Massachusetts Institute of Technology
Department of Economics
Working Paper Series

# Perspectives, Opinions, and Information Flows

**Rajiv Sethi**
**Muhamet Yildiz**

Working Paper 13-23
December 6, 2013

Room E52-251
50 Memorial Drive
Cambridge, MA 02142

# Perspectives, Opinions, and Information Flows*

Rajiv Sethi† Muhamet Yildiz‡

December 6, 2013

**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). He also chooses a target individual whose opinion is then observed. This choice involves a fundamental trade-off between *well-informed* targets, whose signals are precise, and *well-understood* targets, whose perspectives are well known by the observer. Observing an opinion provides information not just about the current state, but also about the target's perspective; hence observed individuals become better-understood over time. This leads to path dependence and the possibly that some individuals never observe certain others in the long run. We identify a simple condition under which long-run behavior is efficient and history-independent. When this condition fails, with positive probability, a single individual emerges as an opinion leader in the long-run. Moreover, the extent to which an individual learns about a target's perspective depends on how well-informed both agents are in the period of observation. This gives rise to symmetry breaking, and can result in observational networks involving information segregation, or static graphs with rich and complex structures.

*We thank Daron Acemoglu and Sanjeev Goyal for helpful suggestions, and the Institute for Advanced Study at Princeton 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 judgements 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 then faces a fundamental 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 the source's information from her opinion. One may therefore be able to extract more information from the opinion of a less-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 with 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.[1] Sophisticated poll aggregators not only adjust for house effects, they also put more weight on polls when these effects are more confidently known.

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 of all individuals are identical to begin with. It is these effects with which the present paper is concerned, with particular focus on long-run efficiency (or lack thereof), opinion leadership, and information segregation.

[1] Since response rates for opinion polls are extremely low, pollsters weight their data based on the demographic characteristics of respondents, in order to match the sample space with the voting population expected to turn out on election day. These expectations are based in part on subjective judgements, which introduces systematic biases in favor of one party or another. For the 2012 presidential elections, Pew Research was found after the fact to have had a 3.2% Democratic bias while Gallup had a 2.5% Republican bias.

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 exogneously fixed network, and focuses on the possibility of double counting and related inference problems.[2] 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.

Specifically, we model a finite set of individuals facing a sequence of periods. Corresponding to each period is an 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. Levels of expertise are independently and identically distributed across individuals and periods, 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 individual then chooses a target whose opinion is to be 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 affects the observer's belief about the current period state and allows her to take a better informed action. Second, the observer's belief about the target's perspective itself becomes more precise. Hence there will be a tendency to link to previously observed targets even when they are not the best-informed in the current period. But, importantly, the level of attachment to a previously observed target depends on the expertise realizations of both observer and observed in the period in which the observation occurred. Specifically, 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. But holding constant one's own expertise, one learns more about the perspective of a *poorly* informed target, since the opinion of such a target will be heavily weighted to their prior rather than their signal. This effect implies 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 choices in subsequent periods, despite the fact that all expertise realizations are public information and a given individual's expertise is common to all observers. Several interesting linkage patterns can arise over time as a result.

Our main results concern these patterns of long run linkages. We begin by deriving the long-

[2]See DeMarzo, Vayanos, and Zweibel (2003) for a state-of-the-art model in this tradition.

run frequency of networks on any given history $h$. Let $J_h(i)$ be the set of individuals to whom an individual $i$ links infinitely often on $h$. In the long run, each player $i$ links to the most informed individual in $J_h(i)$, yielding an independently and identically distributed process of networks. Along a given history, some observational links may break. That is, for any given individual $i$, there may be some set of other individuals who are observed only a finite number of times, while the remainder are observed infinitely often. The observer $i$ learns the perspectives of those in the latter group to an arbitrarily high level of precision, and eventually chooses among them on the basis of their expertise levels. Since all choices are made simultaneously, this places sharp restrictions on the linkage patterns that can emerge in the long run.

Our subsequent results identify conditions on the key parameters of the model—the degree of initial uncertainty about the perspectives of others, and the distribution from which expertise is drawn—under which some interesting linkage patterns can arise with positive probability. An important case occurs when none of the links break, so that $J_h(i)$ includes all other individuals. In this case, everyone links to the most informed individual in the population, yielding a uniform distribution on all star-shaped networks in the long run. This corresponds to *long-run efficiency*.[3] We show that when the initial uncertainty about the perspectives is below a certain threshold, long run efficiency arises with probability one. That is, all effects of path-dependence disappear in the long run.

When history independence fails to hold, a particular form of path-dependence emerges with positive probability. An individual $j_1^*$ emerges as the *opinion leader* so that everybody links to $j_1^*$ while $j_1^*$ links to some $j_2^*$—regardless of the expertise levels. The resulting star-shaped network is *static*, in that it arises at all dates in the long run. This is the least efficient long run outcome, as individuals do not differentiate at all on the basis of expertise. Interestingly, such extreme long-run inefficiency is inevitable when the initial uncertainty about perspectives is sufficiently high, because everyone attaches to the first individual observed, leading to opinion leadership with probability one.

Both long-run efficiency and opinion leadership involve only star-shaped networks, but several other patterns of linkages can also arise. For intermediate levels of initial uncertainty about perspectives, we show that *any* given network $g$ emerges as the limiting network with positive probability (i.e., $J_h(i) = \{g(i)\}$ for every individual $i$ on a set of histories $h$ 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. Since such networks are identified with minimal long-run efficiency, this shows that all possible forms of extreme long-run inefficiency emerge with positive probability.

Another interesting linkage pattern is *information segregation*: the population is partitioned into subgroups, and individuals observe only those within their own subgroup. For intermediate

[3] The behavior in our model is always ex-ante efficient.

levels of initial uncertainty and 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. In fact, our result concerning the arbitrariness of static limiting networks immediately implies the possibility of information segregation, but such segregation can arise even in the absence of convergence to a static network.

The remainder of the paper is structured as follows. In Section 2 we specify the information structure, including the distributions from which signals and priors are drawn. Section 3 examines the evolution of beliefs and networks as individuals make observational choices. The set of networks that can arise in the long run are characterized in Section 4, and the conditions for long run history independence are stated in Section 5. Some special structures that arise when history independence fails are described in Section 6, including opinion leadership, segregation, and static networks. Section 7 identifies a sufficient condition for hysteresis. Section 8 presents the extension of our results to the case in which the states are observed with a possible delay. Section 9 discusses the connection between our work and other theoretical research on heterogeneous priors, observational learning, and network formation. It also discusses evidence for the stable variability in individual perspectives that motivates our analysis. Section 10 concludes.

## 2 The Model

Consider a population $N = \{1, \ldots, n\}$, and a sequence of periods, $t = 1, 2, \ldots$. In each period $t$, 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 $\mu_i$ and variance 1:

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

We shall refer to prior mean $\mu_i$ as the *perspective* of individual $i$. An individual's perspective is not directly observable by any other individual, but it is commonly known that the perspectives $\mu_1, ..., \mu_n$ are independently and identically 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 $\pi_{it}$ capture the degree to which any given individual $i$ is well-informed about the state in period $t$. We shall refer to $\pi_{it}$ as the *expertise* level of individual $i$ regarding the period $t$ state, and assume that these expertise levels are public information. Levels of expertise $\pi_{it}$ are independently and identically distributed across individuals and periods, in accordance with an absolutely continuous distribution function $F$ having support $[a, b]$, where $0 < a < b < \infty$. That is, no individual is ever perfectly informed of the state, but all signals carry at least some information.[4]

**Remark 1.** *Since priors are heterogenous, each individual has his 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 $\theta_t$ is normally distributed with mean $\mu_i$ 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 $\theta_t$ from the value $\mu_i$, $j$ does update her belief about $\theta_t$ 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 among 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 in conformity with Bayes' rule.[5] 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 $\theta_t$ 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 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 $j_{it} \in N$, and observes her opinion $y_{j_{it}t}$ about the current state $\theta_t$. This information is useful because $i$ then chooses an action $\hat{\theta}_{it} \in \mathbb{R}$ in order to minimize

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

[4]Since $\pi_{it}$ is observable, myopic individuals need not consider the distribution from which $\pi_{it}$ is drawn. Nevertheless, this distribution affects the pattern of linkages that emerges in the long run.

[5]Specifically, 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

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

and $w = v + r$.

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 this 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 his own noisy signal $x_{it}$ and forms his opinion $y_{it}$.

3. Each $i$ chooses a target $j_{it} \in N \backslash \{i\}$.

4. Each $i$ observes the opinion $y_{j_{it}t}$ of his target.

5. Each $i$ 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 $n \times n$ matrix $L^t := [l_{ij}^t]$ defines a directed graph or network that describes who listens to whom. Consistent with this interpretation, we shall say that $i$ *links to* $j$ in period $t$ if $j$ is the target selected by $i$ in this period. Note that information flows in the reverse direction of the graph. We are interested in the properties of the sequence of networks generated by this process of link formation.

We assume that individuals are myopic, do not observe the actions of others, and do not observe the realization of the state (observability of the past targets of others will turn out to be irrelevant). These are clearly restrictive assumptions, and our results extend to the case in which the states are observed with some delay (see Section 8).[6]

**Remark 2.** *Even though the states, signals and expertise levels are all distributed independently across individuals and time, the inference problems at any two dates $t$ and $t'$ are related. This is because each individual's ex-ante expectation of $\theta_t$ 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 $\mu_j$ 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 the target at any later date $t'$. In particular, the initial symmetry is broken after individuals choose their first target, potentially leading to highly asymmetric outcomes.*

[6] Note that the desire to make good decisions even when the state realization is unobserved is quite common. For instance, one might wish to vote for the least corrupt political candidate, or donate to the charity with the greatest social impact, or support legislation regarding climate change that results in the greatest benefits per unit cost. We actively seek information in order to meet these goals, and act upon our expectations, but never know for certain whether our beliefs were accurate *ex-post*.

# 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 $\theta_t$ and $j$'s perspective $\mu_j$ as a result of this observation. This determines the process for the evolution of beliefs and the network of information flows.

Given the hypothesis that the perspectives are independently drawn from a normal distribution, posterior beliefs held by one individual about the perspectives of any another will continue to be normally distributed throughout the process of belief revision. Write $v_{ij}^t$ for the precision of the distribution of $\mu_j$ according to $i$ at beginning of $t$. Initially, these precisions are identical: for all $i \neq j$,

$$v_{ij}^1 = v_0. \tag{4}$$

The precisions $v_{ij}^t$ in subsequent periods depend on the history of realized expertise $(\pi^1, \ldots, \pi^{t-1})$ and information networks $(L^1, \ldots, L^{t-1})$. These precisions $v_{ij}^t$ of beliefs about the perspectives of others are central to our analyses; 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 $j$, where

$$y_{jt} = \frac{1}{1+\pi_{jt}}\mu_j + \frac{\pi_{jt}}{1+\pi_{jt}}x_{jt}$$

by (2). Since $x_{jt} = \theta_t + \varepsilon_{jt}$, this observation provides the following noisy signal regarding $\theta_t$:

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

The signal is noisy in two respects. First, the information $x_{jt}$ of $j$ is itself noisy, with signal variance $\varepsilon_{jt}$. Furthermore, since the opinion $y_{jt}$ depends on $j$'s unobservable perspective $\mu_j$, the signal observed by $i$ has an additional source of noise, reflected in the term $\mu_j/\pi_{jt}$.

Taken together, the variance of 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 $\varepsilon_{jt}$. 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 $\mu_j$ of $j$, and corresponds to the variance of $\mu_j/\pi_{jt}$ (where $\pi_{jt}$ is public information and hence has zero variance). This component decreases as $i$ becomes better acquainted with the perspective $\mu_j$, that is, as $j$ becomes better understood by $i$.

The variance $\gamma$ reveals that in choosing an 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 $\gamma$ is lowest. Ties arise with zero probability but for completeness we assume that they are broken in favor of the individual with the smallest label. That is,

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

Note that $j_{it}$ and hence $L^t$ have two determinants: the current expertise levels $\pi_{jt}$ and the precision $v_{ij}^t$ of individuals' beliefs regarding the perspectives of others. The first determinant $\pi_{jt}$ is exogenously given and stochastically independent across individuals and times. In contrast, the second component $v_{ij}^t$ is endogenous and depends on the sequence of prior target choices $(L^1, \ldots, L^{t-1})$, which in turn depends on previously realizes 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 $\mu_j$:

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

Observe that the signal contains an additive noise $\pi_{jt}\theta_t + \pi_{jt}\varepsilon_{jt}$. The variance of the noise is

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

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, using the formula in Footnote 5, 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 $\mu_j$ remains unchanged. This leads to the following closed-form solution:

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

**Remark 3.** *This derivation assumes that individuals do not learn from the target choices of others, as described in $L^t$. If fact, under our assumptions, there is no additional information contained in these choices because $i$ can compute $L^t$ using publicly available data even before $L^t$ has been observed.*[7] *This simplifies the analysis dramatically, and is due to the linear formula in Footnote 5 for normal variables. In a more general model, $i$ may be able to obtain useful information by observing $L$. For example, without linearity, $v_{kj}^{t+1} - v_{kj}^t$ could depend on $y_{jt}$ for some $k$ with $j_{kt} = j$. Since $y_{jt}$ provides information about $\mu_j$, and $v_{kj}^{t+1}$ affects $j_{kt'}$ for $t' \geq t+1$, one could then infer useful information about $\mu_j$ from $j_{kt'}$ 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.*

**Remark 4.** *By the argument in the previous remark, assumptions about the observability of the information network $L$ are irrelevant for our analysis. However, assumptions about the observability of the state $\theta_t$ and the actions $\hat{\theta}_{kt}$ of others (including the actions of one's target, which incorporate information from her own target) are clearly relevant.*

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 $\pi_{it}$ and $\pi_{jt}$ in the period of observation, in accordance with the following.

**Observation 1.** *$\delta(\pi_{it}, \pi_{jt})$ is strictly increasing $\pi_{it}$ and strictly decreasing $\pi_{jt}$. 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 the signal and is therefore relatively uninformative about $j$'s 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. Finally, there is

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

a positive lower bound $\underline{\delta}$ on the amount of increase in precision, making beliefs about observed individuals more and more precise as time passes.

Given the precisions $v_{ij}^t$ at the start of period $t$, and the realizations of the levels of expertise $\pi_{it}$, the links chosen by each individual in period $t$ are given by (6). This then determines the precisions $v_{ij}^{t+1}$ at the start of the subsequent period in accordance with (8), with initial precisions given by (4). For completeness, we set $v_{ii}^t = 0$ for all individuals $i$ and all periods $t$. This defines a Markov process, where the sample space is the set of nonnegative $n \times n$ matrices and the period $t$ realization is $V^t := [v_{ij}^t]$.

For any period $t$, let $h_t := \{v_{ij}^1, ..., v_{ij}^t\}$ denote the history of beliefs (regarding perspectives) up to the start of period $t$. Any such history induces a probability distribution over networks, with the period $t$ network being determined by the realized values of $\pi_{it}$. It also induces a distribution over the next period beliefs $v_{ij}^{t+1}$. It is the long run properties of this sequence of networks and beliefs that we wish to characterize.

### 3.3 Network Dynamics

Recall from (6) that at any given date $t$, each individual $i$ chooses a target $j_{it}$ with the goal of minimizing the perceived variance $\gamma(\pi_{jt}, v_{ij}^t)$. At the start of this process, since the precisions $v_{ij}^1$ are all equal, the expertise levels $\pi_{jt}$ are the only determinants of this choice. Hence the criterion (6) reduces to

$$j_{i1} = \min \left\{ \arg\max_{j \neq i} \pi_{j1} \right\}.$$

That is, the best informed individual in the initial period is linked to by all others, and herself links to the second-best informed.

This pattern of information flows need not hold in subsequent periods. By Observation 1, individual beliefs about the perspectives of their past targets become strictly more precise over time. Since $\gamma$ is strictly decreasing in such precision, an individual may continue to observe a past target even if the latter is no longer the best informed. And since better informed individuals learn more about the perspectives of their targets, they may stick to past targets with greater likelihood than poorly informed individuals, adding another layer of asymmetry.

This trade-off between being well informed and being well understood can prevent the formation of networks in which all individuals link to the best informed, and can give rise to history dependence. One of the key questions of interest in this paper is whether this is a temporary effect, or whether it can arise even in the long run.

In order to explore this question, we introduce some notation. We say that the link $ij$ is *active* in period $t$ if $l_{ij}^t = 1$. Given any history $h_t$, we say that the link $ij$ is *broken* in period $t$ if, conditional on this history, the probability of the link being active in period $t$ is zero. That is, the link $ij$ is broken in period $t$ conditional on history $h_t$ if

$$\Pr(l_{ij}^t = 1 \,|\, h_t) = 0.$$

If a link is broken in period $t$ we write $b_{ij}^t = 1$. It is easily verified that if a link is broken in period $t$ then it is broken in all subsequent periods.[8] 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. That is, link $ij$ is free in period $t$ if

$$\Pr(b_{ij}^{t+s} = 1 \,|\, h_t) = 0$$

for all non-negative integers $s$. If a link is free at time $t$, there is a positive probability that it will be active in the current period as well as in each subsequent period.

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 $\beta : (0, \overline{v}) \to R_+$, by setting

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

This function 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 $\beta\left(v\right)$.

In our analysis, we shall ignore histories that result in ties and arise with zero probability. Accordingly, define

$$\mathcal{V} = \left\{ \left(v_{ij}\right)_{i \in N, j \in N \setminus \{i\}} \,|\, v_{ij} \neq \beta\left(v_{ik}\right) \text{ and } v_{ij} \neq \overline{v} \text{ for all distinct } i, j, k \in N \right\}$$

and

$$H = \left\{ h_t \,|\, v^t\left(h_t\right) \in \mathcal{V} \right\}.$$

We shall consider only histories $h_t \in H$.

Our first result characterizes histories after which a link is broken.

[8] This follows from the fact that the process $\{v_{ij}^t\}$ is non-decreasing, and $v_{ij}$ increases in period $t$ if and only if $l_{ij} = 1$.

**Lemma 1.** *For any history $h_t \in H$, a link $ij$ is broken at $h_t$ if and only if $v_{ik}^t(h_t) > \beta(v_{ij}^t(h_t))$ for some $k \in N \backslash \{i,j\}$.*

When $v_{ik}^t > \beta(v_{ij}^t)$, individual $i$ never links to $j$ because the cost $\gamma(\pi_{kt}, v_{ik}^t)$ of linking to $k$ is always lower than the cost $\gamma(\pi_{jt}, v_{ij}^t)$ of linking to $j$. 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 the inequality is reversed, $i$ links to $j$ when $j$ is sufficiently well-informed and all others are sufficiently poorly informed.

The next result characterizes histories after which a link becomes free.

**Lemma 2.** *A link $ij$ is free after history $h_t \in H$ if and only if*

$$v_{ij}^t(h_t) > \min\left\{\overline{v}, \max_{k \in N \backslash \{i,j\}} \beta\left(v_{ik}^t(h_t)\right)\right\}.$$

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 by Lemma 1, and hence $i$ links to $j$ in all subsequent periods, and $ij$ is therefore free. Moreover, 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. (By Observation 1, this happens when $i$ links to $k$ at least $(\beta(v_{ij}^t(h_t)) - v_{ik}^t(h_t))/\underline{\delta}$ times.) Note that the above lemmas imply that along every infinite history, every link eventually either breaks or becomes free.

To illustrate these ideas, consider a simple example with $N = \{1,2,3\}$. Figure 1 plots regions of the state space in which the links 31 and 32 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). It is assumed that $a = 1$ and $b = 2$ so $\overline{v} = 0.5$. In the orthant above $(\overline{v}, \overline{v})$ links to both nodes are free by Lemma 2. Individual 3 links to each of these nodes with positive probability thereafter, eventually becoming arbitrarily close to learning both their perspectives. Hence, in the long run, she links with likelihood approaching 1 to whichever individual is better informed in any given period. This limiting behavior is therefore independent of past realizations, and illustrates our characterization of history independence.

When $v_{32} > \beta(v_{31})$, the region above the steeper curve in the figure, the link 31 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

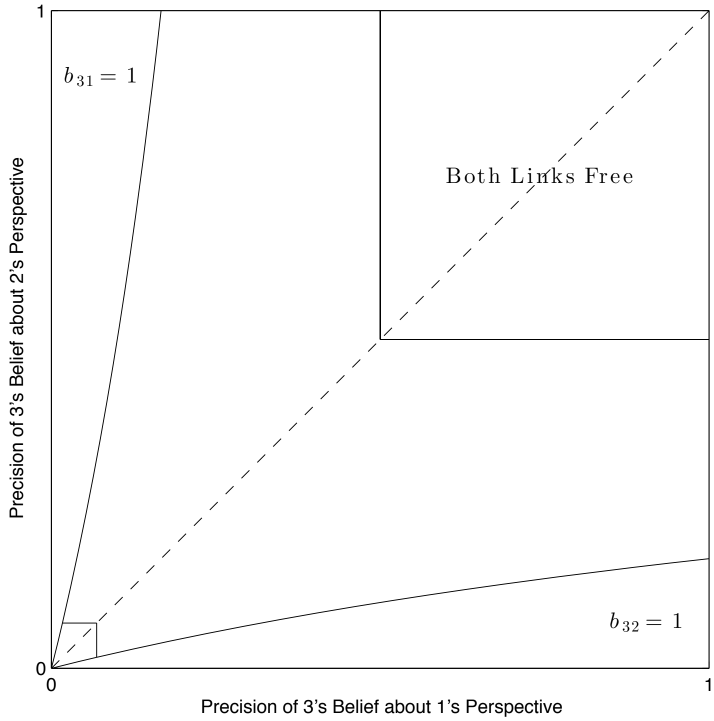


Figure 1: Regions of State Space with Broken and Free Links

link 32 is free, the link 31 has not been resolved. Depending on subsequent expertise realizations, either both links will become free or 31 will break. Symmetrically, when $\overline{v} < v_{31} < \beta(v_{32})$, the link 31 is free while the other link 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 or enter the orthant in which both links are free. Note that the probability of reaching the orthant in which both links are free is zero for sufficiently small values of $(v_{31}, v_{32})$. For example, when $\beta(v_0) - v_0 < \underline{\delta}$, regardless of the initial expertise levels, 3 will attach to the very first individual to whom she links. The critical value of $v_0$ in this example is approximately 0.07, and the relevant region is shown at the bottom left of the figure.

Since the initial precisions of beliefs about perspectives lie on the 45 degree line by assumption, the size of this common precision $v_0$ determines whether history independence is ensured, is possible but not ensured, or is not possible.[9] In the first of these cases, individuals almost always link to the best informed person in the long run, and the history of realizations eventually ceases to matter.

[9]It is tempting to conclude that in the three person case, these three regimes correspond to the three segments of the diagonal in Figure 1. But this is not correct, since the condition $\beta(v_0) - v_0 < \underline{\delta}$ is sufficient but not necessary for at least on link to break. Specifically, there are values of $v_0$ outside the region on the bottom left of the figure such that both links can become free in the long run for any one observer, but not for all three. A fuller characterization is provided below.

In the second case, this outcome is possible but not guaranteed: there is a positive probability that some links will be broken. And in the third case, history matters perpetually and initial realizations have permanent effects.

The fact that every link either breaks or becomes free along any infinite history allows us to place sharp restrictions on the long run frequency of networks, which we turn to next.

# 4 Long-run Frequency of Networks

In this section, we characterize the long-run frequency of each communication network that can emerge in our model. This allows us to provide a simple expression for long-run payoffs and long-run efficiency.

Let $G$ denote the set of functions $g : N \to N$ that satisfy $g(i) \neq i$ for each $i \in N$. Each element of $G$ thus corresponds to a directed graph in which each node is linked to one target. This is the set of all feasible networks that can arise. Our main goal in this section is to find the frequency with which each $g \in G$ is realized in the long run. To this end, for each infinite history $h$, each $t$, and each $g$, define

$$\phi_t\left(g \mid h\right) = \frac{\#\left\{s \leq t \mid j_{is}\left(h\right) = g\left(i\right) \; \forall i \in N\right\}}{t}$$

as the empirical frequency of the graph $g$ up to date $t$ at history $h$. When $\phi_t$ has a limit, this is denoted

$$\phi_\infty\left(g \mid h\right) \equiv \lim_{t \to \infty} \phi_t\left(g \mid h\right).$$

We call this the *long-run frequency* of graph $g$ at history $h$.

The long-run frequencies are determined by the free links. Towards establishing this, for any mapping $J : N \to 2^N$ with $i \notin J\left(i\right)$ and $J(i)$ nonempty for each $i$, define

$$p_J(g) = \Pr\left(g(i) = \arg\max_{j \in J(i)} \pi_j \; \forall i \in N\right) \tag{10}$$

at each $g \in G$. Note that $p_J(g) = 0$ if $g(i) \notin J(i)$ for some $i$. If each individual $i$ were restricted to choose the most informed individual in $J(i)$ as the target, each graph $g$ would be realized with probability $p_J(g)$.

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

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

Here $J_h\left(i\right)$ is the (nonempty) set of individuals to whom $i$ links to infinitely many times along the history $h$. On this path, eventually, the links $ij$ with $j \in J_h\left(i\right)$ become free, and all other links break. The following result states that, in the long run, each individual $i$ links to the most informed target in $J_h\left(i\right)$.

**Proposition 1.** *Almost surely, the long-run frequency* $\phi_\infty(\cdot|h)$ *exists, and*

$$\phi_\infty(\cdot|h) = p_{J_h}.$$

Proposition 1 provides a sharp, testable prediction regarding the joint distribution of behavior. For each individual, consider the set $J_h(i)$ of targets that each individual $i$ links to with positive long-run frequency. Then, the frequency in which a graph $g$ is realized is the probability that $g(i)$ is the most informed individual in $J_h(i)$ for each $i$ simultaneously. This simultaneity requirement sharply restricts the set of possible graphs. For example, if two individuals $i$ and $i'$ each links to both $j$ and $j'$ with positive frequency, then, in the long run, $i$ cannot link to $j$ while $i'$ links to $j'$.

As a special case, consider histories along which each individual links to each other individual infinitely often, i.e., $J_h(i) = N \setminus \{i\}$ for every $i$. Then Proposition 1 implies that the set of networks with positive long-run frequency consists of the graphs $g_{i_1,i_2}$ in which $i_1$ links to $i_2$ (i.e., $g_{i_1,i_2}(i_1) = i_2$) and all other individuals link to $i_1$ (i.e., $g_{i_1,i_2}(i) = i_1$ for all $i \neq i_1$). By symmetry, it further predicts that each such graph occurs with equal frequency, yielding a uniform distribution on the set of such graphs.[10]

More generally, Proposition 1 implies that each individual $i$ eventually uses each of his long-run targets $J_h(i)$ with equal frequency. Let

$$\phi_{t,i}(j \mid h) = \frac{\#\{s \leq t \mid j_{is}(h) = j\}}{t}$$

denote the frequency with which $i$ links to $j$ over the first $t$ periods of history $h$. Then we have

**Corollary 1.** *For each* $j \in J_h(i), \phi_{t,i}(j \mid h) \to 1/|J_h(i)|$ *almost surely.*

In the long run, each individual $i$ observes the most informed member of $J_h(i)$ in any given period. Using this fact, one can show that his expected payoff at the start of of each period $t$ converges to

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

We call $u_{\infty,i,h}$ the *long-run payoff* of $i$ at history $h$. Note that the long run payoff is simply a function of the number of active links, and it is increasing in that number. In particular, the highest long-run payoff is obtained when $J_h(i) = N \setminus \{i\}$, yielding $u(n-1)$. Long-run efficiency is obtained when all links are free and each individual's payoff is $u(n-1)$. In this case long-run behavior is history independent, in that each individual observes the most informed individual at each date, yielding an approximately i.i.d. sequence of star shaped graphs $g_{i_1 i_2}$. At the other

[10] Note that Proposition 1 does not require that the expertise realizations $\pi_1, ..., \pi_n$ be independently and identically distributed. Even with asymmetric distributions of expertise, one can obtain sharp predictions regarding the long run network structure. For instance, along histories where all individuals link to all others infinitely often, a graph $g$ has positive long-run frequency if and only if $g = g_{i_1 i_2}$ for some distinct $i_1, i_2 \in N$, although all such graphs need not arise with equal frequency.

extreme, the lowest long-run payoff is obtained when the individual ends up with just a single target (i.e. $\#J_h(i) = 1$), obtaining $u(1)$. Accordingly, the least efficient long-run behavior arises when a single graph $g$ repeats itself forever in the long run; we call such a network $g$ *static*.

We next provide a simple necessary and sufficient condition under which long-run efficiency obtains at all histories. We then show that, when the condition fails, many interesting communication structures such as information segregation and opinion leadership emerge with positive probability in the long run. In particular, the least efficient long run-behavior also arises with positive probability, and any arbitrary $g \in G$ can emerge as a limiting static network.

## 5 History Independence

In this section, we characterize the conditions under which the long-run behavior is necessarily efficient and (equivalently) history independent. Long-run efficiency is characterized by $J_h(i) = N \backslash \{i\}$ for every $i$ with probability 1. A more direct definition is as follows. When the process of network formation is history independent in the long run, each individual will *eventually* observe the best informed individual with high probability. Specifically, this probability can be made arbitrarily close to 1 if a sufficiently large number of realizations is considered:

**Definition 1.** *For any given history $h_t$, the process $\{V^t\}_{t=1}^{\infty}$ is said to be history independent at $h_t$ if, for all $\varepsilon > 0$, there exists $t^* > t$ such that*

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

*for all $t' > t^*$ and $i \in N$. The process $\{V^t\}_{t=1}^{\infty}$ is said to be history independent if it is history independent at the initial history $h_1$.*

This definition is equivalent to the above characterization. Clearly the process cannot be history independent in this sense if there is a positive probability that one or more links will be broken at any point in time. Moreover, history independence is obtained whenever all links become free and have uniform positive bound on probability of occurrence throughout. Building on this fact and Lemma 2, the next result provides a simple characterization for history independence.

**Proposition 2.** *For any $h_t \in H$, the process $\{V^t\}_{t=1}^{\infty}$ is history independent at $h_t$ if and only if $v_{ij}^t(h_t) > \overline{v}$ for all distinct $i, j \in N$. In particular, the process $\{V^t\}_{t=1}^{\infty}$ is history independent if and only if $v_0 > \overline{v}$.*

The condition for history independence may be interpreted as follows. For any given value of the support $[a, b]$ from which levels of expertise are drawn, history independence arises if beliefs about the perspectives of others are sufficiently precise. That is, if each individual is sufficiently well-understood by others even before any opinions have been observed. Conversely, when there is substantial initial uncertainty about the individuals' perspectives, the long-run behavior is history dependent with positive probability.

Depending on the extreme values $a$ and $b$ of possible expertise levels, the threshold $\overline{v}$ can take any value. When expertise is highly variable in absolute or relative terms (i.e. $b - a$ or $b/a$ are large), $\overline{v}$ is small, leading to history independence for a broad range of $v_0$ values. Conversely, when expertise is not sufficiently variable in the same sense, the threshold $\overline{v}$ becomes large, and history independence is more likely to fail. This makes intuitive sense, since it matters less to whom one links under these conditions, and hysteresis is therefore less costly in informational terms.

The logic of the argument is as follows. When $v_{ij}^0 = v_0 > \overline{v}$, there is a positive lower bound on the probability that a $i$ links to $j$ at the outset, regardless of her beliefs about others. Since $v_{ij}^t$ is nondecreasing in $t$, this lower bound is valid at all dates and histories, so $i$ links to $j$ infinitely often with probability 1. But every time $i$ links to $j$, $v_{ij}^t$ increases by at least $\underline{\delta}$. Hence, after a finite number of periods, $i$ knows the perspective of $j$ with arbitrarily high precision. This of course applies to all other individuals, so $i$ comes to know all perspectives very well, and chooses targets largely on the basis of their expertise level. Conversely, when $v_{ij}^0 = v_0 < \overline{v}$, it is possible that $i$ ends up linking to another individual $j'$ sufficiently many times, learning his perspective with such high precision that the link $ij$ breaks. After this point, $i$ no longer observes $j$ no matter how well informed the latter may be.

Proposition 2 identifies a necessary and sufficient condition for history independence at the initial history. If this condition fails to hold, then the process $\{V^t\}_{t=1}^{\infty}$ exhibits *hysteresis*: there exists a date $t$ by which at least one link is broken with positive probability. History independence (at the initial history) and hysteresis are complements because in our model any link either becomes free or breaks along every path, and history independence is equivalent to all links becoming eventually free with probability 1. Proposition 2 therefore can be restated as follows: $\{V^t\}_{t=1}^{\infty}$ exhibits hysteresis if and only if $v_0 < \overline{v}$.

When history independence fails, a number of interesting network structures can arise. We shall consider three of these: opinion leadership, informational segregation, and static communication networks.

# 6 Network Structures

Before describing some of the long run communication structures that can arise, we develop a simple example to illustrate the phenomenon of symmetry breaking. This plays a key role in allowing for complex structures such as information segregation to arise.

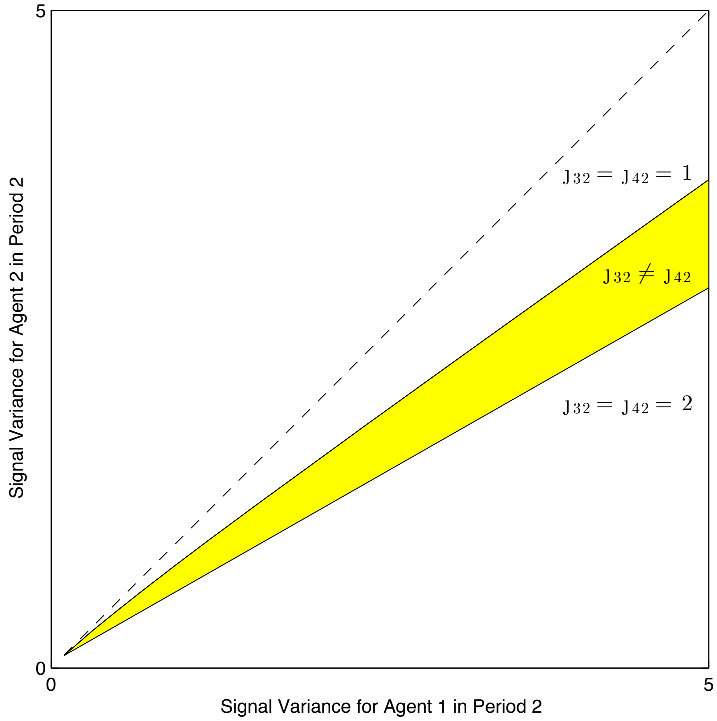


Figure 2: Agents 3 and 4 choose different targets when variances lie in the shaded region

### 6.1 Symmetry Breaking

Consider the simple case of $n = 4$, and suppose (without loss of generality) that $\pi_{1t} > \pi_{2t} > \pi_{4t} > \pi_{3t}$ at $t = 1$. Then individual 1 links to 2 (i.e. $j_{1t} = 2$) and all the others link to 1 (i.e. $j_{2t} = j_{3t} = j_{4t} = 1$). Individuals 2, 3, and 4 all learn something about the perspective of individual 1. The precisions $v_{i1}^2$ of their beliefs about $\mu_1$ at the start of the next period are all at least $v_0 + \underline{\delta}$, while the precisions of their beliefs about the perspectives of other individuals remain at $v_0$. Moreover, they update their beliefs *to different degrees*, with those who are better informed about the state ending up with more precise beliefs about 1's perspective: $v_{21}^2 > v_{41}^2 > v_{31}^2 \geq v_0 + \underline{\delta}$.

Now consider the second period, 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 who are best informed. On the other hand, there is an open set of expertise realizations for which 3 and 4 remain linked to 1 despite the fact that 2 is now better informed. In Figure 2, this event ($j_{32} = j_{42} = 1$) occurs for expertise realizations between the shaded region and the 45-degree line.[11] In this region, while 2 is better informed than 1 ($\pi_{2t} > \pi_{1t}$), the difference between their expertise levels is not

[11]The figure has variances of $\varepsilon_1$ and $\varepsilon_2$ on the horizontal and vertical axes respectively, and is based on the specification $v_0 = 1, v_{31} = 2$, and $v_{41} = 4$. Since 2 is assumed to be better informed than 1 in period 2, all expertise realizations must lie below the 45-degree line.

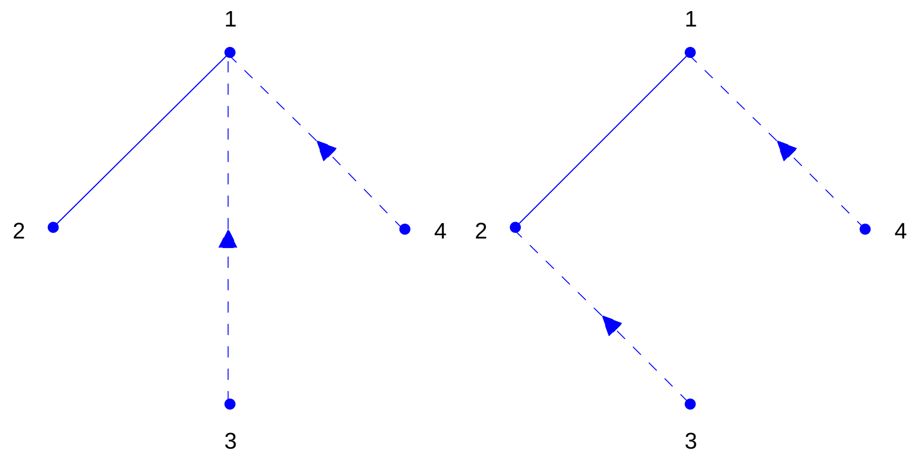


Figure 3: Asymmetric effects of first period observations on second period links.

large enough to overcome the stronger attachment of individuals 3 and 4 to their common past target ($v_{i1}^2 > v_{i2}^2$ for $i \in \{3, 4\}$). Below the shaded region, the difference in expertise levels between 1 and 2 is large enough to induce both individuals 3 and 4 to switch to the best informed target in the second period ($j_{32} = j_{42} = 2$).

Within the shaded region, however, symmetry is broken and individuals 3 and 4 choose different targets: 3 switches to the best informed individual ($j_{3t} = 2$) while 4 remains linked to her previous target ($j_{4t} = 1$). In this region, the difference between the expertise levels of 1 and 2 is large enough to overcome the preference of 3 towards 1, but not large enough to overcome the stronger preference of individual 4, who was more precisely informed of the state in the initial period, and hence learned more about the perspective of her target.

A particular set of realizations that generates this effect is shown in Figure 3, where a solid line indicates that links are formed in both directions and a dashed line indicates a single link in one direction. Nodes (corresponding to individuals) are numbered in increasing order anti-clockwise, starting from the top. Nodes 1 and 2 link to each other in both periods. Nodes 3 and 4 link to node 1 (the best informed) in the first period. In the second period node 3 switches to node 2, who is now the best informed, but node 4 continues to observe node 1. This is because the perspective of 1 is better known to 4 than to 3, since 4 was better informed than 3 about the state in the initial period.

This example illustrates how two individuals with a common observational history can start to make different choices at some period of time, even though expertise levels are public information in all periods. We now explore some of the long run implications of this.

## 6.2 Opinion Leadership

One network structure that can arise is *opinion leadership*, with some subset of individuals being observed with high frequency even when their levels of expertise are known to be low, while others are never observed regardless of their levels of expertise. This can happen because repeated observation of a leader allows her perspective to become well understood by others, and hence her opinion can be more easily interpreted even when her information is poor.

We say that a sample path exhibits opinion leadership if there is some period $t$ and some nonempty subset $S \subset N$ such that $b_{ij} = 1$ for all $(i, j) \in N \times S$. That is, opinion leadership exists if some individuals are never observed (regardless of expertise realizations) after time $t$ along the sample path in question.

A special case of opinion leadership arises when $n$ links are free while the rest are all broken. In this case, all individuals are locked into a particular target, regardless of expertise realizations. In an extreme case, there may be a single leader to whom all others link, and a second individual to whom the leader alone links in all periods. We refer to this property of sample paths as *extreme opinion leadership.*

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

$$\beta(\tilde{v}) - \tilde{v} = \underline{\delta}. \tag{12}$$

The following result establishes that unless we have history independence (in which case hysteresis is impossible) there is a positive probability of extreme opinion leadership, and such extreme leadership is inevitable when $v_0$ is sufficiently small:

**Proposition 3.** *For $h_1 \in H$, $\{V^t\}_{t=1}^{\infty}$ exhibits extreme opinion leadership (i) with positive probability if and only if $v_0 < \overline{v}$, and (ii) with probability 1 if and only if $v_0 < \tilde{v}$.*

The intuition for this result is straightforward: 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. Moreover, when $v_0 < \tilde{v}$, we have $v_0 + \underline{\delta} > \beta(v_0)$ and, by Lemma 1, each individual adheres to their very first target regardless of subsequent expertise levels. 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.

More generally, two or more information leaders may emerge, who might themselves have different sets of targets. An example is shown in Figure 4, where nodes 1 and 4 emerge as leaders, and themselves link to 5 and 3 respectively. By the sixth period all links that target a member of

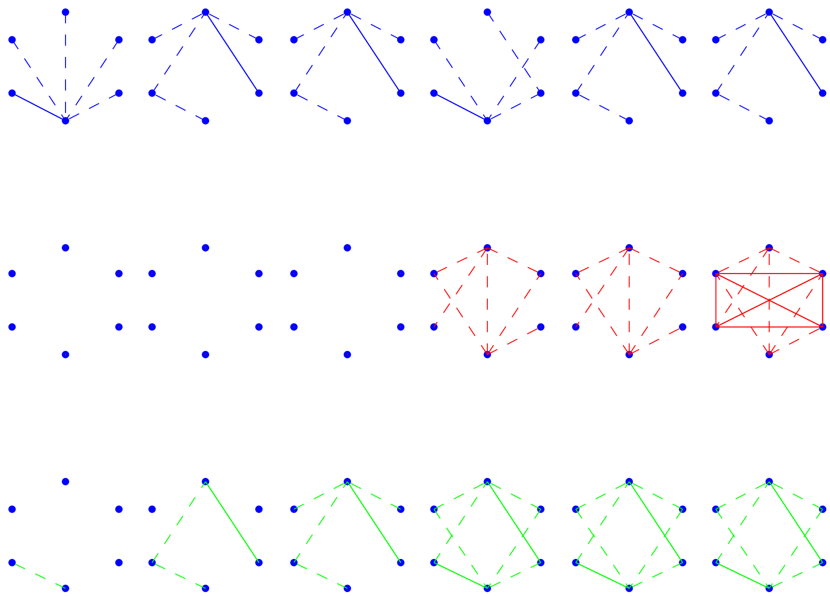

Figure 4: Emergence of Opinion Leadership

the set $\{2, 6\}$ are broken, and these two individuals are never subsequently observed. Furthermore, the two information leaders are each locked in to a single target, while the remaining individuals observe both information leaders with positive probability in all periods.

### 6.3 Information Segregation

Despite the *ex ante* symmetry of the model, it is possible for clusters to 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 history independence within clusters, so that individuals tend to link to the best informed in their own group, but avoid linkages that cross group boundaries.

We say that a sample path exhibits *segregation over a partition* $\{S_1, S_2, \ldots, S_m\}$ of $N$ if there is a period $t$ such that $b_{ij}^t = 1$ for all $(i, j) \in S_k \times S_l$ with $k \neq l$. That is, segregation over a partition $\{S_1, S_2, \ldots, S_m\}$ is said to arise if no link involving elements of different clusters can form after some period is reached, and members of each cluster $S_k$ communicate only with fellow members of their own cluster. We say that a sample path exhibits *segregation* if it exhibits segregation over some partition with at least two disjoint clusters.

The first few periods of a sample path that exhibits segregation is illustrated in Figure 5. In this case the disjoint clusters $\{1, 2, 3\}$ and $\{4, 5, 6\}$ emerge with positive probability. Although this

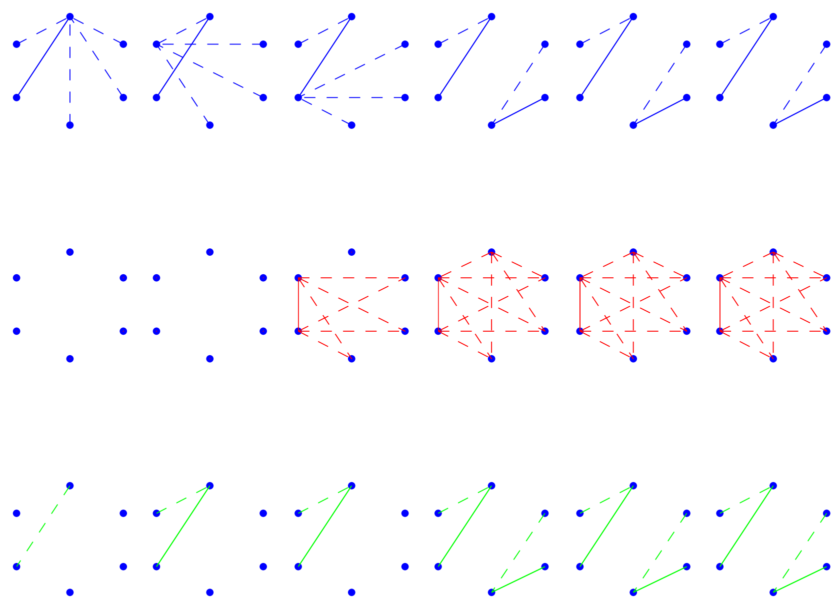

Figure 5: Emergence of Segregated Clusters

network is not resolved by the end of the last period depicted, it is easily seen that there as a positive probability of segregation after this history since no link that connects individuals in two different clusters is free.

In order for a segregation to arise over a partition $\{S_1, S_2, \ldots, S_m\}$, each $S_k$ must have at least two elements. Excluding the trivial partition $\{N\}$, write $\mathcal{P}$ for the set of all partitions $\{S_1, S_2, \ldots, S_m\}$ with $m \geq 2$ and $|S_k| \geq 2$ for all $k$. This is the set of all partitions over which segregation could conceivably arise.

Segregation can arise only if initial precision level $v_0$ are small enough to rule out history independence. 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 $v_0 + \underline{\delta} > \overline{v}$. These links are then free by Proposition 2, which clearly rules out segregation. So $v_0$ cannot be too large if segregation is to arise. And it cannot be too small either, otherwise individuals get locked into common early targets. For example, extreme opinion leadership, in which a single information leader is observed repeatedly by all others, is inconsistent with segregation and arises with certainty when $v_0 < \tilde{v}$ (Proposition 3). The following result establishes that in all the other cases, segregation arises with positive probability over *any* partition in $\mathcal{P}$:

**Proposition 4.** *Suppose $n \geq 4$. For any $h_1 \in H$ and any partition $\{S_1, S_2, \ldots, S_m\} \in \mathcal{P}$, the process $\{V^t\}_{t=1}^{\infty}$ exhibits segregation over $\{S_1, S_2, \ldots, S_m\}$ with positive probability if and only if*

$v_0 \in (\tilde{v}, \overline{v} - \underline{\delta})$.

The forces that give rise to segregation can be understood by reconsidering the example depicted in Figure 5, where two segregated clusters of equal size emerge in a population of size 6. 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 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 ends 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. Hence a very particular pattern of expertise realizations is required to generate segregation, but any partition of the population into segregated clusters can arise with positive probability.

## 6.4 Static Networks

When $v_0 > \overline{v}$, all links are free to begin with. At the other extreme, when $v_0 < \tilde{v}$, the long run outcome is necessarily extreme opinion leadership, resulting in the lowest possible level of information aggregation. For intermediate values of $v_0$, while extreme opinion leadership remains possible, other structures can also arise. As shown above, individuals can be partitioned into any arbitrary set of clusters of at least two individuals, with no cross-cluster communication at all.

This indeterminacy of network structures extends further. We shown next that each individual may be locked into a single, arbitrarily given target in the long run. This implies that every worst case scenario (with respect to information aggregation) can arise with positive probability.

Let $G$ denote the set of functions $g : N \to N$ that satisfy $g(i) \neq i$. Each element of $G$ thus corresponds to a directed graph in which each node is linked to one (not necessarily unique) target. We say that a sample path *converges* to $g \in G$ if there exists a period $t^*$ such that, for all $i \in N$ and all $t > t^*$, $j_{it} = g(i)$. The process $\{V^t\}_{t=1}^{\infty}$ converges to $g$ with positive probability if the probability that a sample path will converge to $g$ is positive. In this case there is a positive probability that each individual eventually links only to the target prescribed for her by $g$.

In order to identify the range of parameter values for which any given network $g \in G$ can emerge with positive probability as an outcome of the process, we make the following assumption.

**Assumption 1.** *There exists $\pi \in (a, b)$ such that $\gamma(\pi, v_0) < \gamma(a, v_0 + \delta(\pi, b))$ and $\gamma(b, v_0) < \gamma(\pi, v_0 + \delta(\pi, b))$.*

Note that this assumption is satisfied whenever $v_0 > v^*$ where $v^*$ is defined by

$$\beta(v^*) - v^* = 2\delta(b, b).$$

In addition to Assumption 1, convergence to an arbitrary network $g \in G$ requires that $v_0$ be sufficiently small:

**Proposition 5.** *Assume that* $v_0 < \overline{v} - \delta(b, b)$ *and satisfies Assumption 1. Then, for any graph* $g \in G$*, the process* $\{V^t\}_{t=1}^{\infty}$ *converges to* $g$ *with positive probability.*

A sufficient condition for such convergence to occur is $v_0 \in (v^*, \overline{v} - \delta(b, b))$, and it is easily verified that this set is nonempty. For instance if $(a, b) = (1, 2)$, then $(v^*, \overline{v} - \delta(b, b)) = (0.13, 0.20)$.

While the emergence of opinion leadership is intuitive, the possibility of convergence to an arbitrary graph is much less so. Since all observers face the same distribution of expertise in the population, and almost all link to the same target in the initial period, the possibility that they may all choose different targets in the long run is counter-intuitive. Nevertheless, there exist sequences of expertise realizations that result in such strong asymmetries.

## 7 Strong Hysteresis

The three classes of networks discussed in the previous section are not by any means exhaustive, and a variety of other outcomes are possible when the condition for history independence does not hold at the initial history. Recall that the process $\{V^t\}_{t=1}^{\infty}$ exhibits *hysteresis* if there exists a date $t$ by which at least one link is broken with positive probability. Note that this is consistent with the possibility that all links become free with positive probability. Hysteresis rules out history independence at the initial history, but allows for history independence to arise after some histories with positive probability.

We now introduce a stronger notion of hysteresis, which rules out the possibility that all links will eventually be free. For any given history $h_t$, the process $\{V^t\}_{t=1}^{\infty}$ is said to exhibit *strong hysteresis* at $h_t$ if the probability that no links will break in period $t+1$ is zero. It is said to exhibit *strong hysteresis* if it exhibits strong hysteresis at the initial history $h_0$.

An immediate implication of Proposition 3 is that the process exhibits strong hysteresis if $v_0 < \tilde{v}$, since this is sufficient for opinion leadership to arise with probability 1. In this case each individual links perpetually to the first person they observe. However, $v_0 < \tilde{v}$ is not necessary for strong hysteresis. To see why, consider the three agent example described in Section 3.3. Here $v_0 < \tilde{v}$ corresponds to the segment of the 45 degree line in the bottom left section of Figure 1. If $v_0$ lies within this range, one of the two links originating at 3 will break after the fist observation is made. If $v_0$ lies outside this range, then there is a positive probability that both links 31 and 32 will eventually be free. But this does not mean that there is a positive probability that *all* links in the network will be free: sample paths that result in both 31 and 32 being free might *require* that some other link be broken. This is in fact the case.

To identify a necessary and sufficient condition for strong hysteresis, define $\hat{v}$ as the unique solution to

$$\beta(\hat{v}) - \hat{v} = \delta(b, b). \tag{13}$$

We then have:

**Proposition 6.** *For any $h_1 \in H$, the process $\{V^t\}_{t=1}^{\infty}$ exhibits strong hysteresis if and only if $v_0 < \hat{v}$.*

It is easily verified that $\hat{v} > \tilde{v}$, as expected. The condition $v_0 < \tilde{v}$ is necessary and sufficient for all links to break in the initial period except for the ones that are active, resulting in opinion leadership. The weaker condition $v_0 < \hat{v}$ is necessary and sufficient for at least one link to break. This rules out history independence at any future period, but allows for a broad range of network structures to emerge in the long run, including segregation and static networks.

# 8 Observable States

For simplicity, our main model assumes that $\theta_t$ is not observable. In this section, we extend our results to the case in which states are publicly observable with some delay.

**Assumption 2.** *For all $t$, $\theta_t$ becomes publicly observable at the end of period $t + \tau$ where $\tau \geq 0$ is a constant (independent of $t$).*

Note that $\tau = 0$ corresponds to observability of $\theta_t$ 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 main model), which corresponds to the limit $\tau = \infty$.

Under Assumption 2, 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

$$v_{ij\tau}^t = v_{ij}^0 + \sum_{\{t' < t - \tau : j_{it'} = j\}} 1/\pi_{jt'} + \sum_{\{t - \tau \leq t' < t : j_{it'} = j\}} \delta\left(\pi_{it'}, \pi_{jt'}\right). \tag{14}$$

For $t' < t - \tau$, individual $i$ retrospectively updates his belief about the perspective of his target $j$ at $t'$ by using the true value of $\theta_{t'}$ instead of his private signal $x_{it'}$. This adds to the precision of his belief $1/\pi_{jt'}$, instead of $\delta(\pi_{it'}, \pi_{jt'})$, increasing the precision of his belief. 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. Indeed, the improvement in precision due to observability of past states is

$$v^t_{ij\tau} - v^t_{ij} = \sum_{\{t'<t-\tau : j_{it'}=j\}} 1/\left(1 + \pi_{it'} + \pi_{jt'}\right).$$

Such an improvement only enhances the attachment to previously observed individuals. This makes opinion leadership more likely to arise, but it does not affect our results about the long-run frequency of networks or long-run efficiency.

**Proposition 7.** *Under Assumption 2, for $v_0 \notin \{\tilde{v}, \overline{v}\}$ and for any $\tau \geq 0$, the following are true.*

1. *Almost surely, the long-run frequency $\phi_\infty(\cdot|h)$ exists, and $\phi_\infty(\cdot|h) = p_{J_h}$ (cf. Proposition 1).*
2. *The process $\{V^t\}_{t=1}^{\infty}$ is history independent if and only if $v_0 > \overline{v}$ (cf. Proposition 2).*
3. *The process $\{V^t\}_{t=1}^{\infty}$ exhibits extreme opinion leadership (i) with positive probability if $v_0 < \overline{v}$, and (ii) with probability 1 if $v_0 < \tilde{v}$ (cf. Proposition 3).*

Part 1 states that the long-run behavior along a given history does not depend on the observability of states: each individual's beliefs about the targets that have been observed infinitely often are arbitrarily precise, and hence he observes the most informed one among them. (Since the path is given, observability simply improves this already high precision.) Part 2 states that we necessarily have long-run efficiency (or history independence) whenever $v_0 > \overline{v}$. In principle, observability of past states could make long-run efficiency more difficult to attain since it increases the level of attachment to past targets. Nevertheless, the proof of Proposition 2 uses the worst-case scenario in which beliefs about past targets are infinitely precise. Improved precision due to observability does not make any difference in this case. In the alternative case of $v_0 < \overline{v}$, opinion leadership emerges with positive probability, as stated in Part 3. Since our proof of Proposition 3 is based on repeated observation of an early leader, observability of states only helps, as it can only increase the attachment to that leader. The same applies for the necessity of opinion leadership when $v_0 < \tilde{v}$. On the other hand, with observable states, 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 $\beta(\tilde{v}') - \tilde{v}' = 1/b$. Our proofs of the original results also extend to these extensions *mutatis mutandis* (by replacing $v^t_{ij}$ with $v^t_{ij\tau}$), and we will not repeat them.

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 $\theta_{t'}$, his signal $x_{it'}$ does not affect his beliefs at any *fixed* history, as seen in (14). 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 them to possibly different paths of observed targets. Indeed, our results about information segregation and static networks extends to the case of delayed observability for a sufficiently long delay $\tau$.

**Proposition 8.** *Assume that $v_0 < \overline{v} - 1/b$ and satisfies Assumptions 1 and 2. Then, there exists $\overline{\tau}$ such that the following are true for all $\tau \geq \overline{\tau}$.*

1. *For any graph $g \in G$, the process $\{V^t\}_{t=1}^{\infty}$ converges to $g$ with positive probability (cf. Proposition 5).*
2. *In particular, for any partition $\{S_1, S_2, \ldots, S_m\} \in \mathcal{P}$, the process $\{V^t\}_{t=1}^{\infty}$ exhibits segregation over $\{S_1, S_2, \ldots, S_m\}$ with positive probability (cf. Proposition 4).*

For sufficiently large delay $\tau$, the first part of this result extends Proposition 5, concluding that every network emerges as the static network with positive probability. Moreover, for any $\{S_1, S_2, \ldots, S_m\} \in \mathcal{P}$, there exists $g \in G$ that maps each player $i$ to a member in his own group (i.e. $g(i) \in S_k \iff i \in S_k$). Under such a static network $g$, we have information segregation with the given partition. The second part states this, extending Proposition 4.[12] The idea of the proof is rather simple. 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.

Although asymmetric linkage patterns are more difficult under observability of states, similar results still hold even under immediate observability of states. This is because the players' level of attachment still depends on the expertise levels of their targets, and the most-informed player observes a different individual than others at the beginning. (Clearly, unlike our main results, such results rely heavily on our modeling assumptions.)

**Proposition 9.** *Assume that $v_0 < \overline{v} - 1/b$ and there exists $\pi \in (a, b)$ such that $\gamma(\pi, v_0) < \gamma(a, v_0 + 1/b)$ and $\gamma(b, v_0) < \gamma(\pi, v_0 + 1/b)$. Assume also that $\theta_t$ becomes publicly observable at the end of each period $t$. Then, for any $g \in G$, the process $\{V^t\}_{t=1}^{\infty}$ converges to $g$ with positive probability.*

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.

[12]The assumptions in our extension differ from those of Proposition 5 only by requiring that $v_0 < \overline{v} - 1/b$ instead of requiring $v_0 < \overline{v} - \delta(b, b)$. While Proposition 4 identifies a broader range of $v_0$ as the domain of information segregation, we present information segregation as a special case of a static network here for simplicity.

## 9 Related Literature

A key idea underlying our work is that there is some aspect of cognition that is variable across individuals and stable over time, and that affects the manner in which information pertaining to a broad range of issues is filtered. 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.[13] Two especially relevant contributions from the perspective of our work are by Galeotti and Goyal (2010) and Acemoglu et al. (2011a). Galeotti and Goyal (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.

[13]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. (2011b), 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), Bloch et al. (2008) and Calvó-Armengol et al. (2011). We follow Bala and Goyal in focusing on the noncooperative formation of directed links.

Acemoglu et al. (2011a) 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.

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. (2009) 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, and this constitutes our main contribution to the literature.

## 10 Conclusions

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 prior 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 informations segregation.

Our analysis has been based on a number of simplifying assumptions. We have assumed that just one target can be observed in each period rather than several, and this could be relaxed by allowing for costs of observation that increase with the number of targets selected. Observation of the actions of others, and observation of the state itself could also be informative and affect beliefs about perspectives. It would also be worth relaxing the assumption of myopic choice, which would allow for some experimentation. We suspect that perfectly patient players will choose targets in a manner that implies history independence, but that our qualitative results will survive as long as players are sufficiently impatient. But these and other extensions are left for future research.

# Appendix

## Evolution of Beliefs and Information Networks

*Proof of Lemma 1.* To prove sufficiency, take $v_{ik}^t(h_t) > \beta\left(v_{ij}^t(h_t)\right)$. By definition of $\beta$,

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

where the inequality is by monotonicity of $\gamma$ and the equality is by definition of $\beta$. 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 $h_t$. Conversely, suppose that $v_{ik}^t(h_t) < \beta\left(v_{ij}^t(h_t)\right)$ for every $k \in N\setminus\{i,j\}$. Then, by definition of $\beta$, for all $k \notin \{i,j\}$,

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

where the inequality is by $\gamma$ being decreasing in $v$. Hence, by continuity of $\gamma$, 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 $\pi_{jt} \in [b-\eta, b]$ and $\pi_{kt} \in [a, a+\eta]$ for all $k \neq j$. This has positive probability, and on this event $l_{ij}^t = 1$, showing that link $ij$ is not broken at $h_t$. □

*Proof of Lemma 2.* To prove sufficiency, first take any $i,j$ with $v_{ij}^t(h_t) > \overline{v}$. Then, by definition of $\overline{v}$, for any $k \notin \{i,j\}$,

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

where the first inequality is because $\gamma$ is decreasing in $v$ and the second inequality is by definition of $\overline{v}$. Hence, by continuity of $\gamma$, 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 $\pi_{jt} \in [b-\eta, b]$ and $\pi_{kt} \in [a, a+\eta]$ for all $k \neq j$. This has positive probability, and on this event $l_{ij}^t = 1$. Hence $\Pr(b_{ij}^t = 1) = 0$. For any $s \geq t$, since $v_{ij}^s \geq v_{ij}^t \geq \overline{v}$, we have $\Pr(l_{ij}^s = 1) > 0$, showing that the link $ij$ is free. On the other hand, if $v_{ij}^t(h_t) \geq \max_{k\in N\setminus\{i,j\}} \beta\left(v_{ik}^t(h_t)\right)$, then, by Lemma 1, all the links $ik$ with $k \in N\setminus\{i,j\}$ are broken at $h_t$, and hence $i$ links to $j$ with probability one thereafter. Therefore, the link $ij$ is free. This proves sufficiency.

For the converse, take $v_{ij}^t(h_t) < \min\left\{\overline{v}, \max_{k\in N\setminus\{i,j\}} \beta\left(v_{ik}^t(h_t)\right)\right\}$. We will show that the link $ij$ will break with positive probability by some $t^* > t$. Since $v_{ij}^t(h_t) < \overline{v}$, $\beta\left(v_{ij}^t(h_t)\right)$ is finite. Moreover, since $v_{ij}^t(h_t) < \max_{k\in N\setminus\{i,j\}} \beta\left(v_{ik}^t(h_t)\right)$, there exists $k \neq j$ such that $\gamma\left(b, v_{ik}^t(h_t)\right) >$

$\gamma\left(a, v_{ik'}^t(h_t)\right)$ for every $k'$. If $v_{ik}^t(h_t) > \beta\left(v_{ij}^t(h_t)\right)$, by Lemma 1, the link $ij$ is broken at $h_t$, as desired. Assume that $v_{ik}^t(h_t) < \beta\left(v_{ij}^t(h_t)\right)$. By continuity of $\gamma$, there exists $\eta > 0$ such that $\gamma\left(\pi_{kt}, v_{ik}^t(h_t)\right) > \gamma\left(\pi_{k't}, v_{ik'}^t(h_t)\right)$ on the positive probability event that $\pi_{kt} \in [b-\eta, b]$ and $\pi_{k't} \in [a, a+\eta]$ for all $k' \neq k$. In that case, $i$ links to $k$, increasing $v_{ik}^t$ and keeping $v_{ik'}^t$ as is. Hence, $i$ keeps linking to $k$ on the positive probability event that $\pi_{ks} \in [b-\eta, b]$ and $\pi_{k's} \in [a, a+\eta]$ for all $k' \neq k$ and $s \in \{t, t+1, \ldots, t^*\}$ where $t^* = t + \left\lceil \left(\beta\left(v_{ij}^t(h_t)\right) - v_{ik}^t(h_t)\right)/\underline{\delta} \right\rceil$.[14] Then, on that event, by (9),

$$v_{ik}^{t^*} = v_{ik}^t(h_t) + \sum_{s=t}^{t^*} \delta(\pi_{is}, \pi_{ks}) \geq v_{ik}^t(h_t) + \left\lceil \left(\beta\left(v_{ij}^t(h_t)\right) - v_{ik}^t(h_t)\right)/\underline{\delta} \right\rceil \underline{\delta} > \beta\left(v_{ij}^t(h_t)\right),$$

where the inequality is by Observation 1. Therefore, the link $ij$ breaks by $t^*$ on this event. □

## Long-Run Frequency of Networks

In this subsection, we prove Proposition 1. The following definitions and notation will be useful. Let

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

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

$$p_{J,\lambda}(g) = \Pr\left(p_J(g) \mid \pi \notin D^\lambda\right)$$

denote the conditional probability distribution on $g$ obtained by restricting expertise realizations to lie outside the set $D^\lambda$, and $p_J$ is as defined in (10). Finally, for any probability distribution 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 $\varepsilon$ 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 proof.

**Observation 2.** *The following are true.*

1. *For every $\varepsilon > 0$, there exists $\lambda > 0$ such that $\Pr\left(D^\lambda\right) < \varepsilon$.*
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). The following lemma is the main step in our proof.

[14] Here, $\lceil x \rceil$ denotes the smallest integer larger than $x$.

**Lemma 3.** *Let $\lambda \in (0,1)$, $t_0$, $J$, and $h_{t_0}$ be such that*

$$v_{ij}^{t_0}(h_{t_0}) > \overline{v}_\lambda \text{ and } b_{ij'}(h_{t_0}) = 1 \qquad \left(\forall i \in N, \forall j \in J(i), \forall j' \notin J(i)\right),$$

*where $\overline{v}_\lambda$ is as in Observation 2. Then, for any $\varepsilon > \Pr\left(D^\lambda\right)$,*

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

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

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

where

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

and

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

Here, the last equality is by the hypothesis in the lemma and by the definition of $\overline{v}_\lambda$ in Observation 2. Hence, by the strong law of large numbers, as $t \to \infty$,

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

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 Proposition 1.* For every $h \in H$ and $\varepsilon > 0$, there exist $\lambda > 0$ and $h_{t_0}$ such that $\Pr(D^\lambda) < \varepsilon$, $|p_{J,\lambda}(g) - p_{J_h}(g)| < \varepsilon$ for all $g \in G$, and the hypothesis of Lemma 3 holds for $J = J_h$. Hence, writing

$$H^\varepsilon = \{h \in H \mid \phi_t(\cdot \mid h) \in B_{2\varepsilon}(p_J) \text{ eventually}\},$$

we conclude from Lemma 3 via the law of iterated expectations that

$$\Pr(H^\varepsilon) = 1.$$

Clearly, $H^\varepsilon$ is decreasing in $\varepsilon$, and as $\varepsilon \to 0$,

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

Therefore,

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

□

## History Independence

*Proof of Proposition 2.* First take $v_{ij}^t(h_t) < \overline{v}$ for some distinct $i, j \in N$. If

$$v_{ij}^t(h_t) \geq \max_{k \in N \setminus \{i,j\}} \beta\left(v_{ik}^t(h_t)\right),$$

then all the links $ik$ with $k \neq j$ are broken at $h_t$. Otherwise, as shown in the proof of Lemma 2, the link $ij$ is broken with positive probability by some $t^* > t$. In either case, $\Pr(j_{is} \in \arg\max_k \pi_{ks} | h_t)$ is bounded away from 1, showing that $\left\{V^t\right\}_{t=1}^\infty$ is *not* history independent at $h_t$.

Assume now $v_{ij}^t(h_t) > \overline{v}$ for all distinct $i, j \in N$. Of course, $v_{ij}^s(h_s) \geq v_{ij}^t(h_t) > \overline{v}$ for all distinct $i, j \in N$ and for every history after $h_t$. Now, since $\gamma(\pi, v)$ is continuous in $\pi$ and $1/v$ and $F$ is continuous over $[a, b]$, for every $\varepsilon > 0$, there exists $\tilde{v} < \infty$ such that $\Pr\left(j_{is} \in \arg\max_{j \neq i} \pi_{js}\right) > 1 - \varepsilon$ whenever $v_{ij}^s > \tilde{v}$ for all distinct $i$ and $j$. Hence, it suffices to show that, conditional on $h_t$, $v_{ij}^s \to \infty$ as $s \to \infty$ for all distinct $i$ and $j$ almost surely. To this end, observe that

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

where the first equality is because $\gamma$ is decreasing in $v_{ij}^t$ and the second inequality is by definition of $\overline{v}$. Hence, by continuity of $\gamma$, there exists $\eta > 0$ such that

$$\gamma\left(b - \eta, v_{ij}^t(h_t)\right) < \gamma(a + \eta, v) \qquad (\forall v, i, j).$$

Since $v_{ij}^s(h_s) \geq v_{ij}^t(h_t) > \overline{v}$, this further implies that

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

for every history that follows $h_t$, for every distinct $i$, $j$, $k$, and for every $s$. Consequently, $l_{ij}^{s+1} = 1$ whenever $\pi_{js} > b - \eta$ and $\pi_{ks} \leq a + \eta$ for all other $k$. Thus,

$$\Pr(l_{ij}^{s+1} = 1) \geq \lambda$$

after any history that follows $h_t$ and any date $s \geq t$ where

$$\lambda = F\left(a+\eta\right)^{n-2}\left(1 - F\left(b-\eta\right)\right) > 0.$$

Therefore, $l_{ij}^{s+1} = 1$ occurs infinitely often for all distinct $i, j \in N$ almost surely conditional on $h_t$. But whenever $l_{ij}^{s+1} = 1$, $v_{ij}^{s+1} \geq v_{ij}^{s} + \underline{\delta}$, where $\underline{\delta} = \delta\left(a, b\right) > 0$, showing that $v_{ij}^{s} \to \infty$ as $s \to \infty$ for all distinct $i, j \in N$ almost surely conditional on $h_t$. This completes the proof. □

## Network Structures

*Proof of Proposition 3.* Clearly, when $v_0 > \overline{v}$, the long-run outcome is history independent by Proposition 2, 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 $t^* > \left(\beta\left(v_0\right) - v_0\right)/\underline{\delta}$. 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 $ij \in S \equiv \{12, 21, 31, \ldots, n1\}$,

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

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

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

$$v_{ij_{i1}}^{2} = v_0 + \delta\left(\pi_{i1}, \pi_{ij_1}\right) \geq v_0 + \underline{\delta} \geq \beta\left(v_0\right)$$

while $v_{ik}^{2} = v_0$ for all $k \neq j_{i1}$, showing by Lemma 1 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 Proposition 4, which is proved below. □

*Proof of Proposition 4.* 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

$$v_0 + \delta\left(a + \varepsilon, b - \varepsilon\right) < \min\left\{\beta\left(v_0\right), \overline{v}\right\} \tag{15}$$

and

$$\delta(b-\varepsilon, b) > \delta(a+\varepsilon, b-\varepsilon). \tag{16}$$

By (16) and by continuity and monotonicity properties of $\gamma$, there also exist $\pi^* \in (a,b)$ and $\varepsilon' > 0$ such that

$$\begin{aligned} \gamma\left(\pi^* - \varepsilon', v_0 + \delta(b-\varepsilon, b)\right) &< \gamma(b, v_0) \\ \gamma\left(\pi^* + \varepsilon', v_0 + \delta(a+\varepsilon, b-\varepsilon)\right) &> \gamma(b-\varepsilon, v_0). \end{aligned} \tag{17}$$

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

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

Fixing

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

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 (16).Therefore, by (17), at $t = 2$, each $i \in S_1$ sticks to his previous link

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

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

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

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 (17),

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

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

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

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

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

Hence, by (17), $i$ sticks to $j_{im}$ by date $m + t^*$, yielding

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

for each $j \neq j_{im}$. By Lemma 1, this shows that the link $ij$ is broken. Since $j_{im} \in S_k$, this proves the result.

□

*Proof of Proposition 5.* 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 $\pi$ be as in Assumption 1. By continuity of $\delta$ and $\gamma$, there exists a small but positive $\varepsilon$ such that

$$\gamma(\pi, v_0) < \gamma(a, v_0 + \delta(b - \varepsilon, \pi + \varepsilon)) \tag{18}$$

$$\gamma(b - \varepsilon, v_0) < \gamma(\pi + \varepsilon, v_0 + \delta(\pi + \varepsilon, b - \varepsilon)) \tag{19}$$

$$\delta(b - \varepsilon, \pi + \varepsilon) > \delta(\pi + \varepsilon, b - \varepsilon). \tag{20}$$

Fix some

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

and consider the following positive probability event:

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

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 at every date $t \in N$, the individual $t$ becomes an ultimate expert (with precision nearly $b$), and his 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 $n + t^* + 1$. Towards this goal, we will first make the following observation:

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

At $t = 1$, 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 $t$, for any $i \geq t$, $i$ has observed each $j \in \{1, \ldots, t-1\}$ once, when his own precision was in $[a, \pi + \varepsilon]$ and the precision of $j$ was in $[b - b, b]$. Hence, by Observation 1, $v_{ij}^t \leq v_0 + \delta(\pi + \varepsilon, b - \varepsilon)$. He has not observed any other individual, and hence $v_{ij}^t = v_0$ for all $j \geq t$. Thus, by (19), for any $i > t$, $\gamma(\pi_{tt}, v_{it}^t) < \gamma\left(\pi_{jt}, v_{ij}^t\right)$ for every $j \in N \setminus \{i, t\}$, showing that $i$ observes $t$, i.e., $j_{it} = t$. Likewise, by (18), for $i = t$, $\gamma\left(\pi_{g(t)t}, v_{tg(t)}^t\right) < \gamma\left(\pi_{jt}, v_{tj}^t\right)$ for every $j \in N \setminus \{t, g(t)\}$, showing that $t$ observes $g(t)$, i.e., $j_{tt} = g(t)$. Finally, for any $i < t$, by the inductive hypothesis, $i$ has observed any $j \neq g(i)$ at most once, yielding $v_{ij}^t \leq v_0 + \delta(\pi + \varepsilon, b - \varepsilon)$. Hence, as above, for any $j \in N \setminus \{i, t, g(i)\}$, $\gamma(\pi_{tt}, v_{it}^t) < \gamma\left(\pi_{jt}, v_{ij}^t\right)$, showing that $i$ does not observe $j$, i.e., $j_{it} \in \{g(i), t\}$.

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

$$v_{ij}^{n+1} \leq v_0 + \delta(\pi + \varepsilon, b - \varepsilon) \qquad (\forall j \neq g(i)). \tag{21}$$

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

$$v_{ig(i)}^{n+1} \geq v_0 + \delta(b - \varepsilon, \pi + \varepsilon). \tag{22}$$

By definition of $A$, inequalities (21) and (22) imply that each $i$ observes $g(i)$ at $n+1$. Consequently, the inequalities (21) and (22) also hold at date $n+2$, leading each $i$ again to observe $g(i)$ at $n+2$, and so on. Hence, at dates $t \in \{n+1, \ldots, t^* + n\}$, each $i$ observes $g(i)$, yielding

$$\begin{aligned} v_{ig(i)}^{n+t^*+1} &\geq v_{ig(i)}^{n+1} + t^* \underline{\delta} > v_0 + \delta(b - \varepsilon, \pi + \varepsilon) + \beta(v_0 + \delta(\pi + \varepsilon, b - \varepsilon)) - v_0 \\ &> \beta(v_0 + \delta(\pi + \varepsilon, b - \varepsilon)). \end{aligned}$$

For any $j \neq g(i)$, since $v_{ij}^{n+t^*+1} = v_{ij}^{n+1}$, together with (21), this implies that

$$v_{ig(i)}^{n+t^*+1} > \beta\left(v_{ij}^{n+t^*+1}\right).$$

Therefore, by Lemma 1, the link $ij$ is broken at date $t^* + n + 1$. □

*Proof of Proposition 6.* Take $v_0 \leq \hat{v}$, so that $v_0 + \delta(b, b) \geq \beta(v_0)$. Write $i^* = \arg\max_i \pi_{i1}$ and $j^* = \arg\max_{i \neq i^*} \pi_{i1}$. With probability 1, $\pi_{i^*1} > \pi_{j^*1}$. Hence,

$$v_{i^*j^*}^2 = v_0 + \delta(\pi_{i^*1}, \pi_{j^*1}) > v_0 + \delta(\pi_{i^*1}, \pi_{i^*1}) \geq v_0 + \delta(b, b) \geq \beta(v_0),$$

showing that the link $i^*j^*$ is broken by Lemma 1. To see the penultimate equality, note that $\delta(\pi, \pi)$ is decreasing in $\pi$. Conversely, when $v_0 > \hat{v}$, there exists $\varepsilon > 0$ such that $v_0 + \delta(b, b - \varepsilon) < \beta(v_0)$. Then, no link is broken in the first period when $(\pi_{11}, \ldots, \pi_{n1}) \in [b - \varepsilon, b]^N$. □

*Proof of Proposition 8.* In the proof of Proposition 5, for sufficiently small $\varepsilon$, take

$$t^* > \left(\beta\left(v_0 + 1/\left(b - \varepsilon\right)\right) - v_0\right)/\underline{\delta},$$

and set

$$\overline{\tau} = t^* + n.$$

As shown there, on the open set $A$, each player $i$ observes $g\left(i\right)$ at date $i$ and at all dates $\{n+1, \ldots, n+t^*\}$, while observing any other player $j$ at most once—at date $j$ when $\pi_{jj} \geq b - \varepsilon$. Under delayed observation, the same behavior emerges at those dates. As in the unobservable case,

$$v_{ig(i)}^{n+t^*+1} \geq v_{ig(i)}^{n+1} + t^*\underline{\delta} > \beta\left(v_0 + 1/\left(b - \varepsilon\right)\right),$$

and for any $j \neq g\left(i\right)$,

$$v_{ij}^{n+t^*+1} \leq v_0 + \delta\left(a + \varepsilon, b - \varepsilon\right) < v_0 + 1/\left(b - \varepsilon\right).$$

Therefore, $i$ does not observe $j$ under any realization on dates $t \in \{n + t^* + 1, \ldots, \tau + j\}$. At the end of date $\tau + j$, $\theta_j$ becomes observable. If $i$ observed $j$ on date $j$, he updates his belief about $\mu_j$, and $v_{ij}^{\tau+j+1}$ becomes higher than $v_{ij}^{n+t^*+1}$ but we still have

$$v_{ij}^{\tau+j+1} = v_0 + 1/\pi_{jj} \leq v_0 + 1/\left(b - \varepsilon\right).$$

Since $v_{ig(i)}^{\tau+j+1} \geq v_{ig(i)}^{n+t^*+1} > \beta\left(v_0 + 1/\left(b - \varepsilon\right)\right)$, the link $ij$ is still broken. □

*Proof of Proposition 9.* In the proof of Proposition 5, simply change each $\delta\left(\pi_i, \pi_j\right)$ to $1/\pi_j$.

## References


[1] Acemoglu, Daron, Victor Chernozhukov, and Muhamet Yildiz. 2009. “Fragility of Asymptotic Agreement under Bayesian Learning.” Massachusetts Institute of Technology Department of Economics Working Paper No. 08-09.

[2] Acemoglu Daron, Kostas Bimpikis, Asuman Ozdaglar (2011a), “Dynamics of Information Exchange in Endogenous Social Networks.” Unpublished Manuscript, MIT.

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

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

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

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

[7] 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.

[8] 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.

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

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

[11] Calvó-Armengol, Antoni, Joan de Martí and Andrea Prat (2011). Communication and Influence, Unpublished Manuscript, London School of Economics

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

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

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

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

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

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

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

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

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

[21] 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.

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

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

[24] 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.

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

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

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

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

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

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

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