NBER WORKING PAPER SERIES

SHARED MODELS IN NETWORKS, ORGANIZATIONS, AND GROUPS

Joshua Schwartzstein
Adi Sunderam

Working Paper 30642
http://www.nber.org/papers/w30642

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
November 2022

We thank Chiara Aina, Francesca Bastianello, Alessandro Bonatti, Tristan Gagnon-Bartsch, Ben Enke, Simone Galiperti, Robert Gibbons, Sam Hanson, 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, and UC Santa Barbara for helpful comments. We thank John-Henry Pezzuto and Emma Ronzetti for outstanding research assistance. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Shared Models in Networks, Organizations, and Groups
Joshua Schwartzstein and Adi Sunderam
NBER Working Paper No. 30642
November 2022
JEL No. D83,D85,D9,G40

ABSTRACT

To understand new information, we exchange models or interpretations with others. This paper provides a framework for thinking about such social exchanges of models. The key assumption is that people adopt the interpretation in their network that best explains the data, given their prior beliefs. An implication is that interpretations evolve within a network. For many network structures, social learning mutes reactions to data: the exchange of models leaves beliefs closer to priors than they were before. Our results shed light on why disagreements persist as new information arrives, as well as the goal and structure of meetings in organizations.

Joshua Schwartzstein
Baker Library 461
Harvard Business School
Soldiers Field
Boston, MA 02163
jschwartzstein@hbs.edu

Adi Sunderam
Harvard Business School
Baker Library 359
Soldiers Field
Boston, MA 02163
and NBER
asunderam@hbs.edu

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 social exchanges of interpretations. 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 network and settle on the interpretation they are exposed to that is most compelling. 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. That is, she picks the model she is exposed to 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); Weick (1995)); (iii) make the past feel more predictable (Schulz and Sommerville (2006); Gershman (2019)); and (iv) have the most "explanatory power" (Lombrozo (2016)).[1]

To see some key implications of this formulation, consider an example in the spirit of one introduced and analyzed by Aina (2021), where right-leaning voters are trying to assess the outcome of an election—both in terms of who technically won (i.e., received the highest certified vote tally) and whether the election was fair.[2] Right-leaning voters' priors are that the left-leaning and right-leaning candidates are equally likely to win, but the left-leaning candidate is more likely to win unfairly.[3] An example of such a prior is given in the following table:

μ_0	u	f
l	.375	.125
r	.125	.375

,

where l stands for the left-leaning candidate winning, r for the right-leaning candidate winning, u for the election being unfairly won, and f for the election being fairly won. The number in each cell corresponds to the prior likelihood of the row-column combination. After the election, data comes out: $h =$ "the left-leaning candidate won the certified vote tally with no official evidence of fraud". In reality, the data perfectly reveals that the state is $\omega = (l, f)$: the left-leaning candidate won a fair election. The data is closed to interpretation on who won the election—there is only one possible interpretation because the winner is by definition the candidate who received the highest certified vote tally. However, the data is open to interpretation on the election's fairness; there are many different ways to think about the implications for fairness of the lack of evidence of fraud.

Following the release of data then, right-leaning voters agree that the left-leaning candidate technically won the election, but may disagree about whether the election was fairly won because they use different models to interpret the data. Some right-leaning voters initially stick with the default interpretation that the data reveals that the left-leaning candidate won the election fairly.

[1]Recent work (e.g., Barron and Fries (2022), Kwon et al. (2022)) experimentally tests and find support for the assumption made by Schwartzstein and Sunderam (2021) that people find better-fitting models more persuasive.

[2]The analysis in Aina (2021) complements ours. It asks how a person (e.g., a political candidate) could send a menu of models on how to interpret data *ahead of its release* (e.g., the outcome of an election) to control agents' (e.g., voters') ultimate interpretations of the data. For example, a candidate's campaign could suggest the possibility that an election could be rigged to prime voters to consider this interpretation in the event that the candidate loses the election. In contrast, we focus on the spread of interpretations after the release of data.

[3]The analysis of left-leaning voters is symmetric.

Others, however, come up with different interpretations. Some entertain the possibility that the election system is imperfect: sometimes there is no evidence of fraud even though the election was unfair. Others think the election system is poorly administered, so that the fact that there is no official evidence of fraud is completely uninformative about the fairness of the election. Still others think the election system is rigged so that there will never be evidence of fraud if the left-leaning candidate wins.[4] Assuming the population is sufficiently large that roughly every interpretation consistent with the left-leaning candidate winning is someone's initial reaction, and that the network is sufficiently connected that the most compelling interpretation spreads throughout the population, we ask: which take goes viral?

The rigged-election model, which eventually all right-leaning voters will end up holding. This is the best-fitting model consistent with the left-leaning candidate winning, and it implies a low probability that the election was fair: under this model m^r (defined in footnote 4), $\Pr(f|h, m^r) = .25$. Social learning spreads an interpretation of the data under which right-wing voters' posterior on election fairness is the same as their prior, despite the fact that "no evidence of fraud" is explicitly in the data. As shown in Schwartzstein and Sunderam (2021), models that fit well tend to result in posterior beliefs close to prior beliefs. Intuitively, models that fit well imply the data is unsurprising, which means beliefs should not move much in response to it. In this example, right-leaning voters' prior is that the left-leaning candidate is unlikely to win fairly. The model that best fits the voters' knowledge (i.e., their prior and the data) leads them not to move from their prior at all. Thus, following social learning, right-leaning voters agree on the interpretation that the left-leaning candidate's victory is unsurprising because the election was likely unfair. In contrast, if the data had been $h =$ "the right-leaning candidate won the certified vote tally with no official evidence of fraud," these same voters would have adopted a different model. In this case, right-leaning voters all end up holding an interpretation of the data implying a high probability the election was fair because that interpretation makes the data as unsurprising as possible.

This example illustrates three main points. First, social learning *hardens* reactions to data that is open to interpretation: 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 would not find on their own. The fit of the model voters converge to is $\Pr(h|m^r, \mu_0) = 0.5$, four times the fit of the default model (.125). Second, interpretations *evolve* in ways that often make final interpretations

[4]Formally, we assume that under the default (and correct) interpretation, the data has a probability of one if the left-leaning candidate won fairly and zero otherwise: $\pi_d(h|(l, f)) = 1$ and $\pi_d(h|\omega) = 0$ for all $\omega \neq (l, f)$. Under "imperfect system" models, m^i, the data has positive probability if the left-leaning candidate won unfairly: $\pi_{m^i}(h|(l, u)) > 0$. Under "poorly administered system" models, m^p, the data has the same probability regardless of whether the election was fair: $\pi_{m^p}(h|(l, u)) = \pi_{m^p}(h|(l, f))$. And under the "rigged election" model, m^r, the data has a probability of one if the left-leaning candidate won, regardless of whether the election was fair: $\pi_{m^r}(h|(l, u)) = \pi_{m^r}(h|(l, f)) = 1$.

less accurate than initial reactions. In the example, many right-leaning voters initially have the correct interpretation, which results in the belief that the election was likely fairly won. However, the social exchange of models leads them to converge on interpretations that result in the belief that the election was unlikely to have been fair. 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 those built on motivated reasoning or preferences over beliefs. In these alternative formulations, if, for example, right-leaning voters prefer accounts that a left-leaning candidate could only win unfairly, then their *initial* reactions will exhibit that preference. A third point is that social learning not only has a tendency to harden reactions but also to *mute* them—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 relevant data.

Section 2 introduces the model. We say social learning hardens a person's reaction to the data when it leaves her with a model that better explains the data than she previously could. We say that social learning mutes a person's reaction to the data when it leaves her with a model under which her posterior beliefs are closer to her prior than they were before social learning. We first establish a basic result for the case where people are willing to entertain roughly every possible interpretation of the data and everyone is exposed to everyone else's model (i.e., the network is complete). In this case, everyone adopts a model that perfectly explains the data, which implies there is nothing to learn from it. Thus, social learning maximally hardens and mutes a person's reaction to the data—people end up convinced that they perfectly understand the data and that their prior beliefs are consistent with this understanding.

This stylized case captures one important feature of reality, highlighted in the voting example above: when new data is released, interpretations initially diverge before converging as people are exposed to other interpretations within their network, and the interpretations that the network converges on tend to pull beliefs back towards views people held before seeing the data. For instance, commentators have noted the stability of political polls in recent years. Views have also been very stable on other issues like the health risks of Covid-19: despite wide swings in the data (e.g., case rates, hospitalizations, and deaths), 82-85% of Democrats consistently felt Covid was a significant health risk between May 2020-February 2021, while only 41-46% of Republicans felt so over the same time period.[5] 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 discussing reactions to news about Covid-19, New York Times writer Charlie Warzel observed: "a story comes out about a study/specific spreader event/ whatever & it's like 1) immediate intense reactions followed by 2) 36 hrs of long threads by

[5]See, e.g., Pew surveys here and here.

smart & not smart/qualified & not qualified people picking apart/casting doubt & 3) usually calm consensus later in the week".[6] In our framework, this disconnect between the data and long-run beliefs is driven by the adoption of models through social learning.

We next turn to the impact of network structure on interpretations and beliefs. Section 3 studies networks formed on the basis of shared beliefs, where people exchange models with others who had similar initial reactions to the data. In the voting example above, suppose voters who initially interpret the data as suggesting the election was unfair all talk to each other, while voters who initially interpret the data as suggesting the election was fair all talk to each other. We show that within each network social learning leads beliefs to converge to the initial reaction in the network that is closest to the prior. Members of the "election was fairly decided" and the "election was unfairly decided" networks will continue to disagree, but less so over time as all right-leaning voters converge on models that bring their beliefs closer to the 25% prior probability they attached to the left-leaning candidate winning fairly. We also show that shared belief networks can lead to polarization of beliefs across multiple issues. If networks are formed based on one issue (e.g., the environmental impact of genetically modified crops), exchange of interpretations leads to the convergence of within-network beliefs on a second issue (e.g., the safety of genetically-modified crops). In other words, beliefs across issues become more uni-dimensional after social learning.

This analysis begs a question, which we tackle in Section 4: how can differences in beliefs across networks persist when there is some communication across networks? We draw 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. Weak exposure is more effective in moving beliefs before social learning than after. By supplying people with interpretations that fit the data well, social learning inoculates them against finding compelling models that support alternative beliefs.

Armed with these results, Section 5 then considers how someone could shape the communication network to her advantage. For example, a manager can shape communication within a firm by forming specific teams or holding meetings that invite a select group. Activists may form grass-roots groups based on certain shared beliefs. Twitter users shape networks in their choice of which voices to amplify by re-tweeting. A key result is that the optimal structure of the network depends on the shaper's objective. If the shaper cares more about reaching consensus than about the specific conclusion reached, then she wants everyone to share interpretations with each other. But this

[6] https://mobile.twitter.com/cwarzel/status/1421177475111931904

approach will favor the conclusion that there is little to learn from the data. If the shaper wants people to take a particular action, then she wants to expose people to all models that support that action. She should crowdsource arguments and see what resonates instead of using one specific argument. Put differently, a network shaper who supports a particular action is better off using a collection of individuals (i.e., a platform) to articulate arguments for taking that action rather than using any given individual. The shaper also wants to prevent the audience from hearing arguments that the data is unsurprising. Such arguments will be compelling because they fit the data given priors well and will lead people to conclude the status quo should prevail.

We then spell out some applications of our results. In Section 6, we consider why disagreement persists in the face of new information. Why do misconceptions survive in some groups when people have access to so much high-quality information? What role do ideological bubbles play, given that people have diverse news diets and do not appear to systematically avoid counter-attitudinal information (Gentzkow and Shapiro (2011); Guess et al. (2018))? In financial markets, why are there groups of persistent optimists ("bulls") and pessimists ("bears") about the same stock (Diether et al. (2002); Hong and Stein (2007))? Our framework offers a simple explanation, complementing recent models that instead highlight the role of social media echo chambers (Bowen et al. (2021)): Within a network or bubble, people are exposed to crowdsourced models that evolve to fit the data better and better, making them more compelling and resistant to change. In our framework, bubbles do not prevent people from being exposed to the right take on an event, but, by hardening reactions, 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. We document stylized evidence from a social-media network for stock market investors consistent with our model's predictions: Investors who are bullish about a company become less bullish in the immediate aftermath of a negative earnings surprise, but quickly revert back to being bullish; bearish investors similarly become less bearish in the immediate aftermath of a positive surprise, but quickly revert back.

Section 7 studies 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 results on shaping communication have strong implications for meeting structure. A manager focused on keeping workers on the same page, for example if there is a strong coordination motive, wants to have a very open flow of communication. When new data arrives (e.g., someone is surprisingly denied a promotion or earnings are lower than expected), the manager should allow everyone to share interpretations because they will settle on the view that there is little to learn from the event. On the other hand, a manager focused on shifting workers' beliefs in response to an event, for example if she seeks to manage change, must control the flow of communication. She wants to call a meeting where only interpretations supporting desired conclusions are voiced. Even with such strong control, however, the manager will not be able to get everyone on the same page, perhaps shedding light on why organizations find it difficult to reach shared understandings that differ from the status quo (e.g., Gibbons and Henderson (2012a)).

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. Increasing connectedness tends to untether beliefs from relevant data 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 such as Shiller (2017) and Hirshleifer (2020) have called for studying the social transmission of narratives in economics and finance.[7] These

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

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. 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)).[8] 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.

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

[8]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).

each agent to all models held by other agents in her social networks. After social learning, each agent adopts the model that best explains the data from the full set of models she has been opposed 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.[9] 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 $\{\pi_d(\cdot|\omega)\}_{\omega\in\Omega}$,[10] (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 social network, where we let $M_i \subseteq M$ denote the set of models proposed by someone in i's social 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 goes with 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 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 $\bar{M}(h, \mu_0, d, M)$.[11] Further assume that that population is large enough that, for each model

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

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

[11] 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

$m \in \bar{M}(h, \mu_0, d, M)$, someone in the population generates that model herself.

In the typical case, we set the default interpretation to be the true-model interpretation, $d = m^T$. We also typically let M be the set of all possible models M^a—i.e., 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 $\bar{M}(h, \mu_0, m^T, M^a)$.[12]

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 which 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 members. 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, including both directed networks, where the flow of communication goes one way, and undirected networks, where it goes two ways.

Third, implicit in the idea that a person is exposed only to the models of those 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)), thinking there is not much reason to

social learning, it would not influence the distribution of beliefs following social learning.

[12] One technical issue arises when M^a is the set of all models. In this case, even assuming a continuum of individuals, the space $\bar{M}(h, \mu_0, m^T, M^a)$ may be too large to guarantee that, for every model in $\bar{M}$, there exists a person who holds that model before social learning. For readers concerned about such cardinality issues, we note that 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.

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 think once she has heard one such argument she has heard them all, perhaps underappreciating the diversity of these arguments.

2.3 Examples

Example 1 (Interpreting data about policy issues)**.** We now sketch two brief examples, which we will return to throughout the paper.

The first involves interpreting data about a binary state space, $\Omega = \{l, r\}$. The history h can take on two values, (h^l, h^r), and the data fully reveals the state under the true model: $\pi_{m^T}(h^l|l) = \pi_{m^T}(h^r|r) = 1$. Further assume people are maximally open to persuasion, $M = M^a$, and the default model is the true model, $d = m^T$. The payoffs U_i are such that $a = L$ is optimal if $\mu(l) \geq .75$, $a = M$ is optimal if $\mu(l) \in (.25, .75)$, and $a = R$ is optimal if $\mu(l) \leq .25$.

One application of this example relates to 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. The prior over states is prior $\mu_0(l) = 1/2$. Further suppose that people can take three possible actions, $a \in \{L, M, R\}$,where action $a = L$ is to vote Democrat, $a = M$ is to abstain from voting, and $a = R$ is to vote Republican. 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 ($a = M$).

Another application applies to firms. In state $\omega = l$, it is optimal for the firm to cut costs, and $\omega = r$ for it is optimal to invest in growth. Action $a = L$ is to fire employees, $a = M$ to stay the course, and $a = R$ to hire employees.

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, in the introduction, people updated both about who won the election and whether the election was fair. Similarly, 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).

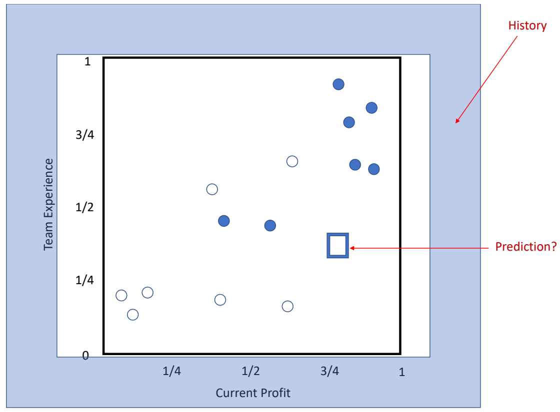
Example 2 (Interpreting data about startups)**.** Our second example involves investing. Relative to the first example, there are two qualitative differences. First, this example highlights the role of the data more clearly. Second, it illustrates how restricting the model space impacts our results.

Consider a community of venture capitalists trying to predict the success of a startup in a new sector (e.g., cryptocurrency) 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 1a 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.

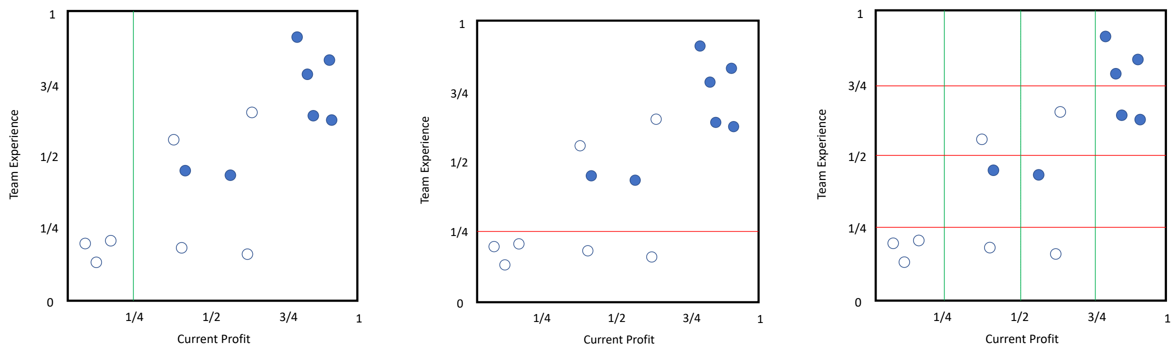
We assume there are four types of models in the model space M. First, all venture capitalists start with the default model that all startups in the new sector have the same success probability regardless of their characteristics. Second, there are models that are cutoff rules in profit: all startups with profit below the cutoff share the same success probability and all startups with profit above the cutoff share the same success probability.[13] For instance, the vertical green line in Figure 1b 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 1c 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. Formally, under the model m^{ind}, $\Pr(y|x_3, m^{ind}, \mu_0) = 1$ for $y \equiv (y_j)_j$ and $x_3 \equiv (x_{3j})_j$.

Prior to any social learning, venture capitalists consider the default and one other model randomly selected from the other three model types so long as it fits better than the default. As shown in Figure 1d, 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.

[13]Formally, success probabilities below and above the cutoff are independently drawn—once and for all—from the uniform distribution.



(a) Setup



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

Figure 1: Predicting the success of a startup

2.4 Basic 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 $\mu(h, m')$. Following social learning, the person adopts the model

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

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* the person's reaction. We say that social learning *mutes* a person's reaction to data when it moves the person's beliefs closer to her prior. Formally, following Schwartzstein and Sunderam (2021), let $\text{Movement}(\tilde{\mu}; \mu_0) \equiv \max_{\omega \in \Omega} \tilde{\mu}(\omega)/\mu_0(\omega)$ be a measure of the change in beliefs from prior μ_0 to posterior $\tilde{\mu}$. Social learning mutes reactions to the data when $\text{Movement}(\mu_i; \mu_0) \leq \text{Movement}(\mu_i'; \mu_0)$. When social learning does not mute a person's reaction to data, we say it *intensifies* the person's reaction.

A simple observation is that social learning must harden a person's reaction 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.

3 Social Exchange of Models

3.1 Initial Reactions

We start by briefly describing people's initial reactions before social learning. There are two key points. First, before social learning, people have a variety of reactions to the data. Second, there will be a greater variety of initial reactions when people are more surprised by the data in the sense that the default model provides a poorer fit to the data, fixing prior beliefs.

Let $\bar{\Delta}(h, \mu_0, d, M) = \{\mu \in \Delta(\Omega) : \mu = \mu(h, m) \text{ for some } m \in \bar{M}(h, \mu_0, d, M)\}$ 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 1. *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 $\bar{\Delta}(h, \mu_0, d, M) = \bar{\bar{\Delta}}(h, \mu_0, d, M)$.

Proposition 1, which is essentially a restatement of Proposition 1 in Schwartzstein and Sunderam (2021), characterizes the set of initial reactions to the data.[14] From this result, it is immedi-

[14]To derive the distribution over initial reactions, we need to additionally specify the distribution over models people come up with on the fly. When people instead talk to each other, many of our results are independent of the choice of this distribution.

ate that 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.

To illustrate, return to the voting example from the introduction. Proposition 1 implies that some right leaning voters initially react to $h =$ "the left-leaning candidate won the certified vote tally with no official evidence of fraud" by concluding that the election outcome was surely unfair. Right-leaning voters are so surprised by the idea that a left-leaning candidate would win fairly that their initial reactions are all over the place, i.e., they are unconstrained by Proposition 1. Formally,

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

where the equality follows from $\mu_0(l,u)/\Pr(h|d,\mu_0) = .375/.125 = 3$ and $\mu_0(l,f)/\Pr(h|d,\mu_0) = .125/.125 = 1$. In other words, the inequality constraint never binds because the right-hand-side is above 1 for all states consistent with the left-leaning candidate winning.

On the other hand, if the data were $h =$ "the right-leaning candidate won the certified vote tally with no official evidence of fraud," Proposition 1 says that *no* right-leaning voter would initially react by concluding the election outcome was surely unfair—right-leaning voters are *not* surprised by the idea that a right-leaning candidate would win fairly, so their initial reactions are more concentrated around their default interpretation. Formally,

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

Given h, the highest probability right-leaning voters could initially attach to the right-leaning candidate winning unfairly given that she won equals $(1/3)/(1/2) = 2/3$.

Since the set of beliefs following social learning are a subset of initial beliefs, reactions to the data following social learning are subject to the same constraints in Proposition 1 as initial reactions. The next section begins a more in-depth analysis of how social learning shapes beliefs.

3.2 Complete Networks

We begin by establishing a result for the case where the network is "complete" in the sense that every person is exposed to everyone else's model. (All proofs are in Appendix A.)

Proposition 2. *Suppose everyone is maximally open to persuasion and talks to every person:* $M = M^a$ *and* $M_i = M^a$ *for all* i*. Then social learning mutes everyone's reaction to the data: for every person* i*, Movement*$(\mu_i; \mu_0) \leq$ *Movement*$(\mu_i'; \mu_0)$*. Indeed, social learning maximally mutes and hardens everyone's reaction to the data in the sense that each person sticks with their prior belief with a model that perfectly explains the data: for every person* i*,* $\mu_i = \mu_0$ *and* $\Pr(h|m_i, \mu_0) = 1$.

This result says that if everyone talks to each other, then following social learning they do not react to the data at all because they view it as inevitable in hindsight. Given that the world is large, someone will come up with the model that explains why whatever happened was bound to happen. That model will spread throughout the network, and since the network connects everyone, everyone will adopt it. The fact that a model that fits the data well will be broadly adopted in turn means that beliefs will move very little. Intuitively, models that fit well imply the data is unsurprising, which means beliefs should not move much in response to it.

In the context of the policy example above, this result says that a given person's reaction to an event might push her to favor taking actions $a = L$ or R, but after being exposed to many different arguments she will go back to favoring the status quo of $a = M$. For instance, a school shooting might initially lead people to think we need a change in gun-control policies, but they will eventually favor interpretations that say we did not learn much from the shooting. Indeed, there is empirical evidence of exactly 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)). This pattern is harder to explain in models of motivated beliefs, which would instead suggest that people who are initially against gun control would immediately (i.e., before social learning) come up with ways to view the mass shooting as confirming prior beliefs against gun control.

More generally, our model predicts that following a realization of new data that is open to interpretation, there is initially a broad divergence of opinion followed by convergence as people share their interpretations and settle on commonly believing they learned little from the data. Models evolve though social learning to fit the data better, leading people's beliefs to move less and less.

While Proposition 2 considers the impact of social learning when *everyone* talks to each other, people are often embedded in smaller networks.

3.3 Shared Belief Networks

We next turn to networks formed on the basis of shared beliefs.[15] For instance, networks are formed based on beliefs that one political party typically governs better than others, that vaccines are harmful, and that free markets lead to prosperous societies.

To analyze such networks, suppose 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*, $M_i = \left\{m \in \bar{M}(h, \mu_0, d, M) : \mu(h, m) \in s(\mu(h, m_i'))\right\}$ for every person i.

Given our assumption of common priors, this definition, taken literally, 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. While this literal interpretation is a reasonable approximation of reality for certain events, such as the earnings announcements we consider in Section 6, in many instances networks based on shared beliefs are probably 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. In such cases, a broader interpretation setup is appropriate—shared-belief networks are formed among people who initially hold similar beliefs about some question of interest, whether or not these similar beliefs arise literally from having a common initial reaction to the most recent event.[16]

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

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

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.

[15] Appendix B analyzes networks instead formed on the basis of shared models.

[16] We will more formally capture this idea in briefly studying dynamics in Section 6.

Proposition 3. *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 to share the initial belief within their network that is closest to the prior: for every person i, $\mu_i \in \arg\min_{\mu \in s(\mu_i')}$ Movement$(\mu; \mu_0)$.*

This result says that a person who only exchanges models with others who react similarly to data ends up at a belief that reacts least to the data among those that are shared with her. By the earlier lemma, such a belief is supported by a better-fitting model than any other she is exposed to.

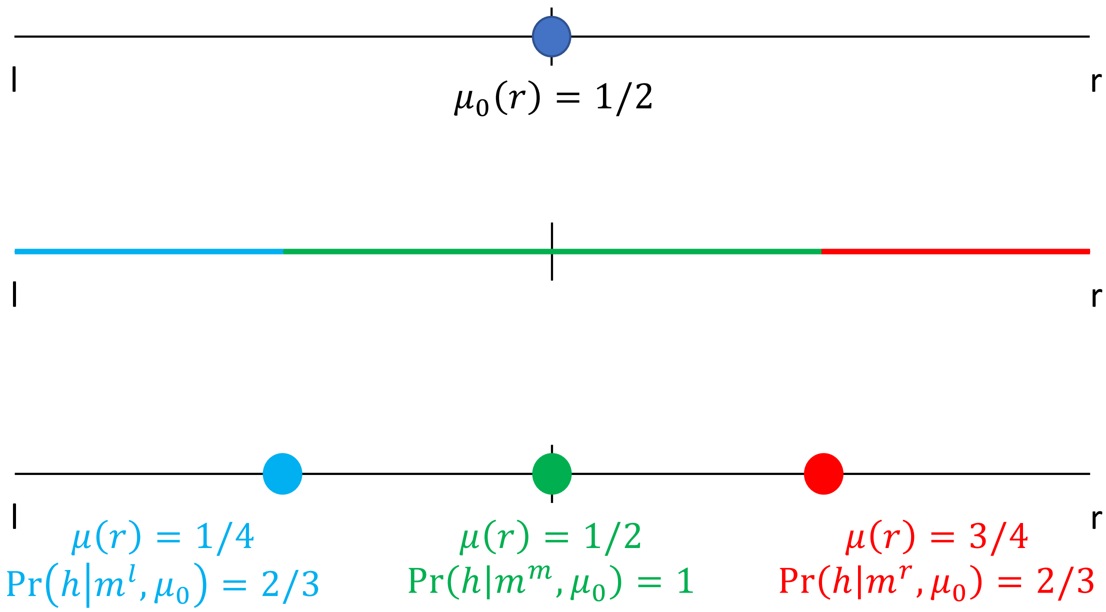
An implication of these results is that while people in different networks will end up with different interpretations of the same data, no one will believe in coincidences (i.e., that the data are random). For instance, suppose an athlete performs poorly in an event and it was raining during the event. While in truth rain and performance may be unrelated, a network formed around the shared belief the athlete is good will tend to settle on the interpretation that the poor performance should be ignored because it was raining. In contrast, a network formed around the shared belief the athlete is bad will tend to settle on the interpretation that poor performance in the rain is perfectly consistent with the athlete being bad. Thus, our results in some sense generalize the idea of conspiratorial thinking in settings with social learning: Narratives that spread through a network will tend to link events that under many gut reactions (and the true model) are unconnected in a way that is consistent with the group's initial views.

3.4 Examples

Interpreting Data About Policy Issues

As an illustration of Proposition 3, suppose that in the public-policy example, $\mu_0(l) = 1/2$ and shared-belief networks are formed based on views of the optimal action: Everyone with an initial reaction supporting a right-leaning action ($\mu_j'(r) \in [.75, 1]$) is in one network, everyone with an initial reaction supporting the neutral action ($\mu_j'(r) \in (.25, .75)$) is in another, and everyone with an initial reaction supporting the left-leaning action ($\mu_j'(r) \in [0, .25]$) is in the final network. Then Proposition 3 says, and Figure 2 illustrates, that everyone in a given network will end up at the belief that is closest to the prior within her network. 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 will end up muting her reaction to $\mu(r) = .75$. Each person's reaction will also be hardened: everyone in the right-leaning network ends up adopting a best-fitting model supporting the right-leaning action, namely a model satisfying $\Pr(h|m^r, \mu_0) = 2/3$ (by Lemma 1).

Figure 2: Evolution of Beliefs Across Shared-Belief Networks Surrounding a Single Policy Issue



As in the special case where a person talks to everyone, with social learning in shared-belief networks, people's final beliefs move less than their initial reactions, but they become more certain that their interpretations of the data are correct. These dynamics may help explain the recent stability of polls on political issues, as discussed in the introduction. If voters are exchanging interpretations of data within shared-belief networks, their beliefs will not respond much to that data.

In addition, Proposition 3 illustrates that, within any given network, social learning leads beliefs to converge. That is, beliefs within any network become more homogeneous. However, across networks, beliefs remain divergent with everyone becoming more confident in their reaction than before social learning. 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.[17]

[17] 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 3 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.

Proposition 3 also has implications for political polarization. Consider an extension of the example where there are multiple issues, but networks are formed based on shared beliefs about one of them. Let $\Omega = \Omega^1 \times \Omega^2$ and describe marginal beliefs over Ω^j by μ^j. Then networks are formed based on shared beliefs over issue 1 but not issue 2 when $s(\mu)$ depends only on μ^1.

In particular, let $\Omega^1 = \{l, r\}$ be whether a left- or right-leaning candidate governs better and $\Omega^2 = \{n, y\}$ be whether we are (y) or are not (n) in the sort of crisis that requires the expertise of scientists. Networks are formed given beliefs over $\{l, r\}$ but not $\{n, y\}$: suppose people with initial beliefs $\mu_i^{1'}(l) \geq .75$ are in one network (the "left-leaning network"), those with initial beliefs $\mu_i^{1'}(l) \in (.25, .75)$ are in another (the "centrist network"), and those with initial beliefs $\mu_i^{1'}(l) \leq .25$ are in another (the "right-leaning network").

Even though beliefs over the second issue do not influence network formation, final beliefs over that issue differ across networks when prior beliefs are correlated across the issues. For example, people might believe that left-leaning candidates tend to govern better at times when a crisis requires the expertise of scientists:

μ_0	n	y
l	.125	.375
r	.375	.125

.

In this case, the movement-minimizing belief among members of the left-leaning network is

$\mu^{\text{left-leaning}}$	n	y
l	$.75 \cdot .25 = .1875$	$.75 \cdot .75 = .5625$
r	$.25 \cdot .75 = .1875$	$.25 \cdot .25 = .0625$

,

the movement-minimizing belief among members of the centrist network is the prior, while the movement-minimizing belief among members of the right-leaning network is:

$\mu^{\text{right-leaning}}$	n	y
l	$.25 \cdot .25 = .0625$	$.25 \cdot .75 = .1875$
r	$.75 \cdot .75 = .5625$	$.75 \cdot .25 = .1875$

.

By Proposition 3, $\mu^{\text{left-leaning}}$ is the shared final belief among members of the left-leaning network, μ_0 is the shared final belief among members of the centrist network, and $\mu^{\text{right-leaning}}$ is the shared final belief among members of the right-leaning network. While members of the left-leaning network will view the data as suggesting the likelihood of a crisis is $\mu^{\text{left-leaning}}(y) = .5625 + .0625 = .625$, members of the right-leaning network will view this same data as suggesting the likelihood of a crisis is $\mu^{\text{right-leaning}}(y) = .1875 + .1875 = .375$. 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 it is likely there is a crisis that requires the expertise of scientists. Conversely, sharing models that suggest the right-leaning candidate is better at governing leads members of the network to also interpret the same data as suggesting that it is unlikely there is a crisis. In other words, the belief that there is a crisis becomes a "spurious justification" for the belief that the left-leaning candidate will govern better. Agents who interpret the data as supporting the left-leaning candidate will diagnose the data along other dimensions in a way that justifies that candidate.

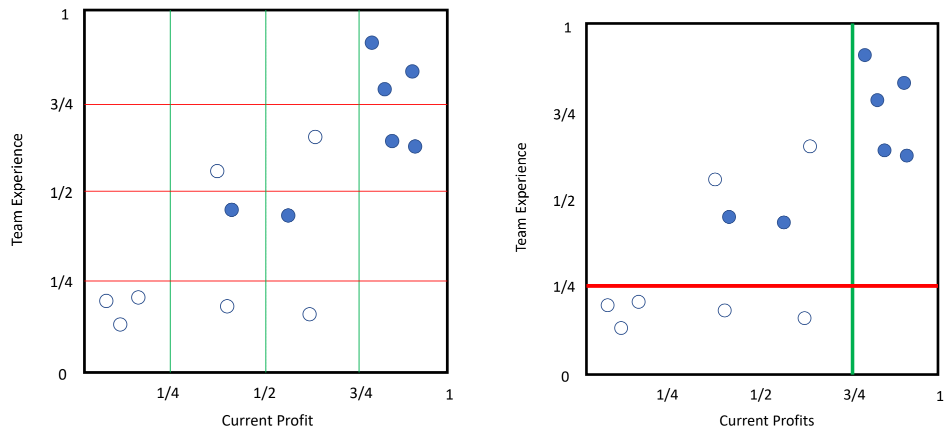
These results illustrate how networks based on one issue shape views on connected issues, perhaps shedding light on the so-called "polarization of reality" documented by Alesina et al. (2020). They show how the political left and right differ in their perceptions of factual issues, for example on the 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 is likely to govern better, as is naturally the case with social mobility since it connects to policy. 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. For instance, since views on sports are not naturally connected with politics, views on which teams are most promising do not become more correlated with political leanings through social learning.

Interpreting Data About Startups

Next return to the startup example and 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 believe 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 that best fits the data within that network. 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 1b and 1c. The experience-based model in Figure 1c more effectively separates successes from failures than does the profit-based model in Figure 1b.[18] Thus, after social learning, all optimists adopt the experience-based model,

[18]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 $(\int_0^1 (1-\theta)^3 d\theta) \cdot (\int_0^1 \theta^7(1-\theta)^4) \approx .000063$.



(a) Initial models (b) Models after social learning

Figure 3: Evolution of Beliefs Across Shared-Belief Networks Surrounding Startup Success

depicted by thick-red horizontal line in Figure 3b. 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.[19]

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 3b. Given the data, members in the pessimist network forecast average startup success to be $3/4 \cdot ((2+1)/(9+2)) + 1/4 \cdot ((5+1)/(5+2)) \approx .42$.

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 may evolve in very different ways across networks. Members of the optimist network initially disagree about whether experience or profit matter; some believe experience matters, while others believe profit matters. Yet all come to believe that startup

[19]This example illustrates that with a limited model space, muting does not always occur. In this case, optimists' beliefs are more optimistic after social learning than before. However, we show in simulations in Appendix C that even in this example social learning likely mutes the data on average (i.e., across different draws of startup characteristics).

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. Members of the neutral network come to believe that success is unpredictable *ex ante* because individuating characteristics are all that matter. Thus, the network structure influences the models and beliefs people in each network end up with.

4 Making People More Connected

This section analyzes the impact of making people more connected, addressing two questions. First, how can differences in beliefs across networks persist when there is some communication across networks? Our main result here is 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. The second question we ask is whether beliefs become more accurate when people are exposed to a richer set of models through network expansion. In contrast to standard information-based theories of social learning, we show that larger networks do not generally lead to more accurate beliefs.

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 $\tilde{\mu}$. 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-vax" networks who believe otherwise. However, this person is likely only aware of a thin slice of anti-vax arguments.

Proposition 4. *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, there are an infinite number of ways for this weak exposure to have no impact on the person's final belief. 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 this impacts the person's final beliefs 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 4 implies that weak exposure to beliefs outside a person's network is never guaranteed to impact her beliefs. The second part of Proposition 4 implies that strongly exposing a person to an alternative belief has at least as much impact on her ultimate beliefs and behavior as weakly exposing her to that alternative belief. 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, despite the fact that there is communication across networks. We think of cross-network communication as weak exposure. People might exchange models in addition to 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 4 says that weak exposure to anti-gun control arguments is unlikely to move the beliefs in the pro-gun control network. While a person might become convinced by listening to a broad set of arguments for a position, she is unlikely to be convinced by a narrow subset of the arguments (or simply a statement of the position itself).[20]

Another way of thinking about Proposition 4 concerns what it means to get someone out of an ideological bubble. Strongly exposing a person to a belief could be thought of as getting her outside of a bubble by immersing her in the arguments of people with different beliefs, while weakly exposing a person to a belief could be thought of as just making her aware that someone outside of her network holds that belief. Under this interpretation, Proposition 4 suggests that the former is more effective at changing minds because it increases the diversity of arguments a person is exposed to that support a belief.[21]

Exposure to alternative beliefs is also more effective when it comes before social learning, since it may impact which network a person joins. To see this, consider shared-belief networks. Imagine

[20] A different reason why beliefs might not converge across networks is that, after engaging in social learning within a network, a person's posteriors may become her priors for evaluating arguments outside the network—i.e., she might evaluate model m based on $\Pr(h|m,\mu(h,m_i))$, rather than $\Pr(h|m,\mu_0)$. This is a sort of confirmation bias, which would advantage models supporting beliefs close to $\mu(h,m_i)$. We explore a related possibility when briefly considering dynamics in Section 6.

[21] 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)). 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).

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 5. *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 being weakly (strongly) exposed to a belief after social learning impacts person i's final beliefs, $\mu_i^e \neq \mu_i$, then being weakly (strongly) exposed to a belief before social learning impacts person i's final beliefs, $\mu_i^p \neq \mu_i$. However, the converse does not hold.*

This result says that exposing someone to an alternative belief is more likely to have an impact on her final beliefs if this exposure comes before the person exchanges models with others in a shared-belief network. The reason is simple: as we saw before, social learning hardens a person's reaction to data. As we see from this result, such hardening inoculates the person against finding models supporting alternative beliefs compelling. This force may shed light on why firms often have "culture training" for new employees, for example, presumably in an attempt to influence what they come to believe after having a chance to talk to their colleagues.

We next turn to analyzing the impact of network expansion more broadly.

4.2 Expanding Networks

Next consider network expansions through merging networks: Expanding person i's network by merging it with $\tilde{M}$ enlarges the set of models that are shared with person i to $M_i \cup \tilde{M}$. This exercise helps assess the impact of increasing social connectedness.

Proposition 6. *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$(\mu_i; \mu_0)$.*

The first part of Proposition 6 shows that expanding a network always (weakly) hardens a network member's beliefs. The most basic impact of increasing connectedness in our model, for example by expanding the set of people a person discusses or debates a set of data with, is increasing the person's view that she can explain the data.

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. In the limit where the person is exposed to all models, we saw from Proposition 2 that 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 highlight the differences between our framework and typical information-based theories of social learning, in which increasing connectedness tends to lead to more accurate beliefs. In our setting, larger networks increase the chances of hearing an interpretation that suggests the data is perfectly consistent with a person's prior and hence there is no need to update. These results speak to concerns about the increased connectedness between people generated by social media, which we discuss in Section 6 below.

5 Managing Networks

Here we ask how someone could try to shape communication to her advantage. For instance, the shaper could write a book, form groups based on certain shared beliefs/experiences/interests, or hold meetings that invite a select group of people. The shaper 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 manager might insist on being in all meetings with certain subordinates.

5.1 Promoting Specific Actions

Suppose first that the network shaper wants to encourage people to take some action in response to the data. For example, in response to a school shooting, the shaper might want to encourage interpretations that either support tightening gun control, the status quo, or loosening gun control. Or, following unexpectedly low earnings, a leader 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 shaper'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 shaper 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 7. *Suppose each person has a finite action space and the network shaper's objective is a strictly monotonically increasing function of the fraction of people who choose her ideal action*

$a^s \in A$. The network shaper 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 network shaper's objective is maximized by setting

$$M_i = \left\{ m \in \bar{M}(h, \mu_0, d, M) : a(\mu(h, m)) = a^s \right\} \tag{1}$$

for all i. The network shaper's objective continues to be maximized by adding to M_i specified in Eq. (1) *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 network shaper wants to expose people to all models that support taking action a^s and no other models, except perhaps ones that fit the data plus people's priors worse than the best-fitting model supporting a^s. That is, the shaper 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 may want to control interpretations of earnings announcements by disproportionately calling on bullish analysts in earnings calls (Cohen et al. (2020)). If the exact person who would communicate the best-fitting model supporting action a^s were known, the shaper would do as well by having everyone just listen to this person. However, the shaper may not be able to identify this person ahead of time. The network shaper does no worse by exposing everyone to the arguments of people who support action a^s.

This result suggests that a network shaper who supports a particular action is better off by using a collection of individuals—a platform—to articulate arguments for taking that action than any single individual. A person who wants people to react to recent election results by concluding there is election fraud does better by crowdsourcing arguments from people who have reached this conclusion than by tapping a single person to argue. Communities of anti-vaxxers or conspiracy theorists are more persuasive than (almost) any single person. A platform of different contributors (say Breitbart) will tend to be more influential than any single contributor (say Steve Bannon), even if that contributor reaches the same audience. The reason is that increasing a person's exposure to a broad range of arguments supporting a given conclusion makes it more likely that she will find one of those arguments compelling than if she is exposed to only a few of those arguments.

We can say more if people are maximally open to persuasion, which we will assume for the rest of this section. Under the optimal network from the perspective of the network shaper, she is only able to get everyone to take her desired action if it is the action people would take in the absence of data—that is, if a^s is the status-quo action $a(\mu_0)$. If a^s is this status-quo action then the best-fitting model among actions that support a^s is the one that says the history was inevitable, $\Pr(h|m, \mu_0) = 1$, which everybody will adopt. If a^s is not this status-quo action, then the best-fitting model that supports a^s has an associated likelihood $\Pr(h|m, \mu_0)$ that is bounded away from one, so a positive measure of individuals will stick with models they came up with in the absence of social learning. These models have greater associated likelihoods and support sticking with the status quo. Thus, the network shaper is at an advantage if she wants everyone to stick with the status quo in response to the data. This is true even if the right interpretation of the data is that it supports taking a different action.

To illustrate these results, take the binary example above with $\mu_0(l) = 1/2$ and $h = h^l$. The network shaper who wants people to choose $a = L$ would want to form networks of small numbers of people who, in the absence of conversation, would choose $a = R$ in a sea of people who would, in the absence of conversation, choose $a = L$. For example, if the networks were of the form "all $\mu(l) \geq .75$ talk to a single person with $\mu(r) > .75$", then that person would end up believing $\mu(l) = .75$. An equally effective network would be one where everybody just listens to anyone who, in the absence of conversation, would choose $a = L$. The most important thing for the network shaper is that people are not exposed to arguments from many people who support the status quo—i.e., those with $\mu(l) \in (.25, .75)$.

Or imagine the status quo is not to take action, the network shaper wants to promote action, and there is an open-to-interpretation event that could lead to action one way or the other. For example, the status quo is some amount of gun control and a school shooting could lead to loosening or tightening gun-control restrictions. Imagine further that the network shaper supports left-leaning action, e.g., gun control. The people the shaper most wants to silence are moderates who argue for inaction, whether or not they are left- or right-leaning. The shaper wants people arguing for left-leaning action to speak and everyone else to listen. And, continuing this logic in a trivial dynamic extension, once all the people arguing for left-leaning action have discussed issues with each other enough to harden beliefs, the shaper is not worried about them having bilateral conversations with reactionaries on the other side—but they would still be wary of them having bilateral conversations with those who support the status quo.

5.2 Promoting Shared Models and Actions

We see from the discussion above that, unless the desired action is the status-quo, promoting specific actions typically conflicts with promoting shared models and actions. And a network shaper will sometimes benefit from promoting shared models and actions, for example if she derives benefits from people coordinating on their actions.

To analyze the case where the network shaper wants to promote shared models and actions, suppose the shaper'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 shaper prefers 75% of individuals to hold one model and 25% the other over 60% holding one model and 40% the other, over

50% of individuals holding one model and 50% the other, etc.

Proposition 8. *Suppose the network shaper'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 network shaper cannot do better than, for every person i, exposing her to all models: That is, the network shaper's objective is maximized by setting for all i*

$$M_i = \bar{M}(h, \mu_0, d, M). \tag{2}$$

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

6 Application 1: Sustained Disagreement in the Face of New Information

Why do people disagree when the Internet and social media give them access to high-quality information? Echo chambers are a common answer to this question. Under the echo-chamber view, while people could access high-quality information, their actual media diets and social networks only expose them to information consistent with their existing beliefs.

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. Thus, 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 and thus more resistant to change. In sharp contrast to the echo-chamber view, social networks are not insulating people from certain information; they are exposing people to interpretations of that data that favor certain beliefs and inoculating them against finding alternative beliefs compelling. Thus, in our framework, the primary impact of social networks is not to further polarize beliefs, but to harden and make them resistant to change. In particular, even if beliefs react a lot in the immediate aftermath of a big event, networks lead members to adopt interpretations that mute and harden their reactions. 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.[22]

We explore this idea in two settings. First, using a dynamic extension of the baseline model, we discuss a particularly extreme and persistent type of disagreement: beliefs in misconceptions (e.g., genetically modified crops and vaccines are dangerous) and conspiracy theories (e.g., QAnon), which are often tenaciously held by a subset of the population. Second, we discuss why disagreement between optimists ("bulls") and pessimists ("bears") persists in financial markets, highlighting a set of stylized empirical facts from a social network for investors.

6.1 The Evolution and Spread of Misconceptions Through Networks

We start by discussing why some people persistently believe in misconceptions and conspiracy theories. 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, endogenous network formation creates strong path dependence in the way people interpret information. Once a person "goes down the rabbit hole", it is difficult to use new information to convince them to come out.

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,

[22] 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.

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 bubbles have lasting consequences on how people interpret subsequent events. By Proposition 3, 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, Proposition 2 applies. 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 networks formed based on shared beliefs may result in beliefs being persistently untethered from data that is open to interpretation. 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 take on an event can be very hard to reverse.

6.2 Persistent Disagreement in Finance

We next turn to persistent disagreement between bullish and bearish investors in the stock market. A large literature in finance, going back to Miller (1977) and Harrison and Kreps (1978), studies the consequences of disagreement for asset prices.[23] Disagreement has proven useful in explaining several phenomena in the stock market, including high trading volumes and overpriced stocks. Yet, a key question remains: why does disagreement persist in the face of information that should resolve uncertainty?

Our framework delivers an answer: if bullish investors tend to talk to other bullish investors, interpretations of new data that lead them to remain bullish will evolve in the bulls' network. Similarly, bearish interpretations of new data will evolve in the bears' network. We provide stylized empirical facts consistent with this idea using data from StockTwits, a social network for investors that has been studied in recent papers on disagreement in financial markets, including Cookson and Niessner (2020), Divernois and Filipovic (2022), and Cookson et al. (2022). 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.[24] The platform is geared towards allowing investors to share with each other information and analysis about individual stocks. In particular, it allows

[23] See Hong and Stein (2007) and Barberis (2018) for comprehensive reviews.
[24] See this Bloomberg article.

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, Cookson et al. (2022) study the network formation behavior of users to shed light on disagreement in the stock market. They 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, StockTwits users act in a way consistent with our modeling of shared-belief networks in Section 3. Following a news announcement, they are more likely to form networks with others who share their views on a particular stock.

Second, we analyze the dynamics of user sentiment around earnings announcements in StockTwits messages between January 2011 and July 2018.[25] 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 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.[26] 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.[27]

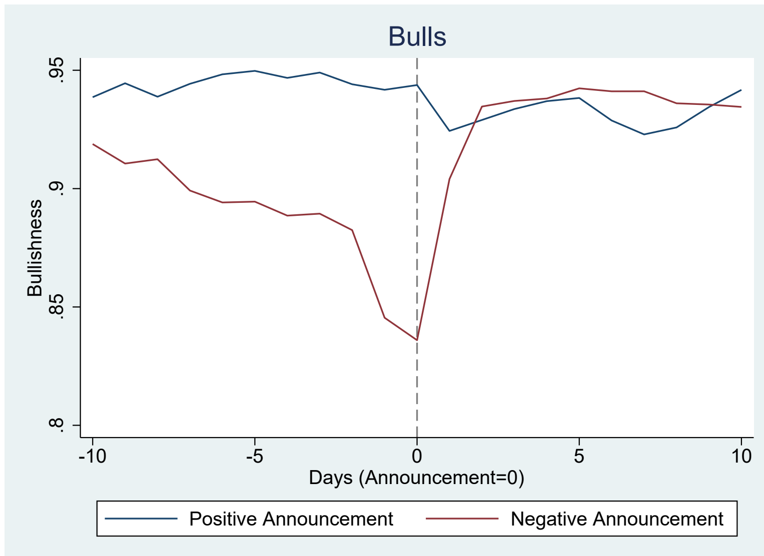
[25]We thank Marc-Aurèle Divernois and Damir Filipović for very generously sharing their data with us. Divernois and Filipovic (2022) study this data, showing that sentiment measured from StockTwits can be used to forecast stock returns on high-message volume days.

[26]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.

[27]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 announce is restrictive.

Figure 4: Evolution of Beliefs Around Earnings Announcements

(a) Bulls



(b) Bears

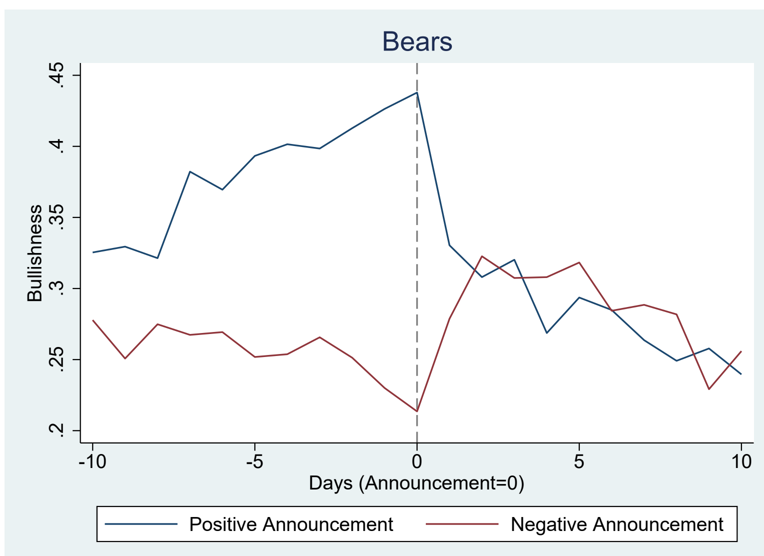


Figure 4a shows the evolution of sentiment for bulls around positive and negative earnings announcements. Corresponding regressions are reported in Appendix D. Around positive announcements, sentiment remains unchanged: generally 94-95% of messages are labeled as bullish. The pattern around negative 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.[28] In the days following the announcement, however, sentiment rapidly improves, 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 4b shows similar patterns for bears around positive news announcements. Sentiment first improves but then converges back to its original level. 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.

These patterns are consistent with our results in Section 3. 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.

7 Application 2: 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. In the model, 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 structure and goals of meetings are not fixed but depend on

[28] 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.

workers' flow of communication outside meetings and how the organization prioritizes adaptation versus coordination. In particular, meetings may help workers get on the same page by commonly muting their reaction to data instead of better adapting 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 function

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

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 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.[29] That is, almost all followers care about taking an action that is well-adapted to the state, 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:[30]

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

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

[29] 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.

[30] For simplicity, we assume the leader evaluates her expected payoff according to her own expectation and not followers' subjective expectations. For example, 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 to the environment. Introducing the latter force could provide an additional reason why leaders want 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.

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. 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 9. *In the leader-follower example of this section:*

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 9 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, in the leader's mind, followers are reacting to data when they should not be or if the organization places a sufficiently large weight on coordination, then the leader 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 leaders do 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 muted relative to her private response: if she believes that she cannot persuade enough followers in her desired course of action through a meeting, 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 organization places a sufficiently large weight on adaptation, then the leader calls a meeting which features a *persuasive campaign.* The leader ensures that the loudest voices are those that interpret the event in a way 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 5 there is also desire to hold the meeting as soon as possible, before workers can share interpretations with each other on their own.

8 Discussion

This paper is a first step to studying the social transmission of models. 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 6.1, 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.

The framework 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)? 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. 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

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

□

Proof of Proposition 2. That social learning hardens every person's reaction to the data is immediate from how models are selected. That social learning leads everyone to end up at their prior follows from the fact that someone will come up with and communicate the model m that $\pi_m(h|\omega) = 1$ for all $\omega \in \Omega$, which will beat all other models (since $\Pr(h|m, \mu_0) = 1$) and leads to $\mu(h, m) = \mu_0$.

□

Proof of Proposition 3. 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 have the best fit of all models represented in i's network, so everyone in i's network will adopt it.

□

Proof of Proposition 4.

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 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 Proposition 5. 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

$$\Pr(h|m(\tilde{\mu}), \mu_0) > \Pr(h|m_i', \mu_0). \tag{3}$$

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

$$\Pr(h|m(\tilde{\mu}), \mu_0) > \max_{m' \in \bigcup_{\mu \in s(\mu_i')} M(\mu)} \Pr(h|m', \mu_0). \tag{4}$$

The result follows from the right-hand-side of inequality (4) being larger than the right-hand-side of inequality (3).

A similar proof applies to the case of strong exposure to beliefs, replacing the left-hand-sides of inequalities (3) and (4) with $\max_{m' \in M(\tilde{\mu})} \Pr(h|m', \mu_0)$. □

Proof of Proposition 6.

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).

□

Proof of Proposition 7. The network-shaper'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 network-shaper does no worse by exposing people to all models specified in Eq. (1) (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. (1). This remains true if we add to models specified in (1) 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 network-shaper's payoff is strictly worse if we add to models specified in (1) 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 8. 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 network-shaper cannot do better, since everyone will end up sharing the same model.

□

Proof of Proposition 9. 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

$$-\alpha \mathbb{E}_{\mu(h,d)} \int_i (a_i - \omega)^2 di = -\alpha \mathbb{E}_{\mu(h,d)} \int_i (\omega_0 - \omega)^2 di, \tag{5}$$

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

$$-\alpha \left[\mathbb{E}_{\mu(h,d)} \rho \int_i (\mathbb{E}_{\mu(h,d)}[\omega] - \omega)^2 di + (1-\rho) \int_i (\omega_0 - \omega)^2 di\right] - \kappa \int_j (a_j - \rho \mathbb{E}_{\mu(h,d)}[\omega] - (1-\rho)\omega_0)^2 dj, \tag{6}$$

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 (6) is larger than (5) 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.[31]

□

[31] To see when else the leader wants to hold such a meeting, (6) minus (5) 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.

Appendices for Online Publication

B Shared Model Networks

Some networks are based not on shared beliefs, but shared models. 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 fundamentals (e.g., the economy) in predicting who will win an election, while others focus on polls. How do networks shape views in such cases?

To analyze shared-model networks, consider a partition $\mathcal{C}$ over the set of admissable 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 $\Pr(h|\cdot, \mu_0)$.

We analyze 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.

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:

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

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. 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 as being inevitable no matter the state, and similarly for people on the right.

Proposition 10. *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 10. 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 $\pi_m(h^a, h^b|\omega) = 1 \cdot \pi_{m^{f^a}}(h^a|\omega) = \pi_{m^{f^a}}(h^a|\omega)$ for all $\omega \in \Omega$. Similarly, the best fitting model in c^B is $\pi_m(h^a, h^b|\omega) = 1 \cdot \pi_{m^{f^b}}(h^b|\omega) = \pi_{m^{f^b}}(h^b|\omega)$ 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, $h^a \in \{\underline{h}^a, \bar{h}^a\}$, and soft, $h^b \in \{\underline{h}^b, \bar{h}^b\}$, 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 10, 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.[32]

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

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

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.

C Muting in the VC Example

In the VC example in Section 3.4 of the main paper, 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 in Section 3.4. 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 Results

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.[33] 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:

$$1[Bullish]_{i,u,t,s,q} = \alpha + \sum_{l=-10}^{10} \beta_l 1[l=t] + \sum_{l=-10}^{10} \gamma_l 1[l=t] 1[NegativeSurprise_{s,q}] + \varepsilon_{i,u,t,s,q}, \quad (7)$$

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.

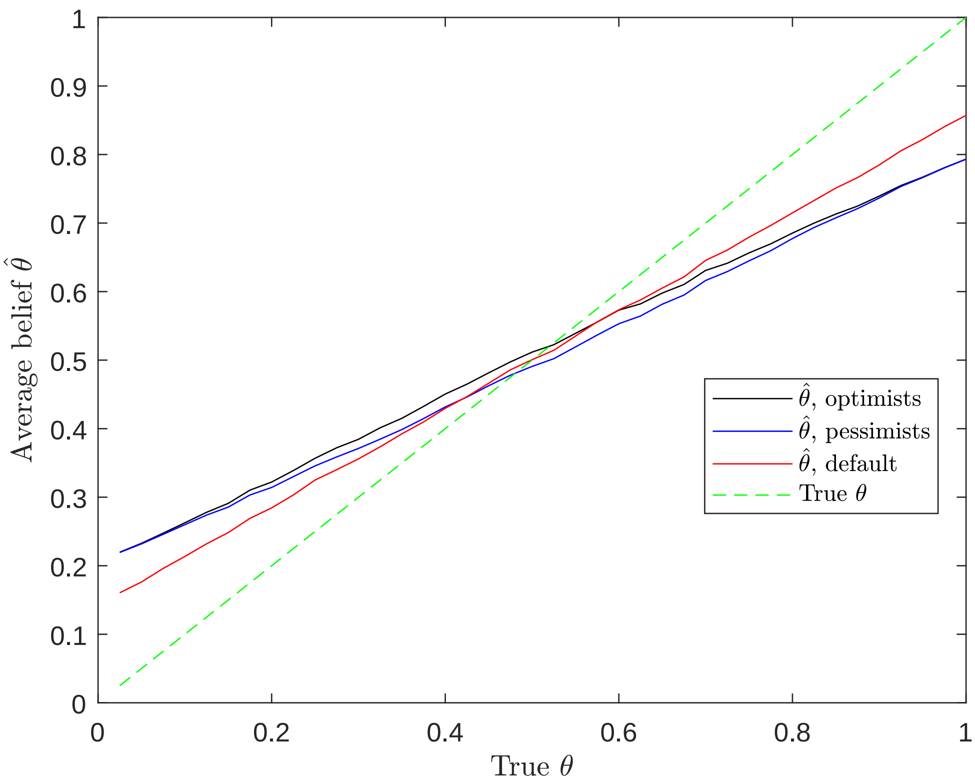
[33] Announcement days are therefore defined as the first day that the stock can be traded following the announcement. For announcements that occur after the market close on a given day, the announcement day is then coded as the following day.

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.

Table A2 presents regression evidence corresponding to Figure 4b in the main paper. Estimating Eq. (7) 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 $\pm$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.

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.

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.

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.*

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.

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

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.

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.