Sharing Models to Interpret Data

Joshua Schwartzstein and Adi Sunderam[*]
Harvard Business School

March 3, 2024

Abstract

To understand new data, we share models or interpretations with others. This paper studies such exchanges of models in a community. The key assumption is that people adopt the interpretation in their community that best explains the data, given their prior beliefs. An implication is that interpretations evolve within communities to better fit prior knowledge, potentially making final reactions less accurate than initial reactions. When people entertain a rich set of possible interpretations, social learning often mutes reactions to data: the exchange of models leaves beliefs closer to priors than they were before, untethering beliefs from data. Our results shed light on why disagreements persist as new information arrives, why popular theories link seemingly unrelated events, why ideological bubbles need not be hermetically sealed, and why firms and politicians can benefit from preemptively framing news.

[*] Contact: jschwartzstein@hbs.edu and asunderam@hbs.edu. We previously circulated this paper under the title "Shared Models in Networks, Organizations, and Groups". We thank Chiara Aina, Francesca Bastianello, Roland Benabou, Alessandro Bonatti, Tristan Gagnon-Bartsch, Ben Enke, Simone Galiperti, Robert Gibbons, Sam Hanson, Alex Imas, Sendhil Mullainathan, Matthew Rabin, Kunal Sangani, Andrei Shleifer, Mario Small, Jeremy Stein, and seminar participants at UC San Diego, Stanford GSB, Stanford Economics, the NBER Organizational Economics Meetings, the University of Zurich, SITE, Harvard, UC Santa Barbara, NYU, Chicago Booth, and briq for helpful comments. We thank John-Henry Pezzuto and Emma Ronzetti for outstanding research assistance.

1 Introduction

We make sense of the world together. Why is the unemployment rate lower than expected? Why did one employee receive a promotion while another did not? Why did a political candidate underperform her polls? Why is the stock market skyrocketing? In response to such questions, we share not only information but also interpretations. Unemployment numbers are lower than expected because “economic growth is strong” or because “there was a one-time blip in manufacturing”. One candidate received a promotion over another because “she is uniquely qualified” or “the firm is signaling that it values a particular set of skills”. The stock market is rising because of “fundamentals” or “dumb money”. What is the outcome of this exchange of interpretations? Does it push us towards the truth? How does with whom we talk affect the interpretation we come to believe? And how might an interested party like a firm manager influence patterns of communication to shape ultimate interpretations?

This paper presents a formal framework for thinking about such exchanges of interpretations in a community. The basic ingredients of the model follow Schwartzstein and Sunderam (2021). Everyone shares a common prior μ_0 over states of the world ω and observes a common, public history h. Aspects of the history are open to interpretation, meaning that people are willing to entertain different interpretations of the same data. Interpretations are represented by models, which we formalize as likelihood functions that link the history to states. In other words, interpretations capture different ways people can use the history to update their beliefs. When people are exposed to multiple interpretations, they adopt the one that best fits the data, fixing prior beliefs. People have a default interpretation d, represented by likelihood function $\pi_d(h|\omega)$, and come up with a single alternative interpretation—their initial reaction to the data—that they adopt if it is more compelling than their default interpretation, i.e., it better fits the data plus their prior.

In contrast to standard models where social learning is driven by the desire to learn others’ private information (e.g., reviewed in Golub and Sadler (2016)), in our framework everyone shares the same information but learns from others’ interpretations. People are exposed to the interpretations of others in their community or network and settle on the one that is most compelling.[1] Formally, person i adopts the model she is exposed to m (represented by likelihood function $\pi_m(h|\omega)$) if

$$m \in \arg \max_{\tilde{m} \in \left\{d, m_i'\right\} \cup M_i} \underbrace{\Pr(h|\tilde{m}, \mu_0)}_{=\int \pi_{\tilde{m}}(h|\omega) d\mu_0(\omega)},$$

where m_i' represents the model person i comes up with initially and M_i is the set of models the person is exposed to in her network. Thus, the key role of the community in our framework is to determine the set of models she is exposed to. Given this set of models, she picks the one

[1] We will use the terms network and community interchangeably.

that maximizes the probability of the history given her prior, $\Pr(h|m, \mu_0)$. In Bayesian terms, the person acts as if she has a flat prior over the models she is exposed to and then selects the model that best fits the data and her prior, which is equivalent to selecting the model with the highest associated posterior probability. More intuitively, this assumption loosely captures ideas from the social sciences about what people find persuasive, including that people favor models that (i) have high "fidelity" to the data as emphasized in work on narratives (Fisher 1985); (ii) help with "sensemaking" as discussed in work on organizational behavior and psychology (Chater and Loewenstein (2016)); (iii) make the past feel more predictable (Schulz and Sommerville (2006); Gershman (2019)); and (iv) have the most "explanatory power" (Lombrozo (2016)).[2]

To see some key implications of this formulation, consider an example where a community of investors assesses a technology firm. The firm's fundamentals are either good or bad, $\omega \in \{g, b\}$, and investors in this community are optimistic, attaching prior probability $\mu_0(g) = 0.75$ to fundamentals being good. Data comes out: h = "the firm's earnings this quarter were lower than expected and the aggregate economy entered a recession".

Investors consequently disagree about the firm's fundamentals because they use different models to interpret the data. Suppose that under the true model m^T, lower than expected earnings are a negative signal about fundamentals— $\pi_{m^T}(\text{low earnings}|b) = 0.75 > 0.25 = \pi_{m^T}(\text{low earnings}|g)$— but the aggregate economy is irrelevant for updating about the firm's fundamentals. Under these assumptions, the true posterior probability that the firm's fundamentals are good is $\Pr(g|h, m^T) = 0.5$. However, the investors are willing to entertain additional interpretations. Some initially come up with the true interpretation that the aggregate economy is irrelevant, but low earnings reduce the probability that the firm's fundamentals are good. Others think that low earnings are a particularly negative signal in a recession, following famed investor Warren Buffett's adage that "only when the tide goes out do you discover who's been swimming naked." Yet others believe that these kinds of technology firms always have low earnings in a recession, regardless of fundamentals. Suppose, in fact, that the only restriction on the set of models investors are willing to entertain is that they agree with the true model on the marginal probabilities of a recession and of low earnings, and the population is sufficiently large that roughly every such interpretation is someone's initial reaction. Assuming that the network is also sufficiently connected that the most compelling interpretation spreads throughout the population, we ask: which take goes viral?

Not the right one. When low earnings are more likely than recessions, the "always low in a recession" interpretation will eventually be held by all investors in this community because it is the best-fitting model consistent with the marginal probability of a recession. Social learning spreads an interpretation of the data under which investors' posterior on the firm's fundamentals equals

[2] Recent work (e.g., Barron and Fries (2022), Kwon et al. (2022)) experimentally tests and finds support for the assumption that people find better-fitting models more persuasive.

their prior probability of 0.75, instead of the true posterior of 0.5. As shown in Schwartzstein and Sunderam (2021), models that fit well imply the data is unsurprising, which means beliefs should not move much away from priors in response to it. In this example, investors' prior is that the firm likely has good fundamentals. The model that best fits their knowledge (i.e., their prior and the data) leads them to not move away from their prior.

This example illustrates four main points. First, explanations that link events that are in truth unrelated can be more persuasive than the true model. For instance, the "always low in a recession" model, which links recessions and low earnings, fits the data better than the true model that they are independent events. In our framework, such "conspiratorial" models will tend to spread more easily than the truth. Indeed, the model that investors end up holding is "maximally conspiratorial": it connects low earnings and the recession as strongly as possible in a way we make precise below. Second, social learning *hardens* reactions to data: following the exchange of models, people are more convinced they have the right explanation for the data in the sense of having a model with a higher value of $\Pr(h|m, \mu_0)$. Exposure to others' models helps them find ways to explain the data that they may not find on their own. Thus, there are investors who could have been persuaded of the true model prior to social learning but who cannot be persuaded after because social learning provides them with a better explanation for the data. Third, interpretations *evolve* in ways that often make final interpretations *less* accurate than initial reactions. Before social learning, some investors correctly interpret low earnings and put a low posterior probability on the state in which the firm's fundamentals are good. However, the sharing of models leads them to interpretations resulting in the belief that the fundamentals are likely to be good. In other words, the marketplace of models pushes people away from the right interpretation. This evolution of beliefs highlights a key distinction between our formulation and models where people simply believe what they want to believe. In these alternative formulations, if investors prefer accounts that the firm's fundamentals are good, then their *initial* reactions will exhibit that preference. A fourth point is that social learning has a tendency to *mute* reactions—bringing posterior beliefs closer to prior beliefs—by increasing the chances that people are exposed to models that explain why the data is unsurprising and hence beliefs should not move. Put differently, the exchange of models untethers beliefs from data that is open to interpretation.

Untethering appears to be an important feature of many real-world settings. For instance, views on issues like the health risks of Covid-19 are stable over time: despite wide swings in the data (e.g., case rates, hospitalizations, and deaths), the fraction of Republicans who felt Covid was a significant health risk in 2020 and 2021 was consistently 41-46%.[3] Consistent with our model, this stability does not mean that people do not react to news. They do react, but the impact of news tends to fade quickly, with people returning to their previous views. In our framework, this

[3] See, e.g., Pew surveys here (link) and here (link).

disconnect between data that is open to interpretation and long-run beliefs is driven by the adoption of models through social learning.

Section 2 introduces the model and basic definitions. We then turn to the impact of community structure on interpretations and beliefs. Section 3 studies networks formed on the basis of shared beliefs, where people who have similar initial reactions to the data exchange models. Such networks are common—for instance, groups often form based on beliefs that one political party governs better than others—and have likely become easier to form over time due to technology like social media. To illustrate the intuitions that emerge, consider the example above, and suppose investors who initially interpret the data as suggesting firm fundamentals are good all talk to each other, while those whose initial interpretations suggest fundamentals are bad all talk to each other. We show that social learning then generates disagreement: while they all held the same prior beliefs, members of the "good fundamentals" and "bad fundamentals" networks end up disagreeing despite being exposed to the same information. Investors in the "good fundamentals" network converge on models that bring their beliefs closer to the 75% conditional prior probability they attached to the firm being good, while those in the "bad fundamentals" network converge to a probability less than 50%. As they exchange models, members of each network become less persuadable to arguments made outside their network (beliefs are hardened) and hold beliefs that are less tethered to the data (beliefs are muted).

We next document stylized evidence consistent with these predictions from a social-media network for stock market investors. Previous work has documented that optimistic investors (i.e., "bulls") tend to form networks with other optimists, while pessimistic investors (i.e., "bears") tend to form networks with other pessimists (Cookson et al. (2022)). We study the dynamics of optimism following the release earnings announcements. We show that investors who are bullish about a company become less bullish in the immediate aftermath of a negative earnings surprise. However, they quickly revert back to being bullish. Similarly, bearish investors become less bearish after a positive earnings surprise, but quickly revert back. In our framework, these dynamics are driven by the spread of interpretations within the bullish and bearish networks. Within each network, investors are exposed to interpretations of the data that make it less surprising, allowing disagreement to persist in the face of new information.

In Section 4, we show that inaccurate beliefs and disagreement can persist even as people communicate across networks and issues. We first consider communication across networks, drawing a distinction between weak and strong exposure to beliefs outside a person's network. We say a person is weakly exposed to a belief if she is aware of a single model that when combined with the data implies that belief. She is strongly exposed to a belief if she is aware of all models implying that belief. We think of communication within networks as strong exposure and communication across networks as weak exposure. Under this view, members of a network can be aware that people outside their network have different beliefs, but they will be unpersuaded by the interpretations of the data they know in favor of those different beliefs. In other words, ideological bubbles or echo chambers need not be hermetically sealed. People in a network can be exposed to a few interpretations from outside the network without finding those interpretations compelling.

We then study communication across issues, showing that it can lead to polarization. If networks are formed based on one issue, the exchange of interpretations leads to a divergence across networks of beliefs on a second issue. For instance, a person in a network formed based on the shared belief that genetically-modified crops are unsafe is more likely to end up believing that they are also bad for the environment. Thus, members of two communities can end up disagreeing on issues that were not central to the formation of their communities, despite the fact they are interpreting the same data. Such disagreement is consistent with the "polarization of reality" documented by Alesina et al. (2020). The influence of the network on beliefs about the second issue is governed by the extent to which people connect with others because of their views on the first issue and the correlation in prior beliefs across the two issues.

Section 5 then studies the implications of these results for how someone should manage communication. We first show that providing people with a favored interpretation before social learning—i.e., "getting in front of the news"—is valuable in our framework. It helps the communications manager inoculate people against finding compelling models that support alternative beliefs. We then consider situations where the manager directly influences what communication takes place, for instance by restricting meeting attendance. We show that the communications manager faces a tradeoff between getting people to agree on a model and promoting a specific action. To create agreement on a model, the manager wants everyone to share interpretations, which leads to the widespread adoption of a model suggesting that there is little to learn from the data. In contrast, if the manager wants people to take a particular action, then she wants only models that support that action to be shared. In this case, however, not everyone will end up with the same model, as some people will find their initial reactions more compelling than the models that are shared.

We then sketch some applications of our results in Section 6. We first study implications for how firm managers should run meetings. The traditional view in economics is that meetings enable information exchange (e.g., Dessein and Santos (2006)). In contrast, in our framework meetings serve to help workers interpret shared information, a view that builds on a large literature in organizational studies arguing that sensemaking is a central activity of organizations (e.g., Weick (1995)). Our key result is that leaders can find it optimal to use meetings to coordinate workers on interpretations that suggest there is little to learn from the data, even if the leaders themselves interpret the data as suggesting the organization needs to better adapt to the environment.

We then consider why disagreement persists in the face of new information. Why do misconceptions survive in some groups, given that people have access to high-quality information and diverse news diets (Gentzkow and Shapiro (2011); Guess et al. (2018))? We offer a simple explanation, complementing recent models that instead highlight the role of social media echo chambers (Bowen et al. (2021)): Within a network or ideological bubble, people are exposed to crowdsourced models that evolve to better and better fit data that is open to interpretation, making them less persuadable. In our framework, bubbles do not prevent people from being exposed to the right take on an event, but they inoculate against finding that take compelling. Vaccine skeptics are aware that many people say vaccines are safe and know some pro-vaccine arguments (Larson et al. (2011)), but they have been exposed to a broad diversity of arguments for why vaccines are unsafe and find some such arguments more persuasive. This incoluation may shed light on why reducing exposure to like-minded information on social media does not appear to significantly impact beliefs (e.g., Nyhan et al. (2023)).

Related Literature

There is a large literature on social learning reviewed in Golub and Sadler (2016), with influential early contributions in economics such as Banerjee (1992), Bikhchandani et al. (1992), and Smith and Sørensen (2000). While much of this work assumes Bayesian updating of beliefs, important recent contributions study naive social learning by building on the simple DeGroot (1974) model of linear updating (Golub and Jackson (2010)) or on more psychologically microfounded updating rules (e.g., Eyster and Rabin (2010, 2014); Enke and Zimmermann (2019); DeMarzo et al. (2003); Gagnon-Bartsch and Rabin (2016)). This work focuses on people sharing information (e.g., how much they enjoyed meals at a restaurant) or observing each others' actions (e.g., seeing that a restaurant is popular), and studies questions like whether social learning successfully aggregates individuals' private information in the long run. Our focus is instead on the many situations where people share essentially the same information, and social learning primarily involves exchanging interpretations to make sense of that information.

While frameworks featuring social learning of information tend to predict long-run consensus and relatively effective information aggregation, in our framework the marketplace for models naturally generates long-run disagreement and the persistence of false beliefs. As such, it may help explain why disagreement persists in so many domains from stock prices to public health to evolution, despite an abundance of data. Increasing connectedness tends to untether beliefs from data that is open to interpretation by increasing the chances of being exposed to a model that provides a compelling case that the data is unsurprising. Wrong interpretations are adopted in our framework not because they are repeatedly heard, but because social learning selects interpretations that compellingly fit people's prior knowledge.

A smaller literature on social learning examines how people could leverage networks to their advantage in spreading information. Much of this work considers how to best seed a network with information to boost its diffusion (e.g., Akbarpour et al. (2020)). Murphy and Shleifer (2004) present a model of the creation of social networks based on shared beliefs in the context of political persuasion. This work considers social learning of information or beliefs rather than of models.

Closer to our work, recent presidential addresses in finance, such as Shiller (2017) and Hirshleifer (2020), have called for studying the social transmission of narratives in economics and finance.[4] These addresses, as well as a related book (Shiller (2020)), have laid the groundwork for this study by providing vivid illustrations of the importance of socially-emergent narratives as drivers of economic and financial events. They also sketch models of narrative transmission that liken the spread of narratives to the spread of viruses. Bénabou et al. (2018) model the spread of moral narratives (e.g., "thou shall not do this because") by strategic actors, while Bursztyn et al. (2023) and Bursztyn et al. (2022) argue that providing rationales and narratives importantly influences people's willingness to voice certain beliefs. Our work adds to this line of study by formally modeling social forces that shape the narratives themselves and highlighting that good explanatory power helps narratives "go viral".

We build on our earlier work on model persuasion (Schwartzstein and Sunderam (2021)), which itself built on behavioral models of persuasion based on coarse or associational thinking (e.g., Mullainathan et al. (2008)).[5] Froeb et al. (2016) present an earlier related model in the context of studying adversarial decision making in law, Levy and Razin (2020) present a related model speaking to the problem of combining expert forecasts, Aina (2021) builds on the model persuasion framework by considering what happens when persuaders need to commit to models before seeing all the data, and Ichihashi and Meng (2021) considers the interaction between Bayesian persuasion (Kamenica and Gentzkow (2011)) and model persuasion. Other recent work (Eliaz and Spiegler (2020); Bénabou et al. (2018); Yang (2022); Eliaz et al. (2022)) take somewhat different approaches to formalizing models or narratives and what makes them persuasive. For example, Eliaz and Spiegler (2020) assume that people favor "hopeful narratives", Eliaz et al. (2022) assume that narratives emerge competitively to increase political mobilization, and Yang (2022) assumes that people favor "decisive models". A growing empirical and experimental literature measures people's models or narratives, as well as how they influence expectations and decisions (e.g., Barron and Fries (2022); Andre et al. (2022); Flynn and Sastry (2022); Hüning et al. (2022)). We add to this work by formalizing how social learning influences which models emerge and persist.

[4] While not all narratives are models and vice-versa, they are closely related and we sometimes interchangeably use the terms narratives, stories, and models.

[5] Our framework also connects to the literature on learning under misspecified models (e.g., Esponda and Pouzo (2016); Acemoglu et al. (2016); Heidhues et al. (2018); Montiel Olea et al. (2022); Mailath and Samuelson (2020); Haghtalab et al. (2021)), which sometimes feature agents who statistically test their models and abandon them in favor of alternatives which fit better. Examples include Fudenberg and Kreps (1994); Hong et al. (2007); Gagnon-Bartsch et al. (2021); Fudenberg and Lanzani (2021); Ba (2021).

2 Model

2.1 Setup

The basic setup follows Schwartzstein and Sunderam (2021). Broadly, individual agents take the following steps in interpreting data. All agents share a common default model for interpreting data, and in addition each agent comes up with a model of their own. Prior to social learning, each agent selects from these two models the one that best explains the data. Social learning then exposes each agent to all models held by other agents in her community or network. After social learning, each agent adopts the model that best explains the data from the set of models she has been exposed to: the default, the model she comes up with on their own, and the models others in her social network have come up with.

Formally, there are a continuum of agents $i \in [0,1]$ who hold beliefs μ_i over states of the world ω in finite set Ω.[6] Agent i takes an action a from compact set A to maximize the expectation under μ_i of $U_i(a,\omega)$. In the baseline setup, agents share a common prior $\mu_0 \in \text{int}(\Delta(\Omega))$ over Ω and observe a public history of past outcomes, h, drawn from finite outcome space H. Agents can end up with different posteriors if they use different models to interpret this history. Given state ω, the likelihood of h is given by $\pi(\cdot|\omega)$. The true model m^T is the likelihood function $\{\pi_{m^T}(\cdot|\omega)\}_{\omega\in\Omega} = \{\pi(\cdot|\omega)\}_{\omega\in\Omega}$. We assume that every history $h \in H$ has positive probability given the prior and true model.

Agents do not know the true model. A given agent updates her beliefs based on either (i) the default model $\{\pi_d(\cdot|\omega)\}_{\omega\in\Omega}$,[7] (ii) the model m'_i that she generates herself to explain the history, where m'_i is taken from compact set M and indexes a likelihood function $\left\{\pi_{m'_i}(\cdot|\omega)\right\}_{\omega\in\Omega}$, or (iii) a model she learns from someone in her network, where we let $M_i \subseteq M$ denote the set of models proposed by someone in i's network.

Given the history and the set of models the agent is exposed to, she adopts the one that best explains the history. Formally, let $\mu(h,\tilde{m})$ denote the posterior distribution over Ω given h and model $\tilde{m} \in M \cup \{d\}$, as derived by Bayes' rule. We assume the receiver adopts the model m and hence posterior $\mu(h,m)$ if

$$m \in \arg\max_{\tilde{m}\in\left\{d,m'_i\right\}\cup M_i} \underbrace{\Pr(h|\tilde{m},\mu_0)}_{=\int \pi_{\tilde{m}}(h|\omega)d\mu_0(\omega)} .$$

That is, the person picks the model she is exposed to that best fits the data. Upon adopting a model $\tilde{m}$, the person uses Bayes' rule to form posterior $\mu(h,\tilde{m})$ and takes an action that maximizes her

[6] In examples we sometimes relax the assumption that Ω is finite.

[7] The default can be a function of h. We suppress the dependence of d on h when it does not cause confusion.

expected utility given that posterior belief: $a(h, \tilde{m}) \in \arg\max_{a \in A} \mathbb{E}_{\mu(h,\tilde{m})}[U_i(a, \omega)]$.

To close the baseline model, we need to specify the model a person generates herself. Let $\bar{M}(h, \mu_0, d, M) = \{m \in M : \Pr(h|m, \mu_0) \geq \Pr(h|d, \mu_0)\}$ denote the set of models in M that explain the history as well as the person's default interpretation given her prior over states. Assume that measure δ of the population generates the default model and measure $(1 - \delta)$ generates a model in $\bar{M}(h, \mu_0, d, M)$.[8] Further assume that population is large enough that, for each model $m \in \bar{M}(h, \mu_0, d, M)$, someone in the population generates that model herself.

In the typical case, we set the default interpretation to be the true model, $d = m^T$ and focus on situations where data are open to interpretation—i.e., where people are in fact sharing interpretations of data. To discipline the analysis, we also typically let M be the set of all possible models M^a: for any likelihood function $\{\tilde{\pi}(\cdot|\omega)\}_{\omega \in \Omega}$ there is an $m \in M^a$ with $\{\pi_m(\cdot|\omega)\}_{\omega \in \Omega} = \{\tilde{\pi}(\cdot|\omega)\}_{\omega \in \Omega}$. We refer to this as the case where people are *maximally open to persuasion.* We simply write $\bar{M}(h, \mu_0)$ as shorthand for $\bar{M}(h, \mu_0, m^T, M^a)$.[9]

2.2 Discussion of Model Assumptions

The building blocks of the model come from Schwartzstein and Sunderam (2021), and we refer to that paper for a detailed discussion of the basic assumptions. We depart from that paper in a few crucial ways. First, we allow some receivers by themselves to generate a model other than the default. In the notation of our current framework, our previous paper assumes $\delta = 1$ (receivers stick with the default before being exposed to persuasion), while the analysis in this paper focuses on the case where $\delta < 1$. For many topics, it is plausible that some people generate an initial interpretation of the data, prior to sharing interpretations with others. Many of us have gut reactions about why the stock market moved yesterday, who is responsible for the storming of a government building, or what the latest school shooting implies about the merits of gun control. These gut reactions may be constructed spontaneously in response to the data and differ across people (see, e.g., Andre et al. (2022) for evidence of heterogeneity in households' and experts' models of the causes of inflation). Crucially, however, we assume that a given person does not come up with all models she is willing to entertain, so she is influenced by the set of models she is exposed to.

Second, the focus of this paper's analysis is on the social exchange of models, not on the behavior of a strategic persuader who attempts to influence the beliefs and behavior of audience

[8] Alternatively, we could endogenize δ by assuming that people sometimes generate models outside of $\bar{M}(h, \mu_0, d, M)$ in which case they stick with the default model. This would suggest that δ is larger when the default does a good job explaining the data h. While this change would influence the distribution of beliefs prior to social learning, it would not influence the distribution of beliefs following social learning.

[9] All our results and intuitions stated for the case of $M = M^a$ continue to hold if we instead make the following assumption on M: For every belief $\tilde{\mu}$ that is a posterior for some model in M^a given data h, prior μ_0, and default d, M includes the best-fitting model inducing that posterior as well as one worse-fitting model inducing that posterior.

members. The role of the community or network in our framework is simply to influence the set of models a person is exposed to. By taking as primitive the set of models a given person i is exposed to, M_i, our framework accommodates a variety of network structures.

Third, implicit in the idea that a person is exposed only to the models within her network is an assumption that she does not actively seek out the models proposed by members of other networks. One way of thinking about this assumption is that people exhibit a sort of out-group homogeneity bias (e.g., Quattrone and Jones (1980); Bursztyn and Yang (2023)), thinking there is not much reason to investigate the models in other networks because they are "all the same". A person who favors gun control may be aware of some arguments for why shootings suggest weaker gun control (e.g., "we need more guns in the hands of good guys") and may think once she has heard one such argument she has heard them all, perhaps underappreciating the diversity of these arguments.

2.3 Examples

We now sketch two brief examples, which we will return to throughout the paper.

Example 1 (Interpreting data about investments)**.** Extend the example from the introduction to let $\omega \in \{g, b\}$ and $h = (h_1, h_2)$ with $h_1 \in \{l, r\}$ and $h_2 \in \{u, d\}$. For example, h_1 could stand for whether the aggregate economy has entered a recession, h_2 for whether company earnings are high or low, and ω for whether company fundamentals are good or bad. Letting $\pi_1(h_1) \equiv \sum_{\omega'} \pi_{m^T,1}(h_1|\omega') \cdot \mu_0(\omega')$ and $\pi_2(h_2) \equiv \sum_{\omega'} \pi_{m^T,2}(h_2|\omega') \cdot \mu_0(\omega')$, suppose that, as in the introductory example, the set of models the person is willing to entertain agree with $\pi_{m^T}(h|\omega)$ on the marginal probability of h_1 and h_2:

$$M = \left\{ m \,\middle|\, \sum_{h_2' \in \{u,d\}} \Pr(h_1, h_2'|m, \mu_0) = \pi_1(h_1)\ \forall h_1 \text{ and } \sum_{h_1' \in \{l,r\}} \Pr(h_1', h_2|m, \mu_0) = \pi_2(h_2)\ \forall h_2 \right\}.$$

Example 2 (Interpreting data about policy issues)**.** Our second example involves interpreting data about a binary state space, $\Omega = \{l, r\}$. Unlike the first example, here the space of models is unrestricted. The prior over states is $\mu_0(l) = 1/2$. Further assume people are maximally open to persuasion, $M = M^a$, given the data h and the default model is the true model, $d = m^T$. People can take three possible actions, $a \in \{L, M, R\}$. Payoffs U_i are such that the optimal action is $a = L$ if $\mu(l) \geq .75$, $a = M$ if $\mu(l) \in (.25, .75)$, and $a = R$ if $\mu(l) \leq .25$.

This example can capture optimal public-policy choices. In state $\omega = l$, a Democrat would make a better US president, and in state $\omega = r$ a Republican would make a better US president. Actions correspond to voting Democrat ($a = L$), abstaining ($a = M$), and voting Republican

($a = R$). Alternatively, one can think of the states as corresponding to whether some left- or right-leaning policy (e.g., involving gun control, climate change, pandemic policy) would be effective, and the actions as corresponding to supporting such policies ($a = L$, R) or the status quo ($a = M$). The example can also capture choices of firms, for instance to cut costs, grow, or stay the course.

We will sometimes extend this example to cases where people may use the same data to update beliefs about a variety of issues. For instance, people may interpret data about genetically-modified crops using models that have implications for both their safety and impact on the environment (e.g., how their adoption influences pesticide use). To accommodate such examples, let $\Omega = \Omega^1 \times \Omega^2$. We will consider how network members' beliefs over Ω^1 (e.g., what the data implies about the environmental impact of genetically-modified crops) spill over to influence beliefs over Ω^2 (e.g., what the data implies about their safety).

2.4 Basic Observations and Definitions

Prior to social learning, a person adopts the model

$$m' \in \arg \max_{\tilde{m} \in \left\{d, m'_i\right\}} \Pr(h|\tilde{m}, \mu_0)$$

and holds beliefs $\mu(h, m')$, which we call their "initial reaction."

Appendix Section B.1 analyzes initial reactions before social learning, adapting Proposition 1 in Schwartzstein and Sunderam (2021) to the present context. Two key points follow. First, before social learning, people have a variety of reactions to the data. Second, there are constraints on initial reactions, which in turn imply constraints on final beliefs. In particular, the set of initial reactions is constrained by prior beliefs, $\mu_0(\omega)$, as well as the ability of the default to explain the data given those prior beliefs, $\Pr(h|d, \mu_0)$. Intuitively, the better the default model fits the data, the harder it is for an initial reaction to fit the data even better. And the more unlikely a state under peoples' prior, the less likely it is that their beliefs following their initial reaction put a lot of weight on that state. If the data is maximally open to interpretation, sticking with prior beliefs is always an initial reaction to the data and the range of initial reactions is greater when people are more surprised by the data, i.e., when $\Pr(h|d, \mu_0)$ is lower.[10]

Following social learning, the person adopts the model

$$m \in \arg \max_{\tilde{m} \in \left\{d, m'_i\right\} \cup M_i} \Pr(h|\tilde{m}, \mu_0)$$

[10] The range of initial reactions will be smaller when prior beliefs are more informed, for example because they reflect a long history of closed-to-interpretation data (see Proposition 2 in Schwartzstein and Sunderam (2021)).

and holds beliefs $\mu(h, m)$ when such maximizers exist—assume throughout the paper that M_i is indeed such that such maximizers exist. As shorthand, write μ_i' (m_i') as person i's beliefs (adopted model) prior to social learning and μ_i (m_i) as her beliefs (adopted model) following social learning.

We say that social learning *hardens* a person's reaction to data when she can better explain the data following social learning than before: that is, when $\Pr(h|m_i, \mu_0) \geq \Pr(h|m_i', \mu_0)$. When social learning does not harden the person's reaction, we say it *softens* her reaction. We say that social learning *mutes* a person's reaction to data when it moves her beliefs closer to her prior. Formally, following Schwartzstein and Sunderam (2021), let Movement$(\tilde{\mu}; \mu_0) \equiv \max_{\omega\in\Omega} \tilde{\mu}(\omega)/\mu_0(\omega)$ be a measure of the change in beliefs from prior μ_0 to posterior $\tilde{\mu}$. Social learning mutes reactions to the data when Movement$(\mu_i; \mu_0) \leq$ Movement$(\mu_i'; \mu_0)$. When social learning does not mute a person's reaction to data, we say it *intensifies* her reaction. Note that these terms compare final beliefs to people's initial reactions to data before social learning, not to their prior.

A simple observation is that our assumptions imply that social learning must harden reactions to data: being exposed to more explanations of the data enables the person to better explain the data. Social learning leads a person to become more convinced she understands why the market moved as it did, why an unexpected political event occurred, or the daily movement in pandemic deaths. Following social learning, any event seems more explainable.

The next section begins a more in-depth analysis of how social learning shapes beliefs.

3 Social Exchange of Models

3.1 Shared-Belief Networks

In our framework, the network determines the set of models that people are exposed to. Network formation therefore plays a crucial role in determining ultimate beliefs. Throughout the paper, we will frequently analyze the case where networks are formed on the basis of shared beliefs. Such networks are quite common (e.g., McPherson et al. (2001)). For instance, networks are formed based on beliefs that one political party typically governs better than others, that vaccines are harmful, and even whether a specific stock is likely to rise (Cookson et al. (2022)). Moreover, technology like social media has made it easier for such networks to form and for members of these networks to exchange interpretations. Of course, not all networks are formed on the basis of shared beliefs. People may connect with others for other reasons, for instance because they live in the same physical neighborhood or because they share the same models. We briefly discuss implications of alternative network structures in Section 7.

The key feature of a *shared-belief network* is that the beliefs a person holds prior to talking to others influences who she talks to. Formally, consider a partition $\mathcal{S}$ over the set of beliefs $\Delta(\Omega)$, where we denote $s(\mu)$ as the element in $\mathcal{S}$ that belief $\mu \in \Delta(\Omega)$ belongs in. In a shared-belief network, a person i exchanges models with another person j if and only if their initial beliefs are similar, in the sense that they fall in the same element of $\mathcal{S}$.

Definition 1. In a *shared-belief network*, $M_i = \big\{m \in \bar{M}(h, \mu_0, d, M) : \mu(h, m) \in s(\mu(h, m_i'))\big\}$ for every person i.

Given our assumption of common priors, this definition says that a shared-belief network forms based on a common reaction to a specific event. For example, a shared-belief network could form among people who react similarly to a shooting in their beliefs on the need for gun control. This literal interpretation is a reasonable approximation of reality for certain events, such as the earnings announcements we consider in Section 3.2. However, in other instances networks based on shared beliefs are formed based on common reactions to a broader set of events. For example, people who lean left in their interpretations might share views on the most recent event, even if their initial views on that most recent event are quite different. We will more formally capture this idea in briefly studying dynamics in Section 6.

As an illustration, consider a complete network and return to the investment example (Example 1).

Proposition 1. *In the setup of the investment example (Example 1), suppose $\tilde{h} = (\tilde{h}_1, \tilde{h}_2) = (r, d)$ is realized with $\pi_1(r) \leq \pi_2(d)$. Even if the relationship between the two variables under the true model is $\Pr(d|r, m^T, \mu_0) - \Pr(d|l, m^T, \mu_0) = 0$, social learning in a complete network leads people to hold a model $m' \in M$ that maximally connects the variables:*

$$\Pr(d|r, m', \mu_0) - \Pr(d|l, m', \mu_0) = \max_{m \in M} \Pr(d|r, m, \mu_0) - \Pr(d|l, m, \mu_0) = \frac{1 - \pi_2(d)}{1 - \pi_1(r)}.$$

Under any such model m' that has the property that $\pi_{m'}(r|b) = \pi_{m'}(r|g)$, people's posterior beliefs equal their prior beliefs μ_0.

As in the introductory example, natural restrictions on the model space clarify how social learning may lead people to draw connections between events that they would otherwise be surprised by. The proposition shows that people end up holding models that view the rarer realized h_i as implying the more common realized h_j with certainty. For example, when recessions are rarer than low earnings, low earnings are viewed as being inevitable given recessions. When high earnings are viewed as more likely than a good economy, then high earnings are viewed as inevitable given a good economy, and a bad economy is viewed as unsurprising given low earnings. Put differently, people have a tendency to say "of course X (more common outcome) given Y (rarer outcome)". Explanations and sense-making center around the rarer outcome. Thus, our framework puts structure on the kinds of "conspiratorial" links that groups of people will tend to draw between unrelated events. The proposition additionally shows that such explanations neutralize the data, leaving posterior beliefs the same as prior beliefs, so long as people do not view the rare outcome as itself signaling something about ω, e.g., they do not think a recession by itself provides news about whether a company is good.

Neutralization of the data is in fact the norm if people are maximally open to persuasion. We first recall a lemma from Schwartzstein and Sunderam (2021).

Lemma 1 (Schwartzstein and Sunderam (2021))**.** *Fix history h and let*

$$\textit{Fit}(\tilde{\mu}; h, \mu_0) \equiv \max_m \Pr(h|m, \mu_0) \textit{ such that } \mu(h, m) = \tilde{\mu}$$

be the maximal fit of any model that induces posterior $\tilde{\mu}$ given the history h and a person's prior μ_0. Then

$$\textit{Fit}(\tilde{\mu}; h, \mu_0) = 1/\textit{Movement}(\tilde{\mu}; \mu_0),$$

where Movement$(\tilde{\mu}; \mu_0) \equiv \max_{\omega \in \Omega} \tilde{\mu}(\omega)/\mu_0(\omega)$ is the movement to $\tilde{\mu}$ from μ_0.

Intuitively, fit and movement are inversely related because models that fit the history well say it is unsurprising in hindsight, which then implies that beliefs should move little. So, for any given belief μ, the maximal fit of a model inducing that belief is greater the closer this belief is to μ_0.

Proposition 2. *Suppose everyone is in a shared-belief network and is maximally open to persuasion, $M = M^a$. Then social learning mutes every person's reaction to the data: for every person i, Movement$(\mu_i; \mu_0) \leq$ Movement$(\mu_i'; \mu_0)$. In fact, social learning leads everyone's final beliefs to be in the set of initial beliefs within the network that are closest to the prior: for every person i, $\mu_i \in \arg\min_{\mu \in s(\mu_i')}$ Movement$(\mu; \mu_0)$.*

Proof. All proofs are in Appendix A.

□

This result says that if a person only exchanges models with others who react similarly to data, they end up with the belief that reacts least to the data *within their network*. This result follows from Lemma 1 and the following feature of shared-belief networks: For every belief represented in a network, the best-fitting model supporting that belief is represented in the network.

Proposition 2 has a key implication for the evolution of beliefs within a network. As an illustration, a school shooting might initially lead people to support a change in gun-control policies, but they will eventually favor interpretations that say we did not learn much from the shooting. There is empirical evidence suggestive of such dynamics. Following mass shootings, Twitter users who are initially against gun control temporarily become more open to the idea. However, as narratives evolve and spread in the weeks following a mass shooting, these Twitter users slowly revert back towards their original beliefs (Lin and Chung (2020)).

To further illustrate Proposition 2, suppose that in the public-policy example (Ex. 2), $\mu_0(l) = 1/2$ and the data is surprising—i.e., under the default model, it has very low probability. For example, suppose $h =$ "a natural disaster struck and GDP growth this quarter was low." Suppose further that shared-belief networks are formed based on views of the optimal action: Everyone with an initial reaction supporting a right-leaning action like a tax cut is in one network ($\mu'_j(r) \in [k, 1]$) for $k \in (.5, 1)$), everyone with an initial reaction supporting the neutral action is in another ($\mu'_j(r) \in (1 - k, k)$), and everyone with an initial reaction supporting a left-leaning action like stimulus checks is in the final network ($\mu'_j(r) \in [0, 1 - k]$). The variable k could be viewed as parameterizing the degree to which people form shared-belief networks based on the intensity of their views on the issue. Under this interpretation, it is larger when people connect with others based on their views about the issue (e.g., through social media) and smaller when people connect with others for other reasons (e.g., because they live in the same physical neighborhood).

Proposition 2 says that everyone in a given network ends up at the belief that is closest to the prior within her network (Figure 1 illustrates this for $k = .75$). For example, someone whose initial reaction to the data moves her belief from $\mu_0(r) = .5$ to $\mu'(r) = .9$ will exchange models with others whose initial reactions support the right-leaning action (pictured in red in the figure), which mutes her reaction to $\mu(r) = k$. Polarization in views across the right- and left-leaning networks equals $k - (1 - k) = 2k - 1$: it is increasing in the extent to which people talk to others based on their views along the issue. In other words, the network formation process (indexed by k) drives polarization across networks, while the exchange of models drives the homogeneity of beliefs within the network.[11]

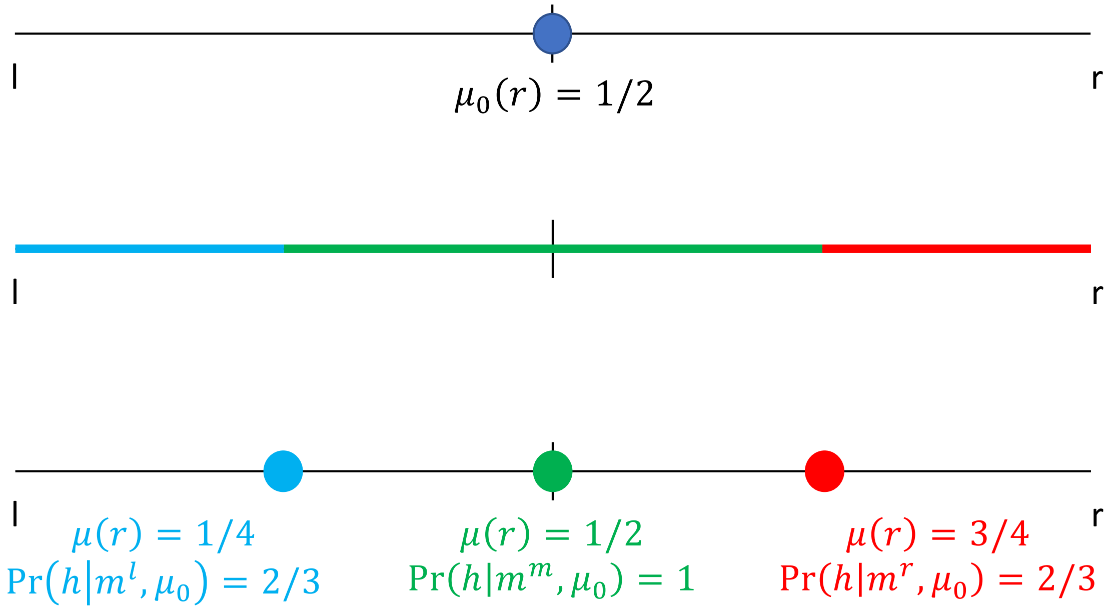
Taken together, these results highlight the differences between our framework and typical information-based theories of social learning, in which social exchanges of information tend to lead to more accurate beliefs. In our setting, social exchanges of models increase the chances of hearing an interpretation that suggests the data are relatively consistent with a person's prior and hence there is little need to update. In other words, the "marketplace of ideas" need not result in beliefs that are closer to the truth.

3.2 Empirical Evidence

We next provide empirical evidence consistent with the key mechanisms in Proposition 2. We use data from StockTwits, a social network for investors that has been studied in recent papers, including Cookson and Niessner (2020), Divernois and Filipovic (2022), and Cookson et al. (2022).

[11] Section C provides an example that shows how interpretations may evolve differently across shared-belief networks, illustrating an additional form of polarization.

Figure 1: Evolution of Beliefs Across Shared-Belief Networks Surrounding a Single Policy Issue



StockTwits is a social media platform similar to Twitter: users choose other users to follow and post short messages visible to their followers. Founded in 2008, the platform had 6 million total users and 1 million active monthly users at the end of 2021.[12] The platform is geared towards allowing investors to share with each other information and analysis about individual stocks. In particular, it allows users to (i) tag their messages with individual stock tickers and (ii) label their messages with a flag for bullish or bearish sentiment. These features make it straightforward to track a particular user's sentiment towards a particular stock over time. Cookson and Niessner (2020) perform a variety of exercises to validate the data's quality for measuring sentiment and disagreement.

Two stylized facts from StockTwits are relevant to our framework. First, based on the evidence in Cookson et al. (2022), the way we model the formation of a shared-belief network is consistent with how StockTwits members actually form their networks. Cookson et al. (2022) show that users who are bullish on a particular stock are more likely to start following other users who are also bullish on the same stock. Similarly, bearish users are more likely to start following other bearish users. Moreover, this behavior is more pronounced immediately following earnings announcements. In other words, following a news announcement, StockTwits users are more likely to form networks with others who share their views on a particular stock.

Second, beliefs of StockTwits users around earnings announcements evolve as Proposition 2

[12] See this (link) Bloomberg article.

predicts. We analyze the dynamics of user sentiment around earnings announcements in StockTwits messages between January 2011 and July 2018.[13] For each message about a particular stock, we code sentiment as 1 if the user labels the message as bullish and 0 if the user labels the message as bearish. We drop messages that users do not label. We consider windows from 10 days before an earnings announcement to 10 days after for a given stock. We restrict attention to users who have ever posted a message about that stock prior to 10 days before the earnings announcement. We code a user as a bull on the stock if the user labeled as bullish at least 50% of their messages about the stock prior to 10 days before the earnings announcement. We code the user as a bear if they labeled as bearish more than 50% of their messages about the stock. We then track how sentiment evolves in response to different earnings announcements over the surrounding windows. We code an announcement as positive news if the announcement day return is greater or equal to zero and as negative news if the announcement day return is negative.[14] Because different stocks receive different amounts of attention, we weight the data so that each earnings announcement is equally weighted. The final sample consists of roughly 1.8 million messages across 40,000 earnings announcements from 65,000 unique users.[15]

Figure 2a shows the evolution of sentiment for bulls around positive and negative earnings announcements. Corresponding regressions are reported in Appendix D and show that the patterns in the figure are statistically significant. Note that our data do not allow us to directly demonstrate that these patterns are driven by the network itself—they could reflect evolving interpretations people come up with by themselves. However, in traditional social learning models based on sharing information, the network should push against such tendencies.

Figure 2a shows that around positive announcements, sentiment remains unchanged: generally 94-95% of messages are labeled as bullish. The pattern around negative earnings announcements contrasts sharply. Sentiment deteriorates in the days leading up to the earnings announcement. There is then a sharp decline in sentiment at the earnings announcement, with the fraction of messages labeled as bullish falling to under 85%, statistically and economically different from its baseline value.[16] In the days following the announcement, however, sentiment rapidly improves,

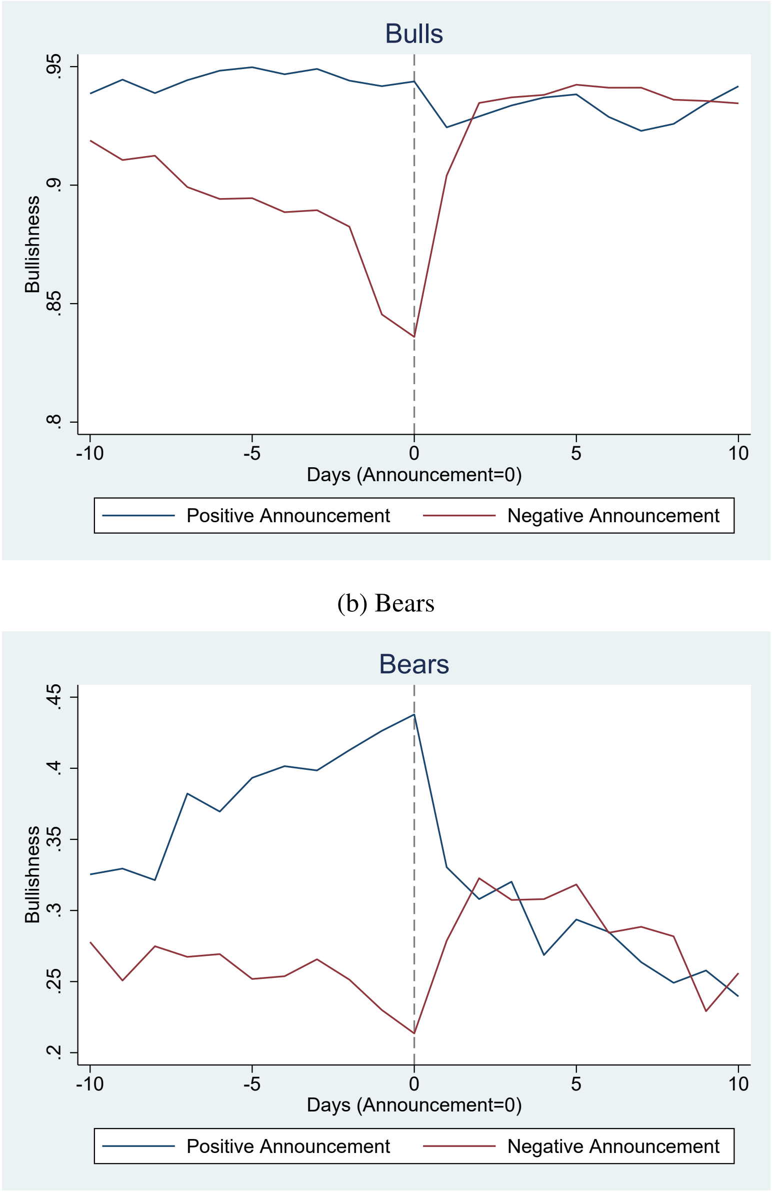
[13] We thank Marc-Aurèle Divernois and Damir Filipović for very generously sharing their data with us. Divernois and Filipovic (2022) study this data, showing that sentiment measured from StockTwits can be used to forecast stock returns on high-message volume days.

[14] Announcement days are defined as the first day that the stock can be traded after the announcement. For announcements that occur after the market close on a given day, the announcement day is thus coded as the next day.

[15] The sample is smaller than the overall scale of StockTwits for several reasons. First, we focus on messages about individual stocks, not indices like the S&P 500. Second, we restrict to attention to messages labeled as bullish or bearish by users. Third, we focus on windows around earnings announcements, which account for less than one-third of trading days. Finally, the requirement that the user has posted a message with a bullish/bearish label about the stock prior to the earnings announcement is restrictive.

[16] Average sentiment for bulls during these event windows is 0.94 with a standard deviation of 0.24. For bears, the corresponding numbers are 0.28 and 0.45.

Figure 2: Evolution of Beliefs Around Earnings Announcements


(a) Bulls

(b) Bears

converging back to 94-95% bullish. Within two days of the announcement, sentiment is the same, regardless of whether the announcement was positive or negative news. Figure 2b shows similar patterns for bears around positive news announcements. Sentiment first improves but then converges back to its original level.

These patterns are consistent with our theoretical results. Following a news announcement, users form networks with users of similar beliefs. Within those networks, they are exposed to interpretations of the data that make it less surprising. Thus, while users' initial reactions may push them away from their prior beliefs, the network will expose them to interpretations of the data that pull them back. Bulls about a stock become more bearish following a negative announcement, for example, but they return to being bullish once they are exposed to the best-fitting interpretations that evolve within the bullish network.

However, a key question remains: How do bulls and bears persistently disagree when bulls are likely exposed to at least *some* of the arguments of bears and vice-versa? The next section takes up this question.

4 Communication Across Networks and Issues

Can inaccurate beliefs and disagreement persist as people communicate across networks and issues? This section analyzes this question and presents two main results. First, we show that differences in beliefs can persist when members of one network only hear some arguments made by members of another network. In other words, ideological bubbles need not be hermetically sealed. So long as only some arguments are transmitted across networks, differences in beliefs can persist. Second, we show that communication across issues leads to cross-issue polarization.

4.1 Communication Across Networks

We first consider the impact of two different types of communication across networks. Person i is *weakly exposed to belief* $\tilde{\mu}$ if the set of models she is exposed to expands from M_i to $M_i \cup \{m(\tilde{\mu})\}$, where $m(\tilde{\mu})$ is a specific model that supports belief $\tilde{\mu}$. On the other hand, a person is *strongly exposed to belief* $\tilde{\mu}$ if the set of models she is exposed to expands from M_i to $M_i \cup M(\tilde{\mu})$, where $M(\tilde{\mu})$ is the set of all models that induce $\tilde{\mu}$. A person is *exposed to belief* $\tilde{\mu}$ if she is either weakly or strongly exposed to $\tilde{\mu}$. We think of weak exposure as capturing most communication across networks. For instance, a person who views evidence as suggesting that a new vaccine is safe is likely aware that there are people in "anti-vaccine" networks who believe otherwise. However, this person is likely only aware of a thin slice of anti-vaccine arguments.

We say a person is *persuaded* through exposure to belief $\tilde{\mu}$ if such exposure changes her final beliefs. Person i is more persuadable than person j if i is persuaded through exposure to belief $\tilde{\mu}$ whenever j is.

Proposition 3. *Suppose everyone is maximally open to persuasion, $M = M^a$.*

1. *Suppose person i is weakly exposed to a belief $\tilde{\mu}$ not represented in her network. Independent of her network and the alternative belief, the person need not be persuaded through this exposure: For every set of models M_i and belief $\tilde{\mu}$ not supported by any model in M_i, there exists positive measure of models $\tilde{m} = m(\tilde{\mu})$ supporting $\tilde{\mu}$ that fit less well than the best-fitting model in M_i.*

2. *Suppose person i is strongly exposed to a belief $\tilde{\mu}$ not represented in her network. Then the person is persuaded by this exposure if $\tilde{\mu}$ is closer to her prior, as measured by Movement$(\cdot; \mu_0)$, than any belief supported by a model in M_i.*

The first part of Proposition 3 implies that weak exposure to beliefs outside a person's network is never guaranteed to impact her beliefs. The second part of Proposition 3 implies that strong exposure to an alternative belief has at least as much impact on ultimate beliefs and behavior as weak exposure. While weak exposure to an alternative belief is never guaranteed to move final beliefs, strong exposure will move final beliefs whenever the alternative belief is closer to the person's prior than other beliefs represented in her network.

These results provide a simple way of understanding why different beliefs persist across networks, such as StockTwits, despite the fact that there is communication across networks. We think of cross-network communication as weak exposure. People might exchange both models and beliefs when interacting with others in the same network, while only exchanging beliefs (and perhaps a subset of models supporting those beliefs) when interacting with members of different networks. For instance, a person who believes a school shooting indicates the need for stricter gun-control measures is likely aware that there are others who conclude the opposite without being intimately familiar with all of their arguments. Proposition 3 says that weak exposure to anti-gun control arguments need not move the beliefs in the pro-gun control network. While a person could become convinced by listening to a broad set of arguments for a position, she is less likely to be convinced by a narrow subset of the arguments (or simply a statement of the position itself).[17] This result highlights a key difference between our framework and information-based approaches. In information-based approaches, if two networks start from the same prior and see the same data, they will end up with the same beliefs. In contrast, in our setting only strong exposure will tend to lead to convergence in beliefs.

[17] In highlighting the importance of the breadth of arguments a person is exposed to, our model relates to "persuasive-arguments theory" from psychology (e.g., Burnstein and Vinokur (1977) and also see Hüning et al. (2022) for recent related evidence). However, persuasive-arguments theory emphasizes the number of distinct arguments a person is exposed to, while we emphasize the compellingness of arguments (in terms of fit).

4.2 Communication Across Issues

It is of course not always the case that new networks form around every issue, even in the age of social media. In many instances, a network formed on one issue becomes a venue for discussing a second issue. For instance, a group formed based on a shared skepticism of vaccine safety might also discuss the merits of public education. Or, a group formed on the basis of a shared support for environmental regulation might also discuss the safety of genetically-modified foods. In this subsection, we consider how networks formed on shared beliefs about one issue influence beliefs on a second issue. We show that Proposition 2 has implications for *cross-issue* polarization.

Formally, consider an extension of the policy example where there are two issues: $\Omega = \Omega^1 \times \Omega^2$ and describe marginal beliefs over Ω^j by μ^j. For concreteness, let $\Omega^1 = \{l, r\}$ be whether a left- or right-leaning candidate governs better and $\Omega^2 = \{n, y\}$ be whether the economy will grow next quarter (y) or not (n). Networks are formed based on initial beliefs over $\{l, r\}$ but not $\{n, y\}$: $s(\mu)$ depends only on $\mu^{1'}$. Suppose that $\Omega^1$and Ω^2 are binary, with priors given by the following table:

μ_0	n	y
l	a	b
r	b	a

.

Suppose further that $\mu_0(l) = .5$ and, for $k \geq \mu_0(l)$, shared belief networks are formed based on whether $\mu^{1'}(l) \geq k$ (the left-leaning network), $\mu^{1'}(l) \in (1-k, k)$ (the centrist network), and $\mu^{1'}(l) \leq 1-k$ (the right-leaning network). Refer to this generalization of the multiple-issues policy example as the *multiple issues setup.*

Corollary 1. *In the multiple issues setup, shared-belief networks of intensity $k \geq \mu_0(l)/(\mu_0(l) + \mu_0(l, y))$ formed based on views on issue l versus r spill over to influence views of issue y versus n: The simple average between the lowest and highest potential values of $\mu_i^{\textit{left-leaning}}(y) - \mu_j^{\textit{right-leaning}}(y)$ across people i (in the left-leaning network) and j (in the right-leaning network) equals*

$$\begin{cases} \frac{1}{2} \times k \times (\mu_0(y|l) - \mu_0(y|r) + 3) - 1 & \textit{if } k < \frac{\mu_0(l)}{\mu_0(l)+\mu_0(r,y)} \\ k \times (\mu_0(y|l) - \mu_0(y|r)) & \textit{if } k \geq \frac{\mu_0(l)}{\mu_0(l)+\mu_0(r,y)} \end{cases}. \tag{1}$$

Moreover, for $k \geq \frac{1}{2\times\mu_0(y|l)}$, the lower bound of $\mu_i^{\textit{left-leaning}}(y) - \mu_j^{\textit{right-leaning}}(y)$ is greater than 0 when $\mu_0(y|l) > \mu_0(y|r)$.

Even though beliefs over the second issue do not influence network formation, final beliefs over that issue differ across members of the left-leaning and right-leaning networks. Corollary 1 (of Proposition 2) identifies two factors that influence belief polarization in views about y versus n, given by Equation (1). First, polarization is increasing in k, which could be interpreted as the degree to which networks are formed based on shared beliefs along issue l versus r—e.g., the extent to which people connect with others because of their views about that issue. Second, such polarization is increasing in the degree $(\mu_0(y|l) - \mu_0(y|r))$ to which issue y versus n is prior-connected to issue l versus r. For instance, since views on sports are not naturally connected with politics (i.e., $\mu_0(y|l) \approx \mu_0(y|r)$), views on which teams are most promising do not become more correlated with political leanings through social learning.[18] While there are potentially a range of beliefs within each network, for k sufficiently large and $\mu_0(y|l) > \mu_0(y|r)$, *every* member of the left-leaning network is more optimistic about y than all members of the right-leaning network.

To illustrate these results, consider the example from Section 3.1 where people interpret the data $h =$ "a natural disaster struck and GDP growth this quarter was low." In our two-issue extension, sharing models that suggest the left-leaning candidate is better at governing leads members of the network to also interpret the data as suggesting that the economy is likely to grow next quarter. Conversely, sharing models that suggest the right-leaning candidate is better at governing leads members of the network to interpret the same data as suggesting that the economy is unlikely to grow.[19] The right-leaning network's interpretation is that a natural disaster and low current growth are likely if the right-leaning candidate was better at governing (e.g., because the left-leaning party in power is incompetent). The left-leaning network's interpretation is that the data is likely if the left-leaning candidate is better at governing (e.g., because GDP growth could have been much worse in the face of a natural disaster). In other words, beliefs about the economy become "spurious implications" of beliefs about the candidate that governs better. Thus, shared belief networks lead to polarization and disagreement on issues beyond the issue driving network formation.

Consistent with this idea, there is sharp disagreement between Democrats and Republicans about economic conditions that has emerged over the last 10 years. For instance, Democrats report much higher consumer confidence before the 2016 election and after the 2020 election (when Democrats were president), while Republicans report much higher confidence between the two

[18] For $\mu_0(y|l) \approx \mu_0(y|r)$, $\mu_0(l, y) = a \approx b = \mu_0(r, y)$, so the simple average between the lowest and highest potential values of $\mu_i^{\text{left-leaning}}(y) - \mu_j^{\text{right-leaning}}(y)$ across people i (in the left-leaning network) and j (in the right-leaning network) is given by the second line of Equation (1), given the assumption that $k \geq \mu_0(l)/(\mu_0(l) + \mu_0(l, y))$.

[19] More formally, suppose $a = 0.125$ and $b = 0.375$ so the issues are prior connected. Members of the left-leaning network in this example adopt the model: $\pi_m^{\text{left-leaning}}(h|l, n) = \pi_m^{\text{left-leaning}}(h|l, y) = 1$ and $\pi_m^{\text{left-leaning}}(h|r, n) = \pi_m^{\text{left-leaning}}(h|r, y) = 1/3$. Members of the centrist network adopt the model: $\pi_m^{\text{centrist}}(h|l, n) = \pi_m^{\text{centrist}}(h|l, y) = 1$ and $\pi_m^{\text{centrist}}(h|r, n) = \pi_m^{\text{centrist}}(h|r, y) = 1$. Members of the right-leaning network in the example adopt the model: $\pi_m^{\text{right-leaning}}(h|l, n) = \pi_m^{\text{right-leaning}}(h|l, y) = 1/3$ and $\pi_m^{\text{right-leaning}}(h|r, n) = \pi_m^{\text{right-leaning}}(h|r, y) = 1$. Under these models, members of the left-leaning network in this example view the data as suggesting a high likelihood that the economy will grow next quarter: $\mu^{\text{left-leaning}}(y) = .5625 + .0625 = .625$. In contrast, members of the right-leaning network view this same data as suggesting a low likelihood: $\mu^{\text{right-leaning}}(y) = .1875 + .1875 = .375$. While in Corollary 1 that there are potentially many beliefs represented in each network, a natural refinement selects these model. Letting $\mu^{\text{left-leaning}}$ be a belief that minimizes $\max_{\omega \in \Omega} \mu_0(\omega)/\tilde{\mu}(\omega)$ among beliefs $\tilde{\mu}$ in $\arg\min_{\mu \in s(\mu_i')} \text{Movement}(\mu; \mu_0)$ and letting $\mu^{\text{right-leaning}}$ be defined analogously, then the particular models described at the beginning of this footnote are selected.

elections.[20] This pattern emerges in our framework if people discuss economic data within networks primarily formed based on partisan affiliation.

These results illustrate how networks based on one issue shape views on connected issues, perhaps shedding light on the so-called "polarization of reality" documented by Alesina et al. (2020). They show how the political left and right differ in their perceptions of, for example, the probability of upward social mobility. Our model suggests that such polarization is likely to occur along issues that the electorate believes are connected to whether the left or right governs better, as is naturally the case with social mobility since it connects to policy. Insofar as social media facilitates the formation of networks based on shared beliefs (increasing k), our model also suggests that it may play a role in increasing the polarization of reality along such issues. However, our model also suggests that such polarization is *un*likely to occur with issues that the electorate believes are *not* connected to whether the left or right is likely to govern better (e.g., sports).

5 Managing Communication

We now turn to the implications of our framework for how someone could try to manage communication to her advantage. We call this person a "communications manager."

5.1 Managing Communication Through Messaging

We first consider managing communication through messaging. The key result is that messaging is most effective as soon as data that is open to interpretation is released—that is, before social learning—since it may impact which network a person joins. In other words, our framework provides a reason for a communications manager to "get in front of the news" by providing a favored interpretation.

To see this, consider shared-belief networks. Imagine that before joining such a network, person i with belief μ_i' is weakly exposed to belief $\tilde{\mu} \notin s(\mu_i')$ with supporting model $m(\tilde{\mu})$. Following exposure to model $m(\tilde{\mu})$, the person potentially updates her beliefs and joins the shared-belief network associated with her posterior.

Proposition 4. *Suppose everyone is maximally open to persuasion, $M = M^a$, and is in a shared-belief network. Let μ_i denote a person's belief following social learning without being exposed to a belief $\tilde{\mu} \notin s(\mu_i')$, μ_i^e denote her belief following social learning after being exposed to belief $\tilde{\mu}$, and μ_i^p denote her belief following social learning when exposed to belief $\tilde{\mu}$ before social learning. If a person is persuaded through (weak or strong) exposure to $\tilde{\mu}$ after social learning, $\mu_i^e \neq \mu_i$,*

[20] See evidence on consumer confidence from the Michigan Survey of Consumers at this link.

then she is also persuaded through exposure to $\tilde{\mu}$ before social learning, $\mu_i^p \neq \mu_i$. However, the converse does not hold.

This result says that exposing people to messaging (i.e., models supporting an alternative belief) is more likely to impact which shared-belief network they join and hence their final beliefs if exposure comes before they exchange models with others. As we see from this result, the reason is simple: social learning hardens reactions to data, which inoculates people against finding models supporting alternative beliefs compelling. This may shed light on organizations' attempts to preemptively frame surprising news to avoid "losing control of the narrative." For example, firms often announce that a CEO unexpectedly resigned to "spend time with family." Firms also often have "culture training" for new employees before they socialize with existing employees. In our framework, these actions serve to provide early interpretations that imply favorable beliefs—e.g., the CEO's resignation does not imply the firm is in trouble—which in turn prevents people from hardening views within networks centered around less favorable beliefs. After social learning, people are less persuadable because their beliefs are supported by a better-fitting models. In other words, social learning makes people more certain that their interpretations of the data are correct and hence less open to other interpretations. A person who only talks to others who share the reaction that the latest school shooting indicates the need for stricter gun-control measures will become more confident in the rationale for drawing this conclusion from the data; a person who only talks to others who share the reaction that the shooting indicates the need for looser gun-control measures will similarly become more confident in drawing this conclusion from the data.[21]

5.2 Managing Communication by Influencing Networks

We next consider how a communications manager might act by directly influencing the network structure. For instance, she could hold meetings that invite a select group of people or form groups based on certain shared beliefs, experiences, or interests. She might also try to prevent certain groups from forming, actively trying to discourage people in one group from speaking to people in another. For example, a CEO might insist on being in all meetings with certain subordinates. A key point is that the communications manager often faces a tradeoff between inducing people to take a specific action and inducing people to agree on the same model.

[21] In a sense, this is consistent with Schkade et al. (2007), which found that after group interactions views on climate change, affirmative action, and civil unions became more homogeneous and more confident. Some studies on such "group polarization" find that beliefs also become "more extreme" after group interactions. Proposition 2 is consistent with those findings insofar as extremity is measured by confidence and inconsistent with those findings insofar as extremity is measured by how strongly beliefs react to data (if groups are formed based on shared beliefs and people have common priors). On this last point, Roux and Sobel (2015) shows how group polarization naturally arises in models of rational information aggregation.

5.2.1 Promoting Specific Actions

Suppose first that the communications manager wants to encourage people to take some action in response to the data. For example, in response to a school shooting, the manager might want to encourage interpretations that either support tightening gun control, the status quo, or loosening gun control. Or, following unexpectedly low earnings, a CEO may want to encourage followers to interpret the data as supporting cutting costs, staying the course, or investing in growth.

Formally, consider the case where each person has a finite action space and the communications manager's objective is a strictly monotonically increasing function of the fraction of people who choose her ideal action $a^s \in A$. How would the communications manager want to structure the network—i.e., the set of models M_i a given person i is exposed to—to maximize this objective?

Proposition 5. *Suppose each person has a finite action space and the communications manager's objective is a strictly monotonically increasing function of the fraction of people who choose her ideal action $a^s \in A$. The communications manager cannot do better than, for every person i, exposing her to all people who would choose a^s in the absence of social learning, and exposing her to nobody else: That is, the communications manager's objective is maximized by setting*

$$M_i = \left\{m \in \bar{M}(h, \mu_0, d, M) : a(\mu(h, m)) = a^s\right\} \tag{2}$$

for all i. The communications manager's objective continues to be maximized by adding to M_i specified in Eq. (2) *any model m with* $\Pr(h|m, \mu_0) < \max_{\tilde{m} \in M_i} \Pr(h|\tilde{m}, \mu_0)$*, but it is no longer maximized by adding a model m with* $\Pr(h|m, \mu_0) > \max_{\tilde{m} \in M_i} \Pr(h|\tilde{m}, \mu_0)$.

This result says that the communications manager wants to expose people to all models that support taking action a^s and no other models. That is, the communications manager wants to form a directed network where everybody listens to people who support action a^s and does not want people to hear good-fitting arguments supporting other actions. For example, a firm CEO may want to control interpretations of earnings announcements by disproportionately calling on bullish analysts in earnings calls (Cohen et al. (2020)).

To illustrate these results, take the public-policy example (Ex. 2) above with $\mu_0(l) = 1/2$ and a history h that under the default is perfectly diagnostic of the underlying state being l. Suppose the communications manager wants people to choose $a = L$. She should take individuals who would choose $a = R$ in the absence of communication and surround them with people who would choose $a = L$ in the absence of communication. For example, she should form networks where all people whose initial reactions are left-leaning ($\mu(l) \geq .75$) talk to each person whose initial reaction is right-leaning ($\mu(r) > .75$). In this case, the right-leaning people would end up believing $\mu(l) = .75$. A key implication of our framework concerns whom the communications manager most wants to silence: people who support the status quo—i.e., those with $\mu(l) \in (.25, .75)$. These people will have arguments that fit the data given priors very well and support inaction.

For example, suppose a school shooting could lead to a loosening or tightening of gun-control restrictions and the communications manager supports tighter gun control. The communications manager wants people arguing for tighter gun control to speak and everyone else to listen. The people the communications manager most wants to silence are moderates who argue for inaction, whether or not they are left- or right-leaning.[22]

5.2.2 Promoting Shared Models

By this logic, expanding people's networks could reduce polarization but also mute reactions to data that are open to interpretation. In the limit where the person is exposed to all possible models, the person will adopt a model that completely neutralizes the data: when data is open-to-interpretation and relevant for updating beliefs about ω under the true model, expanding a person's shared-belief network further untethers her beliefs from reality. These results speak to concerns about the increased connectedness between people generated by social media.

An implication of the above discussion is that promoting specific actions typically conflicts with promoting shared models. Because the status quo is favored by interpretations that fit the data perfectly and hence imply beliefs do not need to update, it can be hard to move people whose initial reaction is that the data supports the status quo. In contrast, if the communications manager simply wants people to end up with the same model—for instance, because she benefits when their actions are coordinated—then she should encourage open communication.

To see this formally, suppose the communications manager's objective is a strictly monotonically-increasing function of the fraction of people who share what ends up being the most popular model.

Proposition 6. *Suppose the communications manager's objective is a strictly monotonically-increasing function of the fraction of people who share what ends up to be the most popular model. The communications manager cannot do better than, for every person i, exposing her to all models: That is, the communications manager's objective is maximized by setting for all i*

$$M_i = \bar{M}(h, \mu_0, d, M). \tag{3}$$

This result says that if the goal is for everyone to end up sharing the same model, the communications manager cannot do better than encouraging everyone to talk to each other and share their

[22] Continuing this logic in a trivial dynamic extension, once all the people arguing for tighter gun control have spoken enough to harden beliefs, the communications manager is not worried about them having bilateral conversations with people favoring looser gun control—but they would still be wary of them having bilateral conversations with those who support the status quo.

models. When receivers are maximally open to persuasion, this means that the desire for everyone to end up sharing the same model will lead everyone to end up with interpretations that neutralize the data and promote the status quo. Note that this last point does not rely on the assumption that people hold common priors: When people are maximally open to persuasion with non-common priors, then it is still the case that the "h was inevitable" model is the only one that fits better than every other model represented in the population. And when everyone adopts this model, they stick with their prior beliefs and take the same action they would have taken in the absence of seeing the data h—i.e., they stick with their status-quo actions.

6 Applications

6.1 When and How to Hold a Meeting

Why do organizations hold so many meetings? Economic models typically assume meetings are fundamentally about information exchange: One worker holds a piece of information that another does not and exchanging information helps workers adapt to the environment and coordinate their actions (e.g., Dessein and Santos (2006)). Under this view, meetings are essentially no different from other communication technologies (e.g., emails) and are called when workers do not share the same information set. After meetings, workers all agree on the optimal action, which is better adapted to the full information set.

Organizational scholars view meetings much more broadly. They come in different forms, such as town halls or all hands. They are sometimes about information exchange, but they are also about diagnosing problems, communicating organizational priorities, and exchanging or amplifying views on the right course of action.

This section formalizes such a role for meetings, building on the view put forward in Weick (1995) that sensemaking is a fundamental activity of organizations. Following the logic of Section 5 costly meetings are called to help workers make sense of shared information. Meetings allow leaders to control interpretations workers share with each other, and they are called even when workers do not have any new private information. The key result that emerges is that the structure of meetings is not fixed but depends on workers' flow of communication outside meetings and how the organization prioritizes adapting to the environment versus coordinating among workers. In particular, leaders can find it optimal to use meetings to coordinate workers by muting their reactions to data, even if the leaders themselves interpret the data as suggesting the organization needs to better adapt to the environment.

We consider a similar setting to Dessein and Santos (2006) and Bolton et al. (2013), closely following the latter paper's language and formulation. The environment is parameterized by $\omega \in$

$[0, 1]$, which is not known by the leader or a continuum of followers. Instead, they have a uniform prior over ω and interpret data h in terms of what it implies about ω.

The timing of the game is: (1) everyone observes h, (2) the leader announces the organization's strategy $a_L \in [0, 1]$ and perhaps holds a meeting to discuss it in light of h, (3) each follower $i \in [0, 1]$ chooses an action $a_i \in [0, 1]$, and (4) payoffs are realized. Each follower i has payoff:

$$-\alpha \cdot (a_i - [l_i \cdot a_L + (1 - l_i) \cdot \omega])^2 - \kappa \int_j (a_j - \bar{a})^2 dj,$$

where $\alpha > 0$, $\kappa > 0$, $l_i \in [0, 1]$ and $\bar{a} \equiv \int a_j dj$. That is, each follower values (i) taking an action that is aligned with a weighted average of the organization's strategy a_L and the environment ω and (ii) coordinating with others. To limit the number of cases, assume that $l_i = 0$ for almost all followers and $l_i = 1$ for positive fraction $\varepsilon \to 0$ of followers.[23] That is, almost all followers care about taking an action that is well-adapted to the environment, rather than than taking an action that is aligned with the organization's strategy, and the rest of the followers mechanically follow the organization's strategy. Since it focuses on the case where $l_i = 0$ for fraction $(1 - \varepsilon) \approx 1$ of followers, the analysis better applies to situations where workers care more about getting things right than about following the leader. The leader's payoff simply aggregates the followers' payoffs:[24]

$$-\alpha \int_i (a_i - [l_i \cdot a_L + (1 - l_i) \cdot \omega])^2 di - \kappa \int_j (a_j - \bar{a})^2 dj.$$

The leader and followers share the same default model. While the leader is dogmatic the default is correct, followers may move away from it by sensemaking with fellow followers.

Because Ω in this example is the full unit interval, we for simplicity limit the set of models M followers could consider to be finite. We assume M always includes (i) the default model d, (ii) the best-fitting model m^{bf} that induces the same beliefs as d (i.e., $\mu(h, m^{bf}) = \mu(h, d)$), (iii) a model that says the history is inevitable in hindsight (i.e., a model m such that $\Pr(h|m, \mu_0) = 1$), and (iv) at least one model m with a fit between the default's and the best-fitting model's: $\Pr(h|m, \mu_0) \in (\Pr(h|d, \mu_0), \Pr(h|m^{bf}, \mu_0))$ and $\mu(h, m) \neq \mu(h, d)$. For simplicity, we also assume that m^{bf} fits better than all models in M except for the model that says the history is inevitable in hindsight.

If the leader does not hold a meeting, then workers make sense of h in their own networks.

[23] Having some followers mechanically follow the organization's strategy induces a cost to the leader of announcing a different strategy from what she thinks is subjectively optimal. There are other ways to generate such a cost, e.g., by assuming that followers and leaders value an organization that is well-adapted to its environment as Bolton et al. (2013) do. Our approach is analytically simple, but our qualitative results do not hinge on our precise formulation.

[24] For simplicity, we assume the leader evaluates her expected payoff according to her own expectation and not followers' subjective expectations. The leader has an incentive for followers' actions to be well-adapted to the leader's view of the environment, but does not directly care whether the followers believe their actions are well-adapted. Introducing the latter force could provide another reason to hold meetings in our framework: to get followers on board with the direction of the organization, even when getting followers on board does not influence their actions.

Holding a meeting costs the leader a positive amount c that is vanishingly small. By holding a meeting, the leader is able to perfectly control the set of models each worker is exposed to, M_i, by influencing the flow of communication between followers.

Proposition 7. *In the leader-follower example:*

1. *If information h is closed to interpretation or followers always stick with their default interpretation of the information absent persuasion ($\delta = 1$), the leader never holds a meeting. In this case, $a_L = \mathbb{E}_{\mu(h,d)}[\omega]$ for all h, and $a_i = a_L$ for all i.*

2. *Otherwise, the leader may hold a meeting.*

   (a) *If the weight placed on coordination (κ) is sufficiently large or if h is uninformative under the default model in the sense that $\mathbb{E}_{\mu(h,d)}[\omega] = \mathbb{E}_{\mu_0}[\omega] \equiv \omega_0$, then the leader calls a meeting whenever some followers take an action other than ω_0 absent a meeting. In this case (i) an optimal meeting features open communication ($M_i = M$ for all i), (ii) $a_L = \omega_0$, and (iii) $a_i = \omega_0$ for all i.*

   (b) *If the weight placed on adaptation (α) is sufficiently large and followers should react to the information under the default model in the sense that $\mathbb{E}_{\mu(h,d)}[\omega] \neq \mathbb{E}_{\mu_0}[\omega]$, then the leader calls a meeting whenever too many followers take an action other than $\mathbb{E}_{\mu(h,d)}[\omega]$ absent a meeting. In this case (i) an optimal meeting features directed communication with $M_i \neq M$, (ii) $a_L \neq \omega_0$, and (iii) not all followers take the same action.*

The first part of Proposition 7 says that, when data is closed to interpretation or followers do not try to make sense of the data on their own, then there is no need for the leader to call a meeting to discuss the organization's strategic response to publicly available data. The leader just announces her strategic response, which varies one-for-one with the leader's reaction to the data.

The second part of the proposition shows that the leader's reaction is very different when data is open to interpretation and followers try to make sense of it on their own. Meetings then allow leaders to better control interpretations followers share with each other. If the leader thinks followers are reacting to data when they should not be, or if the leader highly values coordination, then she calls a meeting which features open communication: everyone shares their view of what the event means for the organization. While opinions will be voiced that the leader does not agree with, at the end of the day everyone will share a view that the event teaches them little that they did not already know. Thus, the status quo will prevail. In this case, the leader's strategic response to publicly available data may be different than her private response: if she believes that she cannot persuade enough followers of her desired course of action, her best alternative is to ensure coordination by structuring the meeting to neutralize the data. This may be one reason why informal (e.g., relational) contracts are "hard to build *and change*" (emphasis added, Gibbons and Henderson (2012b)).

On the other hand, if too many followers are underreacting to the data or the leader strongly values adaptation, then the leader calls a meeting featuring a *persuasive campaign.* The leader ensures that the loudest voices are those with interpretations consistent with her view of the optimal action $\mathbb{E}_{\mu(h,d)}[\omega]$. While not everyone ends up on board with the shift in strategy from the status quo ω_0, as many as possible will be on board. Per Proposition 4 there is also a motive to hold the meeting as soon as possible, before workers can share interpretations with each other on their own.

6.2 The Evolution and Spread of Misconceptions Through Networks

Why do people believe in misconceptions (e.g., GMOs and vaccines are dangerous) and conspiracy theories (e.g., QAnon) when the Internet and social media also give them access to high-quality information? Echo chambers are a common answer to this question. While people have access to high-quality information, their media diets and social networks only expose them to misinformation and falsehoods. Under this view, falsehoods spread like viruses and crowd out the truth. People hear the same falsehood repeatedly and perhaps then overweight it.

An emerging literature suggests that this echo-chamber view is incomplete. Guess et al. (2018) argue that most Americans have diverse media diets, and that social media like Twitter tend to increase the diversity of viewpoints that people are exposed to. Similarly, Bertrand and Kamenica (2020) find that while social attitudes have become stronger predictors of political ideology over time, they have not become stronger predictors of media diet. In addition, Boxell et al. (2017, 2020) find that while political polarization is increasing, it is not increasing faster for people who extensively use the Internet and social media. Thus, while echo chambers could be a concern, they may not be as widespread a problem as conventional wisdom portrays. The persistence of disagreement remains a puzzle not fully explained by echo chambers.

Our framework offers a different explanation, highlighting the difference between interpretations and information. Within a network, people are exposed to crowdsourced models that evolve to fit the data better and better, which makes them more certain their interpretation of the data is correct. Insofar as social media and the Internet make it easier to form shared-belief networks, our framework predicts that their primary impact will be to make people's beliefs resistant to change.

Proposition 8. *Suppose person i and j hold the same beliefs μ. If person i formed those beliefs through social learning in a shared-belief network and j formed those beliefs in some other way, then person j is (weakly) more persuadable than person i.*

Proposition 8 says that shared-belief networks inoculate people against finding alternative beliefs compelling. Combined with the earlier result that networks formed based on shared beliefs mute reactions (Proposition 2), this implies that shared-belief networks cause beliefs to be persistently untethered from data that is open to interpretation. To make an analogy to viruses, networks lead interpretations to "mutate" to achieve better fit within the network—and people are exposed to more "variants" within than across networks.[25]

While we could illustrate these results by applying the baseline model we presented above, it is more revealing to consider a simple two-period dynamic extension of the analysis under the assumption that everyone is maximally open to persuasion. The key idea is that if networks form endogenously in response to one set of information, those networks will tend to encourage different interpretations of all future information. In other words, network formation based on shared-prior beliefs creates strong path dependence in the way people interpret information.

Formally, suppose people begin with the same priors, react to data h_1, and form shared belief networks based on their reactions to h_1. Further suppose that after exchanging models through the network, people's posterior beliefs after interpreting h_1 become their priors in interpreting new data h_2. In interpreting h_2, people share models with others in the shared-belief network that was formed based on common reactions to earlier data h_1. That is, networks are sticky across the two periods: people stay in the shared-belief network that was formed in period 1. For example, people may talk to others who share a similar reaction to evidence purporting to show a relationship between vaccines and autism and continue to talk to the same people when new data arrives.

The key result from this dynamic extension is that networks have lasting consequences on how people interpret subsequent events. By Proposition 2, everyone within a given shared-belief network ends up holding the initial belief closest to the prior within that network in response to data h_1. So everyone within a shared-belief network begins with the same prior entering into the second period where they interpret data h_2. Call this prior belief μ_1^s, which differs across networks s. Since people use the same network to exchange interpretations of h_2, social learning maximally mutes and hardens a person's reaction to the data. In other words, everyone ends up at the belief they held prior to seeing h_2 with a model that perfectly explains the data: for every person i in shared belief network s, $\mu_i = \mu_1^s$ and $\Pr(h_2|m_i, \mu_1^s) = 1$.

This analysis suggests that once misconceptions evolve and harden within a network through crowdsourced interpretations of a high-profile event, members of that network explain subsequent events in a way that makes them consistent with the original interpretation. In other words, a bad

[25] Bowen et al. (2021) provide an alternative model where belief polarization is driven by misperceptions about selective sharing of second-hand information within an echo chamber. In Bowen et al. (2021), disagreement and polarization are driven by different people holding different information (having heterogeneous "information diets" of second-hand information) and not properly accounting for that fact; in our model, disagreement arises even when people share the same information. Their framework sheds light on situations where a lot of news is coming out each day and it is hard to keep track of it all (e.g., if there is a war or people are forming beliefs about a new political candidate). We shed light on situations where the basic facts are essentially common knowledge and people are primarily exchanging interpretations of those facts.

take on an event can be very hard to reverse: Once a person "goes down the rabbit hole", it is difficult to use new information or interpretations to convince them to come out. Indeed, recent papers (Nyhan et al. (2023); Guess et al. (2023)) find that temporarily (e.g., around elections) reducing people's exposure to like-minded content and increasing their exposure to counterattitudinal content does not appreciably impact their beliefs. In our framework, this resistance to change emerges because before the intervention the network already exposed people to interpretations that fit their prior knowledge well.

7 Discussion

This paper is a first step to studying the social transmission of models. There are several potential avenues for future work. For instance, while we assume people costlessly exchange models with others, in many cases people devote effort, attention, and time to exposing themselves to new models for reasons of curiosity, identity, and instrumentality. How does incorporating a realistic demand function for models influence, for example, the way networks are structured?

We also show the importance of people's prior beliefs for how they initially react to the data, form networks with others, and ultimately react after exchanging interpretations. But, other than in our dynamic extension in Section 6.2, we say little about where these priors come from. They could come from motivated reasoning, a desire to foster relationships with family and friends, etc. What we do shed light on is why beliefs often appear stable in the face of contradictory, but open-to-interpretation, data. And we make the novel (to our knowledge) prediction that initial beliefs will respond to such data before reverting towards the prior.

Much of our analysis focuses on shared-belief networks because they capture important features of common networks. But sometimes networks are instead formed based on shared models, rather than shared beliefs. Astrologers consider the movement of celestial bodies in making sense of what happened yesterday. Closer to earth, some communities of venture capitalists primarily evaluate startups based on attributes of their products, while others focus on attributes of their founders. In finance, there are contrarians and trend followers. Some political analysts focus on economic fundamentals in predicting election outcomes, while others focus on polls. How do networks shape views in such cases?

Appendix B.3 analyzes a special class of shared models based on shared inflexibility: people may be commonly dogmatic on how to interpret certain types of information. This may arise from shared expertise, shared beliefs about what sort of data is uninformative, shared trust in taking some data at face value, or even a shared convention that some discussions are taboo. The key result is that, supposing data is maximally open to persuasion, people will only end up reacting to the data that they are inflexible in interpreting. Quantitative analysts will talk to other quantitative analysts about how to interpret qualitative information and end up agreeing that, while it initially seemed relevant, it is not useful. Conversely, qualitative analysts will talk to other qualitative analysts about how to interpret quantitative information and end up agreeing that, while it initially seemed relevant, it is not useful. Similarly, people on the left will end up adopting models that neutralize data communicated by right-leaning outlets, and similarly for people on the right. When networks are based on shared models, social learning may intensify some opinions in addition to hardening them.

The framework also admits further applications. For example, if a manager wants to organize teams to help her arrive at a realistic interpretation of the data, how would she do it? Would she like to construct teams who tend to reach similar conclusions (i.e., shared-belief networks)? Or teams who look at the data in similar ways (i.e., shared-model networks, analyzed in Appendix B.3)? A loose intuition reminiscent of Hong and Page (2001) that arises from our framework is that a manager may benefit from aggregating across teams that have different ways of looking at the data (i.e., different models). A manager is less likely to benefit from trying to aggregate across teams that are systematically trying to come to different *conclusions* from the data. For instance, in the venture capital context, it may be helpful to have people who focus on management team experience and people who focus on current profits. It is unlikely to be helpful to have people who always want to invest and people who never want to invest, each of whom comes up with the interpretation of the data that best supports their (pre-specified) conclusion.

A Proofs

Proof of Proposition 1. When (r, d) is realized, any model in M must satisfy $\Pr(r, d|m', \mu_0) \leq \Pr(r|m', \mu_0) = \pi_1(r)$, where the inequality comes from the logic of probability and the equality from m' being in M. This means that if there's a model $m' \in M$ that satisfies $\Pr(r, d|m', \mu_0) = \pi_1(r)$, then this model is a best-fitting model in M given h and μ_0 and is a model that could be adopted by everyone in a complete network. The following is such a model:

$$\begin{aligned}
\pi_{m'}(r, d|\omega) &= \pi_1(r)\ \forall\omega \\
\pi_{m'}(r, u|\omega) &= 0\ \forall\omega \\
\pi_{m'}(l, d|\omega) &= \pi_2(d) - \pi_1(r)\ \forall\omega \\
\pi_{m'}(l, u|\omega) &= 1 - \pi_2(d)\ \forall\omega.
\end{aligned}$$

It is indeed the case that $m' \in M$: $\pi_{m'}(r, d) + \pi_{m'}(r, u) = \pi_1(r)$ and $\pi_{m'}(r, d) + \pi_{m'}(l, d) = \pi_2(d)$. It is also the case that $\Pr(r, d|m', \mu_0) = \pi_1(r)$. In addition, posterior beliefs under m' equal prior beliefs because likelihoods under m' are independent of ω. And, under m', $\Pr(d|r, m', \mu_0) - \Pr(d|l, m', \mu_0) = \frac{1-\pi_2(d)}{1-\pi_1(r)}$.

Now that we've established there is such a best-fitting model, we know that any best-fitting model m' must share the features that posterior beliefs under m' equal prior beliefs and $\Pr(d|r, m', \mu_0) - \Pr(d|l, m', \mu_0) = \frac{1-\pi_2(d)}{1-\pi_1(r)}$. Indeed, any best-fitting model must satisfy $\Pr(r, d|m', \mu_0) = \pi_1(r)$ and then $\Pr(r, u|m', \mu_0) = 0$ to keep $m' \in M$. For $m' \in M$, it must also be true that $\Pr(d|l, m', \mu_0) \cdot (1 - \pi_1(r)) + \pi_1(r) = \pi_2(d)$, which implies that $\Pr(d|l, m', \mu_0) = \frac{\pi_2(d)-\pi_1(r)}{1-\pi_1(r)}$. And, under any such model m' that has the property that $\pi_{m'}(r|b) = \pi_{m'}(r|g)$, then $\pi_{m'}(r, d|\omega) = \pi_1(r)$ for all ω, so posterior beliefs under m' equal prior beliefs given $h = (r, d)$.

It remains to show that

$$\max_{m \in M} \Pr(d|r, m, \mu_0) - \Pr(d|l, m, \mu_0) = \frac{1 - \pi_2(d)}{1 - \pi_1(r)}. \tag{4}$$

By Bayes' rule, $\Pr(d|m, \mu_0) = \Pr(d|r, m, \mu_0) \cdot \Pr(r|m, \mu_0) + \Pr(d|l, m, \mu_0) \cdot \Pr(l|m, \mu_0)$. For $m \in M$, this means that $\pi_2(d) = \Pr(d|r, m, \mu_0) \cdot \pi_1(r) + \Pr(d|l, m, \mu_0) \cdot \pi_1(l)$. Substituting $\pi_1(l) = (1 - \pi_1(r))$ and re-arranging gives

$$\Pr(d|r, m, \mu_0) - \Pr(d|l, m, \mu_0) = \frac{\pi_2(d) - \Pr(d|l, m, \mu_0)}{\pi_1(r)}. \tag{5}$$

We also have that

$$\begin{aligned}
\Pr(d|l, m, \mu_0) &= \frac{\Pr(d, l|m, \mu_0)}{\pi_1(l)} \text{ (by Bayes' rule)} \\
&= \frac{\pi_2(d) - \Pr(d, r|m, \mu_0)}{\pi_1(l)} \text{ (b/c } m \in M) \\
&\geq \frac{\pi_2(d) - \pi_1(r)}{\pi_1(l)} \text{ (b/c, as argued above, } \Pr(d, r|m, \mu_0) \leq \pi_1(r)). \qquad (6)
\end{aligned}$$

Inequality (6) and equality (5) together imply equation (4).

□

Proof of Proposition 2. Consider an arbitrary person i and let

$$\text{MovementMin}_i \equiv \arg\min_{\mu \in s(\mu_i')} \text{Movement}(\mu; \mu_0).$$

Someone in i's network will propose a model $\tilde{m}$ that maximizes $\Pr(h|\cdot, \mu_0)$ subject to $\mu(h, \tilde{m}) \in \text{MovementMin}_i$. By Lemma 1, this model will fit strictly better than all models represented in i's network that imply beliefs outside of MovementMin_i, so everyone in i's network will adopt models that imply beliefs in MovementMin_i.

□

Proof of Proposition 3.

1. For every $\tilde{\mu}$, there exists a positive measure of models $m(\tilde{\mu})$ supporting that belief that are less compelling than the model m_i a person would adopt prior to weak exposure to that belief: for example, take models

   $$\pi_{m(\tilde{\mu})}(h|\omega) = \frac{\tilde{\mu}(\omega)}{\mu_0(\omega)} \cdot (\Pr(h|m_i, \mu_0) - \varepsilon)$$

   for all $\omega \in \Omega$ and for $\varepsilon > 0$ small.

2. When $\tilde{\mu}$ is closer to the person's prior, as measured by $\text{Movement}(\cdot; \mu_0)$, than any belief supported by a model in M_i, then the best-fitting model supporting $\tilde{\mu}$ fits better than any model in M_i (by Lemma 1).

□

Proof of Corollary 1. Let

$\mu^{\text{left-leaning}}$	n	y
l	a^L	b^L
r	c^L	d^L

denote a movement-minimizing belief of members of the left-leaning network and

$\mu^{\text{right-leaning}}$	n	y
l	a^R	b^R
r	c^R	d^R

denote a movement-minimizing belief among members of the right-leaning network.

For the left-leaning network, $\mu^{\text{left-leaning}}(l) = a^L + b^L \geq k$. Assuming, as we will later establish, that $\text{Movement}(\mu^{\text{left-leaning}}; \mu_0) = \max\left\{a^L/a, b^L/b\right\}$, then the inequality must bind, so $b^L = k - a^L$. This then implies that $\text{Movement}(\mu^{\text{left-leaning}}; \mu_0) = \max\left\{a^L/a, (k - a^L)/b\right\}$, which will be minimized when $a^L/a = (k - a^L)/b \Rightarrow a^L = k \times a/(a+b) = k \times \mu_0(n|l)$ and $b^L = k \times b/(a+b) = k \times \mu_0(y|l)$.

Summarizing what we know so far:

$$\begin{array}{|c|c|c|}\hline \mu^{\text{left-leaning}} & n & y \\ \hline l & k \times \frac{a}{a+b} & k \times \frac{b}{a+b} \\ \hline r & c^L & d^L \\ \hline \end{array} = \begin{array}{|c|c|c|}\hline \mu^{\text{left-leaning}} & n & y \\ \hline l & k \times \mu_0(n|l) & k \times \mu_0(y|l) \\ \hline r & c^L & d^L \\ \hline \end{array}$$

and we can similarly establish that

$$\begin{array}{|c|c|c|}\hline \mu^{\text{right-leaning}} & n & y \\ \hline l & a^R & b^R \\ \hline r & k \times \frac{c}{c+d} & k \times \frac{d}{c+d} \\ \hline \end{array} = \begin{array}{|c|c|c|}\hline \mu^{\text{right-leaning}} & n & y \\ \hline l & a^R & b^R \\ \hline r & k \times \mu_0(n|r) & k \times \mu_0(y|r) \\ \hline \end{array}.$$

Proceeding further, c^L and d^L must satisfy the following constraints if the premise above is true that Movement($\mu^{\text{left-leaning}}; \mu_0) = \max\left\{a^L/a, b^L/b\right\}$:

$$\begin{aligned} c^L + d^L = (1-k) \Rightarrow d^L = (1-k) - c^L, &\text{ and} \\ \max\left\{c^L/c, d^L/d\right\} = \max\left\{\frac{c^L}{c}, \frac{(1-k)-c^L}{d}\right\} &\leq \max\left\{a^L/a, (k-a^L)/b\right\} = \frac{k}{\mu_0(l)}, \end{aligned}$$

where the last equality follows from plugging in a^L from the earlier calculations and, here, $c = \mu_0(r,n), d = \mu_0(r,y)$. The range of c^L that satisfy these conditions is given by

$$\max\left\{0, (1-k) - k \times \frac{\mu_0(r,y)}{\mu_0(l)}\right\} \leq c^L \leq \min\left\{1-k, k \times \frac{\mu_0(r,n)}{\mu_0(l)}\right\},$$

which is non-empty given the assumptions that $k \geq \mu_0(l) = 1/2$. This last fact also establishes the premise that Movement($\mu^{\text{left-leaning}}; \mu_0) = \max\left\{a^L/a, b^L/b\right\}$.

We can similarly derive that

$$\begin{aligned} \max\left\{0, (1-k) - k \times \frac{\mu_0(r,n)}{\mu_0(l)}\right\} &\leq d^L \leq \min\left\{1-k, k \times \frac{\mu_0(r,y)}{\mu_0(l)}\right\} \\ \max\left\{0, (1-k) - k \times \frac{\mu_0(l,y)}{\mu_0(r)}\right\} &\leq a^R \leq \min\left\{1-k, k \times \frac{\mu_0(l,n)}{\mu_0(r)}\right\} \\ \max\left\{0, (1-k) - k \times \frac{\mu_0(l,n)}{\mu_0(r)}\right\} &\leq b^R \leq \min\left\{1-k, k \times \frac{\mu_0(l,y)}{\mu_0(r)}\right\}. \end{aligned}$$

Taken together,

$$
\begin{aligned}
\mu^{\text{left-leaning}}(y) &\geq \max\left\{k \times \mu_0(y|l), k \times \mu_0(y|l) + (1-k) - k \times \frac{\mu_0(r,n)}{\mu_0(l)}\right\} \\
\mu^{\text{left-leaning}}(y) &\leq \min\left\{k \times \mu_0(y|l) + 1 - k, k \times \mu_0(y|l) + k \times \frac{\mu_0(r,y)}{\mu_0(l)}\right\} \\
\mu^{\text{center-leaning}}(y) &= \mu_0(y) \\
\mu^{\text{right-leaning}}(y) &\geq \max\left\{k \times \mu_0(y|r), k \times \mu_0(y|r) + (1-k) - k \times \frac{\mu_0(l,n)}{\mu_0(r)}\right\} \\
\mu^{\text{right-leaning}}(y) &\leq \min\left\{k \times \mu_0(y|r) + 1 - k, k \times \mu_0(y|r) + k \times \frac{\mu_0(l,y)}{\mu_0(r)}\right\}.
\end{aligned}
$$

This then means that

$$
\begin{aligned}
\mu^{\text{left-leaning}}(y) - \mu^{\text{right-leaning}}(y) \geq \max&\left\{k \times \mu_0(y|l), k \times \mu_0(y|l) + (1-k) - k \times \frac{\mu_0(r,n)}{\mu_0(l)}\right\} \\
- \min&\left\{k \times \mu_0(y|r) + 1 - k, k \times \mu_0(y|r) + k \times \frac{\mu_0(l,y)}{\mu_0(r)}\right\}.
\end{aligned}
$$

Assuming that $\mu_0(l) = 1/2$ and $\mu_0(l,y) = \mu_0(r,n) \Rightarrow \mu_0(y|l) = \mu_0(n|r)$, then this inequality reduces to

$$
\mu^{\text{left-leaning}}(y) - \mu^{\text{right-leaning}}(y) \geq \begin{cases} 2k \times \mu_0(y|l) - 1 & \text{if } k \geq \frac{\mu_0(r)}{\mu_0(r)+\mu_0(l,y)} \\ (1-2k) & \text{if } k < \frac{\mu_0(r)}{\mu_0(r)+\mu_0(l,y)}. \end{cases}
$$

We also have

$$
\begin{aligned}
\mu^{\text{left-leaning}}(y) - \mu^{\text{right-leaning}}(y) \leq \min&\left\{k \times \mu_0(y|l) + 1 - k, k \times \mu_0(y|l) + k \times \frac{\mu_0(r,y)}{\mu_0(l)}\right\} \\
- \max&\left\{k \times \mu_0(y|r), k \times \mu_0(y|r) + (1-k) - k \times \frac{\mu_0(l,n)}{\mu_0(r)}\right\}.
\end{aligned}
$$

Assuming that $\mu_0(l) = 1/2$ and $\mu_0(l,y) = \mu_0(r,n) \Rightarrow \mu_0(y|l) = \mu_0(n|r)$, then this inequality reduces to

$$
\mu^{\text{left-leaning}}(y) - \mu^{\text{right-leaning}}(y) \leq \begin{cases} 2k \times (\mu_0(y|l) - 1) + 1 & \text{if } k \geq \frac{\mu_0(l)}{\mu_0(l)+\mu_0(r,y)} \\ k \times (1 + \mu_0(y|l) + \mu_0(y|r)) - 1 & \text{if } k < \frac{\mu_0(l)}{\mu_0(l)+\mu_0(r,y)}. \end{cases}
$$

If $k \geq \frac{\mu_0(l)}{\mu_0(l)+\mu_0(r,y)}$, then the middle point between the lower and upper bound of $\mu^{\text{left-leaning}}(y) - \mu^{\text{right-leaning}}(y)$equals

$$
k \times (2 \times \mu_0(y|l) - 1) = k \times (\mu_0(y|l) - \mu_0(y|r)),
$$

while for $\frac{\mu_0(r)}{\mu_0(r)+\mu_0(l,y)} \leq k \leq \frac{\mu_0(l)}{\mu_0(l)+\mu_0(r,y)}$ this middle point equals

$$k \times (\mu_0(y|l) + 1) - 1 = \frac{1}{2} \times k \times (\mu_0(y|l) - \mu_0(y|r) + 3) - 1.$$

Moreover, for $k \geq \frac{1}{2\times\mu_0(y|l)} \geq \frac{\mu_0(l)}{\mu_0(l)+\mu_0(l,y)}$, simple algebra reveals that the lower bound of $\mu^{\text{left-leaning}}(y) - \mu^{\text{right-leaning}}(y)$ is greater than 0.

□

Proof of Proposition 4. Weak exposure to belief $\tilde{\mu}$ prior to social learning impacts the person's final beliefs if and only if the person finds $m(\tilde{\mu})$ more compelling than the model m'_i she currently has in mind supporting belief μ'_i: that is, if and only if

$$\Pr(h|m(\tilde{\mu}), \mu_0) > \Pr(h|m'_i, \mu_0). \tag{7}$$

Weak exposure to belief $\tilde{\mu}$ following social learning impacts the person's final beliefs if and only if the person finds $m(\tilde{\mu})$ more compelling than the best-fitting model among those represented in shared-belief network $s(\mu'_i)$: that is, if and only if

$$\Pr(h|m(\tilde{\mu}), \mu_0) > \max_{m' \in \bigcup_{\mu \in s(\mu'_i)} M(\mu)} \Pr(h|m', \mu_0). \tag{8}$$

The result follows from the right-hand-side of inequality (8) being larger than the right-hand-side of inequality (7).

A similar proof applies to the case of strong exposure to beliefs, replacing the left-hand-sides of inequalities (7) and (8) with $\max_{m' \in M(\tilde{\mu})} \Pr(h|m', \mu_0)$. □

Proof of Proposition 5. The communication manager's objective is clearly maximized by exposing everybody to the best-fitting model that supports action a^s and exposing them to no other models. The communication manager does no worse by exposing people to all models specified in Eq. (2) (i.e., all models that support action a^s), since this includes the best-fitting one and no models that support other actions. That is, everybody's behavior is the same whether they are only exposed to the best-fitting model that supports a^s or models specified in Eq. (2). This remains true if we add to models specified in (2) any model m with $\Pr(h|m, \mu_0) < \max_{\tilde{m} \in M_i} \Pr(h|\tilde{m}, \mu_0)$, since nobody will adopt such a model. However, the communication manager's payoff is strictly worse if we add to models specified in (2) any model m with $\Pr(h|m, \mu_0) > \max_{\tilde{m} \in M_i} \Pr(h|\tilde{m}, \mu_0)$, since anybody who would've adopted a model in M_i will instead adopt this model which supports taking an action other than a^s.

□

Proof of Proposition 6. If everybody is exposed to $\bar{M}(h, \mu_0, d, M)$, then everybody will also end up adopting the model in that set that maximizes $\Pr(h|\cdot, \mu_0)$. The communication manager cannot do better, since everyone will end up sharing the same model.

□

Proof of Proposition 7. For the first case, it's obvious that the leader never holds a meeting because holding a meeting costs $c > 0$ and does not influence beliefs and decisions when information is closed to interpretation or when followers always stick with their default interpretation of the information absent persuasion. Since $a_L = \mathbb{E}_{\mu(h,d)}[\omega]$ implies $a_i = a_L$ for all i (this is obvious for followers who blindly follow a_L and other followers set $a_i = l \cdot a_L + (1-l) \cdot \mathbb{E}_{\mu(h,d)}[\omega] = a_L$), it remains to show in this case that $a_L = \mathbb{E}_{\mu(h,d)}[\omega]$. Setting $a_L = \mathbb{E}_{\mu(h,d)}[\omega]$ uniquely maximizes the coordination term, $-\int_j (a_j - \bar{a})dj$, of the leader's payoff since everyone coordinates on a_L. Since simple algebra shows that a_L doesn't influence the adaptation term, $\int_i -(a_i - [l_i \cdot a_L + (1-l_i) \cdot \omega])^2 di$, it is optimal for the leader to set $a_L = \mathbb{E}_{\mu(h,d)}[\omega]$.

For the first part of the second case, optimizing the leader's payoff becomes equivalent to maximizing the coordination term, $\int_j (a_j - \bar{a})^2 dj$, when the weight placed on coordination κ is sufficiently large. Given that a positive fraction of followers initially adopt the perfectly-fitting neutralizing model, the only way for all followers to perfectly coordinate their actions is for them all to take $a_i = \omega_0$. This is implemented by followers being exposed to all models, either with open communication absent a meeting or with open communication in a meeting. This is also optimal from the point of view of the leader when h is uninformative under the default model in the sense that $\mathbb{E}_{\mu(h,d)}[\omega] = \omega_0$. The leader does better by holding a meeting than not whenever some followers would adopt a model that implies a belief other than μ_0 absent a meeting.

For the last part, if followers are exposed to all models ($M_i = M$ for all i), then they perfectly coordinate their actions and the leader's payoff approximately equals

$$-\alpha \mathbb{E}_{\mu(h,d)} \int_i (a_i - \omega)^2 di = -\alpha \mathbb{E}_{\mu(h,d)} \int_i (\omega_0 - \omega)^2 di, \tag{9}$$

since $l_i = 0$ for almost all followers. If followers are instead all exposed to only models supporting $a_i = \mathbb{E}_{\mu(h,d)}[\omega]$ (i.e., $M_i = \left\{d, m^{bf}\right\}$ for all i), then the leader's payoff approximately equals

$$-\alpha \left[\mathbb{E}_{\mu(h,d)} \rho \int_i (\mathbb{E}_{\mu(h,d)}[\omega] - \omega)^2 di + (1-\rho) \int_i (\omega_0 - \omega)^2 di \right] - \kappa \int_j (a_j - \rho \mathbb{E}_{\mu(h,d)}[\omega] - (1-\rho)\omega_0)^2 dj, \tag{10}$$

where ρ equals the fraction of followers who are persuadable by m^{bf} (i.e., fraction $1-\rho$ are the fraction with the initial reaction to adopt the perfectly-fitting neutralizing model). Since the first term of (10) is larger than (9) when $\mathbb{E}_{\mu(h,d)}[\omega] \neq \omega_0$, in this case the leader holds a meeting that features directed communication whenever α is sufficiently large. Such a meeting will clearly be better than not holding a meeting whenever followers whose initial reaction to the data differs from $\mu(h,d)$ are not exposed to m^{bf} absent a meeting or are exposed to the model that says the history is inevitable in hindsight.[26]

□

[26] To see when else the leader wants to hold such a meeting, (10) minus (9) equals:

$$-\alpha\rho \mathbb{E}_{\mu(h,d)} \left[(\mathbb{E}_{\mu(h,d)}[\omega] - \omega)^2 - (\omega_0 - \omega)^2 \right] - \kappa \left[\int_j (a_j - \rho \mathbb{E}_{\mu(h,d)}[\omega] - (1-\rho)\omega_0)^2 dj \right],$$

which, after some algebra, equals $\alpha\rho(\mathbb{E}_{\mu(h,d)}[\omega] - \omega_0)^2 - \kappa\rho(1-\rho)(\mathbb{E}_{\mu(h,d)}[\omega] - \omega_0)^2$. So a meeting featuring directed communication is optimal whenever $\alpha > \kappa(1-\rho)$. This reveals that a leader is more likely to call a meeting to encourage followers to take an action different from ω_0 the greater the fraction of followers who are persuadable to take such an action—that is, the smaller the fraction of followers who, prior to the meeting, are hardened in their views that the data tells them little they didn't already know.

Proof of Proposition 8. If person i forms her beliefs μ in a shared-belief network, then she adopts the best fitting model in M that supports those beliefs. The fit of person j's model supporting those beliefs must then fit weakly less well than person i's, which makes her weakly more persuadable. □

Appendices for Online Publication

B Model Details

B.1 Initial Reactions

Let $\bar{\Delta}(h, \mu_0, d, M) = \{\mu \in \Delta(\Omega) : \mu = \mu(h, m)$ for some $m \in \bar{M}(h, \mu_0, d, M)\}$ denote the set of initial beliefs in reaction to the data. By assumption, fraction δ of the population sticks with the default and holds beliefs $\mu(h, d)$ and fraction $(1 - \delta)$ holds beliefs in $\bar{\Delta}(h, \mu_0, d, M)$.

Proposition 9. *The set of initial beliefs in reaction to the data is a subset of*

$$\bar{\bar{\Delta}}(h, \mu_0, d, M) = \left\{\mu \in \Delta(\Omega) : \mu(\omega) \leq \frac{\mu_0(\omega)}{\Pr(h|d, \mu_0)} \; \forall \, \omega \in \Omega\right\}.$$

Further, when people are maximally open to persuasion given the data, $M = M^a$, we have $\bar{\Delta}(h, \mu_0, d, M) = \bar{\bar{\Delta}}(h, \mu_0, d, M)$.

Proof of Proposition 9. This proof is essentially the same as the proof of Proposition 1 in Schwartzstein and Sunderam (2021). We repeat it here for completeness.

Note that

$$\mu(\omega|h, m) = \frac{\pi_m(h|\omega) \cdot \mu_0(\omega)}{\Pr(h|m, \mu_0)}$$

by Bayes' Rule. Since $\pi_m(h|\omega) \leq 1$ and, by definition of $\bar{M}(h, \mu_0, d, M)$, $\Pr(h|m, \mu_0) \geq \Pr(h|d, \mu_0)$ for all $m \in \bar{M}(h, \mu_0, d, M)$, beliefs that do not lie in $\bar{\bar{\Delta}}(h, \mu_0, d, M)$ cannot be included in $\bar{\Delta}(h, \mu_0, d, M)$. To see that for rich enough M, all beliefs in $\bar{\bar{\Delta}}(h, \mu_0, d, M)$ are also in $\bar{\Delta}(h, \mu_0, d, M)$, define m by

$$\pi_m(h|\omega) = \frac{\mu(\omega|h, m)}{\mu_0(\omega)} \times \Pr(h|d, \mu_0) \; \forall \omega \in \Omega.$$

□

Proposition 9, which is essentially a restatement of Proposition 1 in Schwartzstein and Sunderam (2021), characterizes the set of initial reactions to the data.[27] As an illustration, return to the

[27] To derive the distribution over initial reactions, we need to additionally specify the distribution of models people initially come up with. Many of our results on beliefs after social learning are independent of this distribution.

cryptocurrency example from the introduction. In this case,

$$\begin{aligned}\bar{\Delta}(h,\mu_0,d,M^a) &= \left\{\mu\in\Delta(\Omega) : \mu(s,g)+\mu(s,b)=1 \text{ and } \mu(\omega)\leq\frac{\mu_0(\omega)}{\Pr(h|d,\mu_0)}\ \forall\,\omega\in\Omega\right\}\\ &= \{\mu\in\Delta(\Omega) : \mu(s,g)+\mu(s,b)=1 \text{ and } \mu(s,b)\leq 2/3 \text{ and } \mu(s,g)\leq 1\},\end{aligned}$$

where the equality follows from $\mu_0(s,g)/\Pr(h|d,\mu_0) = .375/.1875 = 2$ and $\mu_0(s,b)/\Pr(h|d,\mu_0) = .125/.1875 = 2/3$. Given h, the highest probability venture capitalists could initially attach to the founder being bad given that the startup succeeded is $2/3$.

B.2 Expanding Networks

Expanding person i's network by merging it with $\tilde{M}$ enlarges the set of models that are shared with person i to $M_i \cup \tilde{M}$.

Proposition 10. *Suppose everyone is maximally open to persuasion, $M = M^a$. Let μ_i (m_i) denote a person's belief (model) following social learning prior to a network expansion, and μ_i^e (m_i^e) denote her belief (model) following social learning with the expanded network.*

1. *Expanding person i's network in any way weakly hardens her reaction to the data: for any expansion of M_i to $M_i \cup \tilde{M}$ with $\tilde{M} \subset M$, $\Pr(h|m_i^e,\mu_0) \geq \Pr(h|m_i,\mu_0)$.*

2. *If, in addition, everyone is in a shared-belief network, then expanding person i's network in any way also weakly mutes her reaction to the data: for any expansion of M_i to $M_i \cup \tilde{M}$ with $\tilde{M} \subset M$, Movement$(\mu_i^e;\mu_0) \leq$ Movement$(\mu_i;\mu_0)$.*

Proof of Proposition 10.

1. That expanding person i's network hardens her reaction to data follows from the simple fact that $\max_{m\in M^e}\Pr(h|m,\mu_0) \geq \max_{m\in M}\Pr(h|m,\mu_0)$ whenever $M^e \supset M$.

2. That expanding person i's network if anything mutes her reaction to the data when she's in a shared-belief network follows from the fact that m_i is the best-fitting model inducing μ_i, which fits better than any model inducing a belief further from her prior according to Movement$(\cdot;\mu_0)$ (by Lemma 1).

□

The first part of Proposition 10 shows that expanding a network always (weakly) hardens a network member's beliefs and makes them less persuadable. The most basic impact of increasing connectedness in our model is increasing a person's view that she can explain the data and making her resistant to changing her mind. The second part of the proposition shows that when networks are based on shared beliefs, expanding the network always additionally mutes members' beliefs. Being exposed to more models increases fit and consequently reduces movement.

B.3 Shared Model Networks

Some networks are based not on shared beliefs, but shared models. To analyze shared-model networks, consider a partition $\mathcal{C}$ over the set of admissible models M, where we denote $c(m)$ as the element in $\mathcal{M}$ that model $m \in M$ belongs in. In a *shared-model network*, a person i exchanges models with another person j if and only if their initial models are similar, in the sense that they fall in the same element of $\mathcal{M}$.

Definition 2. In a *shared-model network*, $M_i = \left\{m \in \bar{M}(h, \mu_0, d, M) : m \in c(m_i')\right\}$ for every person i.

People in a given shared model network will end up agreeing on whichever model in $c(m)$ maximizes $\Pr(h|\cdot, \mu_0)$.

Decompose h into two types of data, h^a and h^b. In predicting the success of a project, stock, or politician, for example, there may be both quantitative or hard information, as well as qualitative or soft information. In interpreting whether a left- or right-leaning policy is better, there may be data communicated by left-leaning and right-leaning outlets.

Imagine there are networks that view h^a as open to interpretation, but not h^b, and vice-versa. Quantitative analysts may believe they have a good handle on how to interpret hard information but may be more open to different ways of thinking about qualitative information. Symmetrically, qualitative analysts may have a single interpretation of soft interpretations but be open to many interpretations of hard information. People on the left may believe they know how to interpret left-leaning information, e.g., as trustworthy, but may be less sure on how to interpret right-leaning information. More formally, suppose there are three categories of models:

$$\begin{aligned}
c^A &= \left\{m \in \bar{M}(h, \mu_0, d, M) : \pi_m(h^a, h^b|\omega) = \pi_m(h^b|\omega) \cdot \pi_{m^{fa}}(h^a|\omega) \; \forall \; \omega \in \Omega\right\} \\
c^B &= \left\{m \in \bar{M}(h, \mu_0, d, M) : \pi_m(h^a, h^b|\omega) = \pi_m(h^a|\omega) \cdot \pi_{m^{fb}}(h^b|\omega) \; \forall \; \omega \in \Omega\right\} \\
c^O &= \bar{M}(h, \mu_0, d, M) \setminus \left\{c^A, c^B\right\}.
\end{aligned}$$

The first category of models, c^A, has a fixed interpretation m^{fa} of h^a but differing interpretations of h^b. Conversely, category c^B has a fixed interpretation m^{fb} of h^b but differing interpretations of h^a. Finally, category c^O contains all other models. If shared inflexibility stems from shared expertise, it is natural to assume $m^{fa} = m^T$ and $m^{fb} = m^T$; if it stems from shared beliefs that the data is uninformative, it is natural to assume that m^{fa} renders h^a uninformative and m^{fb} renders h^b

uninformative; if it stems from shared trust in knowing the process, it's natural to assume $m^{fa} = d$ and $m^{fb} = d$.

Supposing the data is maximally open to persuasion, $M = M^a$, then people with initial models in c^A will end up convincing themselves that h^b is obvious in hindsight and hence uninformative, while people with initial models in c^B will end up analogously convincing themselves that h^a is uninformative.

Proposition 11. *Suppose everyone is maximally open to persuasion, $M = M^a$, and is in a shared-model network based on shared inflexibility of the form described above, where $c(m) \in \left\{c^A, c^B, c^O\right\}$. Then social learning need not moderate everyone's reaction to the data. In particular, social learning leads members of c^A to view h^b as uninformative, members of c^B to view h^a as uninformative, and members of c^O to view h as uninformative, resulting in final beliefs:*

$$\mu_i = \begin{cases} \mu(h^a, m^{fa}) & \text{if } m_i' \in c^A \\ \mu(h^b, m^{fb}) & \text{if } m_i' \in c^B \\ \mu_0 & \text{if } m_i' \in c^O. \end{cases}$$

Proof of Proposition 11. Recall that

$$c^A = \left\{m \in \bar{M}(h, \mu_0, d, M) : \pi_m(h^a, h^b|\omega) = \pi_m(h^b|\omega) \cdot \pi_{m^{fa}}(h^a|\omega)\ \forall\ \omega \in \Omega\right\}.$$

Clearly, the best fitting model in c^A is $\pi_m(h^a, h^b|\omega) = 1 \cdot \pi_{m^{fa}}(h^a|\omega) = \pi_{m^{fa}}(h^a|\omega)$ for all $\omega \in \Omega$. Similarly, the best fitting model in c^B is $\pi_m(h^a, h^b|\omega) = 1 \cdot \pi_{m^{fb}}(h^b|\omega) = \pi_{m^{fb}}(h^b|\omega)$ for all $\omega \in \Omega$. Finally, the best fitting model in c^O is $\pi_m(h^a, h^b|\omega) = 1$ for all $\omega \in \Omega$. By assumption, someone in each network will propose the associated best-fitting models which all network members will end up adopting. The final beliefs μ_i follow.

□

As an illustration, consider networks based on shared expertise and imagine a company will either be successful ($\omega = 1$) or unsuccessful ($\omega = 0$) with equal probability ex ante. People are trying to forecast the success of the company based on hard, $h^a \in \left\{\underline{h}^a, \bar{h}^a\right\}$, and soft, $h^b \in \left\{\underline{h}^b, \bar{h}^b\right\}$, information. The true probability of h^a being $\bar{h}^a$ or h^b being $\bar{h}^b$ is .75 conditional on future success and .25 conditional on future failure, where hard and soft signals are conditionally independent. Imagine that the hard and soft signals point in opposite directions, with the hard signal being truly good ($h^a = \bar{h}^a$) and the soft signal being bad ($h^b = \underline{h}^b$). Then, the correct response is to predict the probability of future success to be $1/2$.

People's initial reactions to these signals will vary significantly. However, by Proposition 11, the network of soft-information experts will settle on explaining away the hard information and come to believe the likelihood of future success to be $1/4$. Conversely, the network of hard-information experts will settle on explaining away the soft information and come to believe the likelihood of future success to be $3/4$. The non-experts will settle on explaining away all information and believing the likelihood of future success to be $1/2$. Since some people in the hard- and soft-information networks will start with more moderate (and correct) reactions, in this example social learning intensifies some opinions in the hard- and soft-model networks in addition to hardening them.

With re-labeling, a similar example perhaps sheds light on so-called "epistemic closure" in political debates. Political observers argue that, in recent years, many of beliefs held by conservatives and liberals seem divorced from reality. Pundit Jonathan Chait puts it in the following way:

> the problem is that the [conservative] movement has created its own subculture, and within this subculture, only information from sources controlled by the movement is considered trustworthy or even worth paying attention to.[28]

The key problem, as Chait puts it, is *not* necessarily that liberals are unaware of information provided by conservatives and vice-versa, but rather that they hold shared beliefs that information from the other side of the aisle is not worth grappling with. The analysis in this section shows that this would be a consequence of shared inflexibility in believing information from your own side is trustworthy. Under this interpretation, liberals are aware of conservative information. And they begin with quite diverse opinions on how to interpret conservative information. But, in exchanging interpretations, they end up settling on a shared view that they should not update based on that information.

A final example of networks based on shared models is where the measure $(1 - \delta)$ of the population who initially stick with the default are in one network and the rest of the population are in others. For example, some portion of the population may not devote enough attention to an issue to construct their own interpretation of the data beyond the default, nor to exchanging interpretations with others.

When the default is accurate (e.g., in some cases taking scientific consensus at face value), people who adhere to the default end up with more accurate interpretations and beliefs than those in other networks. For example, a 2016 Pew report found that Americans "who care a great deal about GM foods issue expected negative effects from these foods," belying scientific consensus. Similarly, Fernbach et al. (2019) found that people who are extremely opposed to GM foods think they know the most about the safety of those foods, but actually know the least. Such Americans pushed a number of unfounded interpretations of the data, including that eating GM foods caused allergies, cancer, and autism.

[28] https://newrepublic.com/article/74492/what-conservative-epistemic-closure-means

C An Additional Example

This example shows how interpretations may evolve differently across shared-belief networks. Consider a community of venture capitalists trying to predict the success of a startup in a new sector (e.g., generative AI) based on the history of past startups and their characteristics. The characteristics of startup j are its profits (x_{1j}), management team experience (x_{2j}), and an individuating characteristic (x_{3j})—a characteristic that is unique to each startup. The history of past startups is $h = \{(x_{1j}, x_{2j}, x_{3j}, y_j)\}_j$ where $y_j = 1$ if startup j succeeded and $y_j = 0$ if it failed. Figure 3a shows an example history. Each dot represents a previous startup, with profit plotted on the horizontal axis and team experience plotted on the vertical axis. The individuating characteristics are not pictured. A dot is filled in if the startup was successful and is unfilled if it failed. Venture capitalists start with a prior that a given startup's probability of success, θ, is uniformly distributed on $[0, 1]$ and dogmatically believe that (profit) x (experience) characteristics are uniformly distributed in $[0, 1] \times [0, 1]$. They then use the history to make predictions about a new startup k's success probability as a function of its characteristics.

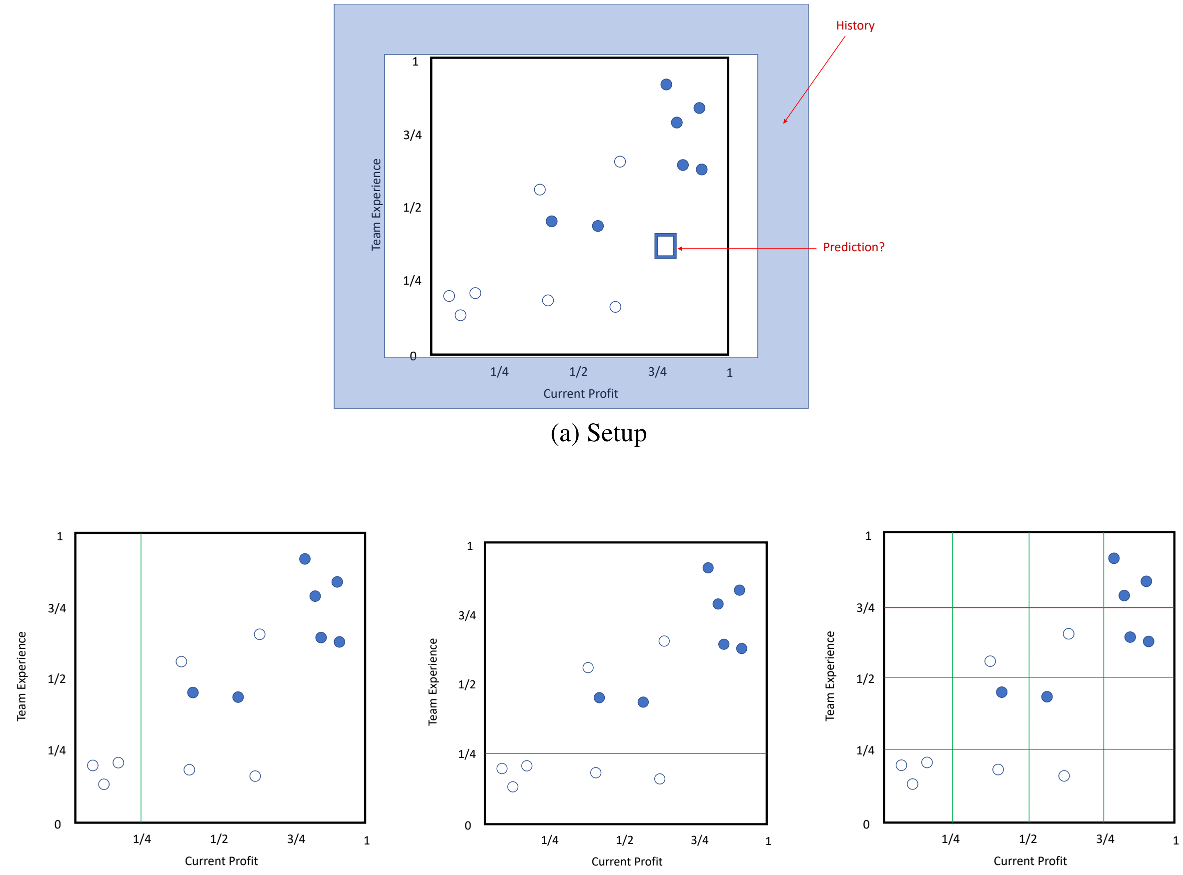
We assume there are four types of models in the model space M. First, the default model is that all startups in the new sector have the same success probability regardless of their characteristics. Second, there are models that are cutoff rules in profit: all startups with profit below the cutoff share the same success probability and all startups with profit above the cutoff share the same success probability.[29] For instance, the vertical green line in Figure 3b depicts the model where the cutoff is the 25th percentile of profits. Third, there are models that are analogous cutoff rules in team experience. For instance, the horizontal red line in Figure 3c depicts the model where the cutoff is the 25th percentile of experience. Fourth, there is a model positing that neither profits nor experience matter. Instead, each startup's outcome is due to its individuating characteristics; in other words, each startup had a unique feature that perfectly determined success or failure. Note that this model perfectly explains each data point.[30]

Prior to social learning, venture capitalists consider the default and one other model randomly selected from the other three model types. As shown in Figure 3d, venture capitalists will have a variety of different interpretations, and thus different beliefs, at this point. In the figure, we depict for simplicity the case where the cutoffs considered are at the 25th, 50th, and 75th percentiles of each dimension. All fit better than the default.

Suppose venture capitalists are in shared-belief networks. Specifically, they share interpretations with others who have similar initial reactions to the data. Optimists who believe the data suggest that the average startup is likely to be successful talk to each other; pessimists who be-

[29] Formally, success probabilities below and above the cutoff are independently drawn from the uniform distribution.

[30] Formally, under the model m^{ind}, $\Pr(y|x_3, m^{ind}, \mu_0) = 1$ for $y \equiv (y_j)_j$ and $x_3 \equiv (x_{3j})_j$.



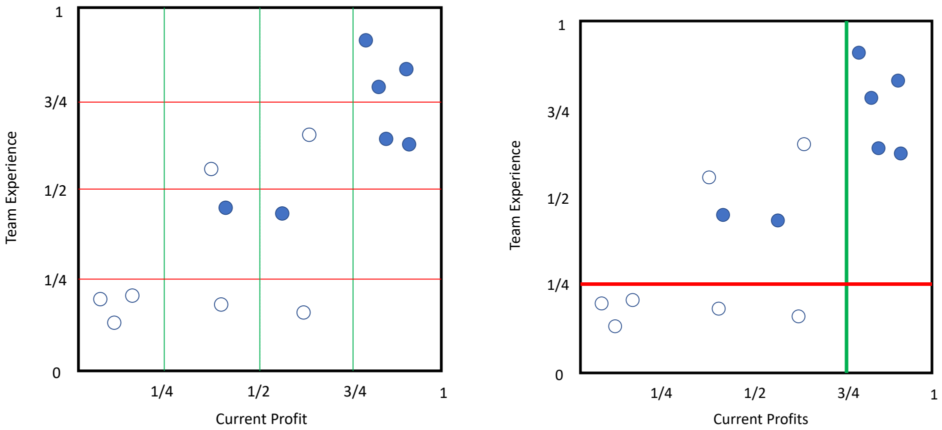
(a) Setup

(b) Model emphasizing current profits (c) Model emphasizing experience (d) Initial models

Figure 3: Predicting the success of a startup

lieve the data suggest that the average startup is likely to be *un*successful talk to each other; and moderates who believe that success of the average startup is 50-50 talk to each other. This network structure may emerge because people with different initial reactions have different objectives going forward. For instance, optimists think they are likely to invest and want to figure out the characteristics that matter most for success, while pessimists want to figure out the most compelling way to explain to their clients why they are not investing.

Social learning will lead beliefs to converge within each network to the model within that network that best fits the data. For instance, consider the optimists. Two models lead to optimistic interpretations of the data: one where the cutoff is at the 25th percentile of experience and one where the cutoff is at the 25th percentile of profits. The former fits the data almost ten times better than the latter. This can be seen in by comparing Figures 3b and 3c. The experience-based model in Figure 3c more effectively separates successes from failures than does the profit-based



(a) Initial models

(b) Models after social learning

Figure 4: Evolution of Beliefs Across Shared-Belief Networks Surrounding Startup Success

model in Figure 3b.[31] Thus, after social learning, all optimists adopt the experience-based model, depicted by thick-red horizontal line in Figure 4b. Given the data and this adopted model, simple application of the standard beta-binomial updating formula tells us that members of the optimist network forecast average startup success to be $3/4 \cdot ((7+1)/(9+2)) + 1/4 \cdot (1/(5+2)) \approx .58$. Essentially, they believe the best way to explain the data is that failure is relatively rare—only the startups with the least experienced management teams fail.

Members of the pessimist network go through a similar evolution. There are two models that lead to pessimistic interpretations: one with a cutoff at the 75th percentile of experience and one with a cutoff at the 75th percentile of profits. In this case, the profit-based model fits approximately ten times better than the experienced-based model, so pessimists converge to the model depicted by the thick-green vertical line in Figure 4b. Given the data, members in the pessimist network forecast average startup success to be $3/4 \cdot ((2+1)/(9+2)) + 1/4 \cdot ((5+1)/(5+2)) \approx .42$, disagreeing strongly with the members of the optimists network.

Finally, consider the neutral network. Prior to social learning, the two models in the neutral network are the default model (that the success probability is the same regardless of characteristics) and the model where the success or failure of each previous startup was inevitable given individuating characteristics. The latter model fits the data perfectly, so members of the neutral network converge to it, while continuing to forecast average startup success to be $.5$.

The example highlights how interpretations evolve differently across networks. Members of

[31] Formally, the likelihood of the data under the experience-based model is proportional to $(\int_0^1 (1-\theta)^5 d\theta) \cdot (\int_0^1 \theta^7 (1-\theta)^2 d\theta) \approx .00046$, while the likelihood of the data under the profit-based model is proportional to $(\int_0^1 (1-\theta)^3 d\theta) \cdot (\int_0^1 \theta^7 (1-\theta)^4) \approx .000063$.

different networks end up not only with different final beliefs about startup success probabilities, but also disagreeing about the characteristics that matter for success. In the optimist network, some initially believe experience matters, while others initially believe profit matters. Yet all come to believe that startup success is predicted by experience and not profit. Members of the pessimist network similarly start out disagreeing, but instead come to believe that startup success is predicted by profits and not experience. In the neutral network, everyone comes to believe that success is unpredictable ex ante because individuating characteristics are all that matter.

In the example, sharing models in shared-belief networks does not result in muting—optimists end up more optimistic after sharing models than they were on average before sharing models and similarly pessimists end up more pessimistic on average. Here we show using simulations that muting does appear to hold on average in the example, just not for the particular realization of the data we consider above. In each simulation, we first choose a value for the true success probability for the startup θ from a set of 40 values evenly distributed on $[0, 1]$. We then randomly draw 5 startups with 3 characteristics: (i) success or failure with success having probability θ, (ii) profit, which is uniformly distributed on $[0, 1]$, and (iii) team experience, which is also uniformly distributed on $[0, 1]$. We consider models to explain success or failure that are cutoff rules in either profit or team experience. Five cutoff rules are considered, evenly spaced on each dimension. We compute the fit for all models, including the default model that the success probability is constant across characteristics. We then select the best-fitting model for optimists, i.e., the best-fitting model that implies an average posterior expected probability of success $\hat{\theta}$ greater than 0.5, and the best-fitting model for pessimists, i.e., the best-fitting model that implies $\hat{\theta}$ less than 0.5. For each value of the true success probability θ, we average the optimists' $\hat{\theta}$ and the pessimists' $\hat{\theta}$ over 5000 simulations.

Figure A1 reports the results. We see that relative to the updating that would have taken place under the default model, there appears to be muting for both optimists and pessimists on average. In other words, for any value of θ, the average optimists' $\hat{\theta}$ and the average pessimists' $\hat{\theta}$ is at least as close to the prior average of 0.5 as the average $\hat{\theta}$ under the default model.

D StockTwits Application Details

In this section, we provide regression evidence to supplement our analysis of StockTwits data in Section 6.2 of the main paper. We start with the universe of StockTwits messages studied by Divernois and Filipovic (2022), covering the period between January 2011 and July 2018. For each message about a particular stock, we code sentiment as 1 if the user labels the message as bullish and 0 if the user labels the message as bearish. We drop messages that users do not label. We restrict the sample to windows from 10 days before an earnings announcement to 10 days after for a given stock and restrict attention to users who have ever posted a message about that stock prior to 10 days before the earnings announcement. We code a user as a bull on the stock if the user labeled as bullish at least 50% of their messages about the stock prior to 10 days before the earnings announcement. We code the user as a bear if they labeled as bearish less than 50% of their messages about the stock. We then track how sentiment evolves in response to different earnings announcements over the surrounding windows. We code an announcement as positive news if the announcement day return is greater or equal to zero and as negative news if the announcement day return is negative.[32] The final sample consists of roughly 1.8 million messages across 40 thousand earnings announcements from 65 thousand unique users.

Table A1 presents regression evidence corresponding to Figure 4a in the main paper. Among the sample of users who are bullish on stock s for earnings announcement q, we estimate the following regression:

$$1[Bullish]_{i,u,t,s,q} = \alpha + \sum_{l=-10}^{10} \beta_l 1[l = t] + \sum_{l=-10}^{10} \gamma_l 1[l = t] 1[NegativeSurprise_{s,q}] + \varepsilon_{i,u,t,s,q}, \quad (11)$$

where

- $1[Bullish]_{i,u,t,s,q}$ is an indicator that tweet i by user u on event-day t is bullish and
- $1[NegativeSurprise_{s,q}]$ is an indicator that the earnings announcement was a negative surprise.

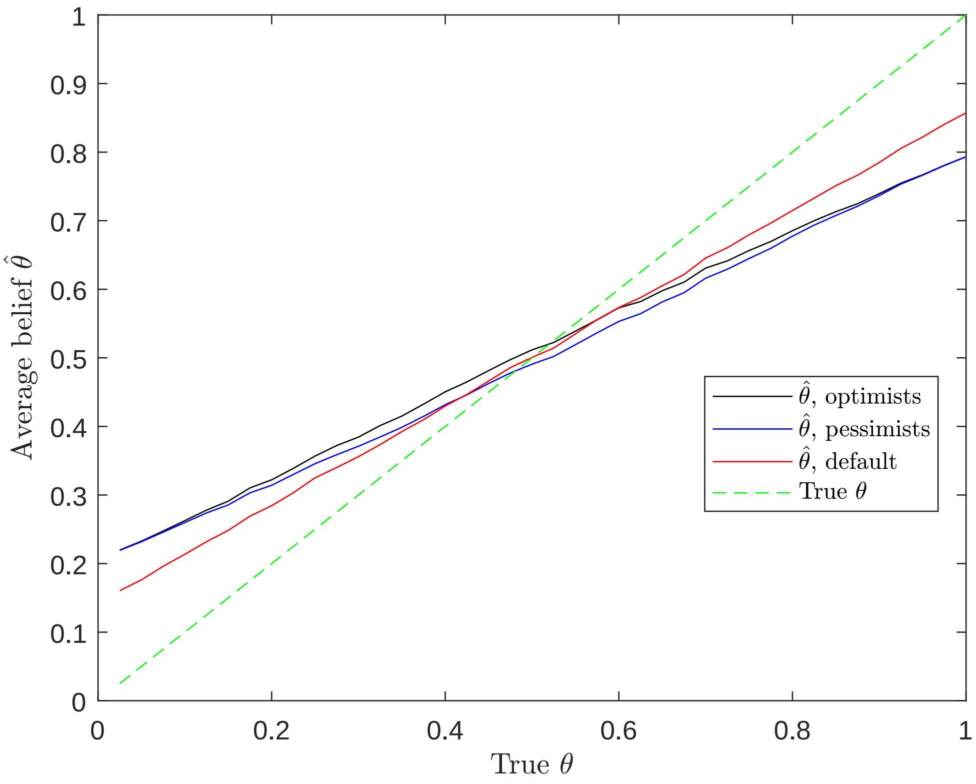
Standard errors are reported in parentheses and clustered by year-month, stock, and user.

The first column replicates Figure 4a, weighting the data so that each earnings announcement is equally weighted. Note that to match the levels of Figure 4a, the constant in the regression must be added back in. The key coefficients are γ_{-1} and γ_0, which show that the sentiment of bullish investors declines substantially around negative earnings announcements, and $\gamma_1$to γ_{10}, which show that this decline is transient. The remaining columns show the robustness of the result. In the second column, we weight tweets equally rather than equal-weighting announcements. This has the effect of placing greater weight on announcements with more tweets. The remaining columns add fixed effects for the year-month of the announcement, the announcement itself, and the user interacted with the announcement. Across these variations, the basic pattern remains, though it weakens somewhat in the last column. The sentiment of bullish investors declines substantially around negative earnings announcements and then rebounds.

[32] Announcement days are therefore defined as the first day that the stock can be traded following the announcement. For announcements that occur after the market close on a given day, the announcement day is then coded as the following day.

Table A2 presents regression evidence corresponding to Figure 4b in the main paper. Estimating Eq. (11) among the sample of users who are bearish on stock s for earnings announcement q, we find that bearish investors become less bearish around positive earnings announcements but then quickly revert.

Figure A1: Muting in the VC Example



Notes: This figure suggests that muting obtains on average in the VC example. For each of 40 different values of the true probability of success, θ, we simulate data on a history of 5 startups, each of which varies in their profit and team experience. VCs entertain models that are cutoff rules on each dimension. The figure plots the average posterior expectation of the success probability, $\hat{\theta}$,of VCs under three models against the true success probability: the default model (red line), the model adopted after optimists share interpretations, and the model adopted after pessimists share. interpretations. The figure averages over 5000 simulations for each value of θ.

Table A1: Bulls' Beliefs around Earnings Announcements

	(1)	(2)	(3)	(4)	(5)
t=-9	0.01	-0.01	0.01	0.00	0.00
	(1.13)	(-0.96)	(1.19)	(0.97)	(1.44)
t=-8	0.00	-0.00	0.00	0.00	0.00
	(0.04)	(-0.62)	(0.16)	(0.27)	(1.15)
t=-7	0.01	-0.00	0.01	0.01*	0.01
	(1.16)	(-0.94)	(1.39)	(1.76)	(1.40)
t=-6	0.01*	0.00	0.01**	0.01**	0.00
	(1.72)	(0.28)	(2.00)	(2.13)	(1.15)
t=-5	0.01**	0.00	0.01**	0.01**	0.01**
	(2.13)	(0.06)	(2.43)	(2.51)	(2.56)
t=-4	0.01	-0.01*	0.01	0.01	0.01*
	(1.20)	(-1.80)	(1.59)	(1.39)	(1.70)
t=-3	0.01	-0.00	0.01**	0.01*	0.01*
	(1.66)	(-0.67)	(2.13)	(1.72)	(1.96)
t=-2	0.01	-0.00	0.01	0.00	0.01**
	(0.96)	(-0.70)	(1.53)	(1.00)	(2.58)
t=-1	0.00	-0.01	0.01	0.01	0.01**
	(0.36)	(-1.46)	(0.95)	(1.42)	(3.27)
t=0	0.01	-0.01**	0.01	0.01*	0.01**
	(0.80)	(-2.02)	(1.58)	(1.91)	(3.39)
t=1	-0.01*	-0.02**	-0.01	-0.01	0.01**
	(-1.87)	(-4.75)	(-1.18)	(-1.24)	(2.33)
t=2	-0.01	-0.01**	-0.00	-0.00	0.01*
	(-1.32)	(-2.10)	(-0.58)	(-0.32)	(1.75)
t=3	-0.01	-0.02**	0.00	-0.00	0.01**
	(-0.74)	(-2.38)	(0.04)	(-0.24)	(2.38)
t=4	-0.00	-0.01*	0.00	0.00	0.01*
	(-0.29)	(-1.68)	(0.60)	(0.42)	(1.72)
t=5	-0.00	-0.01**	0.01	0.00	0.01**
	(-0.05)	(-2.09)	(0.73)	(0.51)	(2.07)
t=6	-0.01	-0.02**	-0.00	-0.00	0.01**
	(-1.58)	(-3.07)	(-0.57)	(-0.85)	(2.20)
t=7	-0.02*	-0.01**	-0.01	-0.00	0.01**
	(-1.97)	(-2.16)	(-1.02)	(-0.59)	(2.54)
t=8	-0.01*	-0.01**	-0.00	-0.00	0.01**
	(-1.79)	(-2.45)	(-0.57)	(-0.40)	(2.62)
t=9	-0.00	-0.02**	0.00	0.01	0.02**
	(-0.58)	(-2.52)	(0.66)	(1.34)	(2.97)
t=10	0.00	-0.00	0.01*	0.01*	0.02**
	(0.46)	(-0.23)	(1.81)	(1.87)	(4.12)
t=-10 × Neg	-0.02**	-0.02**	-0.02**	-0.03**	-0.02**
	(-2.75)	(-2.25)	(-3.03)	(-4.93)	(-4.51)
t=-9 × Neg	-0.03**	-0.02**	-0.03**	-0.04**	-0.03**
	(-3.77)	(-3.22)	(-3.92)	(-6.60)	(-6.83)
t=-8 × Neg	-0.03**	-0.01**	-0.03**	-0.03**	-0.03**
	(-2.86)	(-2.68)	(-2.93)	(-5.56)	(-6.26)
t=-7 × Neg	-0.04**	-0.02**	-0.04**	-0.04**	-0.03**
	(-4.73)	(-4.00)	(-4.62)	(-6.82)	(-5.83)
t=-6 × Neg	-0.04**	-0.03**	-0.04**	-0.04**	-0.03**
	(-5.83)	(-5.55)	(-5.44)	(-7.65)	(-5.69)

t=-5 × Neg	-0.04**	-0.04**	-0.04**	-0.04**	-0.03**
	(-5.71)	(-5.80)	(-5.21)	(-5.78)	(-4.99)
t=-4 × Neg	-0.05**	-0.04**	-0.05**	-0.05**	-0.03**
	(-6.87)	(-5.94)	(-6.26)	(-8.16)	(-6.85)
t=-3 × Neg	-0.05**	-0.04**	-0.05**	-0.05**	-0.03**
	(-5.17)	(-4.93)	(-4.85)	(-6.28)	(-6.44)
t=-2 × Neg	-0.06**	-0.04**	-0.05**	-0.06**	-0.03**
	(-5.80)	(-5.42)	(-5.50)	(-8.39)	(-6.44)
t=-1 × Neg	-0.09**	-0.06**	-0.09**	-0.08**	-0.04**
	(-5.63)	(-7.10)	(-5.55)	(-8.34)	(-8.40)
t=0 × Neg	-0.10**	-0.06**	-0.10**	-0.08**	-0.03**
	(-7.60)	(-8.96)	(-7.25)	(-9.72)	(-7.43)
t=1 × Neg	-0.03**	-0.04**	-0.03**	-0.03**	-0.01**
	(-4.34)	(-5.53)	(-3.63)	(-5.56)	(-3.07)
t=2 × Neg	-0.00	-0.00	0.00	-0.01	-0.01
	(-0.60)	(-0.63)	(0.12)	(-1.32)	(-1.21)
t=3 × Neg	-0.00	0.00	0.00	-0.01	-0.01
	(-0.26)	(0.32)	(0.48)	(-1.53)	(-1.63)
t=4 × Neg	-0.00	0.00	0.00	-0.01	-0.01
	(-0.08)	(0.17)	(0.44)	(-1.29)	(-1.58)
t=5 × Neg	0.00	0.00	0.01	-0.00	-0.01**
	(0.62)	(1.08)	(1.13)	(-0.90)	(-2.05)
t=6 × Neg	0.00	-0.00	0.01	-0.00	-0.00
	(0.41)	(-0.41)	(1.18)	(-0.28)	(-0.51)
t=7 × Neg	0.00	-0.00	0.01	-0.00	-0.00
	(0.39)	(-0.62)	(1.26)	(-0.16)	(-0.91)
t=8 × Neg	-0.00	-0.01*	0.00	-0.00	-0.00
	(-0.40)	(-1.72)	(0.64)	(-0.68)	(-0.83)
t=9 × Neg	-0.00	-0.01	0.00	0.00	-0.00
	(-0.37)	(-1.08)	(0.56)	(0.17)	(-0.95)
t=10 × Neg	-0.00	-0.01	0.00	0.00	0.00
	(-0.57)	(-0.78)	(0.60)	(.)	(.)
Constant	0.94**	0.96**	0.93**	0.94**	0.94**
	(124.89)	(262.61)	(127.29)	(291.92)	(428.90)
Weighting	Event	Tweet	Event	Event	Event
Fixed Effects			YM	Event	User x Event
R^2	.013	.0076	.019	.27	.77
N	1613258	1613258	1613258	1602495	1452718

Notes: This table presents the evolution of bulls' beliefs around positive and negative earnings announcements. The sample is all tweets within 10 days of an earnings announcement by users who have tweeted at least once about the stock before the ±10-day window, with more than 50% of these prior tweets self-labeled as bullish. Let s denote the stock, q denote the announcement event (quarter), t denote the day relative to the event (ranging from -10 to 10 with $t = 0$ corresponding to the event date). The dependent variable is an indicator that a tweet on event-day t for stock f and event q is self-labeled as bullish. The independent variables are dummies for t, interacted with dummies indicating that the earnings announcement was negative, measured by negative announcement day returns. The event date ($t = 0$) is defined as the first day the news is tradeable. The Weighting row indicates whether the regression is weighted to equal-weight each event or unweighted (i.e., each tweet is weighted equally). Standard errors are reported in parentheses and clustered by year-month, stock, and user.

Table A2: Bears' Beliefs around Earnings Announcements

	(1)	(2)	(3)	(4)	(5)
t=-9	0.01	-0.00	0.01	0.00	-0.03
	(0.30)	(-0.06)	(0.34)	(0.11)	(-1.48)
t=-8	-0.02	-0.01	-0.02	-0.01	-0.01
	(-0.78)	(-0.14)	(-0.79)	(-0.51)	(-1.09)
t=-7	0.04	0.01	0.04*	0.03	-0.01
	(1.61)	(0.23)	(1.67)	(1.36)	(-1.10)
t=-6	0.03	0.01	0.03	0.00	0.00
	(1.45)	(0.28)	(1.27)	(0.18)	(0.04)
t=-5	0.06**	0.03	0.05**	0.02	-0.01
	(2.28)	(1.06)	(2.10)	(0.78)	(-0.36)
t=-4	0.07**	0.00	0.06**	0.01	-0.02
	(2.54)	(0.02)	(2.26)	(0.36)	(-0.96)
t=-3	0.06**	0.03	0.05*	0.02	-0.00
	(2.16)	(0.82)	(1.79)	(0.59)	(-0.13)
t=-2	0.08**	0.05	0.07**	0.03	0.00
	(3.08)	(1.13)	(2.71)	(1.53)	(0.36)
t=-1	0.10**	0.04	0.09**	0.02	-0.01
	(3.28)	(1.17)	(2.96)	(0.82)	(-1.17)
t=0	0.11**	0.05	0.09**	0.01	-0.02*
	(3.14)	(1.34)	(2.84)	(0.56)	(-1.92)
t=1	-0.01	-0.02	-0.02	-0.06**	-0.05**
	(-0.34)	(-0.70)	(-0.82)	(-2.66)	(-3.43)
t=2	-0.04	-0.05	-0.05*	-0.07**	-0.05**
	(-1.34)	(-1.61)	(-1.80)	(-2.95)	(-3.51)
t=3	-0.02	-0.05	-0.03	-0.06**	-0.05**
	(-0.78)	(-1.63)	(-1.26)	(-2.52)	(-2.78)
t=4	-0.05*	-0.06	-0.07**	-0.08**	-0.06**
	(-1.88)	(-1.66)	(-2.50)	(-3.25)	(-3.23)
t=5	-0.04	-0.05*	-0.05**	-0.07**	-0.05**
	(-1.65)	(-1.68)	(-2.33)	(-3.47)	(-2.80)
t=6	-0.05*	-0.09**	-0.06**	-0.09**	-0.05**
	(-1.94)	(-2.67)	(-2.66)	(-3.60)	(-2.87)
t=7	-0.08**	-0.11**	-0.10**	-0.10**	-0.07**
	(-2.86)	(-3.13)	(-3.71)	(-3.87)	(-3.80)
t=8	-0.09**	-0.12**	-0.11**	-0.12**	-0.08**
	(-3.27)	(-3.18)	(-4.17)	(-5.07)	(-4.01)
t=9	-0.08**	-0.11**	-0.10**	-0.12**	-0.08**
	(-2.49)	(-3.32)	(-3.34)	(-4.30)	(-3.71)
t=10	-0.10**	-0.12**	-0.12**	-0.14**	-0.09**
	(-3.23)	(-3.28)	(-4.31)	(-5.62)	(-4.09)
t=-10 × Neg	-0.06**	-0.08**	-0.06**	0.05**	0.00
	(-2.36)	(-2.43)	(-2.05)	(2.00)	(0.24)
t=-9 × Neg	-0.09**	-0.05	-0.08**	0.01	-0.00
	(-3.36)	(-1.60)	(-2.96)	(0.56)	(-0.17)
t=-8 × Neg	-0.07**	-0.06	-0.07**	0.02	-0.01
	(-2.85)	(-1.42)	(-2.47)	(1.10)	(-0.32)
t=-7 × Neg	-0.08**	-0.08**	-0.08**	0.03	0.01
	(-3.29)	(-2.19)	(-3.06)	(1.28)	(0.61)
t=-6 × Neg	-0.07**	-0.07*	-0.07**	0.03	0.01
	(-2.82)	(-1.87)	(-2.75)	(1.45)	(0.77)

t=-5 × Neg	-0.09**	-0.08**	-0.09**	0.01	-0.01
	(-3.52)	(-2.03)	(-3.58)	(0.60)	(-0.40)
t=-4 × Neg	-0.10**	-0.10**	-0.10**	-0.00	-0.01
	(-3.50)	(-2.80)	(-3.59)	(-0.10)	(-0.54)
t=-3 × Neg	-0.08**	-0.10**	-0.08**	0.02	-0.00
	(-3.09)	(-2.58)	(-3.13)	(0.91)	(-0.14)
t=-2 × Neg	-0.09**	-0.10**	-0.10**	-0.00	-0.02
	(-3.65)	(-2.74)	(-3.60)	(-0.15)	(-0.96)
t=-1 × Neg	-0.11**	-0.11**	-0.12**	-0.03	-0.03
	(-4.43)	(-3.06)	(-4.76)	(-1.53)	(-1.64)
t=0 × Neg	-0.13**	-0.14**	-0.14**	-0.05**	-0.02
	(-5.32)	(-3.88)	(-5.56)	(-2.87)	(-1.63)
t=1 × Neg	-0.06**	-0.10**	-0.07**	0.01	-0.01
	(-2.13)	(-2.60)	(-2.57)	(0.49)	(-0.29)
t=2 × Neg	-0.01	-0.07*	-0.02	0.07**	0.01
	(-0.45)	(-1.79)	(-0.93)	(3.67)	(0.70)
t=3 × Neg	-0.01	-0.05	-0.02	0.07**	0.02
	(-0.40)	(-1.26)	(-0.78)	(3.97)	(0.94)
t=4 × Neg	-0.02	0.01	-0.02	0.06**	0.00
	(-0.67)	(0.13)	(-0.82)	(2.83)	(0.12)
t=5 × Neg	-0.02	-0.04	-0.03	0.06**	0.01
	(-0.57)	(-1.05)	(-0.98)	(2.42)	(0.47)
t=6 × Neg	-0.04*	-0.08**	-0.06**	0.04**	0.01
	(-1.95)	(-2.22)	(-2.36)	(2.10)	(0.56)
t=7 × Neg	-0.05*	-0.07**	-0.06**	0.03	-0.00
	(-1.86)	(-2.00)	(-2.51)	(1.57)	(-0.08)
t=8 × Neg	-0.05**	-0.09**	-0.07**	0.02	-0.01
	(-2.02)	(-2.63)	(-2.67)	(0.85)	(-0.61)
t=9 × Neg	-0.10**	-0.12**	-0.12**	-0.02	-0.02
	(-3.15)	(-3.35)	(-4.06)	(-0.92)	(-1.42)
t=10 × Neg	-0.08**	-0.16**	-0.11**	0.00	0.00
	(-2.91)	(-4.61)	(-3.65)	(.)	(.)
Constant	0.37**	0.34**	0.38**	0.33**	0.32**
	(14.59)	(10.20)	(16.32)	(27.12)	(33.77)
R^2	0.02	0.02	0.04	0.43	0.81
N	220,280	220,280	220,280	213,854	187,869
Weighting	Event	Tweet	Event	Event	Event
Fixed Effects			YM	Event	User x Event

Notes: This table presents the evolution of bears' beliefs around positive and negative earnings announcements. The sample is all tweets within 10 days of an earnings announcement by users who have tweeted at least once about the firm before the ±10-day window, with more than 50% of these prior tweets self-labeled as bearish. Let f denote the firm, q denote the announcement event (quarter), t denote the day relative to the event (ranging from -10 to 10 with $t = 0$ corresponding to the event date). The dependent variable is an indicator that a tweet on event-day t for firm f and event q is self-labeled as bullish. The independent variables are dummies for t, interacted with dummies indicating that the earnings announcement was negative, measured by negative announcement day returns. The event date ($t = 0$) is defined as the first day the news is tradeable. The Weighting row indicates whether the regression is weighted to equal-weight each event or unweighted (i.e., each tweet is weighted equally). Standard errors are reported in parentheses and clustered by year-month, firm, and user.

References

Acemoglu, Daron, Victor Chernozhukov, and Muhamet Yildiz, “Fragility of Asymptotic Agreement Under Bayesian Learning,” *Theoretical Economics*, 2016, *11* (1), 187–225.

Aina, Chiara, “Tailored Stories,” 2021.

Akbarpour, Mohammad, Suraj Malladi, and Amin Saberi, “Just a Few Seeds More: Value of Network Information for Diffusion,” 2020.

Alesina, Alberto, Armando Miano, and Stefanie Stantcheva, “The Polarization of Reality,” in “AEA Papers and Proceedings,” Vol. 110 2020, pp. 324–28.

Andre, Peter, Ingar Haaland, Christopher Roth, and Wohlfart Johannes, “Narratives about the Macroeconomy,” 2022.

Ba, Cuimin, “Robust Model Misspecification and Paradigm Shifts,” *arXiv preprint arXiv:2106.12727*, 2021.

Banerjee, Abhijit V, “A Simple Model of Herd Behavior,” *The Quarterly Journal of Economics*, 1992, *107* (3), 797–817.

Barberis, Nicholas, “Psychology-Based Models of Asset Prices and Trading Volume,” in B. Douglas Bernheim, Stefano DellaVigna, and David Laibson, eds., , Vol. 1 of *Handbook of Behavioral Economics*, Elsevier, 2018, pp. 79 – 175.

Barron, Kai and Tilman Fries, “Narrative Persuasion,” 2022.

Bénabou, Roland, Armin Falk, and Jean Tirole, “Narratives, Imperatives, and Moral Reasoning,” Technical Report, National Bureau of Economic Research 2018.

Bertrand, Marianne and Emir Kamenica, “Coming Apart? Cultural Distances in the United States Over Time,” 2020.

_ **and Sendhil Mullainathan**, “Are CEOs Rewarded for Luck? The Ones Without Principals Are,” *The Quarterly Journal of Economics*, 2001, *116* (3), 901–932.

Bikhchandani, Sushil, David Hirshleifer, and Ivo Welch, “A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades,” *Journal of Political Economy*, 1992, *100* (5), 992–1026.

Bolton, Patrick, Markus K Brunnermeier, and Laura Veldkamp, “Leadership, Coordination, and Corporate Culture,” *Review of Economic Studies*, 2013, *80* (2), 512–537.

Bowen, Renee, Danil Dmitriev, and Simone Galperti, "Learning from Shared News: When Abundant Information Leads to Belief Polarization," 2021.

Boxell, Levi, Matthew Gentzkow, and Jesse M Shapiro, "Greater Internet Use is Not Associated with Faster Growth in Political Polarization Among US Demographic Groups," *Proceedings of the National Academy of Sciences*, 2017, *114* (40), 10612–10617.

_ , _ , and _ , "Cross-Country Trends in Affective Polarization," Technical Report, National Bureau of Economic Research 2020.

Burnstein, Eugene and Amiram Vinokur, "Persuasive Argumentation and Social Comparison as Determinants of Attitude Polarization," *Journal of Experimental Social Psychology*, 1977, *13* (4), 315–332.

Bursztyn, Leonardo and David Yang, "Misperceptions about Others," *Annual Review of Economics*, 2023, *14*, 425–452.

_ , Georgy Egorov, Ingar Haaland, Aakaash Rao, and Christopher Roth, "Scapegoating During Crises," *American Economic Association Papers and Proceedings*, 2022, *112* (5), 151–155.

_ , _ , _ , _ , and _ , "Justifying Dissent," *The Quarterly Journal of Economics*, 2023.

Chater, Nick and George Loewenstein, "The Under-Appreciated Drive for Sense-Making," *Journal of Economic Behavior & Organization*, 2016, *126*, 137–154.

Cohen, Lauren, Dong Lou, and Christopher J Malloy, "Casting Conference Calls," *Management Science*, 2020, *66* (11), 5015–5039.

Cookson, J. Anthony and Marina Niessner, "Why Don't We Agree? Evidence from a Social Network of Investors," *The Journal of Finance*, 2020, *75* (1), 173–228.

_ , Joseph Engelberg, and William Mullins, "Echo Chambers," *Review of Financial Studies*, 2022.

DeGroot, Morris H, "Reaching a Consensus," *Journal of the American Statistical Association*, 1974, *69* (345), 118–121.

DeMarzo, Peter M, Dimitri Vayanos, and Jeffrey Zwiebel, "Persuasion Bias, Social Influence, and Unidimensional Opinions," *The Quarterly Journal of Economics*, 2003, *118* (3), 909–968.

Dessein, Wouter and Tano Santos, "Adaptive organizations," *Journal of Political Economy*, 2006, *114* (5), 956–995.

Diether, Karl, Christopher Malloy, and Anna Scherbina, "Differences of Opinion and the Cross Section of Stock Returns," *The Journal of Finance*, 2002, *57* (5), 2113–2141.

Divernois, Marc-Aurele and Damir Filipovic, "StockTwits Classified Sentiment and Stock Returns," 2022.

Eliaz, Kfir and Ran Spiegler, "A Model of Competing Narratives," *American Economic Review*, 2020, *110* (12), 3786–3816.

_ , Simone Galperti, and Ran Spiegler, "False Narratives and Political Mobilization," 2022.

Enke, Benjamin and Florian Zimmermann, "Correlation Neglect in Belief Formation," *The Review of Economic Studies*, 2019, *86* (1), 313–332.

Esponda, Ignacio and Demian Pouzo, "Berk–Nash Equilibrium: A Framework for Modeling Agents With Misspecified Models," *Econometrica*, 2016, *84* (3), 1093–1130.

Eyster, Erik and Matthew Rabin, "Naive Herding in Rich-Information Settings," *American Economic Journal: Microeconomics*, 2010, *2* (4), 221–43.

_ and _ , "Extensive Imitation is Irrational and Harmful," *The Quarterly Journal of Economics*, 2014, *129* (4), 1861–1898.

Fernbach, Philip M, Nicholas Light, Sydney E Scott, Yoel Inbar, and Paul Rozin, "Extreme Opponents of Genetically Modified Foods Know the Least but Think They Know the Most," *Nature Human Behaviour*, 2019, *3* (3), 251–256.

Flynn, Joel P and Karthik Sastry, "The Macroeconomics of Narratives," *Available at SSRN 4140751*, 2022.

Froeb, Luke M, Bernhard Ganglmair, and Steven Tschantz, "Adversarial Decision Making: Choosing Between Models Constructed by Interested Parties," *The Journal of Law and Economics*, 2016, *59* (3), 527–548.

Fudenberg, Drew and David M Kreps, "Learning in Extensive Form Games, II: Experimentation and Nash Equilibrium," *Unpublished Paper*, 1994.

_ and Giacomo Lanzani, "Which Misperceptions Persist?," 2021.

Gagnon-Bartsch, Tristan and Matthew Rabin, "Naive social learning, mislearning, and unlearning," 2016.

_ , _ , and Joshua Schwartzstein, "Channeled Attention and Stable Errors," 2021.

Gentzkow, Matthew and Jesse M Shapiro, "Ideological Segregation Online and Offline," *The Quarterly Journal of Economics*, 2011, *126* (4), 1799–1839.

Gershman, Samuel J, "How to Never be Wrong," *Psychonomic Bulletin & Review*, 2019, *26* (1), 13–28.

Gibbons, Robert, "Deals That Start When You Sign Them," 2021.

_ and Laurence Prusak, "Knowledge, Stories, and Culture in Organizations," in "AEA Papers and Proceedings," Vol. 110 2020, pp. 187–92.

_ and Rebecca Henderson, "Relational Contracts and Organizational Capabilities," *Organization science*, 2012, *23* (5), 1350–1364.

_ and _ , "What Do Managers Do? Exploring Persistent Performance Differences among Seemingly Similar Enterprises," 2012.

Golub, Benjamin and Evan Sadler, "Learning in Social Networks," in "The Oxford Handbook of the Economics of Networks" 2016.

_ and Matthew O Jackson, "Naive Learning in Social Networks and the Wisdom of Crowds," *American Economic Journal: Microeconomics*, 2010, *2* (1), 112–49.

Guess, Andrew, Brendan Nyhan, Benjamin Lyons, and Jason Reifler, "Avoiding the Echo Chamber About Echo Chambers," *Knight Foundation*, 2018, *2*.

Guess, Andrew M, Neil Malhotra, Jennifer Pan, Pablo Barberá, Hunt Allcott, Taylor Brown, Adriana Crespo-Tenorio, Drew Dimmery, Deen Freelon, Matthew Gentzkow et al., "How do Social Media Feed Algorithms Affect Attitudes and Behavior in an Election Campaign?," *Science*, 2023, *381* (6656), 398–404.

Haghtalab, Nika, Matthew O Jackson, and Ariel D Procaccia, "Belief Polarization in a Complex World: A Learning Theory Perspective," *Proceedings of the National Academy of Sciences*, 2021, *118* (19), e2010144118.

Harrison, J. Michael and David Kreps, "Speculative Investor Behavior in a Stock Market with Heterogeneous Expectations," *The Quarterly journal of economics*, 1978, *92* (2), 323–336.

Heidhues, Paul, Botond Kőszegi, and Philipp Strack, "Unrealistic Expectations and Misguided Learning," *Econometrica*, 2018, *86* (4), 1159–1214.

Hirshleifer, David, "Presidential Address: Social Transmission Bias in Economics and Finance," *The Journal of Finance*, 2020, *75* (4), 1779–1831.

Hong, Harrison and Jeremy Stein, "Disagreement and the Stock Market," *Journal of Economic Perspectives*, 2007, *21* (2), 109–128.

_ **, Jeremy C Stein, and Jialin Yu**, "Simple Forecasts and Paradigm Shifts," *The Journal of Finance*, 2007, *62* (3), 1207–1242.

Hong, Lu and Scott E Page, "Problem Solving by Heterogeneous Agents," *Journal of Economic Theory*, 2001, *97* (1), 123–163.

Hüning, Hendrik, Lydia Mechtenberg, and Stephanie Wang, "Using Arguments to Persuade: Experimental Evidence," *Available at SSRN 4244989*, 2022.

Ichihashi, Shota and Delong Meng, "The Design and Interpretation of Information," *Available at SSRN 3966003*, 2021.

Kamenica, Emir and Matthew Gentzkow, "Bayesian Persuasion," *American Economic Review*, 2011, *101* (6), 2590–2615.

Kwon, Spencer, Joshua Schwartzstein, and Adi Sunderam, "Experimental Evidence on Model Persuasion," 2022.

Larson, Heidi, Louis Cooper, Juhani Eskola, Samuel Katz, and Scott Ratzan, "Addressing the vaccine confidence gap," *The Lancet*, 2011, *378* (9790), 526–535.

Levy, Gilat and Ronny Razin, "The Drowning Out of Moderate Voices: A Maximum Likelihood Approach to Combining Forecasts," *Theoretical Economics*, 2020.

Lin, Yu-Ru and Wen-Ting Chung, "The Dynamics of Twitter Users' Gun Narratives Across Major Mass Shooting Events," *Humanities and Social Sciences Communications*, 2020, *7* (1), 1–16.

Lombrozo, Tania, "Explanatory Preferences Shape Learning and Inference," *Trends in Cognitive Sciences*, 2016, *20* (10), 748–759.

Loughran, Timothy and Jay Ritter, "The New Issues Puzzle," *The Journal of Finance*, 1995, *50* (1), 23–51.

Mailath, George J and Larry Samuelson, "Learning Under Diverse World Views: Model-Based Inference," *American Economic Review*, 2020, *110* (5), 1464–1501.

McPherson, Miller, Lynn Smith-Lovin, and James M Cook, "Birds of a Feather: Homophily in Social Networks," *Annual Review of Sociology*, 2001, *27* (1), 415–444.

Miller, Edward, "Risk, Uncertainty, and Divergence of Opinion," *The Journal of Finance*, 1977, *32* (4), 1151–68.

Mullainathan, Sendhil and Andrei Shleifer, "The Market for News," *American Economic Review*, 2005, *95* (4), 1031–1053.

_ , Joshua Schwartzstein, and Andrei Shleifer, "Coarse Thinking and Persuasion," *The Quarterly journal of economics*, 2008, *123* (2), 577–619.

Murphy, Kevin M and Andrei Shleifer, "Persuasion in politics," *American Economic Review*, 2004, *94* (2), 435–439.

Nyhan, Brendan, Jaime Settle, Emily Thorson, Magdalena Wojcieszak, Pablo Barberá, Annie Y Chen, Hunt Allcott, Taylor Brown, Adriana Crespo-Tenorio, Drew Dimmery et al., "Like-Minded Sources on Facebook are Prevalent but not Polarizing," *Nature*, 2023, *620* (7972), 137–144.

Olea, José Luis Montiel, Pietro Ortoleva, Mallesh M Pai, and Andrea Prat, "Competing Models," *The Quarterly Journal of Economics*, 2022, *137* (4), 2419–2457.

Quattrone, George A and Edward E Jones, "The Perception of Variability Within In-Groups and Out-Groups: Implications for the Law of Small Numbers.," *Journal of personality and social psychology*, 1980, *38* (1), 141.

Roux, Nicolas and Joel Sobel, "Group Polarization in a Model of Information Aggregation," *American Economic Journal: Microeconomics*, 2015, *7* (4), 202–32.

Schkade, David, Cass R Sunstein, and Daniel Kahneman, "Deliberating About Dollars: The Severity Shift," *Colum. L. Rev.*, 2000, *100*, 1139.

_ , _ , and Reid Hastie, "What Happened on Deliberation Day," *Calif. L. Rev.*, 2007, *95*, 915.

Schulz, Laura E and Jessica Sommerville, "God Does not Play Dice: Causal Determinism and Preschoolers' Causal Inferences," *Child Development*, 2006, *77* (2), 427–442.

Schwartzstein, Joshua and Adi Sunderam, "Using Models to Persuade," *American Economic Review*, 2021, *111* (1), 276–323.

Shiller, Robert J, "Narrative Economics," *American Economic Review*, 2017, *107* (4), 967–1004.

_ , *Narrative Economics: How Stories go Viral and Drive Major Economic Events*, Princeton University Press, 2020.

Smith, Lones and Peter Sørensen, “Pathological Outcomes of Observational Learning,” *Econometrica*, 2000, *68* (2), 371–398.

Weick, Karl E, *Sensemaking in Organizations*, Vol. 3, Sage, 1995.

Yang, Jeffrey, “A Criterion of Model Decisiveness,” 2022.