases and eventually vanishes, while the weight t/α_t of the empirical average approaches 1.

Individuals other than i have two sources of information about the revised perspective μ_{it}: (i) the past opinions of i, which are endogenously and privately observed and are the only sources of information about the initial perspective μ_{i0}, and (ii) the realization of past states, which are exogenously and publicly observed and provide information about the data that individual i uses to update her perspective.

In earlier periods, their main information comes endogenously from the first source, as in our baseline model. However, in the long-run, the accumulated data coming from the public source dominates the privately obtained information, as the perspective approaches the empirical average of the realized states. They eventually learn the perspective of each individual so precisely that they choose their targets based purely on expertise levels, as in the case of known perspectives. Hence efficiency is the only possible outcome in the long run. Nevertheless, the speed of convergence is highly dependent on the initial firmness α_0 of each perspective. The long-run behavior can be postponed indefinitely by considering firmer and firmer initial perspectives. Moreover, under such firm perspectives, the belief dynamics will also be similar to those in our baseline model, which corresponds to infinite initial firmness. Hence, the behavior will be similar to the long-run behavior in the baseline model in those arbitrarily long stretches of time before the perspectives are learned—as we establish below.

Effect of Learning on Choosing Targets

We next describe how individuals choose their targets. We show that, in comparison with the baseline model, there is a stronger motive to listen to better-understood targets vis-a-vis better-informed ones, and this motive decreases over time and approaches the baseline model in the limit. That is, learning *strengthens* path dependence early on.

At the beginning of period t, the belief of any individual j about the state is

$$\theta_t \sim_j N\left(\mu_{j(t-1)}, 1 + 1/\alpha_{t-1}\right).$$

This is as in our baseline model, except that the individual faces additional uncertainty about the underlying distribution, so the variance is $1 + 1/\alpha_{t-1}$ rather than 1. Her opinion is accordingly

$$y_{jt} = \frac{1}{1+\hat{\pi}_{jt}}\mu_j + \frac{\hat{\pi}_{jt}}{1+\hat{\pi}_{jt}}x_{jt}$$

where

$$\hat{\pi}_{jt} = \pi_{jt}/\left(1 + 1/\alpha_{t-1}\right). \tag{52}$$

That is, opinions are formed as in our baseline model, but with individuals having effectively lower expertise, reflecting their uncertainty about the underlying process. The effective expertise level $\hat{\pi}_{jt}$ approaches the

[14] Note also that the perspective μ_{it} is a random variable, as it depends on an empirical average, while its firmness α_t is deterministic and increasing in t.

nominal expertise level π_{jt} of our baseline model as $t \to \infty$. This modification can be incorporated into our earlier analysis by modifying the distribution of expertise levels at each period, taking the bounds of expertise to be

$$a_t = a/\left(1 + 1/\alpha_{t-1}\right) \text{ and } b_t = b/\left(1 + 1/\alpha_{t-1}\right), \tag{53}$$

which converge to the original bounds a and b, respectively, as $t \to \infty$.

Writing v_{ij}^t for the precision of the belief of i about the perspective $\mu_{j(t-1)}$ of player j at the beginning period t, the opinion y_{jt} provides a noisy signal

$$\frac{1 + \hat{\pi}_{jt}}{\hat{\pi}_{jt}} y_{jt} = \theta_t + \varepsilon_{jt} + \frac{1}{\hat{\pi}_{jt}} \mu_{j(t-1)},$$

as in the baseline model. Once again, the variance of the additive noise in the signal observed by i is

$$\gamma(\hat{\pi}_{jt}, v_{ij}^t) \equiv \frac{1}{\hat{\pi}_{jt}} + \frac{1}{\hat{\pi}_{jt}^2} \frac{1}{v_{ij}^t}.$$

This leads to the same behavior as in the baseline model, with effective expertise replacing nominal expertise:

$$j_{it} = \min \left\{ \arg\min_{j \neq i} \gamma(\hat{\pi}_{jt}, v_{ij}^t) \right\}. \tag{54}$$

Here, individual i simply discounts the expertise levels of her potential targets, making expertise less valuable vis-à-vis familiarity, tilting the scale towards better-understood targets. That is, in the short run, learning actually increases the attachment to previously observed targets, leading to *stronger* path-dependence. Towards stating this formally, we define the marginal rate of substitution of expertise level π_{jt} for the precision v_{ij}^t of variance at t as

$$MRS_{\pi,v}^t \equiv -\frac{\partial \gamma(\hat{\pi}_{jt}, v_{ij}^t)/\partial \pi_{jt}}{\partial \gamma(\hat{\pi}_{jt}, v_{ij}^t)/\partial v_{ij}^t} = \frac{1}{1 + \alpha_{t-1}} + 2v_{ij}^t/\partial \pi_{jt}.$$

In the baseline model, the marginal rate of substitution is

$$\overline{MRS}_{\pi,v} \equiv -\frac{\partial \gamma(\pi_{jt}, v_{ij}^t)/\partial \pi_{jt}}{\partial \gamma(\pi_{jt}, v_{ij}^t)/\partial v_{ij}^t} = 1 + 2v_{ij}^t/\partial \pi_{jt}.$$

The following proposition immediately follows from the above expressions.

Proposition 10. *The marginal rate of substitution of π_{jt} for the precision v_{ij}^t of variance is higher in the model with learning:*

$$MRS_{\pi,v}^t > \overline{MRS}_{\pi,v}.$$

Moreover, $MRS_{\pi,v}^t$ is decreasing in t and converges to $\overline{MRS}_{\pi,v}$ as $t \to \infty$. In particular, for any i, j and j' with fixed $v_{ij}^t > v_{ij'}^t$, if i prefers j to j' at t in the baseline model, she also prefers j to j' at t in the model with learning.

That is, learning increases attachment to familiar targets at the expense of more informed ones when we fix beliefs about the other players' perspectives. However, the beliefs about the other players perspectives is different under learning—as we show next.

Belief Dynamics

The updating of beliefs about perspectives is somewhat more interesting. At any history h at the beginning of a date t, write $v_{ij}^t = 1/Var\left(\mu_{j(t-1)}\,|\,h\right)$ and $v_{ij0}^t(h) = 1/Var\left(\mu_{j0}\,|\,h\right)$ for the precisions of the beliefs of i about the current and the initial perspective of j, respectively. By (50), the variance of the current perspective $\mu_{j(t-1)}$ is

$$Var\left(\mu_{j(t-1)}\,|\,h\right) = \frac{1}{v_{ij0}^t(h)}\left(\frac{\alpha_0}{\alpha_t}\right)^2 + \frac{1}{t}\left(\frac{t}{\alpha_t}\right)^2. \tag{55}$$

Here, the first term reflects the uncertainty about the initial perspective and depends on previous observations. The second term reflects the flow of public information, depending only on time and the firmness of beliefs. Then, the precision of beliefs about the current perspective is

$$v_{ij}^t = 1/Var\left(\mu_{j(t-1)}|h\right) = \frac{\alpha_t^2 v_{ij0}^t(h)}{\alpha_0^2 + t v_{ij0}^t(h)}. \tag{56}$$

Since $\alpha_t = \alpha_0 + t$, we observe that v_{ij}^t is an increasing function of both v_{ij0}^t (reflecting the information gathered by observing the opinions directly) and t (reflecting the flow of public information).

To determine v_{ij}^t, we next determine v_{ij0}^t. As in the baseline model, observation of y_{jt} by i provides the following signal about $\mu_{j(t-1)}$:

$$(1+\hat{\pi}_{jt})y_{jt} - \hat{\pi}_{jt}\theta_t = \mu_{j(t-1)} + \hat{\pi}_{jt}\varepsilon_{jt}.$$

Together with (50), this leads to the following signal about μ_{j0}:

$$\frac{\alpha_{t-1}}{\alpha_0}(1+\hat{\pi}_{jt})y_{jt} - \left[\frac{\theta_1+\cdots+\theta_{t-1}}{\alpha_0} + \frac{\alpha_{t-1}}{\alpha_0}\hat{\pi}_{jt}\theta_t\right] = \mu_{j0} + \frac{\alpha_{t-1}}{\alpha_0}\hat{\pi}_{jt}\varepsilon_{jt}.$$

The precision of this signal (i.e., the inverse of the variance of the additive noise term $\frac{a_{t-1}}{\alpha_0}\hat{\pi}_{jt}\varepsilon_{jt}$) is

$$\Delta_t(\pi_{jt}) = \frac{\alpha_0^2\left(1+\alpha_{t-1}\right)^2}{\alpha_{t-1}^4}\frac{1}{\pi_{jt}}.$$

As in the baseline model, this leads to the following closed-form solution:

$$v_{ij0}^t = v_0 + \sum_{s=1}^{t-1}\Delta_s(\pi_{js})l_{ij}^s, \tag{57}$$

where l_{ij}^s is 1 if i links to j at s and 0 otherwise. By substituting (57) into (56), we also obtain a formula for v_{ij}^t.

Note that one applies to each variance $1/\pi_{js}$ a weight that is decreasing in s, where the weight is approximately $\left(\alpha_0/\alpha_{s-1}\right)^2$. That is, earlier observations add more precision to the belief about μ_{j0}—because those opinions reflect the initial perspective more strongly. As time progresses, the opinions become less valuable sources of information about the initial perspective, and their impact on v_{ij0}^{t+1} eventually becomes negligible. Note also that the precisions v_{ij0}^t are uniformly bounded from above.

Long-run Behavior and the Robustness of the Baseline Model

Since the precisions v_{ij0}^t are uniformly bounded, the long-run behavior is driven by the flow of public information. Indeed, by (56), when $t \gg \alpha_0$, v_{ij}^t is approximately t, regardless of the history. Hence, for any fixed α_0, as $t \to \infty$, v_{ij}^t also goes to ∞, leading to long-run efficiency.

Proposition 11. *In the model with learning, for any fixed α_0, we have long-run efficiency.*

Proof. Since $v_{ij0}^t \geq v_0$ and v_{ij}^t is an increasing function of v_{ij0}^t, we can conclude from (56) that

$$v_{ij}^t(h) \geq \frac{\alpha_t^2 v_0}{\alpha_0^2 + t v_0} = \frac{(\alpha_0 + t)^2 v_0}{\alpha_0^2 + t v_0} \qquad (\forall h \in H).$$

Hence, for every history $h \in H$,

$$\lim_{t \to \infty} v_{ij}^t(h) \geq \lim_{t \to \infty} \frac{(\alpha_0 + t)^2 v_0}{\alpha_0^2 + t v_0} = \infty.$$

□

Despite this, the long-run outcome can be postponed indefinitely by choosing firmer initial beliefs, and the medium-run behavior is similar to the long-run behavior of our baseline model. In a nutshell, this can be gleaned from (56) and (57) as follows. In (56), for any fixed $\left(t, v_{ij0}^t\right)$, $\lim_{\alpha_0 \to \infty} v_{ij}^t = v_{ij0}^t$. At the same time, in (57),

$$\lim_{\alpha_0 \to \infty} v_{ij0}^t = v_0 + \sum_{s=1}^{t-1} l_{ij}^s / \pi_{js} \equiv \overline{v}_{ij}^t,$$

where $\overline{v}_{ij}^t$ is the precision of the belief of i about the perspective of j in the baseline model with observable states. Hence,

$$\lim_{\alpha_0 \to \infty} v_{ij}^t = v_0 + \sum_{s=1}^{t-1} l_{ij}^s / \pi_{js} \equiv \overline{v}_{ij}^t. \tag{58}$$

That is, the beliefs as functions of past behavior remain close to those in the baseline model when the initial conviction in perspectives is sufficiently high. Since past behavior is also a function of past beliefs, this further implies that the behavior and the beliefs remain close to their counterparts in the baseline model. This is formalized in the next result. The first part states that, with high probability, the individuals choose their target between dates $\overline{t}$ and $\overline{t} + l$ according to the long-run behavior without learning. The second part states then that the possible patterns of behavior within that (arbitrarily long) time interval coincide with the patterns possible under the long-run behavior of the baseline model.

Proposition 12. *For every $\varepsilon > 0$ and any positive integer l, there exists a date $\overline{t} < \infty$ and $\overline{\alpha} < \infty$ such that*

$$\Pr\left(j_{it}(h) \in \arg\max_{j \in \overline{J}_h(i)} \pi_{jt} \qquad \forall t \in \left\{\overline{t}, \overline{t}+1, \ldots, \overline{t}+l\right\} \right) > 1 - \varepsilon \qquad \forall \alpha_0 > \overline{\alpha},$$

where $\overline{J}_h(i)$ is the set of long-run experts in the baseline model with observable states. Moreover, for any mapping $J : N \to 2^N \backslash \{\varnothing\}$, we have $\Pr\left(\overline{J}_h = J\right) > 0$ if and only if

$$\Pr\left(j_{it}(h) \in \arg\max_{j \in \overline{J}_h(i)} \pi_{jt} \qquad \forall t \in \left\{\overline{t}, \overline{t}+1, \ldots, \overline{t}+l\right\} | h_{\overline{t}} \in H^J \right) > 1 - \varepsilon \qquad \forall \alpha_0 > \overline{\alpha}$$

for some positive probability event H^J.

Proof. Fix positive ε and l as in the proposition. There then exists $\varepsilon' > 0$ such $\Pr(\Pi^{\varepsilon'}) > (1 - \varepsilon/4)^{1/l}$ where $\Pi^{\varepsilon'} = \{\pi | |\pi_i - \pi_j| > \varepsilon'\}$. There also exists finite $\overline{t}$ such that $\Pr(H') > 1 - \varepsilon/4$ for $H' = \left\{h \in H | \tau(h) < \overline{t}\right\}$

where the $\tau(h)$ is defined for $\varepsilon'/2$ in Proposition 1. Note that there exists $\lambda > 0$ such that whenever $(\pi_{1t}, \ldots \pi_{nt}) \in \Pi^{\varepsilon'}$ and $t \geq \bar{t} > \tau(h)$, we have $\gamma(\pi_{j^*_{it}(h)t}, \overline{v}^t_{ij^*_{it}(h)}(h)) < \gamma\left(\pi_{jt}, \overline{v}^t_{ij}(h)\right) - \lambda$ for all $j \in N \backslash \{i, j^*_{it}(h)\}$ where $j^*_{it}(h) = \arg\max_{j \in \overline{J}_h(i)} \pi_{jt}$ and $\overline{v}^t_{ij}(h)$ is the precision of belief of i about the perspective of j in the baseline model under observable states. One can further show that there exists $\lambda' \in (0, \lambda)$ such that the probability of the event

$$H'' = \left\{h \in H | \gamma\left(\pi_{j^*_{it}(h)t}, \overline{v}^t_{ij^*_{it}(h)}(h)\right) < \gamma\left(\pi_{jt}, \overline{v}^t_{ij}(h)\right) - \lambda' \qquad \forall t \leq \bar{t}, j \in N \backslash \{i, j^*_{it}(h)\}\right\}$$

also exceeds $1 - \varepsilon/4$. Consider the set $\hat{H}$ of histories in the intersection of the events H', H'', and that all the realizations of expertise levels between $\bar{t}$ and $t + l$ are in $\Pi^{\varepsilon'}$. Clearly, $\Pr(\hat{H}\ ) > 1 - \varepsilon$, as the probabilities of the excluded events add up to $3\varepsilon/4$. Note however that, since $v^t_{ij}(h) \geq v_0$ throughout, there then exists $\lambda'' > 0$ such that, for all

$$\left\|v^t_i(h) - \overline{v}^t_i(h)\right\| < \lambda'' \Rightarrow \gamma\left(\pi_{j^*_{it}(h)t}, v^t_{ij^*_{it}(h)}(h)\right) < \gamma\left(\pi_{jt}, v^t_{ij}(h)\right) \qquad \forall h \in \hat{H}, t \leq \bar{t} + l, j \in N \backslash \{i, j^*_{it}(h)\}.$$

But, since the limit in (58) is uniform over all hitories with $t \leq \bar{t} + l$, there also exists $\overline{\alpha}$ such that $\left\|v^t_i(h) - \overline{v}^t_i(h)\right\| < \lambda''$ for all histories with $t \leq \bar{t} + l$ whenever $\alpha_0 > \overline{\alpha}$.

The second part of the proposition can be obtained from the first part using Proposition 7. □