Ideological Bias and Trust in Social Networks

PRELIMINARY: PLEASE DO NOT CITE OR CIRCULATE

Matthew Gentzkow*
Stanford University and NBER

Michael B. Wong
MIT

Allen T. Zhang
Harvard University

September 2015

Abstract

Two key features robustly describe ideological differences in society: (i) individuals persistently disagree about objective facts; (ii) individuals also disagree about which sources can be trusted to provide reliable information about these facts. We develop a model in which these patterns arise endogenously as the result of small deviations from Bayesian information processing. Individuals receive information from direct observation, subject to an ideological bias of which they are unaware. They also receive information through social networks and media. Sources differ in their reliability and individuals must learn which sources they can trust. We show that the entry of partisan information sources could generate large ideological disagreements, even when individuals have arbitrarily small biases and observe the same set of sources. Individuals may also come to trust the most biased like-minded sources, leading to ideological segregation in social networks and extreme bias by strategic media outlets.

Keywords: Bias, trust, networks, social media, media bias
JEL: D83, D85, L82

*

1 Introduction

Individuals persistently disagree about objective facts along ideological lines. Fervent political debates over the validity of global warming, evolution, and vaccination have persisted long after the establishment of a scientific consensus. The First Assessment Report of the Intergovernmental Panel on Climate Change (IPCC 1990) crystallized a scientific consensus that global temperatures are rising. However, a 2013 Pew survey found that while over 80 percent of Democrats believe that there is solid evidence to show that the earth is warming, fewer than half of Republicans and fewer than 30 percent of Tea Party Republicans do. Similarly, Darwinian evolutionary theory has been dominant among academicians for nearly a century. Yet, a 2013 Pew survey also found that, while over 70 percent of Democrats believe that humans have evolved over time, fewer half of Republicans do.

Where do these ideological differences come from? A clue is that, while news consumption in the U.S. is not generally segregated or one-sided (Gentzkow and Shapiro 2011; Prior 2013), individuals disagree intensely over which news sources can be trusted to provide reliable information. Public Policy Polling (2013), for instance, found that while over 70 percent of Democrats trust MSNBC, a left-leaning television broadcaster, fewer than 10 percent of Republicans do. By contrast, nearly 80 percent of Republicans trust Fox News, a right-leaning broadcaster, while fewer than 25 percent of Democrats say the same.

We propose a theory in which divergent trust of information sources arises endogenously as the result of small deviations from Bayesian information processing. The theory predicts large and persistent disagreements even when all individuals have access to identical and arbitrarily informative signals. We show that the proliferation of partisan information sources increases ideological polarization even when media consumption is unsegregated. We then endogenize choices by firms and peers in media markets and social networks, and show that endogenous choices amplify political disagreements through ideological segregation in social network formation and media consumption.

In our model, individuals wish to learn a distinct state variable $\omega_t \in \{0,1\}$ drawn independently in each period t. Each state is associated with a binary ideology, L or R, also drawn independently each period. Each individual receives a noisy private signal about ω_t with a fixed precision. We think of private signals as capturing information from personal experience, logical reasoning, or articles of faith. We deviate from pure Bayesian information processing by assuming that an *L*- or *R*-biased agent misreads the private signal with positive probability when the signal disagrees with her ideological leaning, but that she does not realize this mistake. Each individual also receives signals about ω_t through social networks and media. These external information sources differ in their reliability. Individuals entertain the possibility that external signals may be arbitrarily correlated and must learn their joint distribution.

Our first main result is that an agent with any bias fails to fully learn the underlying state, even in the limit where she observes a large number of independent and unbiased neighbors. The intuition for the result is simple: Each agent learns the precision of sources by observing the frequency of agreements between observed signals. Since she believes that her private signal is unbiased, the private signal is the yardstick against which all other signals are judged. However, the agent fails to account for her bias and is hence overconfident in her private signal. She eventually comes to believe that her neighbors are less precise than they in fact are and that their signals are correlated. Consequently, she fails to infer the underlying state from her neighbors, even though she could have had she known the actual precisions of her neighbors.

Our second set of results shows that ideological disagreements about the underlying state ω_t increase with the prevalence of partisan sources even when two agents observe the same set of information sources. These disagreements can be both large and persistent. Suppose, for example, that two individuals with moderate but opposite biases observe a large number of both biased and unbiased information sources. We find that they may disagree as much as half the time in expectation even after a large number of periods — much more than if she had only observe unbiased neighbors. To see why, consider that each agent observes too frequent agreements with her neighbors when the underlying state and her ideological bias are aligned to justify her initial belief about her own precision. Consequently she becomes more overconfident about her private signal and correspondingly less confident about the unbiased neighbors. Therefore the introduction of partisan information sources intensifies ideological disagreements, despite the absence of ideological segregation in social networks or media consumption.

We relax our assumption that networks are fixed and unsegregated to obtain our third set of results: Biased information processing generates ideological homophily. More formally, we show that a biased agent's trust in her neighbor increases with the neighbor's bias when her own signal is sufficiently imprecise. Such agents segregate into ideologically similar groups when the cost of obtaining information is high. Furthermore, media outlets adopt extreme partisan position in order to attract such listeners. Both effects increases the intensity of ideological disagreements in the population.

Formally, our model is most closely related to the literature on asymptotic agreement under Bayesian learning. A number of classic papers argue that when two agents observe a common sequence of exchangeable signals and agree on zero probability events, they will asymptotically agree on the data generating process (Savage 1954; Blackwell and Dubins 1962). More recently, Acemoglu et al. (forthcoming) demonstrate that a small amount of uncertainty and heterogeneous priors could lead to significant asymptotic disagreements due to lack of identification. Our model is a variant of the latter work: In our setup, each individual must learn the trustworthiness of *multiple* information sources from sequences of signals, with a *different* underlying state in each period. We then apply this theoretical framework to study the effect of media market structure and social networks on political polarization. In this sense, our work is also related to models in which media bias arises from preference for like-minded news or reputational concerns (see, e.g., Mullainathan and Shleifer 2005; Gentzkow and Shapiro 2006; Burke 2008; Stone 2011). We deviate from these papers by deriving conditions under which bias and ideological disagreements persist even in the asymptotic limit where aggregate information becomes arbitrarily large.

Topically, our work relates to the large literature on the causes and effects of political polarization (Glaeser and Ward 2006; McCarty et al. 2006). While we theoretically study the relationship between media and polarization, a number of empirical papers consider the effects of media on ideological extremism (Campante and Hojman 2013; Prior 2013). A literature relate ideological extremism to non-Bayesian information processing (Lord et al. 1979; Glaeser and Sunstein 2013; Ortoleva and Snowberg 2015). By contrast, Benoit and Dubra (2014) explain attitude polarization using a rational theory of information processing. A related literature considers ideological segregation in social networks (Gentzkow and Shapiro 2011; Flaxman et al. 2013; Halberstam and Knight 2014). A large body of empirical work also study the effects of media on political beliefs and behavior (DellaVigna and Kaplan 2007; Gerber et al. 2009; Chiang and Knight 2011; Enikolopov et al. 2011; Falck et al. 2014; Martin and Yurukoglu 2014).

The paper proceeds as follows. Section 2 describes the model and discusses the assumptions of the model. Section 3 discusses convergence in the asymptotic limit when agents have observed signal realizations over many periods. Section 4 characterizes when agents are able to learn the underlying state. Section 5 characterizes asymptotic ideological disagreement in exogenously fixed social networks where all agents observe the same set of signals. Section 6 characterizes asymptotic disagreement when agents endogenously choose their neighbors and media outlets strategically choose their biases. Section 7 concludes.

2 Model

In this section, we develop a model of imperfect Bayesian learning from potentially biased information sources. The model explains how large ideological disagreements may arise when agents must learn which information sources to trust or distrust.

2.1 Preliminaries

There are a large number of agents $i \in \{1,\dots,N\}$ and time periods $t \in \{1,\dots,T\}$. In each period t, agents form beliefs about an unknown state of the world $\omega_t \in \{0,1\}$. It is common knowledge that the state is drawn independently each period with $\Pr(\omega_t = 1) = \frac{1}{2}\,\forall t$. Agents value having accurate beliefs, and discount the future at rate δ.

In each period, one of the possible states $\omega_t \in \{0,1\}$ is associated with ideology R, while the other one is associated with ideology L. Let r_t and $l_t = 1 - r_t$ denote the R- and L-associated states, respectively. The ideological assignment r_t does not depend on ω_t and is drawn independently each period with $\Pr(r_t = 1) = \frac{1}{2}$. We assume that r_t is unobserved to agents. This assumption captures the idea that agents do not know a priori which policies or scientific conclusions favor one side of the political spectrum or the other. Rather, ideology emerges endogenously from underlying correlations in signal generation.[1] However, each agent i has a bias $\beta_i \in [-1,1]$ and receives information that favors the state associated with their bias. We will say that an agent i is "unbiased" if $\beta_i = 0$, that she is "R-biased" if $\beta_i > 0$, and that she is "L-biased" if $\beta_i < 0$.

[1] To model a world in which ideology is observed rather than unobserved, we might assume that all agents observe a neighbor j whose signal $s_{jt} = r_t$. This is the case where each agent has neighbors that are fully informative about r_t (as defined in section 5).

Agents receive information about ω_t from two types of sources. First, each agent i observes a noisy signal $s_{it} \in \{0,1\}$ about ω_t in each period t. The signal s_{it} is based on an underlying signal $\tilde{s}_{it} \in \{0,1\}$, which has precision $\alpha_i \in \left(\frac{1}{2},1\right]$ in the sense that $\tilde{s}_{it} = \omega_t$ with probability α_i and $\tilde{s}_{it} = 1-\omega_t$ with probability $1-\alpha_i$. Unbiased agents will always observe $s_{it} = \tilde{s}_{it}$. R-biased (L-biased) agents believe they observe $s_{it} = \tilde{s}_{it}$, but in fact mistake $\tilde{s}_{it}$ for r_t (l_t) with probability $|\beta_i|$.[2]

It will be useful to characterize bias in terms of an agent's likelihood of mistaking $\tilde{s}_{it}$ for r_t or l_t. We say that an agent i has **extreme bias** if $\beta_i \geq \psi(\alpha_i) \equiv 1 - \frac{1}{2\alpha_i}$ and **moderate bias** if $\beta_i < \psi(\alpha_i)$. For example, if i is an R-biased agent with extreme bias, then $\Pr(s_{it} = r_t) \geq \Pr(s_{it} = \omega_t)$. Note that for any $\beta_i \in \left(-\frac{1}{2},\frac{1}{2}\right)$, there exists an $\alpha_i \in \left(\frac{1}{2},1\right)$ such that bias is not extreme.

Second, each agent also observes signals from her social network. Let $\mathscr{I}_{it} \subseteq \{1,\ldots i-1, i+1,\ldots,N\}$ denote the set of agent i's neighbors in period t. Each agent i directly observes s_{jt} for all $j \in \mathscr{I}_{it}$. In assuming direct observation, we abstract from strategic reporting of signals by agents. We will later relax this "mechanical reporting" assumption by introducing "media outlets" who know the distribution of biases in the population and choose β_i strategically. For notational convenience, we write $\mathbf{s}_{\sim it} \equiv \left\{s_{jt}\right\}_{j\in\mathscr{I}_{it}}$ and $S_{it} \equiv \mathbf{s}_{\sim it} \cup \{s_{it}\}$.

2.2 Prior and Posterior Beliefs

We are interested in the evolution of each agent's posterior beliefs. Let $B_{it}(\cdot)$ denote agent i's posterior belief over the probability of a given event after observing $\{S_{i\tau}\}_{\tau=1}^{t}$ and let $B_i(\cdot) = \lim_{t\to\infty} B_{it}(\cdot)$ if the limit exists. We shall summarize each agent's beliefs using the following mathematical objects:

1. Let $\mu_{it} = B_{it}(\omega_t = 1)$. We shall say that μ_{it} is agent i's **belief** of ω_t in period t. By computing μ_{it}, we can infer whether agents learn the underlying state and how frequently agents disagree over time.

2. Let $T_{ijt} = B_{it}\left(s_{j(t+1)} = \omega_{t+1}\right)$. We shall say that T_{ijt} is agent i's **trust** in agent j in period t. Note that agent i's trust in herself is denoted by T_{iit}. By computing T_{ijt}, we can infer which

[2] For an R-biased agent i, $s_{it} = \tilde{s}_{it}$ if $\tilde{s}_{it} = r_t$; if $\tilde{s}_{it} = l_t$, then $s_{it} = \tilde{s}_{it}$ with probability $1-|\beta_i|$ and $s_{it} = r_t$ with probability $|\beta_i|$. The case for a L-biased agent is analogous.

neighbors agent i believes to be more accurate.

Each agent's inference task over ω_t would be straightforward if the joint conditional distribution of all signals given ω_t were known with no uncertainty. However, agents in our model do not know this joint distribution and must learn it by observing agreements between signal realizations over time.

This modeling setup captures the idea that we often trust information sources only because they agree with *other sources* we trust. One might say, for example, that a new NBER working paper seems credible because estimates are similar to results by other well-known authors, or that Republicans are credible on global warming because they agree with *Fox News.* But this chain has to stop somewhere. Ultimately we must have some independent point of reference in order to learn anything at all. The joint distribution of all available signals cannot be identified from observed signal realizations alone. What an agent learns about ω_t depends crucially on the agent's prior beliefs about the joint distribution even as $t \to \infty$.

We think of each agent i's own signal s_{it} as capturing her independent point of reference. This signal s_{it} may come from agent i's personal experiences, logical reasoning, or articles of faith. For instance, suppose the underlying state concerns the social desirability of unionization. An agent's own signal about the underlying state may arise from her personal experience as a union member, her logical reasoning from economic principles about collective bargaining, or her faith in the veracity of Milton Friedman or Karl Marx's writings. Even though the signal may be colored by ideological bias, in our model each agent i thinks of s_{it} as an unbiased signal of the underlying state.

To make the above precise, we introduce the following assumption.

Assumption 1. *We impose the following restrictions on each agent i's prior and posterior beliefs:*

(a) Agent i's prior on her own bias β_i is degenerate (Dirac) at $\hat{\beta}_i = 0$.

(b) Agent i's prior belief of her own precision α_i has non-zero density on $\alpha \in \left(\frac{1}{2}, 1\right]$, and zero density elsewhere.

(c) Agent i's prior belief over the joint distribution of her neighbors' signals $\mathbf{s}_{\sim it}$ given ω_t has full support over every possible signal realization.

Assumption 1(a) implies that agents do not entertain the possibility that they are biased. The immediate implication is that each agent i believes that errors in s_{it} are uncorrelated with errors in s_{jt} for all $j \neq i$. Assumption 1(b) captures the idea that agents do entertain uncertainty about the precisions of their signals. Their priors may be arbitrarily concentrated on the true α_i, but as we shall see from lemma 1 below, a full-support prior over α_i is necessary for the agent to rationalize all possible signal realizations permissible under the model setup.

Finally, assumption 1(c) posits that agents entertain the full range of possibilities about the precision and correlation of their neighbors' signals. They may assign very high prior weight to what is in fact the true distribution. But they do not rule out the possibility that their neighbors' signals are noisier than they are, or even perverse, in the sense that they are more likely to be wrong than right. Similarly, they do not rule out that their neighbors' signals are correlated with each other, even if they are in fact independent.

3 Asymptotic Convergence

We now consider the agent's posterior beliefs as the number of periods grow large. Here we assume that the social network is fixed over time, i.e., $\mathscr{I}_{it} = \mathscr{I}_i \forall t$. In the limit where $t \to \infty$, the law of large numbers disciplines the agent's posterior beliefs. Therefore, under certain conditions it is possible to derive closed-form expressions for the asymptotic limits of T_{ijt} and μ_{it} and use them to study ideological disagreement in a large class of settings.

We first ask whether the above assumptions on the agent's beliefs can be consistent with the data that she observes. We shall say that the **the data violate the agent's model of the world** if the likelihood that the agent assigns to the observed data approaches zero. Under what prior beliefs does this occur?

Lemma 1. *The agent's model of the world is not violated by agent i's observed signals in the limit as* $t \to \infty$ *if and only if for all* $\mathbf{s} \in \{0,1\}^{|\mathscr{I}_i|}$, $\Pr(s_{it} = 1 | \mathbf{s}_{\sim it} = \mathbf{s}) \in [1-\alpha, \alpha]$ *for some* α *in the support of i's prior on* α_i.

Lemma 1 shows that the data could reject the agent's model of the world only if her prior implied her signals precision was relatively low with probability 1. Conditional on believing her signal is sufficiently precise, an agent can always rationalize observed signal realizations. Given assumption 1(b), lemma 1 implies the data never violate the agent's model of the world.

The following lemma provides sufficient conditions for the agent's beliefs to converge to well-defined limits. The conditions are much stronger than necessary, but they simplify the analysis that follows.

Lemma 2. *Suppose that assumption 1 holds. Furthermore, suppose that agent i's prior distribution on her precision* α_i *is* F_n^i*, where* $\{F_k^i\}_{k=1}^{\infty}$ *a sequence of prior distributions each with non-zero density* f_n^i *on* $\left(\frac{1}{2},1\right]$*, zero density elsewhere, and the property that for any* α' *and* α *such that* $|\alpha-\alpha_i| < |\alpha'-\alpha_i|$*, the likelihood ratio* $f_n^i(\alpha')/f_n^i(\alpha)$ *goes to zero in the limit where* $n \to \infty$*. Then as* $n \to \infty$*, in the limit where* $t \to \infty$*,* μ_{it} *converges in distribution to a random variable* μ_i *and* $\{T_{ijt}\}_{j\in\mathscr{I}_i}$ *converges in probability to constants* $\{T_{ij}\}_{j\in\mathscr{I}_i}$*.*

Lemma 2 establishes that if agent i's prior belief over α_i is sufficiently concentrated around the true value, the induced posterior distribution over her neighbors signals is nearly degenerate in the limit as $t \to \infty$. We shall say that the limit T_{ij} is agent i's **asymptotic trust** in j, and μ_i is i's **asymptotic belief**. For the remainder of the paper, we focus on characterizing asymptotic beliefs and trust, which apply approximately to agents whose priors place high weight on values of α_i close to the true value.

4 Asymptotic Learning

We now ask whether it is possible for an agent to correctly learn the underlying state ω_t if she observes a large number of unbiased and independent neighbors. We use the exercise to provide intuition over the role that assumption 1 play in the formation of each agent's beliefs.

We begin with a few useful definitions.

Definition 1. We say that agent i has **exogenous trust** if her prior over the joint distribution of neighbors' signals is degenerate at the truth, and that agent i has **endogenous trust** if her prior over this joint distribution has full support, as posited by assumption 1.

Definition 2. There is **asymptotic learning** for agent i if for any $\varepsilon > 0$, $\Pr(|\mu_i - \omega_t| < \varepsilon) = 1$. We say that asymptotic learning fails otherwise.

To build intuition over how the assumptions drive the results we shall establish in this paper, we first consider a simple example where agent i has a single neighbor j with signal $s_{jt} = \omega_t$. Under exogenous trust, an agent exogenously (and correctly) believes her neighbor' signal was fully accurate, and the agent would fully learn the underlying state from the accurate neighbor and there would be no disagreement.

However, under endogenous trust, an agent with $\beta_i \neq 0$ must learn how much to trust this neighbor by observing signal realizations. Given assumption 1, a biased agent is unaware of her bias and places an inaccurately high prior on the precision of her own signal.[3] As a result she observes too many disagreements between her own private signal and the neighbor's signal to believe that her neighbor simply reports ω_t. She instead infers that her neighbor must be making mistakes with non-zero probability. The agent does not fully trust her accurate neighbor even though the neighbor possesses fully accurate information. Asymptotic learning fails even for arbitrarily small biases.

Now consider what happens when the agent observes a large number of noisy but unbiased neighbors under endogenous trust. The following proposition extends the logic of the above example.

Proposition 1. *Suppose an agent i with $\alpha_i \in \left(\frac{1}{2}, 1\right)$ and $\beta_i \neq 0$ is connected to N identical neighbors with $\alpha_j > \frac{1}{2}$ and $\beta_j = 0$. Under exogenous trust there is asymptotic learning for agent i in limit where $N \to \infty$. However, under endogenous trust asymptotic learning fails, and agent i's posterior distribution over her neighbors' signals has the following properties as $t \to \infty$:*

1. *$T_{ij} < \alpha$ for all j;*

2. *All s_j are positively correlated.*

Proposition 1 shows that, under endogenous trust, agent i perceives her neighbors to be making correlated errors, when in fact their signals are independent and unbiased. Furthermore, agent i never fully trusts them and asymptotic learning fails.

[3]Formally, $T_{ii} \geq \alpha_i > \Pr(s_{it} = 1 \mid \omega_t = 1)$.

5 Disagreement under Unsegregated Networks

We now turn to characterizing the magnitudes of ideological disagreements arising from these social networks. This section considers disagreement when all agents observe the same exogenously fixed neighbors. We show that ideological disagreements increase with partisan information sources even without ideological segregation in social networks. In the next section, we relax the assumption that the network is fixed, and consider how endogenous network formation and strategic reporting affects the magnitudes of ideological disagreements.

Throughout this section, we consider three representative agents with precision $\alpha_i = a$ and different states of bias: an unbiased agent U with $\beta_U = 0$, an R-biased agent R with $\beta_R = b \in \left(0, \frac{1}{2}\right)$, and a L-biased agent L with $\beta_L = -b$. We are interested in the expected asymptotic disagreement between agents of opposite bias; that is, $\mathrm{E}\left[|\mu_R - \mu_L|\right]$. This expression quantifies the extent of ideological disagreement between the representative R-biased agent and the representative L-biased agent about an underlying state ω_t in the limit as $t \to \infty$.

We consider two special cases: one where the representative agent has N identical unbiased neighbors with precision $\alpha_j > \frac{1}{2}$, and another where all neighbors instead have some identical moderate bias $\beta_j \neq 0$. We characterize ideological disagreement in the limit as these networks become arbitrarily large (as $N \to \infty$). This comparison illustrates the effect of increasingly partisan news on ideological disagreement in the absence of endogenous sorting and network formation.

5.1 Simplifying the Problem

We begin by recasting our task of deriving posterior beliefs in complex social networks. While it is possible to derive closed-form expressions for μ_i in terms of the underlying data-generating parameters, $\left\{\alpha_j, \beta_j\right\}_{j \in \mathscr{I}_i \cup \{i\}}$, any such expression becomes unwieldy and difficult to interpret as the number of neighbors grows.

Instead we use the notion of informativeness to summarize the information content available to an agent through her neighbors. Any set of neighbors can be totally ordered in terms of their informativeness about ω_t (or r_t) following Blackwell's (1951) criterion. Here we examine the limiting cases as information becomes arbitrarily informative or uninformative.

Definition 3. We say that a set of neighbors $\mathscr{I}_i$ is **fully informative** about ω_t if $\Pr\left(\omega_t = \omega \mid \mathbf{s}_{\sim it} = \mathbf{s}\right) \in$

$\{0,1\}$ for all $\omega \in \{0,1\}$ and $\mathbf{s} \in \{0,1\}^{|\mathscr{I}_i|}$. We say that a sequence of sets of neighbors **approaches full informativeness** if $\Pr(\omega_t = \omega \mid \mathbf{s}_{\sim it} = \mathbf{s}) \to 1$ or $\Pr(\omega_t = \omega \mid \mathbf{s}_{\sim it} = \mathbf{s}) \to 0$ for all $\omega \in \{0,1\}$ and $\mathbf{s} \in \{0,1\}^{|\mathscr{I}_i|}$. Similarly, $\mathscr{I}_i$ is **uninformative** about ω_t if $\Pr(\omega_t = \omega \mid \mathbf{s}_{\sim it} = \mathbf{s}) = \frac{1}{2}$ for all $\omega \in \{0,1\}$ and $\mathbf{s} \in \{0,1\}^{|\mathscr{I}_i|}$. We define informativeness about r_t analogously.

The above definition establishes equivalence classes of neighbors in which information content available to an agent i along each of the two relevant dimensions becomes either arbitrarily small or large.

These cases apply intuitively to single-agent networks. An agent with a single neighbor j with $s_{jt} = \omega_t$ is fully informative about ω_t but uninformative about r_t. Conversely, a network consisting of a single agent k with $s_{kt} = r_t$ is fully informative about r_t but uninformative about ω_t. A single agent can only be fully informative about ω_t or r_t but never both. However, if an agent i has both j and k as neighbors, then the set of neighbors is fully informative about both ω_t and r_t.

A set of neighbors may also approach full informativeness as the social network grows arbitrarily large and dense, even if none of the agents are themselves fully informative. The following lemma shows that the networks described earlier fall in these informational limits:

Lemma 3. *A network of N identical neighbors with $\alpha_j > \frac{1}{2}$ and $\beta_j = 0$ approaches fully informativeness about ω_t in the limit as $N \to \infty$. A network of N identical neighbors with $\alpha_j > \frac{1}{2}$ and moderate bias $\beta_j \neq 0$ approaches fully informativeness about both ω_t and r_t in the limit as $N \to \infty$.*

This characterization is useful because μ_i simplifies dramatically in these informational limits, yielding straightforward intuitions and results.[4] It is easy to show, for instance, that μ_i is the same for any set of neighbors in the same informational limit (e.g., fully informative about ω_t and uninformative about r_t). While we established proposition 1 only for the case where the agent has

[4]To see why μ_i dramatically simplifies the problem, recall that, by definition, μ_i is a distribution that depends on the signal realizations $s \in \{0,1\}$ and $\mathbf{s} \in \{0,1\}^{\mathscr{I}_i}$ in period t. We can rewrite μ_i as follows:

$$\mu_i = \left(1 + \frac{B_i(s_{it} = s \mid \omega_t = 0)}{B_i(s_{it} = s \mid \omega_t = 1)} \frac{B_i(\mathbf{s}_{\sim it} = \mathbf{s} \mid \omega_t = 0)}{B_i(\mathbf{s}_{\sim it} = \mathbf{s} \mid \omega_t = 1)}\right)^{-1}. \tag{1}$$

Note that $B_i(s_{0t} \mid \omega_t) \in \{T_{ii}, 1 - T_{ii}\}$ and by equation (5), we have that, almost surely,

$$\frac{B_i(\mathbf{s}_{\sim it} = \mathbf{s} \mid \omega_t = 0)}{B_i(\mathbf{s}_{\sim it} = \mathbf{s} \mid \omega_t = 1)} = \frac{T_{ii} - \Pr(s_{it} = 1 \mid \mathbf{s}_{\sim it} = \mathbf{s})}{T_{ii} - 1 + \Pr(s_{it} = 1 \mid \mathbf{s}_{\sim it} = \mathbf{s})}. \tag{2}$$

Suppose $\omega_t = \omega$ and $r_t = r$ in period t. If $\mathbf{s}$ is fully informative about an underlying state, the probability that $s_{it} = 1$ conditioned on the realized signal $\mathbf{s}$ is effectively the same the probability conditioned on the state. More precisely, it a single neighbor j that reports $s_{jt} = \omega_t$, the proposition in fact applies whenever an agent has a set of neighbors that is fully informative about ω_t.

5.2 Comparative Statics

First, consider the case where the representative agents are connected to the arbitrarily large network of unbiased neighbors. We establish the following lemma regarding the extent of disagreements between agents of opposite biases.

Proposition 2. *Suppose representative agents U, R, and L are each connected to N identical neighbors with $\alpha_j > \frac{1}{2}$ and $\beta_j = 0$. For any $a \in \left(\frac{1}{2}, 1\right)$, $b \in \left(0, \frac{1}{2}\right)$, $\text{plim}_{N\to\infty} \text{E}\left[|\mu_R - \mu_L|\right] \leq b$ and $\text{plim}_{N\to\infty} \mu_U = \omega_t$.*

Proposition 2 shows that asymptotic disagreement is bounded by the bias of the agents. Therefore, asymptotic disagreement is small in a setting where individuals only have moderate biases.

Next, suppose the neighbors have moderate bias $\beta_j \neq 0$. We think of this case as capturing a world in which agents have access to both objective and partisan news. The addition of these neighbors does not affect inference for the unbiased agent U, but dramatically increases disagreement between the biased agents.

Proposition 3. *Suppose representative agents U, R, and L are each connected to N identical neighbors with $\alpha_j > \frac{1}{2}$ and non-zero but moderate bias. For any $\varepsilon > 0$ and $b \in \left(0, \frac{1}{2}\right)$, there exists $a \in \left(\frac{1}{2}, 1\right)$ such that each representative agent's bias is moderate and $\text{plim}_{N\to\infty} \text{E}\left[|\mu_R - \mu_L|\right] > \frac{1}{2} - \varepsilon$, while $\text{plim}_{N\to\infty} \Pr\left(|\mu_U - \omega_t| < \varepsilon\right) = 1$ for any $a \in \left(\frac{1}{2}, 1\right)$.*

Proposition 3 establishes two facts. First, despite the addition of moderate bias in all of her neighbors, the unbiased agent nevertheless accurately learns the underlying state with no uncertainty. Second, agents of opposite but moderate biases in expectation disagree as much as half of

follows that, almost surely,

$$\Pr\left(s_{it} = 1 \mid \mathbf{s}_{\sim it} = \mathbf{s}\right) = \begin{cases} \Pr\left(s_{it} = 1 \mid \omega_t = \omega\right) & \text{if } \mathscr{I}_i \text{ is fully informative about } \omega_t \text{ but uninformative about } r_t; \\ \Pr\left(s_{it} = 1 \mid r_t = r\right) & \text{if } \mathscr{I}_i \text{ is fully informative about } r_t \text{ but uninformative about } \omega_t; \\ \Pr\left(s_{it} = 1 \mid \omega_t = \omega, r_t = r\right) & \text{if } \mathscr{I}_i \text{ is fully informative about both } \omega_t \text{ and } r_t. \end{cases} \tag{3}$$

Combining equations (1), (2), and (3) then yields simple expressions for μ_i.

time. The extent of disagreement is significantly larger than when biased agents did not observe any biased public information.

Consider, for example, the case where each agent's private signal is very imprecise. That is, when $a \to \frac{1}{2}$ and $b \to 0$. By proposition 2, there is almost no disagreement between agents of opposite bias when the neighbors are informative about ω_t but uninformative about r_t. However, when neighbors are informative about both ω_t and r_t, agents may disagree approximately half the time.

The intuition for proposition 3 is as follows. First note the neighbors being fully informative about both r_t and ω_t implies that the space of neighbors' signals $\{0,1\}^{|\mathscr{J}_i|}$ can be split into two sets: one with positive probability only when $\omega_t = l_t$, and another with positive probability only when $\omega_t = r_t$. Now consider an R-biased agent with nearly extreme bias. When $\omega_t = l_t$, her private signal is nearly uninformative in the sense that it fluctuates randomly between ω_t and r_t with probability $\frac{1}{2}$. Thus one set of neighbors' signals are judged to be useless. On the other hand, when $\omega_t = r_t$, she finds herself to be more correlated with the set of neighbors' signal realizations than her prior on precision should allow. In order to rationalize the data, she sets $\mu_i = r_t$ and becomes overconfident. Therefore, an R-biased agent's posterior depends entirely on her (useless) private signal when $\omega_t = l_t$, but converges to r_t when $\omega_t = r_t$. The opposite is true for an L-biased agent, resulting in disagreement approximately half the time in expectation.

Together, propositions 2 and 3 explain how the proliferation of partisan media facilitated by the Internet could have intensified partisan disagreements throughout the population, even when individuals are not ideologically segregated in their social networks or media consumption. Due to endogenous trust, beliefs diverge simply when the partisan content of public information sources increases.

6 Disagreement under Endogenous Networks

We now examine asymptotic disagreement when individuals endogenously select their information sources and strategic media outlets choose potentially biased reporting strategies.

6.1 Asymptotic Trust

We begin by characterizing how an agent's trust in her neighbors depends on the precisions and biases of their signals. Recall that asymptotic trust by i in j is given by $T_{ij} = B_i\left(s_{jt} = \omega_t\right)$. In the limit where an agent i samples a large number of signals from neighbor j, the law of large numbers implies that

$$\Pr\left(s_{it} = s_{jt}\right) = \Pr_i\left(s_{it} = s_{jt}\right) = T_{ii}T_{ij} + \left(1 - T_{ii}\right)\left(1 - T_{ij}\right)$$

almost surely. The limiting value of i's trust in j is therefore given by

$$T_{ij} = \frac{1}{2} + \frac{\Pr\left(s_{it} = s_{jt}\right) - \frac{1}{2}}{2T_{ii} - 1}. \tag{4}$$

Equation (4) shows that i's asymptotic trust in a neighbor increases with the frequency that they agree, and decreases with i's trust in her own private signal. Furthermore, the sign of the derivative of T_{ij} with respect to $\Pr\left(s_{it} = s_{jt}\right)$ is invariant to T_{ii}, since by assumption $T_{ii} > \frac{1}{2}$.

As proved in the appendix, we can write out $\Pr\left(s_{it} = s_{jt}\right)$ as a function of $\alpha_i, \alpha_j, \beta_i$, and β_j, and we have the following lemma.

Lemma 4. *Suppose $\beta_j \neq 0$. Agent i's asymptotic trust in j (T_{ij}), holding T_{ii} fixed, is increasing in $\left|\beta_j\right|$ if and only if $sgn\left(\beta_i\right) = sgn\left(\beta_j\right)$ and*

$$\frac{|\beta_i|}{1 - |\beta_i|} \geq 4\left(\alpha_i - \frac{1}{2}\right)\left(\alpha_j - \frac{1}{2}\right),$$

and strictly so if the inequality holds strictly.

Lemma 4 establishes the conditions under which i will come to trust biased neighbors more than unbiased neighbors. Trust is always decreasing in the magnitude of j's bias when it is in the opposite direction to i's. When i and j are biased in the same direction, trust will be increasing in j's bias whenever the precisions of i and j's signals are low relative to the magnitude of i's bias.

It follows immediately from lemma 4 that agents of a certain precision and bias will eventually place their highest trust in the most biased information source of the same ideological leaning. We formalize this result in the following proposition.

Proposition 4. *For any $\beta_i \in \left(0, \frac{1}{2}\right)$ and $\alpha_j \in \left(\frac{1}{2}, 1\right)$ there exists $\alpha_i \in \left(\frac{1}{2}, 1\right]$ such that agent i's bias is moderate and i's asymptotic trust in j is strictly increasing in β_j.*

Proposition 4 demonstrates that endogenous trust creates ideological homophily among individuals. Agents come to trust information sources with similar bias more than unbiased sources, even when their own biases are small.

6.2 Ideological Segregation

We now turn to study the effect of ideological homophily on network structure and belief formation. We employ the following simple setup. As in section 5, we consider a set of moderately biased agents who observe their neighbors' signals directly. We suppose that the population is made up of the three representative types of agents, namely U, R, and L, in equal proportions. Similarly, let $b = \beta_R = -\beta_L$ and $a = \alpha_i$ denote the agents' biases and precisions, respectively. But now suppose that, in each period t, each agent i may only choose a single neighbor $j_{it} \in \{L, R, U\}$ to observe. This setup approximates a world in which the cost of accessing or sharing information is high.

Under what conditions do agents listen to the most informative agents they have access to, and under what conditions do agents with similar ideological biases associate? Suppose that each agent maximizes an undiscounted sum of their payoffs. Since we assume that agents desire accurate information and also believe their own signals are unbiased, each agent i's maximum payoffs are achieved by observing the neighbor j_i^* such that $T_{ij} \leq T_{ij_i^*}$ for all j. However, she does not initially know T_{ij} and must learn T_{ij} through observation over time. Consequently, she must trade off exploration and exploitation in her choice of neighbor each period. As we show in the appendix, individuals sample a large number of signals from every other agent in the network, but eventually converge to getting information exclusively from the agent they trust the most.[5]

As shown in lemma 4, unbiased agents will always come to trust other unbiased agents the most. For biased agents, however, trust depends on the values of a and b. Holding b constant, decreasing a always increases the agent's trust in biased neighbors. Thus, as formalized in propo-

[5] Our result is a slight twist on a classic result for multi-arm bandits (see, e.g., Lai and Robbins 1985). Typically, multi-arm bandit payoffs are assumed to be i.i.d. over time; we derive similar properties of the optimal allocation rule when the payoffs are approximately i.i.d. as $t \to \infty$.

sition 4, there exists $\overline{a} > \psi^{-1}(b)$ such that for all $a < \overline{a}$, biased agents will place the highest trust in like-minded neighbors in the limit.

We can therefore characterize the resulting network as follows:

Proposition 5. *Suppose the population consists of large numbers of representative agents U, R, and L, and each agent chooses one neighbor to observe each period. Further suppose that each agent maximizes the sum of payoffs from all remaining periods, and the payoff strictly increases in the accuracy of the neighbor her observes. Then for any $b \in \left(0, \frac{1}{2}\right)$ there exists $\overline{a} > \psi^{-1}(b)$ such that in the limit as $t \to \infty$, the fraction of periods that agents listen to a like-minded neighbor converges in probability to one if $a < \overline{a}$ and the fraction of periods that agents listen to an unbiased neighbor converges in probability to one otherwise.*

This proposition shows that even small, non-extreme biases can result in ideological segregation: when private information is sufficiently noisy, agents eventually only get information from like-minded sources.

6.3 Strategic Bias by Media Outlets

We now turn to examine the interaction between ideological homophily and strategic reporting. Thus far we have assumed that agents are able to directly observe their neighbors' signals. We now consider the possibility that some agents misreport their signals strategically in hopes of acquiring connections or reach a larger audience. We interpret such agents as profit-maximizing media outlets.

To incorporate strategic reporting behavior into our model, we introduce into the population a small number of media outlets m who choose β_m directly and potentially exhibit extreme bias. Media outlets have access to accurate information ($\alpha_m = 1$), know the distribution of biases in the population, and can be sampled by agents just like other neighbors from the social network. Before the first period of observation (i.e., at $t = 0$), the media choose their own biases $\beta_m \in [-1, 1]$ to maximize the number of agents who eventually choose them as the asymptotic information source.

Since trust increases with frequency of agreement, an unbiased agent's trust increases in the neighbor's precision whenever $a > \frac{1}{2}$. Therefore, an unbiased agent trusts a media outlet with $\beta_m = 0$ more than she trusts all other agents. By contrast, it follows from proposition 5 that a biased agent will instead trust a media outlet with $\beta_m = \pm 1$ the most if $a < \overline{a}$ (with the sign of β_m dependent on the direction of the agent's bias). Therefore, some media outlets strategically choose extreme biases in a world where agents have small biases and noisy private information.

Even moderately biased agents enter ideological "echo chambers" where their own biases are reinforced by completely biased information. Ideological segregation in the presence of strategically extreme media and ideologues induces extreme disagreement. In fact, agents of opposite biases may disagree nearly all the time, even when agents have arbitrarily small biases.

Formally, we have the following proposition:

Proposition 6. *Suppose the population consists of large numbers of representative agents U, R, and L, and each agent chooses one neighbor to observe each period. Further suppose there exist accurate media outlets ($\alpha_m = 1$) who know the distribution of biases in the population and strategically choose their biases β_m to attract listeners. Then:*

1. *All media choose $\beta_m \in \{-1, 0, 1\}$;*

2. *For any $b \in \left(0, \frac{1}{2}\right)$ there exists $\overline{a} > \psi^{-1}(b)$ such that in the limit as $t \to \infty$, the fraction of periods that agents listen to a like-minded media outlet converges in probability to one if $a < \overline{a}$;*

3. *For any $\varepsilon > 0$ and $\overline{b} > 0$, there exists $b < \overline{b}$ and $a \in \left(\frac{1}{2}, 1\right)$ such that each agent's bias is moderate and $\mathrm{E}\left[|\mu_R - \mu_L|\right] > 1 - \varepsilon$. By contrast, $\mu_U = \omega_t$ for any $a \in \left(\frac{1}{2}, 1\right)$.*

Proposition 6 establishes that in a world with strategic media outlets, the unbiased agent is able to learn the underlying state. However, the presence of strategically extreme media outlets also engenders extreme disagreements between even moderately biased agents.

7 Conclusion

We study trust in information sources and ideological disagreements in a world where individuals fail to account for small biases in their private signals. Central to our analysis is the idea that small biases in private signals are amplified because individuals must infer the precisions of external information source using their private signals. As a result, small deviations from Bayesian reasoning can drive large and persistent disagreements even when all individuals have access to the same set of arbitrarily informative signals.

We apply our model to understand media markets and social networks. We show that partisan content in the media could reinforce the magnitude of disagreements, yielding a rationale for why the proliferation of partisan media facilitated by the popularization of cable television and the Internet could have intensified partisan disagreements throughout the population. We also show that it is possible for biased agents to find the most biased like-minded agents most trustworthy. Such ideological homophily generates ideological segregation, and also incentivizes media outlets to adopt more extreme positions in hopes of attracting a larger audience.

References

Acemoglu, Daron, Victor Chernozhukov, and Muhamet Yildiz. Forthcoming. "Fragility of Asymptotic Agreement under Bayesian Learning." *Theoretical Economics*.

Blackwell, David. 1951. "The Comparison of Experiments." *Proceedings, Second Berkeley Symposium on Mathematical Statistics and Probability*: 93-102. University of California Press.

Benoit, Jean-Pierre and Juan Dubra. 2014. "A Theory of Rational Attitude Polarization." Working paper.

Burke, Jeremy. 2008. "Primetime Spin: Media Bias and Belief Confirming Information." *Journal of Economics and Management Strategy* 17(3): 633–665.

Campante, Filipe and Daniel Hojman. 2013. "Media and polarization. Evidence from the introduction of broadcast TV in the United States." *Journal of Public Economics* 100: 79–92.

Chan, Jimmy and Wing Suen. 2008. "A Spatial Theory of News Consumption and Electoral Competition." *The Review of Economic Studies* 75(3): 699–728.

Chiang, Chun-Fang and Brian Knight. 2011. "Media bias and influence: Evidence from newspaper endorsements." *The Review of Economic Studies* 78(3): 795.

Della Vigna, Stefano and Ethan Kaplan. 2007. "The Fox News effect: Media bias and voting." *Quarterly Journal of Economics* 122(3): 1187–1234.

Enikolopov, Ruben, Maria Petrova, and Ekaterina Zhuravskaya. 2011. "Media and political persuasion: Evidence from Russia." *American Economic Review* 101(7): 3253–3285.

Flaxman, Seth R., Sharad Goel, and Justin M. Rao. 2013. "Ideological Segregation and the Effects of Social Media on News Consumption." Working paper.

Falck, Oliver, Robert Gold, and Stephan Heblich. 2014. "E-lections: Voting Behavior and the Internet." *American Economic Review* 104(7): 2238–65.

Festinger, L. (1957). *A Theory of Cognitive Dissonance.* California: Stanford University Press.

Gentzkow, Matthew and Jesse M. Shapiro. 2006. "Media Bias and Reputation." *Journal of Political Economy* 115(2): 280–316.

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

Gerber, Alan S., Dean Karlan, and Daniel Bergan. 2009. "Does the Media Matter? A Field Experiment Measuring the Effect of Newspapers on Voting Behavior and Political Opinions." *American Economic Journal: Applied Economics* 1(2): 35–52.

Glaeser, Edward L. and Bryce A. Ward. 2006. "Myths and Realities of American Political Geography." *Journal of Economic Perspectives* 20: 119–144.

Halberstam, Yosh and Brian Knight. 2014. "Homophily, group size, and the diffusion of political information in social networks: Evidence from Twitter." NBER Working Paper.

Lai, T. L. and Herbert Robbins. 1985. "Asymptotically Efficient Adaptive Allocation Rules." *Advances in Applied Mathematics* 6, 4-22.

Lord, Charles G., Lee Ross, and Mark R. Lepper. 1979. "Biased Assimilation and Attitude Polarization: The Effects of Prior Theories on Subsequently Considered Evidence." *Journal of Personality and Social Psychology* 37: 2098–2109.

Kunda, Ziva. 1987. "Motivation and inference: Self-serving generation and evaluation of evidence." *Journal of Personality and Social Psychology*, 53, 636-647.

Kunda, Ziva. 1990. "The Case for Motivated Reasoning." *Psychological Bulletin* 108(3): 480-498.

Martin, Gregory J. and Ali Yurukoglu. 2014. "Bias in Cable News: Real Effects and Polarization." NBER working paper.

McCarty, Nolan, Keith T. Poole, and Howard Rosenthal. 2006. *Polarized America: The Dance of Ideology and Unequal Riches.* Cambridge, MA: MIT Press.

Mullainathan, Sendhil and Andrei Shleifer. "The Market for News." *American Economics Review* 95(2005): 1031–1053.

Ortoleva, Pietro and Erik Snowberg. 2015. "Overconfidence in Political Behavior." *American Economic Review* 105(2): 504–35.

Prior, Markus. 2013. "Media and political polarization." *Annual Review of Political Science* 16: 101–27.

Savage, Leonard J. 1954. *The Foundations of Statistics*, Dover reprint, New York, 1972.

Shen, Xiaotong and Larry Wasserman. 2001. "Rates of Convergence of Posterior Distributions." *The Annals of Statistics* 29(3): 687-714.

Stone, Daniel F. 2011. "Ideological Media Bias." *Journal of Economic Behavior and Organization* 78(3): 256–271.

Sunstein, Cass R. 2001. *Republic.com.* Princeton, N.J.: Princeton University Press.

Tversky, Amos and Daniel Kahneman. "Availability: A Heuristic for Judging Frequency and Probability." *Cognitive Psychology* 5, 207-232.

Weizsäcker, Georg. 2010. Do we follow others when we should? A simple test of rational expectations." *American Economic Review.*

Appendices

A Proof of lemma 1

Proof. Let $Y_{ijt} \equiv \mathbf{1}\left\{s_{it} = s_{jt}\right\}$. For notational convenience, we write $\mathbf{Y}_{it} \equiv \left\{Y_{ijt}\right\}_{j\in\mathscr{I}_i}$. In the limit as $t \to \infty$, the law of large number implies that $B_i\left(\mathbf{Y}_{it} = \mathbf{s}\right) = \Pr\left(\mathbf{Y}_{it} = \mathbf{s}\right)$ almost surely for any $\mathbf{s} \in \{0,1\}^{|\mathscr{I}_i|}$. Since the agent believes that her private signal is conditionally independent, we can rewrite $B_i\left(\mathbf{Y}_{it} = \mathbf{s}\right)$ in terms of $\hat{\alpha}_i$ and $B_i\left(\mathbf{s}_{\sim it} = \mathbf{s} \mid \omega_t = 1\right)$. A few lines of manipulation using the laws of probability then show that, almost surely,

$$\begin{bmatrix} B_i\left(\mathbf{s}_{\sim it} = \mathbf{s} \mid \omega_t = 1\right) \\ B_i\left(\mathbf{s}_{\sim it} = \neg\mathbf{s} \mid \omega_t = 1\right) \end{bmatrix} = \frac{2\Pr\left(\mathbf{s}_{\sim it} = \mathbf{s}\right)}{2\hat{\alpha}_i - 1} \begin{bmatrix} \hat{\alpha}_i - 1 + \Pr\left(s_{it} = 1 \mid \mathbf{s}_{\sim it} = \mathbf{s}\right) \\ \hat{\alpha}_i - \Pr\left(s_{it} = 1 \mid \mathbf{s}_{\sim it} = \mathbf{s}\right) \end{bmatrix}. \tag{5}$$

Note that the likelihood that the agent assigns to the observed data does not vanish if and only if (i) $B_i\left(\mathbf{s}_{\sim it} = \mathbf{s} \mid \omega_t = 1\right) \in [0,1]$ for any $\mathbf{s} \in \{0,1\}^{|\mathscr{I}_i|}$, and (ii) $\sum_{\mathbf{s}\in\{0,1\}^{|\mathscr{I}_i|}} B_i\left(\mathbf{s}_{\sim it} = \mathbf{s} \mid \omega\right) = 1$. From equation (5), (i) is satisfied if and only if $\Pr\left(s_{it} = 1 \mid \mathbf{s}_{\sim it} = \mathbf{s}\right) \in [1 - \hat{\alpha}_i, \hat{\alpha}_i]$ $\forall\mathbf{s}$; and (ii) is trivially satisfied. □

B Proof of lemma 2

Throughout this proof, we will let $\Pr(X) \equiv \Pr(X = x)$ for notational brevity. Note also that we depart from the $B_{it}(\cdot)$ notation in this section for clarity, and use α_{i0} to denote i's true precision.

Let $p_{\mathbf{s}_{\sim i}}$ denote a parametrization of the joint distribution of $\mathbf{s}_{\sim i}$ for all possible ω_t and r_t, $\theta \equiv \{p_{\mathbf{s}_{\sim i}}, \alpha_i, \beta_i\} \in \Theta$ be a parametrization of the joint distribution of S_i, and
$A_{O.E.} \equiv \{\theta \in \Theta \mid \Pr\left(\mathbf{Y}_{i\tau} \mid \theta\right) = \Pr\left(\mathbf{Y}_{i\tau} \mid \theta_0\right) \forall\tau\}$ denote all θ that are observationally equivalent to θ_0, the true parameter. We will use $\alpha(\theta)$ to denote the α_i associated with θ.

Finally, let $f_\theta(\theta)$ denote the agent's prior on θ and $f_\theta\left(\cdot \mid \{\mathbf{Y}_{i1}, \ldots, \mathbf{Y}_{it}\}\right)$ denote the agent's posterior at time t.

Proof. We show that f_θ converges to a degenerate point mass at

$$\theta^* \equiv \underset{\left\{\theta \in A_{O.E.} | \beta_i = 0, \alpha_i > \frac{1}{2}\right\}}{\arg\min} |\alpha_i - \alpha_{i0}|.$$

By lemma 1 and equation 5, θ^* is unique.) Let $B_{\theta^*} \equiv B(\theta^*, \varepsilon)$ be an open neighborhood around θ^*. Formally, we show that for any B_{θ^*},

$$\lim_{n\to\infty}\lim_{t\to\infty}\Pr(\theta \in B_{\theta^*} \mid \{\mathbf{Y}_{i1}, \ldots, \mathbf{Y}_{it}\}) = 1. \tag{6}$$

Since Θ is compact for any finite number of neighbors N, we can construct a finite open cover C as follows. Start with B_{θ^*} covering θ^* and another set

$$B_{O.E.} \equiv \cup_{\theta \in A_{O.E.} | \theta \notin B_{\theta^*}} B(\theta, \xi)$$

with $\xi < \varepsilon$, which covers the remaining portion of $A_{O.E.}$. Then we can complete the cover by placing small neighborhoods about all remaining points in Θ and taking the finite subcover.

Let Q denote an open set in C. Now we show that equation 6 holds by showing that for any $Q \neq B_{\theta^*}$, $\lim_{n\to\infty}\lim_{t\to\infty}\Pr(\theta \in Q \mid \{\mathbf{Y}_{i1}, \ldots, \mathbf{Y}_{it}\}) = 0$. For any $Q \in C$, we can rewrite this expression as

$$\begin{aligned}
\Pr(\theta \in Q \mid \{\mathbf{Y}_{i1}, \ldots, \mathbf{Y}_{it}\}) &= \frac{\int_Q f_\theta(\theta \mid \{\mathbf{Y}_{i1}, \ldots, \mathbf{Y}_{it}\})\, d\theta}{\int_\Theta f_\theta(\theta \mid \{\mathbf{Y}_{i1}, \ldots, \mathbf{Y}_{it}\})\, d\theta} \\
&= \frac{\int_Q f_\theta(\theta) \prod_{\tau=1}^t \Pr(\mathbf{Y}_{i\tau} \mid \theta)\, d\theta}{\int_\Theta f_\theta(\theta) \prod_{\tau=1}^t \Pr(\mathbf{Y}_{i\tau} \mid \theta)\, d\theta} \\
&= \frac{\int_Q f_\theta(\theta) \exp\left(\sum_{\tau=1}^t \log\left(\frac{\Pr(\mathbf{Y}_{i\tau}|\theta)}{\Pr(\mathbf{Y}_{i\tau}|\theta_0)}\right)\right) d\theta}{\int_\Theta f_\theta(\theta) \exp\left(\sum_{\tau=1}^t \log\left(\frac{\Pr(\mathbf{Y}_{i\tau}|\theta)}{\Pr(\mathbf{Y}_{i\tau}|\theta_0)}\right)\right) d\theta}.
\end{aligned} \tag{7}$$

First we consider $B_{O.E.}$. We bound $\Pr(\theta \in B_{O.E.} \mid \{\mathbf{Y}_{i1}, \ldots, \mathbf{Y}_{it}\})$ from above by showing that the likelihood ratio of $\theta \in B_{O.E.}$ versus $\theta \in B_{\theta^*}$ becomes small as the sequence of prior distributions

$\{f_n^i\}_{n=1}^{\infty}$ from the lemma supposition approaches its limit. The likelihood ratio is

$$\frac{\Pr(\theta \in B_{O.E.} \mid \{\mathbf{Y}_{i1},\ldots,\mathbf{Y}_{it}\})}{\Pr(\theta \in B_{\theta^*} \mid \{\mathbf{Y}_{i1},\ldots,\mathbf{Y}_{it}\})} = \frac{\int_{B_{O.E.}} f_\theta(\theta) \prod_{\tau=1}^{t} \Pr(\mathbf{Y}_{i\tau} \mid \theta)\, d\theta}{\int_{B_{\theta^*}} f_\theta(\theta) \prod_{\tau=1}^{t} \Pr(\mathbf{Y}_{i\tau} \mid \theta)\, d\theta}.$$

We can rewrite this ratio by splitting the domain of each integral into sets of θ that are observationally equivalent

$$\frac{\Pr(\theta \in B_{O.E.} \mid \{\mathbf{Y}_{i1},\ldots,\mathbf{Y}_{it}\})}{\Pr(\theta \in B_{\theta^*} \mid \{\mathbf{Y}_{i1},\ldots,\mathbf{Y}_{it}\})} = \frac{\int_0^1 p \int_{\theta \in B_{O.E.} \mid \left(\prod_{\tau=1}^{t} \Pr(\mathbf{Y}_{i\tau} \mid \theta) = p\right)} f_\theta(\theta)\, d\theta dp}{\int_0^1 p \int_{\theta \in B_{\theta^*} \mid \left(\prod_{\tau=1}^{t} \Pr(\mathbf{Y}_{i\tau} \mid \theta) = p\right)} f_\theta(\theta)\, d\theta dp}.$$

By assumption 1 and the supposition in lemma 2, in the limit as $n \to \infty$, agent's prior is degenerate at $\beta_i = 0$ and her prior density is decreasing in $|\alpha_i - \alpha_{i0}|$ (such that for any α_i and α_i' with $|\alpha_i - \alpha_{i0}| < |\alpha_i' - \alpha_{i0}|$, the likelihood ratio $f_n^i(\alpha_i') / f_n^i(\alpha_i) \to 0$ as $n \to \infty$) with it equal to zero for any θ such that $\alpha_i < \frac{1}{2}$.[6] Recall that by construction, $B_{O.E.}$ does not contain any points near θ^*, which minimizes both $|\beta_i|$ and $|\alpha_i - \alpha_{i0}|$. Thus any ratio of priors from $B_{O.E.}$ to priors from B_{θ^*} converges to 0. Formally, for all p such that $\{\theta \in B_{\theta^*} \mid (\prod_{\tau=1}^{t} \Pr(\mathbf{Y}_{i\tau} \mid \theta) = p)\} \neq \emptyset$, we have[7]

$$\lim_{n\to\infty} \frac{\int_{\theta \in B_{O.E.} \mid \left(\prod_{\tau=1}^{t} \Pr(\mathbf{Y}_{i\tau} \mid \theta) = p\right)} f_\theta(\theta)\, d\theta}{\int_{\theta \in B_{\theta^*} \mid \left(\prod_{\tau=1}^{t} \Pr(\mathbf{Y}_{i\tau} \mid \theta) = p\right)} f_\theta(\theta)\, d\theta} = 0. \tag{8}$$

Thus we find that

$$\lim_{n\to\infty} \frac{\Pr(\theta \in B_{O.E.} \mid \{\mathbf{Y}_{i1},\ldots,\mathbf{Y}_{it}\})}{\Pr(\theta \in B_{\theta^*} \mid \{\mathbf{Y}_{i1},\ldots,\mathbf{Y}_{it}\})} = 0.$$

Now we bound $\Pr(\theta \in Q \mid \{\mathbf{Y}_{i1},\ldots,\mathbf{Y}_{it}\})$ from above for all $Q \in C \setminus \{B_{\theta^*}, B_{O.E.}\}$. Let $\theta \in Q$. We start with the numerator of equation 7. The summation in the numerator consists of t i.i.d. random variables with fixed θ and θ_0. By construction of B_{θ^*} and $B_{O.E.}$, θ is not in the set of parameters observationally equivalent to θ_0, and so $\mathrm{E}\left[\log\left(\frac{\Pr(\mathbf{Y}_{i\tau}|\theta)}{\Pr(\mathbf{Y}_{i\tau}|\theta_0)}\right)\right] < -c_\varepsilon$ for some $c_\varepsilon > 0$. Thus by WLLN $\exp\left(\sum_{\tau=1}^{t} \log\left(\frac{\Pr(\mathbf{Y}_{i\tau}|\theta)}{\Pr(\mathbf{Y}_{i\tau}|\theta_0)}\right)\right)$ converges to 0 at some rate faster than $\exp(-tc_\varepsilon)$.

Now we turn to the denominator of equation 7. Shen and Wasserman (2001) lemma 1 shows that this denominator is bounded below by $\exp(-\sqrt{t}) \int_{B_t} \frac{1}{2} f_\theta(\theta)\, d\theta$, where B_t is a shrinking

[6] The construction of this sequence means that the full support prior on $p_{\mathbf{s}_{\sim i}}$ becomes irrelevant.

[7] For some p and ξ (the "radius" of $B_{O.E.}$), it may be that the numerator of equation 8 is positive while the denominator is 0. In such cases, we can always pick smaller ξ such that the numerator is also 0 (and hence such p becomes irrelevant for our purposes).

neighborhood (in t) of $A_{O.E.}$ with all $\theta \in B_t$ such that $|\alpha(\theta) - \alpha_{i0}| < |\alpha(\theta^*) - \alpha_{i0}|$ removed.[8] Note that due to the increasingly degenerate prior on α_i as $n \to \infty$ and B_t's restriction on values of α_i, the rate at which $\int_{B_t} \frac{1}{2} f_\theta(\theta)\, d\theta$ converges to 0 is decreasing in n and slower than $\exp\left(-\sqrt{t}\right)$.[9] Since the numerator converges to 0 at rate faster than $\exp(-tc_\varepsilon)$ and the denominator converges to 0 at rate slower than $\exp\left(-2\sqrt{t}\right)$, we must have

$$\lim_{n\to\infty} \lim_{t\to\infty} \Pr(\theta \in Q \mid \{\mathbf{Y}_{i1}, \ldots, \mathbf{Y}_{it}\}) = 0.$$

Note that the prior $f_\theta(\theta)$ implicitly depends on our choice of n, therefore our choice of t is dependent on n.

The above result, combined with our result for $B_{O.E.}$, implies that

$$\lim_{n\to\infty} \lim_{t\to\infty} \Pr(\theta \in B_{\theta^*} \mid \{\mathbf{Y}_{i1}, \ldots, \mathbf{Y}_{it}\}) = 1.$$

The lemma then follows immediately by noting that $\left\{T_{ijt}\right\}_{j\in\mathscr{J}_i}$ and μ_{it} can be expressed as continuous functions of θ_{it}, agent i's posterior of θ at time t. □

C Proof of proposition 1

Proof. The proof of this proposition relies heavily on results from later sections.

First, the result of exogenous trust leading to asymptotic learning follows immediately by applying equation 3 (from section 5.1) and the agent's degenerate prior on the joint distribution of neighbors' signals.

For (1), $T_{ij} < \alpha_j$ for all j follows immediately after rewriting T_{ij} in terms of $\alpha_i, \alpha_j, \beta_i$, and β_j (following equation 4).[10] Now consider two distinct neighbors j and k, and define $J_{ijk} =$

[8] To obtain the lower bound, set $t_n = 1/\left(2\sqrt{t}\right)$, where t_n denotes the sequence mentioned in Shen and Wasserman (2001) lemma 1. Then obtain B_t by removing from their shrinking neighborhood ("S_t") all θ such that $|\alpha(\theta) - \alpha_{i0}| < |\alpha(\theta^*) - \alpha_{i0}|$. Since this refinement results in a B_t being a smaller set, the expression remains a lower bound.

[9] This can be achieved by choosing large enough n^* from earlier (note that this choice of n^* does not depend on t).

[10] In this case, we get $T_{ij} = \alpha_j - \left(\alpha_j - \frac{1}{2}\right)\beta_i$.

$B_i\left(s_{jt}=s_{kt}=\omega_t\right)$. By the inclusion-exclusion principle, we find

$$\begin{aligned} J_{ijk} &= \frac{1}{2}\left(B_i\left(s_{jt}=s_{kt}\right)-1+T_{01}+T_{02}\right) \\ &= \alpha_j^2-\left(\alpha_j-\frac{1}{2}\right)\beta_i. \end{aligned}$$

It then follows that $J_{ijk}-T_{ij}T_{ik}=\frac{\beta_i}{2}\left(1-\frac{\beta_i}{2}\right)\left(1-2\alpha_j\right)^2>0$, implying that the agent believes the covariance between s_{jt} and s_{kt} to be positive, proving (2). □

D Proof of lemma 3

Proof. Consider the case with N neighbors with $\alpha_j>\frac{1}{2}$ and $\beta_j=0$. We can rewrite the expression of interest as

$$\Pr\left(\omega_t=\omega \mid \mathbf{s}_{\sim it}=\mathbf{s}\right) = \frac{\Pr\left(\mathbf{s}_{\sim it}=\mathbf{s} \mid \omega_t=\omega\right)}{\Pr\left(\mathbf{s}_{\sim it}=\mathbf{s} \mid \omega_t=\omega\right)+\Pr\left(\mathbf{s}_{\sim it}=\mathbf{s} \mid \omega_t=\neg\omega\right)}.$$

Thus it suffices to show that for any $\mathbf{s}\in\{0,1\}^{|\mathscr{I}_i|}$, $\frac{\Pr(\mathbf{s}_{\sim it}=\mathbf{s}|\omega_t=\omega)}{\Pr(\mathbf{s}_{\sim it}=\mathbf{s}|\omega_t=\neg\omega)}=0$ or $\frac{\Pr(\mathbf{s}_{\sim it}=\mathbf{s}|\omega_t=\omega)}{\Pr(\mathbf{s}_{\sim it}=\mathbf{s}|\omega_t=\neg\omega)}\to\infty$. By model construction, we have that

$$\Pr\left(\mathbf{s}_{\sim it}=\mathbf{s} \mid \omega_t=\omega, r_t=r\right)=\begin{cases}\left(\alpha_j\right)^{\Sigma_j s_j}\left(1-\alpha_j\right)^{N-\Sigma_j s_j} & \text{if } \omega=1 \\ \left(1-\alpha_j\right)^{\Sigma_j s_j}\left(\alpha_j\right)^{N-\Sigma_j s_j} & \text{if } \omega=0\end{cases}.$$

Rearranging, we find that

$$\frac{\Pr\left(\mathbf{s}_{\sim it}=\mathbf{s} \mid \omega_t=0, r_t=r\right)}{\Pr\left(\mathbf{s}_{\sim it}=\mathbf{s} \mid \omega_t=1, r_t=r\right)}=\left(\frac{1-\alpha_j}{\alpha_j}\right)^{\left(\frac{2}{N}\Sigma_j s_j-1\right)N}.$$

Now note that $\operatorname{plim}_{N\to\infty}\left(\frac{1}{N}\sum_{j=1}^N s_{jt}\middle|\omega_t=\omega, r_t=r\right)\in\left\{1-\alpha_j,\alpha_j\right\}$ for any $r\in\{0,1\}$, with the value depending on ω. The CMT then implies that $\operatorname{plim}_{N\to\infty}\frac{\Pr(\mathbf{s}_{\sim it}=\mathbf{s}|\omega_t=0,r_t=r)}{\Pr(\mathbf{s}_{\sim it}=\mathbf{s}|\omega_t=1,r_t=r)}\in\{0,\infty\}$ for any $r\in\{0,1\}$. This proves full informativeness about ω_t. It is easy to show that

$$\operatorname{plim}_{N\to\infty}\frac{\Pr\left(\mathbf{s}_{\sim it}=\mathbf{s} \mid \omega_t=\omega, r_t=0\right)}{\Pr\left(\mathbf{s}_{\sim it}=\mathbf{s} \mid \omega_t=\neg\omega, r_t=0\right)}=\operatorname{plim}_{N\to\infty}\frac{\Pr\left(\mathbf{s}_{\sim it}=\mathbf{s} \mid \omega_t=\omega, r_t=1\right)}{\Pr\left(\mathbf{s}_{\sim it}=\mathbf{s} \mid \omega_t=\neg\omega, r_t=1\right)}$$

for any $\omega \in \{0,1\}$, which shows that the neighbors are uninformative about r_t.

The result that N identical neighbors with $\alpha_j > \frac{1}{2}$ and moderate bias $\beta_j \neq 0$ approaches fully informativeness about both ω_t and r_t in the limit as $N \to \infty$ can be shown using a similar method.

□

E Proof of proposition 2

Proof. WLOG assume $\omega_t = 1$. Suppose the set of neighbors $\mathscr{I}_i$ is fully informative about ω_t but uninformative about r_t, and recall from equation (3) that $\Pr\left(s_{it} = 1 \mid \mathbf{s}_{\sim it} = \mathbf{s}\right) = \Pr\left(s_{it} = 1 \mid \omega_t = 1\right)$. Lemmas 1 and 2 then imply that $T_{ii} = \alpha_i$. It follows from equation (2) that, with probability 1,

$$\frac{B_i\left(\mathbf{s}_{\sim it} = \mathbf{s} \mid \omega_t = 1\right)}{B_i\left(\mathbf{s}_{\sim it} = \mathbf{s} \mid \omega_t = 0\right)} = \frac{\alpha_i - 1 + \Pr\left(s_{it} = 1 \mid \mathbf{s}_{\sim it}\right)}{\alpha_i - \Pr\left(s_{it} = 1 \mid \mathbf{s}_{\sim it}\right)} = \frac{2 - |\beta_i|}{|\beta_i|}. \tag{9}$$

Substituting equation (9) into (1), we have

$$\mu_i = \begin{cases} \left(1 + \frac{\alpha_i}{1-\alpha_i} \frac{|\beta_i|}{2-|\beta_i|}\right)^{-1}, & s_{it} = 0 \\ \left(1 + \frac{1-\alpha_i}{\alpha_i} \frac{|\beta_i|}{2-|\beta_i|}\right)^{-1}, & s_{it} = 1 \end{cases}. \tag{10}$$

Setting $\beta_i = 0$ gives $\Pr\left(|\mu_U - \omega_t| < \varepsilon\right) = 1$.

Now we examine the degree of disagreement. Note that disagreement is 0 when R and L agree: $|\mu_R - \mu_L|_{s_R = s_L} = 0$. When R and L disagree:

$$|\mu_R - \mu_L|_{disagree} = \left(1 + \frac{a}{1-a}\frac{b}{2-b}\right)^{-1} - \left(1 + \frac{1-a}{a}\frac{b}{2-b}\right)^{-1}$$

which implies that $\mathrm{E}\left[|\mu_R - \mu_L|\right] \leq b$. The desired results then follow by taking sufficiently large N.

□

F Proof of proposition 3

Proof. WLOG assume $\beta_i > 0$, $\omega_t = 1$, and $r_t = r$. Suppose the set of neighbors $\mathscr{I}_i$ is fully informative about ω_t and r_t, and recall from equation (3) that $\Pr\left(s_{it} = 1 \mid \mathbf{s}_{\sim it} = \mathbf{s}\right) = \Pr\left(s_{it} = 1 \mid \omega_t = 1, r_t = r\right)$.

Lemmas 1 and 2 then imply that $T_{ii} = \alpha_i + (1-\alpha_i)|\beta_i|$.

Suppose $r = 1$. By equation (2), with probability 1,

$$\frac{B_i(\mathbf{s}_{\sim it} = \mathbf{s} \mid \omega_t = 1)}{B_i(\mathbf{s}_{\sim it} = \mathbf{s} \mid \omega_t = 0)} = \frac{T_{ii} - 1 + \Pr(s_{it} = 1 \mid \omega_t = 1, r_t = 1)}{T_{ii} - \Pr(s_{it} = 1 \mid \omega_t = 1, r_t = 1)}.$$

Since $2T_{ii} > 1$, it follows immediately that, for any $\varepsilon > 0$, $\Pr(|\mu_i - \omega_t| < \varepsilon \mid \omega_t = r_t) = 1$.

Now suppose $r = 0$. Note that $\Pr(s_{it} = 1 \mid \omega_t = 1, r_t = 0) = \alpha_i(1-\beta_i)$. Then with probability 1,

$$\frac{B_i(\mathbf{s}_{\sim it} = \mathbf{s} \mid \omega_t = 1)}{B_i(\mathbf{s}_{\sim it} = \mathbf{s} \mid \omega_t = 0)} = \frac{T_{ii} - 1 + \Pr(s_{it} = 1 \mid \omega_t = 1, r_t = 0)}{T_{ii} - \Pr(s_{it} = 1 \mid \omega_t = 1, r_t = 0)} = \frac{(2\alpha_i - 1)(1-\beta_i)}{\beta_i}.$$

Let $\alpha_i = \frac{1}{2(1-\beta_i)} + \delta$, where $\delta > 0$, so the agent's bias is not extreme. Then, with probability 1, $\lim_{\delta \to 0} \frac{B_i(\mathbf{s}_{\sim it} = \mathbf{s} \mid \omega_t = 1)}{B_i(\mathbf{s}_{\sim it} = \mathbf{s} \mid \omega_t = 0)} = 1$. It follows that, with probability 1,

$$\begin{aligned} \lim_{\delta \to 0} \mu_i \mid \omega_t = l_t &= \begin{cases} 1 - T_{ii} & \text{if } s_{it} = 0 \\ T_{ii} & \text{if } s_{it} = 1 \end{cases} \\ &= \begin{cases} \frac{1}{2} - \beta_i & \text{if } s_{it} = 0 \\ \frac{1}{2} + \beta_i & \text{if } s_{it} = 1 \end{cases} \end{aligned}$$

Note that in the extreme bias limit, $s_{it} = 1$ with probability $\frac{1}{2}$. In expectation, $\lim_{\delta \to 0} \mathrm{E}[\mu_i \mid \omega_t = l_t] = \frac{1}{2}\left(\frac{1}{2} - \beta_i\right) + \frac{1}{2}\left(\frac{1}{2} + \beta_i\right) = \frac{1}{2}$. The desired result follows immediately by taking sufficiently large N.

In the case of agent U, it follows that $\lim_{N \to \infty} \Pr(s_{it} = 1 \mid \omega_t = 1, r_t = 1) = \Pr(s_{it} = 1 \mid \omega_t = 1, r_t = 0) = \alpha_i = T_{ii}$. □

G Proof of lemma 4

Proof. From equation (4), we have that T_{ij} monotonically increases with $\Pr(s_{it} = s_{jt})$. It is thus sufficient to derive conditions under which $\Pr(s_{it} = s_{jt})$ increases in $|\beta_j|$. Writing out $\Pr(s_{it} = s_{jt})$ in terms of the underlying parameters, we have that

$$\Pr(s_{it} = s_{jt}) = \left(1 - |\beta_i| - |\beta_j| + |\beta_i||\beta_j|\right)\left[\alpha_i\alpha_j + (1-\alpha_i)(1-\alpha_j)\right] + \frac{|\beta_i|}{2} + \frac{|\beta_j|}{2} - |\beta_i||\beta_j|\mathbf{1}(\beta_i\beta_j < 0)$$

It follows from taking derivatives and algebraic manipulation that $\frac{d\Pr\left(s_{it}=s_{jt}\right)}{d\left|\beta_j\right|} \geq 0$ if and only if $\beta_i\beta_j \geq 0$ and

$$|\beta_i| \geq \frac{(2\alpha_i-1)\left(2\alpha_j-1\right)}{1+(2\alpha_i-1)\left(2\alpha_j-1\right)}.$$

The desired inequality follows via a monotonic transformation. Strict inequality can be shown in the same manner. □

H Proof of proposition 4

Proof. We wish to show that for any $\alpha_j \in \left(\frac{1}{2},1\right)$ and $\beta_i \in \left(\frac{1}{2},1\right)$, there exists $\alpha_i > \frac{1}{2(1-\beta_i)}$ such that $\frac{d\Pr\left(s_{it}=s_{jt}\right)}{d\left|\beta_j\right|} > 0$. Let $\alpha_i = \frac{1}{2(1-\beta_i)} + \delta$. By lemma 4, it is sufficient to show that there exists $\delta > 0$ such that

$$\frac{\beta_i}{1-\beta_i} > 4\left(\left(\frac{1}{2\left(1-\beta_i\right)}+\delta\right)-\frac{1}{2}\right)\left(\alpha_j-\frac{1}{2}\right).$$

This equation can be rearranged as

$$1 > \left(1+\frac{2\delta\left(1-\beta_i\right)}{\beta_i}\right)\left(2\alpha_j-1\right),$$

which holds for some $\delta > 0$ so long as $\beta_i > 0$ and $\alpha_j < 1$. □

I Proof of proposition 5

We begin by proving a corollary of SLLN used in our proof of proposition 5.

Lemma 5. *Suppose X_i is an i.i.d. sequence of random variables and Y_i is a sequence of random variables where $Y_i \to_p Y$ and Y is a constant. Let $f(\cdot,\cdot)$ be a function continuous in the second argument. If $\mathrm{E}\left|f(X_1,Y)\right| < \infty$, then*

$$\frac{\sum_{i=1}^{n} f\left(X_i,Y_i\right)}{n} \to_p \mathrm{E} f\left(X_1,Y\right).$$

Proof. By SLLN, we have that $\frac{\sum_{i=1}^{n} f(X_i,Y)}{n} \rightarrow_{a.s.} \mathrm{E} f(X_1,Y)$. So it suffices to show that

$$\frac{\sum_{i=1}^{n} f(X_i,Y_i)}{n} \rightarrow_p \frac{\sum_{i=1}^{n} f(X_i,Y)}{n}.$$

Since $Y_i \rightarrow_p Y$ and $f(\cdot,\cdot)$ is continuous in the second argument, by the continuous mapping theorem we have that for any $\varepsilon > 0$ and any $\delta > 0$, there exists n^* so that for all $i \geq n^*$,

$$\Pr(|f(x,Y_i) - f(x,Y)| \geq \varepsilon) < \delta$$

for all $x \in \mathbb{R}$. Then

$$\begin{aligned}
\delta &> \Pr(|f(X_i,Y_i) - f(X_i,Y)| \geq \varepsilon), \quad \forall i \geq n^* \\
&\geq \Pr\left(\textstyle\sum_{i=n^*}^{n} |f(X_i,Y_i) - f(X_i,Y)| \geq (n - n^* + 1)\varepsilon\right) \\
&\geq \Pr\left(\textstyle\sum_{i=n^*}^{n} |f(X_i,Y_i) - f(X_i,Y)| \geq n\varepsilon\right) \\
&\geq \Pr\left(\left|\textstyle\sum_{i=n^*}^{n} f(X_i,Y_i) - \sum_{i=n^*}^{n} f(X_i,Y)\right| \geq n\varepsilon\right) \\
&= \Pr\left(\left|\frac{\sum_{i=n^*}^{n} f(X_i,Y_i)}{n} - \frac{\sum_{i=n^*}^{n} f(X_i,Y)}{n}\right| \geq \varepsilon\right).
\end{aligned}$$

Thus $\frac{\sum_{i=n^*}^{n} f(X_i,Y_i)}{n} \rightarrow_p \frac{\sum_{i=n^*}^{n} f(X_i,Y)}{n}$. Since $\frac{\sum_{i=1}^{n^*} f(X_i,Y_i)}{n} \rightarrow_p 0$, we are done. □

We now turn to our proof of proposition 5.

Proof. In this proof we adopt the perspective of agent i and drop the subscript i except to be consistent with the rest of the paper. Let J be the set of all potential neighbors of i. Let $J^* \subset J$ be the set of neighbors that i trusts the most in the limit where i observes all $j \in J$ many times.

Let $x(T_{ijt}, T_{iit}, Y_{ijt})$ be the payoff that agent i believes she would receive if she chose to observe j in period t. The payoff $x(\cdot,\cdot,\cdot)$ depends on the agent's beliefs in period t (T_{ijt},T_{iit}) as well as on the observed signal agreement $Y_{ijt} \equiv \mathbf{1}\{s_{it} = s_{jt}\}$. To simplify notation, we write $X_t(j) \equiv x(T_{ijt}, T_{iit}, Y_{ijt})$ and $X(j) \equiv x(T_{ij}, T_{ii}, Y_{ijt})$. If T_{iit} and T_{ijt} converge to constants T_{ii} and T_{ij} in the limit where $t \rightarrow \infty$, then we have $\operatorname{plim}_{t\rightarrow\infty} X_t(j) = X(j)$ and we can write $\bar{X}(j) = \mathrm{E} X(j)$. We write $\bar{X}^* \equiv \max_{j \in J} \bar{X}(j)$.

We denote agent i's optimal adaptive decision rule as a random sequence j_t^o which depends on signal realizations observed by agent i. Let

$$N_t^o(j) \equiv \sum_{\tau=1}^{t} \mathbf{1}\{j_\tau^o = j\}$$

be the number of visits by i to j after period t. Since we assume that agents strictly value information, we have $\bar{X}(j) > \bar{X}(j')$ if and only if $T_{ij} > T_{ij'}$. Then by proposition 4, the desired result is equivalent to $\frac{N_t^o(j)}{t} \rightarrow_p 0$ for all $j \notin J^*$.

Suppose for contradiction that the optimal decision rule has the property that $\frac{N_t^o(j^o)}{t} \not\rightarrow_p 0$ for some $j^o \notin J^*$. Let J^o denote the set of all such j^o. Now let $K_t(j) \equiv \sum_{\{\tau \in \{1,\ldots,t\} | j_\tau^o = j\}} X_\tau(j)$. Note that

$$\frac{\sum_{\tau=1}^{t} X_\tau(j_\tau^o)}{t} = \sum_{j \in J} \frac{K_t(j)}{N_t^o(j)} \frac{N_t^o(j)}{t}. \tag{11}$$

We decompose J into four disjoint sets: J^o, $\{j \in J^* \mid N_t^o(j) \rightarrow \infty\}$, $\{j \in J^* \mid N_t^o(j) < \infty \forall t\}$, and $J \backslash (J^* \cup J^o)$. Since $N_t^o(j^o) \rightarrow_p \infty$, and so $T_{ij^ot} \rightarrow_p T_{ij^o}$. Note also that $T_{iit} \rightarrow_p T_{ii}$ and Y_{ijt} is i.i.d across periods. Hence by lemma 5, we have that

$$\frac{K_t(j^o)}{N_t^o(j^o)} \rightarrow_p \bar{X}(j^o) < \bar{X}^* \tag{12}$$

for any $j^o \in J^o$. By assumption there exists $\eta, \gamma > 0$ such that for any t^*, there exists $t > t^*$ such that

$$\Pr\left(\frac{N_t(j^o)}{t} \geq \eta\right) \geq \gamma \tag{13}$$

for any $j^o \in J^o$. By lemma 5, we have

$$\frac{K_t(j)}{N_t^o(j)} \rightarrow_p \bar{X}^* \tag{14}$$

for any $j \in J^*$ where $N_t^o(j) \rightarrow \infty$. Finally, we have

$$\frac{N_t(j)}{t} \rightarrow_p 0 \tag{15}$$

for any $j \notin J^* \cup J^o$ and for any $j \in J^*$ where $N_t^o(j) < \infty$ for all t. It follows from combining (11), (12), (13), (14), and (15) that there exists $\varepsilon, \delta > 0$ such that for any t^*, there exists $t > t^*$ such that

$$\Pr\left(\frac{\sum_{\tau=1}^{t} X_\tau\left(j_\tau^o\right)}{t} + \varepsilon < \bar{X}^*\right) \geq \delta,$$

and for any $\varepsilon', \delta' > 0$, there exists t^* such that for all $t > t^*$,

$$\Pr\left(\frac{\sum_{\tau=1}^{t} X_\tau\left(j_\tau^o\right)}{t} - \varepsilon' > \bar{X}^*\right) < \delta'.$$

We construct an alternate decision rule j_t^a using the following procedure. For each neighbor j, let $\tau_1^j < \tau_2^j < ...$ be disjoint, increasing sequences of positive integers such that $\tau_t^j/t \to \infty$. Agent i observes neighbor j in period t if $t \in \left\{\tau_1^t, \tau_2^t, ...\right\}$ for some j, and if $t \notin \left\{\tau_1^t, \tau_2^t, ...\right\}$ for any j, agent i observes neighbor j in period t with the maximum $\mathrm{E}\, T_{ijt}$. It follows easily that for any $\varepsilon', \delta' > 0$, there exists t^* such that for all $t > t^*$,

$$\Pr\left(\left|\frac{\sum_{\tau=1}^{t} X_\tau\left(j_\tau^a\right)}{t} - \bar{X}^*\right| > \varepsilon'\right) < \delta'.$$

It follows that there exists $\varepsilon, \delta > 0$ such that for any t^*, there exists $t > t^*$ such that

$$\Pr\left(\sum_{\tau=1}^{t} X_\tau\left(j_\tau^a\right) - \sum_{\tau=1}^{t} X_\tau\left(j_\tau^o\right) > t\varepsilon\right) \geq \delta,$$

while for any $\varepsilon', \delta' > 0$, there exists t^* such that for all $t > t^*$,

$$\Pr\left(\sum_{\tau=1}^{t} X_\tau\left(j_\tau^o\right) - \sum_{\tau=1}^{t} X_\tau\left(j_\tau^a\right) > t\varepsilon'\right) < \delta'.$$

We have therefore shown that the alternate rule outperforms the optimal rule, yielding the contradiction we sought. □

J Proof of proposition 6

We begin by deriving the agent's limiting posterior beliefs when her neighbors are fully informative about r_t but not informative about ω_t.

Lemma 6. *Suppose agent* 0*'s neighbors are fully informative about* r_t *but not informative about* ω_t*. Then for any* $\varepsilon > 0$ *and* $\overline{b} > 0$*, there exists* $b < \overline{b}$ *and* $a \in \left(\frac{1}{2}, 1\right)$ *such that agent 0's bias is not extreme and:*

1. *If* $\beta_0 = b$*,* $|\mathrm{E}[\mu_0 - r_t]| < \varepsilon$*.*
2. *If* $\beta_0 = -b$*,* $|\mathrm{E}[\mu_0 - l_t]| < \varepsilon$*.*

Proof. We now derive μ_i for any agent i observing a neighbor who is fully informative about r_t but uninformative about ω_t. WLOG suppose $\omega_t = 1$. We first consider the case where $r_t = 1$. Since agent i's neighbors are fully informative about r_t but uninformative about ω_t, $\Pr(s_{it} = 1 \mid \mathbf{s}_{\sim 0t} = \mathbf{s}) = \Pr(s_{it} = 1 \mid r_t = 1)$ almost surely. It then follows that, with probability 1,

$$\frac{B_i(\mathbf{s}_{\sim it} = \mathbf{s} \mid \omega_t = 1)}{B_i(\mathbf{s}_{\sim 0t} = \mathbf{s} \mid \omega_t = 0)} = \frac{\alpha_0 - 1 + \Pr(s_{0t} = 1 \mid \mathbf{s}_{\sim 0t} = \mathbf{s})}{\alpha_0 - \Pr(s_{0t} = 1 \mid \mathbf{s}_{\sim 0t} = \mathbf{s})} = \frac{\alpha_0 + \frac{1}{2}\beta_0 - \frac{1}{2}}{\alpha_0 - \frac{1}{2}\beta_0 - \frac{1}{2}}.$$

Let $\alpha_0 = \frac{1}{2(1-\beta_0)} + \delta$, where $\delta > 0$, so the agent's bias is not extreme. By continuity, in the limit where $\delta \to 0$, $\frac{\Pr_0(\mathbf{s}_{\sim it}=\mathbf{s}|\omega_t=1)}{\Pr_0(\mathbf{s}_{\sim it}=\mathbf{s}|\omega_t=0)} \to \frac{2-\beta_0}{\beta_0}$. Now recall that $\mu_i = \left(1 + \frac{B_i(s_{0t}=s|\omega_t=0)}{B_i(s_{0t}=s|\omega_t=1)} \frac{B_i(\mathbf{s}_{\sim 0t}=\mathbf{s}|\omega_t=0)}{B_i(\mathbf{s}_{\sim 0t}=\mathbf{s}|\omega_t=1)}\right)^{-1}$. Taking expectations over s_{it}, we find by continuity that

$$\lim_{\delta \to 0} \mathrm{E}[\mu_i \mid \omega_t = r_t] = \left(\frac{1}{2} + \beta_0\right)\left(1 + (1 - 2\beta_0)\frac{\beta_0}{2 - \beta_0}\right)^{-1} + \left(\frac{1}{2} - \beta_0\right)\left(1 + \frac{1}{1 - 2\beta_0}\frac{\beta_0}{2 - \beta_0}\right)^{-1}.$$

Note that $\lim_{\beta_0 \downarrow 0} \lim_{\delta \to 0} \mathrm{E}[\mu_0 \mid \omega_t = r_t] = 1$. Thus, for any $\varepsilon_1 > 0$, there exists $\bar{b}_1 \in \left(0, \frac{1}{2}\right)$ such that, for any $\beta_0 \in (0, \underline{b})$, $|\mathrm{E}[\mu_0 \mid \omega_t = r_t] - 1| < \varepsilon_1$.

Now consider the case where $r_t = 0$. By the same logic, we have that, with probability 1,

$$\frac{B_i(\mathbf{s}_{\sim 0t} = \mathbf{s} \mid \omega_t = 1)}{B_i(\mathbf{s}_{\sim 0t} = \mathbf{s} \mid \omega_t = 0)} = \frac{\alpha_0 - 1 + \Pr(s_{0t} = 1 \mid \mathbf{s}_{\sim 0t} = \mathbf{s})}{\alpha_0 - \Pr(s_{0t} = 1 \mid \mathbf{s}_{\sim 0t} = \mathbf{s})} = \frac{\alpha_0 - \frac{1}{2}\beta_0 - \frac{1}{2}}{\alpha_0 + \frac{1}{2}\beta_0 - \frac{1}{2}}.$$

It follows that

$$\lim_{\delta \to 0} \mathrm{E}[\mu_0 \mid \omega_t = l_t] = \frac{1}{2}\left(1 + (1 - 2\beta_0)\frac{2 - \beta_0}{\beta_0}\right)^{-1} + \frac{1}{2}\left(1 + \frac{1}{1 - 2\beta_0}\frac{2 - \beta_0}{\beta_0}\right)^{-1}.$$

Note that $\lim_{\beta_0 \downarrow 0} \lim_{\delta \to 0} \mathrm{E}[\mu_0 \mid \omega_t = l_t] = 0$. Thus, for any $\varepsilon_2 > 0$, there exists $\bar{b}_2 \in \left(0, \frac{1}{2}\right)$ such that, for any $\beta_0 \in (0, \underline{b})$, $|\mathrm{E}[\mu_0 \mid \omega_t = l_t] - 0| < \varepsilon_2$. The desired result then follows from picking

$\varepsilon = \min\{\varepsilon_1, \varepsilon_2\}$ and $b \in \left(0, \min\left\{\bar{b}_1, \bar{b}_2\right\}\right)$. □

We now prove the main result.

Proof. In the proof of proposition 5, we show that the fraction of periods that agents listen to their most trusted neighbors approaches one in probability as $t \to \infty$. In our setup, media outlets choose their bias to maximize the number of listeners. From lemma 4, we know that the derivative $\frac{dT_{ij}}{d|\beta_j|}$ does not depend on $|\beta_j|$, hence there cannot be an interior solution where $|\beta_j| \in (0,1)$. It follows immediately that all media choose $\beta_m \in \{-1, 0, 1\}$.

By proposition 4, an unbiased agent's trust always increases in α_j. By proposition 5, $\sum_{\tau=1}^{t} \mathbf{1}\left\{j_{U\tau} \in J^U\right\}/t \to_p$ 1 where j_{Ut} is the stochastic sequence of neighbors that U visits and J^U is the set of unbiased media sources. Note that the unbiased media source is fully informative about ω_t but uninformative. By the same logic as proposition 2, $\mu_U = \omega_t$ for any $a \in \left(\frac{1}{2}, 1\right)$.

Similarly, by proposition 4, for any $b \in \left(0, \frac{1}{2}\right)$, there exists $\overline{a} > \psi^{-1}(b)$ such that R-biased agent's trust increases in β_j if $a < \overline{a}$. Such an agent has the highest trust for a media source with $\beta_j = 1$, which is fully informative about r_t but uninformative about ω_t. Therefore, by proposition 5, $\sum_{\tau=1}^{t} \mathbf{1}\left\{j_{R\tau} \in J^R\right\}/t \to_p 1$ where j_{Rt} is the stochastic sequence of neighbors that such an R-biased agent visits and J^R is the set of media outlets with $\beta_j = 1$. Then by lemma 6, we have that for any $\varepsilon > 0$ and $\overline{b} > 0$, there exists $b < \overline{b}$ and $a \in \left(\frac{1}{2}, 1\right)$ such that agent 0's bias is not extreme and $|\mathrm{E}[\mu_0 - r_t]| < \varepsilon$. The analogous result is true for an L-biased agent. □