Narratives, Belief Movements, and Economic Fluctuations*

Tony Q. Fan† Tilman Fries‡

April 2, 2026

Abstract

We study how the construction and transmission of narratives—interpretations that explain observed data—systematically distort belief formation. In a dynamic environment where signals are generated by a potentially evolving state, we show that popular, well-fitting narratives tend to overfit the data and thereby push beliefs toward greater extremeness and volatility. We test this prediction in a controlled experiment that directly compares beliefs formed with and without exposure to a singular, endogenously constructed narrative. Participants exposed to such narratives form substantially more extreme beliefs, overreact to recent data, and exhibit more volatile belief movements over time. These results identify a mechanism through which narrative-based reasoning induces overreaction to news—one that likely matters broadly given the central role of narratives in economic communication.

*We thank Kai Barron, B. Douglas Bernheim, John Conlon, Chad Kendall, Muriel Niederle, Klaus Schmidt, Sevgi Yuksel, Sili Zhang, and audiences at various seminars and conferences for helpful comments. The RCT registry ID is AEARCTR-0015607. This study was approved by the University of Alabama IRB under Protocol 24-03-7429. We are grateful for financial support from the Society of Experimental Finance and the Behavioral Finance e.V. Research Grant. Tilman Fries gratefully acknowledges financial support from the Deutsche Forschungsgemeinschaft through CRC TRR 190 (project number 280092119).

†The University of Alabama. Email: qfan5@ua.edu

‡LMU Munich. Email: tilman.fries@econ.lmu.de

1 Introduction

The economy provides abundant information from which narratives are constructed. For example, a financial news report may frame a recent stock market rally as evidence of renewed confidence rather than random fluctuations, or a newspaper may portray a leader's shift toward hawkish policy as a display of personal toughness. Individuals routinely encounter such narratives in the news, on social media, and in everyday conversations. As a result, they often rely on these narratives as interpretive lenses through which they make sense of the world—and ultimately act upon them. Because such actions can shape future economic outcomes, it is crucial to understand how narratives influence learning and expectations.

In this paper, we examine whether and how the construction of narratives from data—and the diffusion of such narratives through society—generate systematically different belief biases from those formed directly from data. Economic models of narratives emphasize that a popular narrative must cohere with the data: it should fit the observed evidence, subject to constraints (Schwartzstein and Sunderam, 2021, 2024; Eliaz and Spiegler, 2020).[1] This requirement of internal consistency, however, raises concerns about overfitting— especially in settings where individuals rely on narratives about the past to form expectations about the future, such as when deciding where to invest or whether to hire a job candidate. Although this tension is recognized in theory, it remains largely conceptual. To examine it directly, we enrich a standard learning environment to study—both theoretically and experimentally—how exposure to popular, endogenously constructed narratives shapes inference from data and, in turn, beliefs about the future.

The core of our paper is a laboratory experiment designed to isolate the causal effects of *narrative construction and transmission* on beliefs. In the main experimental task, participants face a belief-formation problem framed as an "evaluation": they observe sequential data about a company and assess its current prospects for success. Although framed in such terms, the company is hypothetical, and the data are generated by a fixed statistical process. This feature is crucial, as it gives us full control over participants' priors and the narratives they may construct to interpret the data. We compare a Baseline treatment—where participants evaluate the company immediately after viewing its data—to two narrative-sharing treatments. In the Developer treatment, participants first construct a singular narrative to make sense of the data before making their evaluation, allowing us to isolate the effect of narrative *construction* on beliefs. In the Receiver treatment, participants receive a narrative created by a Developer participant, along with the same company data, and then make their evaluation; this isolates the effect of narrative *reception* on beliefs.

Our study yields three main findings. First, because Developer participants construct narratives that closely fit the company data, these narratives often imply extreme views of the company. In both narrative-sharing treatments—Developer and Receiver—participants rely heavily on these implied views when forming their own evaluations, resulting in more ex-

[1] In Schwartzstein and Sunderam (2024), agents share models with one another, adopting the model that maximizes fit given the prior and the data. In Schwartzstein and Sunderam (2021), a persuader selects a model to shift a receiver's beliefs in a desired direction, under the constraint that the model must fit the data well enough to be adopted by the receiver. In Eliaz and Spiegler (2020), narratives are represented as directed acyclic graphs (DAGs) linking variables causally; once a DAG is adopted, agents fit it to the data to estimate the strength of its causal links.

treme beliefs than in BASELINE. Second, participants evaluate the company repeatedly over time as new performance data arrive. We find that narratives are frequently adapted to fit the new data, which increases the volatility of company evaluations in the narrative-sharing treatments relative to BASELINE. Finally, our design enables comparison to a Bayesian benchmark. Using the BASELINE data, we replicate a well-documented *underinference* bias: participants adjust their beliefs too little relative to the Bayesian benchmark (e.g. Benjamin, 2019). In both narrative-sharing treatments, this pattern reverses to *overinference*—participants adjusting too much. This shift implies that narratives do not improve belief precision but instead amplify reactions to data. As a result, narrative sharing generates excess volatility in beliefs over time.

These findings have broader implications. Individuals often learn about the economy through narratives—encountered in the media, in conversations, and across social networks—so the same mechanism may help explain why belief updating about real-world economic variables frequently exhibits overreaction rather than underreaction (e.g. Bordalo et al., 2021) and why prices in financial markets often display excess volatility, a long-standing puzzle (e.g. Shiller, 1981). In this sense, the experiment isolates a mechanism that plausibly underlies broader belief dynamics.

In our environment, agents observe a *signal history*—a time series of the company's binary outcomes (successes or failures)—generated by an underlying *latent state*. Each period yields one new signal realization. The initial state is drawn from a prior distribution and may be redrawn between any two consecutive periods, introducing uncertainty about regime shifts that agents must track over time. This feature captures the idea of a company whose fundamentals evolve stochastically with occasional shifts, and it reflects a property common to many economic processes—such as inflation, stock-price growth, or unemployment—that exhibit persistence punctuated by occasional regime changes. Given knowledge of the prior distribution and the redraw probabilities, we can characterize how a fully rational Bayesian agent would form beliefs about the *current state* for any signal history. The resulting Bayesian posterior provides a normative benchmark for how beliefs should evolve given the data. Empirically, however, our identification strategy focuses on deviations from the BASELINE treatment—where participants form beliefs directly from the observed signal history—allowing us to isolate the additional distortion introduced by narratives.

Our experiment implements an instance of this environment. Participants take the role of agents: they observe signal histories generated according to the process described above and report their *mean estimate* of the current state. We provide them with all information needed to compute the Bayesian posterior, as well as incentives to report the corresponding mean estimate truthfully. Each participant completes a sequence of evaluation periods for each company they encounter. In each new evaluation period, additional signals are appended to the history, and participants are asked to state an estimate of the current state associated with the most recent signal. Hence, our experiment permits a dynamic analysis of belief changes between evaluation periods.

In this setting, a narrative is a candidate explanation of how the observed signal history came to be. Formally, it is represented by a vector that specifies the path of latent-state transitions from the company's inception to the current evaluation period. In the DEVELOPER treatment, we elicit a singular narrative from participants—framed as their "best guess" of

the state-transition path—through a structured two-step procedure: developers (i) indicate the periods in which they believe redraws occurred and (ii) provide state estimates for each segment between the redraws they identify. After submitting their narrative, they report a mean estimate of the current state, as in BASELINE. In the RECEIVER treatment, participants also provide a mean estimate of the current state but observe a narrative constructed by a DEVELOPER in addition to the signal history.

Our theoretical framework captures belief formation both with and without narrative sharing and rests on two key assumptions. First, we assume that narratives are persuasive—that is, agents adjust their beliefs about the state toward the state implied by the shared narrative. Such persuasion may arise through several psychological mechanisms, including self-persuasion, salience, or the channeling of attention. Second, since beliefs respond to narratives, it is crucial to characterize the features of developers' narratives. We focus on a central property emphasized in the literature: popular narratives achieve a high posterior likelihood. This idea, developed in Schwartzstein and Sunderam (2021, 2024), reflects the lay intuition that a narrative should "make sense," meaning that it is internally consistent with the available evidence. Under these assumptions, the framework predicts that narrative sharing increases the weight agents place on narratives that fit the data well.

Within our framework, we compare the subjective expectation of an agent in the absence of narrative sharing to the expectation formed with narratives, which places greater weight on narratives with high posterior likelihood. In statistical terms, this comparison mirrors the distinction between a Bayesian expectation and the maximum a posteriori estimator (MAP).[2] Common concerns about the MAP—especially those related to overfitting (see, e.g., MacKay, 2003)—apply here as well, creating a wedge between beliefs formed with and without narrative sharing. Specifically, because the MAP neglects model uncertainty, it reacts more strongly to data, shifting estimates excessively toward realized signals relative to the Bayesian expectation. By the same logic, narrative sharing produces more extreme beliefs—that is, beliefs farther from the prior—in our framework. In addition, our framework predicts that belief volatility rises under narrative sharing. The volatility and extremeness results are tightly connected: when beliefs move further from the prior, their variance necessarily increases. Consequently, when measuring belief revisions between evaluation periods, the average magnitude of belief movement is larger with narrative sharing than without it. Because these results arise from overfitting—a general statistical phenomenon—they are not specific to the particular information environment we study. Moreover, our framework shows that these results are robust to several extensions: (i) allowing for common biases in Bayesian updating, (ii) assuming that agents consider only a subset of possible state-transition paths when forming their BASELINE expectation, and (iii) allowing narrative developers to share a *subjective* best-fit narrative that may differ from the *objectively* best-fitting one.

With the main setup in place, we now provide more details on our experimental results. First, while participants in the DEVELOPER treatment correctly identify the objective best-fit narrative in only 7.45% of cases, most of the narratives they construct are highly coherent: roughly 80% fall within the top 1% of narratives—ranked by fit—that could have been con-

[2]The MAP estimator is related to the maximum likelihood estimator (MLE). While the MLE maximizes the likelihood, the MAP maximizes the posterior density—namely, the likelihood multiplied by the prior (see, e.g., MacKay, 2003).

structed given the observed signals. As a result, the narratives constructed in DEVELOPER are *extreme*: they imply a current-state estimate (i.e., the success probability associated with the most recent signal) that lies far from the prior. On average, DEVELOPER narratives imply current-state values that are about 30% more extreme than mean beliefs in BASELINE. This pattern carries through to mean beliefs in both DEVELOPER and RECEIVER, which are between 20-24% more extreme than those in BASELINE.

Second, turning to dynamics, we find that volatility—measured as the absolute change in beliefs between consecutive evaluation periods—is substantially higher with narrative sharing. Specifically, the volatility of current-state values implied by DEVELOPER narratives is 43% greater than that of mean beliefs in BASELINE. A similar pattern emerges when comparing participants' beliefs across treatments: beliefs in DEVELOPER and RECEIVER are 35–39% more volatile than those in BASELINE.

Third, our theoretical framework identifies the conditions under which narrative sharing most strongly amplifies extremeness and volatility. We predict extremeness to rise particularly strongly when the underlying signal history is highly informative about the current state, while volatility should increase the most when the underlying signals vary substantially over time. In other words, our framework predicts that narratives amplify patterns already present in the data. We test these predictions in a preregistered analysis that groups signal histories by their informativeness and variability. The results confirm both predictions: narratives increase belief extremeness and volatility most strongly in precisely those environments where the framework predicts the strongest effects.

Fourth, our experiment allows a direct comparison to the Bayesian benchmark, providing additional insights into the mechanisms at play. One potential concern is that narrative sharing might simply prompt participants to think more carefully about the data. If so, the increased responsiveness to data could reflect greater attention and more accurate belief formation. But the evidence does not support this interpretation. While participants in BASELINE underreact to the data on average, participants in the narrative-sharing treatments overreact relative to the Bayesian benchmark. Furthermore, while beliefs in BASELINE exhibit volatility comparable to Bayesian posteriors, beliefs in the narrative-sharing treatments display significant excess volatility. Taken together, these patterns imply that narrative sharing does not improve belief accuracy. Instead, the observed overreaction and excess volatility align with the overfitting mechanism highlighted by our framework: when beliefs are drawn toward best-fit narratives, they move too much in response to data.

We provide additional evidence on the mechanisms through which narratives generate belief overreaction, by decomposing overreaction into two distinct channels. First, developers may place redraw events at moments that maximize contrast across segments of the signal history. Second, conditional on the chosen redraw events, they may select extreme state estimates. We decompose the overall overreaction embedded in constructed narratives into these two components and find that redraw placement accounts for 38% of the effect. In a separate exercise, we show that narrative receivers respond more strongly to narratives with higher empirical fit. This implies that narrative sharing distorts beliefs not only because developers' narratives overfit the data, but also because receivers place higher weight on narratives that fit well. Finally, when moving from BASELINE to DEVELOPER or RECEIVER, participant evaluations become more sensitive to the most recent signals—a pattern that directly contributes to the increases in belief extremeness and volatility under narrative

transmission.

We complement these reduced-form findings with a structural estimation directly adapted from the theoretical framework. The estimation reveals four key insights. First, belief formation and narrative construction are well described by a sampling framework where, instead of integrating over all possible state-transition paths, participants only consider a limited set of possible paths that they obtain from a sampling distribution. Second, narrative construction attenuates a BASELINE tendency to under-appreciate the likelihood of state transitions in belief formation. This attenuation provides an alternative explanation of the observed treatment effects, as it can also generate an increase in belief volatility from BASELINE to DEVELOPER. Nonetheless, as a third result, a counterfactual decomposition based on the structural estimates shows that the shift in the *perceived* state-transition probability—while potentially influential—cannot fully account for our treatment effects; specifically, it accounts for only 20% of the effect on belief extremeness and at most 56% of the effect on belief volatility. Finally, distortions in the sampling distribution matter much less for narrative construction than for beliefs: in particular, while such distortions generate an underreaction in BASELINE beliefs, they leave the relative ranking of narratives in fit largely unchanged. Consequently, the main driver of the fit of shared narratives is the number of sampled narratives: we estimate that developers consider only 6-8 narratives when deciding which one to share, causing the empirically observed narratives to often diverge from the true best-fit narrative.

We connect to several strands of the literature. An active body of work on narrative economics documents how narratives diffuse through society (Shiller, 2017) and how individuals rely on them when forming economic expectations (Andre et al., 2025). Several theoretical frameworks formalize the link between narratives and expectations (e.g., Eliaz and Spiegler, 2020; Schwartzstein and Sunderam, 2021, 2024; Aina, 2025), and a growing experimental literature tests the core assumptions and predictions of these models. Experiments have shown, for instance, that individuals follow systematic criteria when selecting between competing causal interpretations (Ambuehl and Thysen, 2024); that the empirical fit of a narrative shapes its adoption and can be exploited for persuasion (Barron and Fries, 2025); and that individuals infer causal relationships from data in predictable ways (Fan, 2024; Charles and Kendall, 2025; Fréchette et al., 2025). We contribute to this literature by showing that even without any conflict of interest from narrative developers, the inherent tendency of narratives to overfit the data gives rise to systematic overreaction and excess volatility in belief formation.

We also contribute to the literature on belief updating, particularly on the causes of under- and overreaction to information (Phillips and Edwards, 1966; Hartzmark et al., 2021; Afrouzi et al., 2023; Enke et al., 2024a; Ba et al., 2025; Fan et al., 2025; Liang, 2025). A central finding in this literature—that individuals on average underinfer from signals (Benjamin, 2019)—also holds in our baseline setting. However, we show that introducing narrative construction and transmission into the belief-updating task can reverse this pattern, generating systematic overinference. Existing explanations for overreaction emphasize mechanisms such as associative memory (Enke et al., 2024b), the representativeness heuristic (Bordalo et al., 2018; Ba et al., 2025), noisy perception of signal strength (Augenblick et al., 2025b), or heuristic responses to problem complexity (Fan et al., 2025). We add a complementary mechanism: overreaction arising from individuals' reliance on overfitting narratives.

Closely related to our work, a recent experimental literature examines decision-making under model uncertainty. Aina and Schneider (2025) show that individuals often rely on a single model instead of averaging across models when forming predictions. Musolff and Zimmermann (2025) further document that this reliance on a single model becomes stronger in more complex environments. Augenblick et al. (2025a) examine beliefs of agents who neglect model uncertainty. While their approach is conceptually related to ours, they focus on different biases and offer a different interpretation: because the agents in Augenblick et al. (2025a) form sparse representations directly from data, they underappreciate model uncertainty and exhibit *overprecision*. In our setting, agents acknowledge model uncertainty upon observing data, but exposure to well-fitting narratives systematically distorts their beliefs, generating *overreaction* and *excess volatility*.

Our finding that narratives increase belief volatility connects to the long-standing stock-price volatility puzzle. A large literature, including Shiller (1981, 1987), Cutler et al. (1988), and Pindyck and Rotemberg (1990) documents that aggregate prices fluctuate more than can plausibly be justified by fundamentals. One prominent explanation, proposed in Shiller (1984) and further developed in Shiller (2017, 2019), argues that socially transmitted stories, fads, and fashions drive fluctuations in investor beliefs, thereby amplifying price movements beyond what fundamentals alone can explain. We contribute to this literature by providing controlled laboratory evidence that the diffusion of narratives causally increases belief volatility, offering a microfoundation for how narrative transmission can translate into excess price volatility.

The literature on social finance examines how social processes influence economic and financial behavior (Hirshleifer, 2015, 2020; Kuchler and Stroebel, 2021). This work provides empirical evidence on belief contagion among peers in settings such as house price expectations (Bailey et al., 2018) and mortgage refinancing decisions (Maturana and Nickerson, 2019), and theoretical analyses show how such contagion can generate boom–bust cycles in housing markets (Burnside et al., 2016). Our paper contributes by demonstrating how contagious narratives produce systematic belief biases in a controlled laboratory environment. In the mechanism we highlight, narratives constructed by symmetric peers—whom individuals may trust precisely because they lack incentives to deceive, unlike professional advisors (Kuchler and Stroebel, 2021)—still distort beliefs due to inherent features of narrative communication.[3]

The remainder of the paper is organized as follows. Section 2 introduces the conceptual framework. Section 3 describes the experimental design. Section 4 presents the main findings, documenting how narrative construction and transmission systematically distort beliefs, and reports the structural estimation of the belief formation model. Section 5 concludes.

[3]Several neighboring disciplines have examined the consequences of encoding information in narrative form. Historians have critically reflected on their own practice of recounting events through narratives, highlighting how the narrative form privileges coherence over complexity (Skinner, 1969) and how historical writing often distinguishes itself more through narrative style than through new empirical discovery (White, 2000). In psychology, Bruner (1986) contrasts two modes of thought: *narrative thinking*, which values drama, coherence, and believability, and *logico-scientific thinking*, which seeks analytic consistency and truth.

2 Conceptual Framework

Narratives condense complex data into coherent stories, which tend to overfit data. Because they overfit, narratives induce beliefs that are extreme and volatile. In this section, we formalize this mechanism and derive testable predictions.

We first introduce the dynamic information environment used in our experiment. We then model agents as "constructive thinkers" who generate a limited set of plausible narratives and aggregate them into an estimate about a state of interest. When agents select high-fit narratives to share with one another, we identify a key statistical force—the *fit-state covariance*—that isolates the belief distortions caused by shared narratives. Finally, we derive testable predictions regarding belief extremeness and volatility.

2.1 Information Environment

We study the implications of narrative communication on beliefs by applying this framework to a dynamic environment governed by a hidden Markov state-transition process.[4]

Consider a finite discrete-time environment $t \in \{1, \ldots, T\}$. Each period t is characterized by a state $\theta_t \in \Omega \equiv \{\omega_1, \ldots, \omega_S\} \subset [0, 1]$ with $S > 1$ distinct states in ascending order, $\omega_1 < \cdots < \omega_S$. The state θ_t is distributed by a stationary prior distribution $\mu \in \Delta(\Omega)$. In the main analyses, we impose two additional simplifying assumptions: (i) The state space is *symmetric*: if $\omega_j \in \Omega$, then $1 - \omega_j \in \Omega$. (ii) The prior is *uniform*: $\forall \omega_j \in \Omega,\ \mu(\omega_j) = \frac{1}{S}$.

In each period t, there is a binary signal $\sigma_t \in \{0, 1\}$ with a good news structure.[5] If the state $\theta_t = \omega_j$, the signal σ_t is distributed according to $\Pr(\sigma_t = 1 | \theta_t = \omega_j) = \omega_j$ and $\Pr(\sigma_t = 0 | \theta_t = \omega_j) = 1 - \omega_j$. For example, if $\Omega = \{0.2, 0.8\}$, then the signal realization $\sigma_t = 1$ occurs with probability 0.2 in state $\theta_t = \omega_1 = 0.2$ and with probability 0.8 in state $\theta_t = \omega_2 = 0.8$. Since the probability of $\sigma_t = 1$ increases (and the probability of $\sigma_t = 0$ decreases) in the state, $\sigma_t = 1$ is indicative of higher states (good news) and $\sigma_t = 0$ is indicative of lower states (bad news).

The state transitions according to a fixed redraw probability $r \in (0, 1/2)$. In each period, with probability r, the state is redrawn from the prior; otherwise, it remains unchanged. Therefore, r measures the degree of state *transience*. Conditional on a redraw after period t, the state value may or may not change. We call an actual change in the state value between two periods a "paradigm shift," which occurs with probability $\frac{S-1}{S} r$ in any period.[6] In the main body of the paper, we focus on the case where $r \in (0, \frac{1}{2})$. That is, we focus on the case where a redraw event is relatively unlikely to occur after any given period, reflecting the dynamics of important economic processes—such as asset prices, GDP growth, or inflation—that are persistent yet subject to infrequent shifts.[7]

[4] Hidden Markov models are widely used in micro- and macroeconomics. Similar environments have been studied experimentally by Gallistel et al. (2014) and Khaw et al. (2017).

[5] The good news information structure that we use follows prior theoretical (Milgrom, 1981) and experimental work (Ba et al., 2025).

[6] A redraw event does not necessarily imply a paradigm shift, since the redraw may leave θ_t unchanged if the same value is drawn again. Such "null transitions" vanish as S grows, so that a redraw event almost always coincides with a paradigm shift when S is large (as in our experiment).

[7] With $r = 0$, the environment reduces to standard learning with a fixed state and sequential binary signals, while with $r \geq 1/2$, paradigm shifts occur so frequently that the model no longer resembles most

Evaluation periods In this information environment, agents progress through a sequence of *evaluation periods* indexed by $\mathcal{T} \subseteq \{1, \ldots, T\}$. In each period $\tau \in \mathcal{T}$, an agent forms a belief $\gamma_\tau \in \Delta(\Omega)$ about the contemporaneous state θ_τ, conditional on the realized signal history up to that point, $\boldsymbol{\sigma}_\tau \equiv (\sigma_1, \ldots, \sigma_\tau) \in \{0,1\}^\tau$. Agents proceed sequentially through evaluation periods, starting from the smallest element of $\mathcal{T}$ and continuing in order. Within this framework, we perform both a static analysis—characterizing beliefs in a given evaluation period—and a dynamic analysis—tracing how beliefs evolve across periods.

Narratives in this environment A narrative is a candidate explanation of how the observed signal history came to be. Formally, narrative $\mathbf{n}_\tau$ is a vector of states up to the current period τ, $\mathbf{n}_\tau = (\theta_1, \ldots, \theta_\tau) \in \Omega^\tau$. The "fit" of a narrative is its Bayesian posterior likelihood given the signal history $\boldsymbol{\sigma}_\tau$:

$$\pi(\mathbf{n}_\tau) = \Pr(\mathbf{n}_\tau | \boldsymbol{\sigma}_\tau, r).$$

We use $\theta_\tau(\mathbf{n}_\tau)$ to refer to the last element of the state vector $\mathbf{n}_\tau$ (the current state), which is the variable of interest.

A working example To illustrate and explore the implications of our results, we will frequently refer to a *working example*: a specific parameterization of the information environment introduced above. In this example, there are $S = 101$ states spaced out over a fine grid $\Omega = \{0, 0.01, \ldots, 1\}$, and the redraw probability is set to $r = 0.2$. This parameterization also serves as the basis for our experimental design.

2.2 Belief Formation with Narratives

Consider an agent who observes a signal $\boldsymbol{\sigma}_\tau$ and forms beliefs about the variable of interest θ_τ. The signal could have been generated by one of several processes, each of which we refer to as a "narrative."

Interim beliefs (before narrative sharing) We model belief formation through a sampling process. Upon observing the data $\boldsymbol{\sigma}_\tau$ and prior to any narrative sharing, each agent i draws $K \geq 1$ narratives from Ω^τ with replacement, according to a sampling distribution $f \in \Delta(\Omega^\tau)$. The resulting collection of sampled narratives, denoted $\mathbf{c}_i = (\mathbf{n}_{i,1}, \ldots, \mathbf{n}_{i,K})$, represents agent i's set of candidate explanations for how $\boldsymbol{\sigma}_\tau$ may have been generated.[8] Because sampling is stochastic, two agents who draw the same number of narratives from the same distribution may nonetheless consider different narratives—and, as we show below, form distinct beliefs.

We assume agents sample narratives based on how well they fit the data. Specifically, the probability of sampling a narrative is increasing in its Bayesian posterior likelihood.

Assumption 1 (regularity in sampling)**.** *The sampling distribution is given by* $f(\mathbf{n}_\tau) = g(\pi(\mathbf{n}_\tau))$*, where* $g(\cdot)$ *is a strictly increasing function.*

real-world processes.

[8] We represent sampled narratives as tuples rather than sets to allow for the possibility that an agent samples the same narrative multiple times.

Sampling implies that agents are *constructive thinkers*. Rather than integrating over the entire, possibly high-dimensional space of all narratives (as a Bayesian would), agents generate a collection of plausible candidate explanations to make sense of the observed data. Assumption 1 additionally captures the intuitive notion that high-fit explanations are "more accessible"—they come to mind more easily.[9]

After deriving a set of candidate hypotheses, agent i forms an interim belief about θ_τ by aggregating across their sampled narratives $\mathbf{c}_i$:

$$\gamma_{i,\text{interim}}(\theta_\tau = \omega_j | \mathbf{c}_i) = \frac{1}{K} \sum_{a=1}^{K} \mathbb{1}_{\theta_\tau(\mathbf{n}_{i,a})=\omega_j}.$$

This aggregation rule leads to an unbiased belief if the sampling distribution is unbiased (if $f = \pi$). In particular, denote the expected interim mean belief as m^I. Then, the framework predicts that

$$m^I = \mathbb{E}_{\mathbf{n}_\tau \sim f} \left[\theta_\tau(\mathbf{n}_\tau) \right]. \qquad \text{(Interim mean belief)}$$

This sampling procedure can be interpreted as a boundedly rational approximation of Bayesian updating: agents treat the sampling distribution as their subjective posterior but approximate the belief by sampling a smaller number of narratives from f and averaging over this sampled set. By requiring that f is monotone in but not equal to the Bayesian posterior, the framework can account for common belief distortions, such as under- or overinference.

Narrative sharing We model the sharing of narratives as a selection process. A "Developer" agent selects one narrative from their sample to share with a "Receiver." We assume that developers are "sincere" sharers who communicate the narrative they subjectively regard as the most plausible interpretation of the data. Specifically, the developer shares their *subjective best-fit narrative,* $\mathbf{n}_i^{BF}$, defined as the narrative with the highest subjective posterior in their sample:[10]

$$\mathbf{n}_i^{BF} = \underset{\mathbf{n}_\tau \in \{\mathbf{n}_{i,1}, \ldots, \mathbf{n}_{i,K}\}}{\arg\max} f(\mathbf{n}_\tau).$$

In real-world settings, the role of developers may be played by media outlets, domain experts, social media influencers, and similar actors who provide interpretations of economic events. We assume they share only their best-fit narrative, which appears to be a good first approximation of how individuals try to collectively make sense of the world, by sharing coherent, reduced representations. This assumption is also supported by evidence from Graeber et al. (2025), who empirically document that word-of-mouth transmission tends to

[9] This follows a literature in cognitive psychology arguing that human inference is often well described by sampling hypotheses from the Bayesian posterior and constructing estimates based on these samples (Vul et al., 2014; Sanborn and Chater, 2016; Dasgupta et al., 2017). The sampling approach relates to an older tradition in decision science that conceptualizes belief construction as mental simulation (Kahneman and Tversky, 1982). Examples of papers from the economics literature using a sampling-based belief formation approach are Gennaioli and Shleifer (2010) and Bordalo et al. (2022a).

[10] When there are multiple best-fit narratives, we assume that the developer picks randomly among them. With slight abuse of notation, $\mathbf{n}_i^{BF}$ denotes the particular narrative that is actually shared.

simplify information relative to what transmitters know. (In Appendix B.6, we show that allowing developers to share their top-M best-fit narratives does not change our core results.)

Since developers share their subjective best-fit narrative, a narrative is only shared if it outperforms all other $K-1$ narratives sampled by a developer in fit. This selective sharing of narratives tends to generate predictions that are supported by high fit narratives. Whenever high fit narratives generate predictions that are different from the average narrative, there is a wedge between the interim expectation and the expected best fit estimate.

To state this result formally, we introduce additional notation. Let m_K^{BF} denote the developer's expected best-fit estimate of θ_τ, and let $F_{K-1}(\mathbf{n}_\tau)$ be the probability that narrative $\mathbf{n}_\tau$ beats all other $K-1$ sampled narratives. This probability is monotone in narrative fit, as higher fit narratives get shared more often. We have the following result:[11]

Lemma 1 (Relation between best-fit and interim mean)**.** *Denote the covariance between the probability that a sampled narrative is shared, $F_{K-1}(\mathbf{n}_\tau)$, and the narrative-implied estimate, $\theta_\tau(\mathbf{n}_\tau)$, by:*

$$\psi_K \equiv \text{Cov}_{n_\tau \sim f}(F_{K-1}(\mathbf{n}_\tau), \theta_\tau(\mathbf{n}_\tau)). \qquad \text{(Fit-state covariance)}$$

The developer's best-fit estimate is equal to:

$$m_K^{BF} = m^I + \psi_K. \qquad \text{(Expected best-fit)}$$

We refer to ψ_K as the *fit-state covariance*. A positive ψ_K indicates that high-fit narratives tend to predict higher states, generating optimistic best-fit estimates; a negative ψ_K captures the symmetric pessimistic case. Relative to the interim mean estimate, the estimate implied by the shared narrative is drawn towards estimates embedded in high-fit narratives.

Posterior beliefs (after narrative sharing) Both developers and receivers update their beliefs about θ_τ after narrative sharing; we refer to agent i's updated belief as their *posterior belief*. We consider a simple belief updating rule in which posterior beliefs are weighted averages of the interim beliefs and the beliefs implied by the shared narrative:

$$\gamma_{i,posterior}(\theta_\tau = \omega_j | \mathbf{c}_i) = \alpha_i \cdot \mathbb{1}_{\theta_\tau(\mathbf{n}_j^{BF}) = \omega_j} + (1-\alpha_i) \cdot \gamma_{i,\text{interim}}(\theta_\tau = \omega_j | \mathbf{c}_i),\ \alpha_i \in (0,1].$$

The posterior updating rule formalizes agents who are *coherence seeking*. When a specific narrative is explicitly articulated—either by oneself (as a Developer) or by a peer (as a narrative-Receiver)—it becomes a focal point. The shared narrative suppresses the ambiguity of the underlying data, leading agents to overweight this single "best" story relative to alternative possibilities. On the developer side, such a rule could be microfounded through a psychological *self-persuasion* mechanism Schwardmann and Van Der Weele (2019); Soldà et al. (2020); Schwardmann et al. (2022). On the receiver side, the narrative may present a particular *cue* Conlon and Kwon (2025); might provide a *frame* or help with *sense-making* (Schwartzstein and Sunderam, 2021); increase *salience* (Bordalo et al., 2022b); or serve as a *cognitive shortcut*, as a sufficient statistic for the signal history.

[11] The proofs of this and all other formal results in the paper are included in Appendix B.

These updating rules imply a direct relationship between the expected posterior mean belief and the properties of the shared narrative. Combining the equations for the Interim mean belief and Expected best-fit, the expected posterior mean is equal to:

$$m^P_{i,K} = m^I + \alpha_i \psi_K. \qquad \text{(Posterior mean belief)}$$

Narrative sharing therefore pulls the posterior mean towards states supported by high-fit narratives. It reveals a fundamental mechanism driving belief distortion: if narratives that fit the data well tend to imply high values of θ_τ ($\psi_K > 0$), the posterior belief exceeds the interim belief. Conversely, if high-fit narratives imply low values of θ_τ ($\psi < 0$), the posterior belief is below the interim belief. Narrative sharing amplifies reactions to data: it pulls expectations away from a "middling" compromise towards the high-fit region.

Remark 1: Relation to the Bayesian benchmark While we motivate the framework with boundedly rational sampling, our core results do not hinge on this assumption. Consider the limit where agents sample many narratives ($K \to \infty$) from the true posterior ($f = \pi$). In this case, the interim belief m^I converges to the standard Bayesian Mean, while the shared narrative n_i^{BF} converges to the Maximum A Posteriori (MAP) estimate. The model thus nests a "Bayes + Persuasion" benchmark: the posterior distortion ψ_K persists as long as the "best" narrative (the mode) differs systematically from the "average" narrative (the mean). This highlights that our results are driven by the selection of high-fit narratives for sharing, rather than by the sampling friction itself.

Remark 2: The agent's error Under the standard Bayesian benchmark, agents should not update towards narratives, as these narratives do not contain any information: developers construct them under symmetric information. However, under a boundedly rational benchmark, exchanging beliefs can reduce error variance, as sampling makes agents' estimates noisy. Therefore, updating beliefs upon observing the narrative can be justified if agents are aware of the sampling noise in their estimates. However, such an update should account for the fact that the developer shares a *selected* narrative (the best-fit) rather than a representative one (the mean). The optimal update involves first "de-biasing" the narrative using the fit-state covariance to recover the developer's mean estimate, and then averaging it with one's own belief. By moving their posterior directly toward the unadjusted narrative, agents in our framework commit *selection neglect*: they treat the selected best-fit narrative as if it were an unbiased estimate of the mean.[12] We formalize this benchmark in Appendix B.7.

2.3 Predictions: Extremeness and Volatility

The analysis in Section 2.2 shows that, when going from interim to posterior, the mean belief moves towards states supported by high-fit narratives. Here, we discuss several implications of this result. We focus on two properties of beliefs: (i) their *extremeness*—their distance from the prior—and (ii) their *volatility*, measured as the change in beliefs *across evaluation*

[12] Selection neglect is a robust belief formation bias documented in, e.g., Enke (2020) and Barron et al. (2024).

periods. Our results suggest that narrative sharing makes posterior beliefs both more extreme and more volatile than interim beliefs.

The intuition behind the increase in extremeness rests on the behavior of the fit-state covariance, ψ_K. Consider an optimistic signal history dominated by good news. In such an environment, narratives that imply a high state fit the data well, while those implying a low state fit poorly. This creates a positive correlation between a narrative's fit and its state estimate ($\psi_K > 0$). Since the posterior mean is the interim mean plus this covariance term, beliefs are pulled further upward—into the optimistic tail. Conversely, for a pessimistic signal history dominated by bad news, the best-fitting narratives are those with low state estimates, generating a negative covariance ($\psi_K < 0$) that pulls beliefs further downward. In both cases, the fit-state covariance amplifies the signal, pushing beliefs away from the prior and generating overreaction. Formally, this suggests the following result:

Proposition 1 (extreme signal histories generate overreaction)**.** *Fix $r \in (0, 0.5)$, agents' sampling distribution $f(\mathbf{n}_\tau)$ for narratives, and $K > 1$. There is a $p^* \in (\omega_1, 0.5)$ such that:*
(i) For any signal history $\boldsymbol{\sigma}_\tau$ that is sufficiently unfavorable about θ_τ so that

$$m^I < p^*$$

we have

$$m^P_{i,K} < m^I.$$

(ii) For any signal history $\boldsymbol{\sigma}_\tau$ that is sufficiently favorable about θ_τ so that

$$m^I > 1 - p^*$$

we have

$$m^P_{i,K} > m^I.$$

Proposition 1 establishes sufficient conditions for extremeness but the mechanism it highlights—that narratives amplify reactions to data—operates more broadly. Intuitively, in the specified information environment, most optimistic signal histories (with $m^I > 1/2$) produce a positive fit-state covariance, while most pessimistic histories produce a negative covariance. Under this alignment, narrative sharing amplifies the direction of the interim inference *on average,* increasing the dispersion of posterior estimates. This restriction is extremely mild: it merely requires that high-fit narratives imply high states on average when signals are good and low states when signals are bad.

Proposition 2 (Average Overreaction)**.** *Narrative sharing amplifies belief extremeness on average if the fit-state covariance is positively correlated with the interim mean belief across signal histories:*[13]

$$\mathrm{Cov}_{\boldsymbol{\sigma}_T \sim \mathbb{P}_T}(\psi_K, m^I) > 0 \implies \mathrm{Var}_{\boldsymbol{\sigma}_T \sim \mathbb{P}_T}(m^P) > \mathrm{Var}_{\boldsymbol{\sigma}_T \sim \mathbb{P}_T}(m^I)$$

[13]We use $\mathbb{P}_T$ to denote the distribution of $\boldsymbol{\sigma}_T$ in this information environment. Formally, $\mathbb{P}_T(\boldsymbol{\sigma}_T) = \sum_{\mathbf{n}_T \in \Omega^T} \Pr(\boldsymbol{\sigma}_T | \mathbf{n}_T) \Pr(\tilde{\mathbf{n}}_T = \mathbf{n}_T | r)$.

We provide computational evidence in Appendix B.5 that $\text{Cov}_{\boldsymbol{\sigma}_T \sim \mathbb{P}_T}(\psi_K, m^I) > 0$ across different parameterizations in the information environment. Therefore, in our framework, beliefs are on average more extreme under narrative sharing. The simulations below illustrate this, showing that, in the working example, posterior means are more extreme than interim means for nearly all signal histories, not just those in the tails.

Extremeness in the working example Figure 1 plots the true best-fit estimates against the Bayesian mean belief for each of the 1,024 possible signal histories in our working example. The figure also reports simulated subjective best-fit estimates for different numbers of sampled narratives ($K = 1, 10, 100, 1000$). To establish a clean benchmark, we assume that agents sample from the Bayesian posterior. The results show that for the vast majority of signal histories the true best-fit estimates are more extreme than the Bayesian mean belief, generating a clear form of overreaction. Moreover, as K increases, the subjective best-fit estimates converge to the true best-fit estimates and inherit their extremeness. Consequently, posterior beliefs, which are influenced by these subjective best-fit estimates, are systematically pulled toward the extremes relative to interim beliefs.

We now turn to volatility. Define the volatility of a mean belief, m_τ, as the expected squared change in beliefs between two evaluation periods τ' and $\tau'' > \tau'$:

$$\text{Vol}_{\boldsymbol{\sigma}_T \sim \mathbb{P}_T}(m_{\tau'}, m_{\tau''}) \equiv \mathbb{E}_{\boldsymbol{\sigma}_T \sim \mathbb{P}_T}\left[(m_{\tau''} - m_{\tau'})^2\right].$$

Narrative sharing tends to increase volatility because the selection of a best-fit narrative introduces instability. While the interim belief evolves smoothly by averaging over many hypotheses, the best-fit narrative is discrete. A small perturbation of the signal history—such as a single new observation—can cause a large change in the fit-state covariance (e.g., by making a particular state-transition timing more plausible). These discrete jumps in the underlying "story" translate into large fluctuations in beliefs, generating volatility that exceeds the Bayesian benchmark. We provide a condition for volatility to increase *on average.*

Proposition 3 (Excess Volatility)**.** *Narrative sharing increases belief volatility on average if changes in the fit-state covariance co-move with changes in the interim belief across signal histories:*

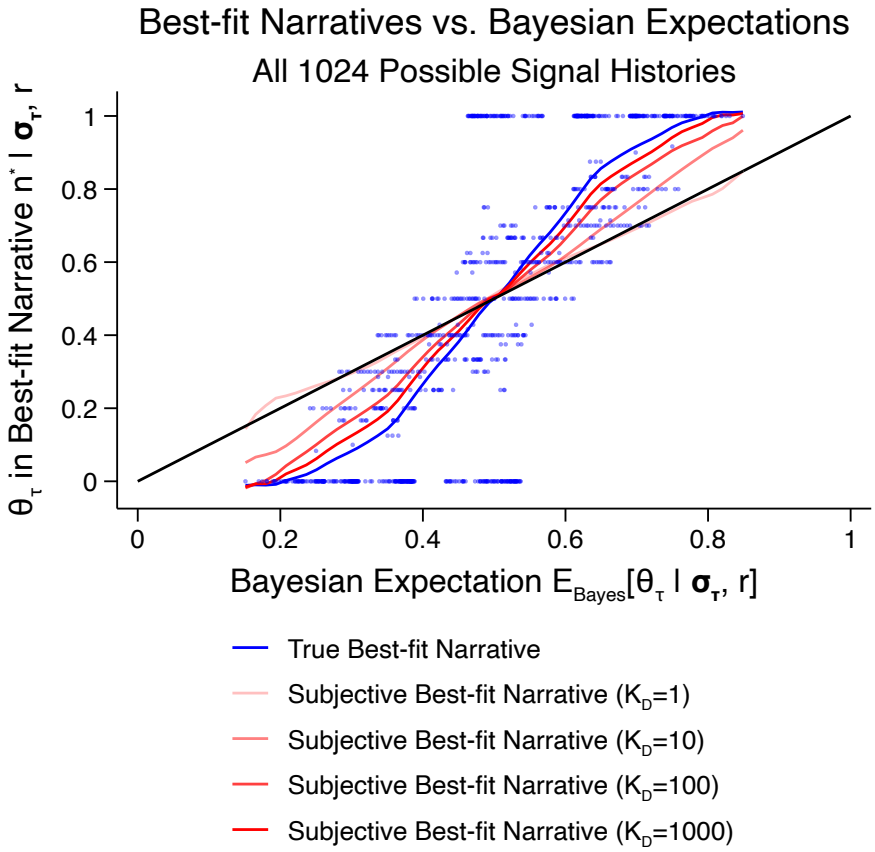
$$\text{Cov}_{\boldsymbol{\sigma}_T \sim \mathbb{P}_T}(\Delta\psi_K, \Delta m^I) > 0 \Rightarrow \text{Vol}_{\boldsymbol{\sigma}_T \sim \mathbb{P}_T}(m^P_{\tau'}, m^P_{\tau''}) > \text{Vol}_{\boldsymbol{\sigma}_T \sim \mathbb{P}_T}(m^I_{\tau'}, m^I_{\tau''}),$$

where

$$\Delta\psi = \psi_K(\boldsymbol{\sigma}_{\tau''}, r) - \psi_K(\boldsymbol{\sigma}_{\tau'}, r), \quad \Delta m^I \equiv m^I(\boldsymbol{\sigma}_{\tau''}, r) - m^I(\boldsymbol{\sigma}_{\tau'}, r).$$

We provide computational evidence in Appendix B.5 that $\text{Cov}_{\boldsymbol{\sigma}_T \sim \mathbb{P}_T}(\Delta\psi_K, \Delta m^I) > 0$ across different parameterizations in the information environment, including the working example. Therefore, when a signal update pushes interim beliefs upward, it tends to also raise the fit-state covariance, thereby amplifying the direction of the belief revision. Such amplification arises naturally in environments like ours due to model-uncertainty neglect. As noted earlier, because a best-fit narrative commits to a particular sequence of states and redraw timings, small changes in the observed signal can dramatically alter which narrative fits best. These shifts induce large movements in ψ_K, reinforcing the belief revision implied by the signal update itself. Posterior beliefs therefore fluctuate more sharply than interim beliefs, generating greater volatility across evaluation periods.

Figure 1: Extremeness of true and subjective best-fit estimates relative to Bayesian benchmark.



Notes: *This figure displays a scatterplot illustrating that, relative to the Bayesian mean estimate, narratives selected based on fit tend to imply more extreme current-state estimates. The blue dots plot the true best-fit estimate of θ_τ against the Bayesian expectation for each of the 1,024 signal histories that can be generated in the working example with $\tau = 10$. The red curves (of varying depth) show local polynomial regressions of subjective best-fit estimates of θ_τ on the Bayesian expectation for varying values of K (the number of sampled narratives).*

Remark on systematic heterogeneity The framework straightforwardly predicts that posterior overreaction and excess volatility occur when interim reaction and volatility are already high: Recall that the posterior mean is equal to $m^I + \alpha_i \psi_K$, so that its bias increases in the absolute size of ψ_K. In our information environment, there is a positive correlation between ψ_K and m^I. Therefore, among optimistic histories with $m^I > 0.5$, increasing m^I on average increases ψ_K, which increases the posterior-interim difference in reaction: Among optimistic signal histories, overreaction is strongest if the interim belief is initially high. Conversely, among pessimistic signal histories, overreaction is strongest if the interim belief is initially low. Heterogeneity in excess volatility follows a similar pattern: Posterior volatility is increasing in $\Delta\psi_K$ and $\Delta\psi_K$ is positively correlated with Δm^I, which is a measure of interim volatility. This generates a positive association between interim volatility and excess posterior volatility. Therefore, the conditions required for narrative sharing to generate

overreaction and excess volatility on average also predict a systematically heterogeneous response, based on the level of interim reaction and volatility.

3 Experimental Design

In the remainder of the paper, we experimentally examine how narrative sharing influences belief biases by implementing a specific instance of the environment outlined in Section 2, corresponding to the *working example* introduced earlier. We design several treatments to induce narrative sharing between a developer and a receiver given a common signal history. Guided by the theoretical framework in Section 2, we formulate empirical hypotheses on how narrative sharing amplifies the extremeness and volatility of beliefs.

3.1 Setup

We frame the belief formation tasks as evaluations of hypothetical "companies" over time. Each company produces a product in each year t, which results in either a **success** ($\sigma_t = 1$) or a **failure** ($\sigma_t = 0$), serving as a binary signal. The underlying state $\theta_t \in \Omega = \{0, 0.01, \ldots, 1\}$ represents the company's **probability of success** in year t, with a uniform prior $\mu(\theta_t) = \frac{1}{101}$ for all $\theta_t \in \Omega$. To describe state redraws or "paradigm shifts," we explain to participants that a company's state is fully determined by its current CEO. In any given year, there is a fixed probability $r = 0.2$ that the company transitions to a new CEO, in which case the state is redrawn from Ω. The set Ω, the prior μ, and the probability r are all known to the participants. This environment corresponds to the "working example" introduced at the end of Section 2.1.

The participants' central task is to report their **mean belief** about the current state θ_τ at each evaluation period $\tau \in \mathcal{T} = \{10, 15, 20, 25, 30\}$, given the visually displayed signal history $\boldsymbol{\sigma}'_\tau = (\sigma_{\tau-9}, \ldots, \sigma_\tau) \in \mathcal{D}$, where $\mathcal{D}$ denotes the set of all possible signal histories. Several comments are in order. First, we focus on the *mean* belief since the full posterior distribution of θ_τ is a highly complex object. Second, to streamline the experimental interface and maintain consistency across evaluation periods, we display only the *ten most recent signals* at each evaluation period rather than the full signal history. Given the relatively high per-period state redraw probability of $r = 0.2$, this simplification has minimal impact on the Bayesian benchmark or the conceptual interpretation of the results.

3.2 Treatments

The experiment consists of three treatments in a between-subject design. The Baseline treatment elicits the *interim* beliefs after observing a signal history, without any narrative sharing. We introduce two additional treatments to examine the impact of narrative sharing on beliefs. In the Developer treatment, each participant is asked to construct a single narrative, which is subsequently shared with another participant in the Receiver treatment. We then elicit beliefs from both Developer and Receiver participants, capturing the *posterior* beliefs after narrative sharing. Comparing beliefs in the Developer and Receiver

treatments to those in the BASELINE treatment allows us to identify the causal impact of narrative sharing on belief formation.

Next, we describe each treatment in detail.

> **Baseline:** For each company, participants begin at evaluation period $\tau = 10$. They observe the company's signal history $\boldsymbol{\sigma}'_{10} = (\sigma_1, \ldots, \sigma_{10})$, as illustrated in Panel (i) of Figure 2. Participants then state their mean belief about θ_{10}—that is, the company's "probability of success" in Year 10. In each subsequent evaluation period, $\tau \in \{15, 20, 25, 30\}$, participants are informed that five additional years have elapsed. They then observe the updated signal history $\boldsymbol{\sigma}'_\tau = (\sigma_{\tau-9}, \ldots, \sigma_\tau)$ and again state their mean belief about θ_τ. Each participant evaluates five companies sequentially, reporting $5 \times 5 = 25$ beliefs in total.

We refer to each evaluation period as one "round." For each participant, one of the 25 rounds is randomly selected for payment. Participants may earn an \$8 bonus based on the accuracy of their reported mean belief, evaluated using the binarized quadratic scoring rule (BQSR, Hossain and Okui, 2013).[14]

Compared with the BASELINE treatment, participants in the DEVELOPER treatment are asked to construct *one specific narrative*—framed as their *best guess* of the state-transition path—before stating their mean belief. This design parallels developers' sharing of *subjective best-fit narratives* in our conceptual framework.

> **Developer:** As in BASELINE, at each evaluation period τ for each company, participants observe the signal history $\boldsymbol{\sigma}'_\tau = (\sigma_{\tau-9}, \ldots, \sigma_\tau)$. They are then asked to provide their *best guess* about whether, when, and how the company's state has changed in the ten most recent years (from year $\tau - 9$ to year τ). Specifically, in a two-step interface, participants first indicate the years (if any) in which they believe a state redraw occurred, and then enter their best guess of the state value before and after each indicated redraw event.[15] After submitting this narrative, participants state their mean belief about θ_τ while their own narrative remains visible on the screen (see Panel (ii) of Figure 2 for an example). This procedure is repeated for all 25 rounds—five evaluation periods for each of five companies.

Participants in DEVELOPER are rewarded for the accuracy of *both* their narratives and their mean beliefs. For each participant, two of the 25 rounds are randomly selected for payment—one for the narrative (the "narrative round") and one for the mean belief (the "belief round").[16] The \$8 bonus for the belief round is determined using the BQSR, as in

[14] Specifically, the probability of receiving the \$8 bonus is

$$100 - \frac{d^2}{100},$$

where d denotes the absolute difference between the reported mean belief about θ_τ and its true value. The BQSR is incentive-compatible for eliciting mean beliefs under a form of compound-lottery reduction known as *Subjective-Objective Reduction*; see Healy and Leo (2024) for details.

[15] See Appendix F for complete experimental instructions.

[16] This procedure prevents hedging motives between narrative and belief reports (Blanco et al., 2010).

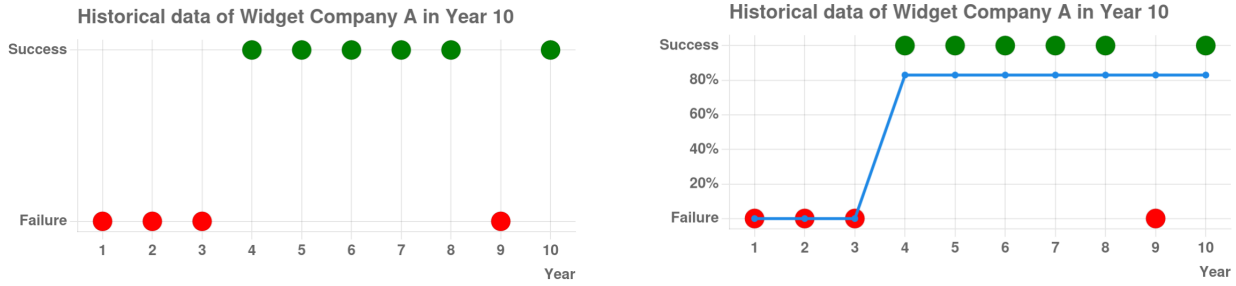
BASELINE. The bonus for the narrative round has two components: Participants receive \$4 if their narrative correctly identifies all state redraws (or their absence), and an additional \$4 based on the accuracy of *the state value in their narrative for a randomly selected year*, again evaluated using the BQSR. We designed this incentivization scheme of narratives with two key considerations in mind. First, participants are *unaware* that their narratives may later be shared with others and are solely incentivized to construct *accurate* narratives. This setup deliberately shuts down strategic motives in narrative construction, consistent with our conceptual framework. Second, the incentive structure is intended to be natural and straightforward for participants to understand.[17] To assess robustness, future extensions may consider alternative incentive structures.

Finally, compared with the BASELINE treatment, participants in the RECEIVER treatment are shown a narrative from a randomly matched participant in DEVELOPER (alongside the signal history) before stating their mean belief about θ_τ in each round.

> **Receiver:** As in BASELINE, at each evaluation period τ for each company, participants observe the signal history $\boldsymbol{\sigma}'_\tau = (\sigma_{\tau-9}, \ldots, \sigma_\tau)$. In contrast to BASELINE, they also view a narrative developed by a randomly matched participant from DEVELOPER for the same signal history, as illustrated in Panel (ii) of Figure 2. Participants are informed that the matched "developers" were incentivized to provide an accurate guess of the state evolution path and that a new match is drawn for each round. Given this information, participants then state their mean belief about θ_τ. This procedure is again repeated for 25 rounds.

Bonus payments in RECEIVER are determined using the BQSR, as in BASELINE.

Figure 2: Illustration of experimental interface for $\tau = 10$ in different treatments



(i) BASELINE (ii) DEVELOPER and RECEIVER

Notes: *The figure shows examples of signal histories (framed as "historical data" of hypothetical companies) as seen by participants in different treatments. The left panel displays the interface in* BASELINE*, where participants report their mean belief about the current state after observing the signal history. The right panel displays the interface in* DEVELOPER *or* RECEIVER*, where participants report their mean belief while a narrative is plotted on top of the signal history. The plotted narrative is developed by the participant themselves in* DEVELOPER *and by another randomly matched participant in* RECEIVER*.*

[17] Strictly speaking, to elicit developers' *subjective best-fit narratives*, they should be paid only if their stated narrative perfectly identifies the true state-evolution path. However, because such exact matches are extremely unlikely, this approach would not be effective in practice.

Relationship of the design to the conceptual framework Our between-subject design employs *separate* participant samples to capture distinct belief objects in the conceptual framework. Specifically, elicited beliefs in BASELINE correspond to *interim beliefs* without narrative sharing; constructed narratives in DEVELOPER measure developers' *subjective best-fit narratives*; and elicited beliefs in DEVELOPER and RECEIVER represent *posterior beliefs* after narrative sharing. In other words, to identify the effect of narrative sharing, we compare beliefs *across* treatments rather than repeatedly eliciting different belief objects *within* a single treatment. We adopt this between-subject approach mainly because of the concern that *explicitly stating* an interim belief may create psychological commitment and reduce participants' willingness to update their beliefs after exposure to new narratives (Falk and Zimmermann, 2018). Accordingly, in the DEVELOPER and RECEIVER treatments, interim beliefs remain implicit rather than being explicitly elicited. This design choice also serves a practical purpose by streamlining the tasks within each round.

Additional Tasks After completing the main belief-elicitation tasks, participants perform an attention check (the "consistency check"), which mirrors the format of the main task. Participants in all treatments evaluate two additional companies with predetermined signal histories, knowing that these evaluations do not affect their payment. Specifically, the first company follows a relatively favorable signal history $(0, 1, 0, 1, 0, 1, 1, 1, 1, 1)$, while the second company has the exact mirror sequence $(1, 0, 1, 0, 1, 0, 0, 0, 0, 0)$. For each company, participants state their mean belief about the company's current state (i.e., the probability of success) in Year 10.

Participants who pay close attention to the task should not update in the wrong direction, and their mean beliefs for the two companies should be roughly symmetric around 0.5 (that is, the two beliefs should approximately sum to 1). We therefore define participants as having "passed" the consistency check if both of the following conditions hold: (i) Their mean belief is not lower than 0.5 for the first company and not higher than 0.5 for the second company; and (ii) The sum of their mean beliefs for the two companies lies between 0.8 and 1.2 (inclusive).[18] This provides us with a subset of "highly consistent" participants, which we employ in robustness tests.

Finally, participants complete a demographic questionnaire before learning their final payment for the experiment.

3.3 Empirical Hypotheses

In this section, we formulate the main empirical hypotheses derived from our experimental design. We begin with the *narratives* constructed in DEVELOPER. We hypothesize that these narratives embed current-state values that are more *extreme* and *volatile* than the beliefs elicited in BASELINE, consistent with their interpretation as *subjective best-fit narratives* in our conceptual framework. Such biases embedded in narratives are then transmitted to beliefs, generating systematic differences in belief patterns across treatments. Specifically, we hypothesize that beliefs in DEVELOPER and RECEIVER are also more *extreme* and *volatile* than those in BASELINE. Finally, we state a set of auxiliary hypotheses concerning the

[18] In Appendix A, we examine the robustness of the results to varying these criteria.

heterogeneity of these effects across different signal histories. All hypotheses introduced in this section are preregistered.

To facilitate exposition, we first introduce notation for the empirical objects elicited in the experiment—namely, narratives and beliefs. Recall that each DEVELOPER narrative, denoted by $\mathbf{n}_\tau^D$, provides a sequence of elicited *state values* covering the ten years of the signal history shown to developers in evaluation period τ: $\mathbf{n}_\tau^D = (\theta_{\tau-9}^D, \ldots, \theta_\tau^D) \in \Omega^{10}$.[19] In such a narrative, θ_τ^D denotes the value of the current state θ_τ. We denote the elicited mean belief about θ_τ, conditional on the signal history $\boldsymbol{\sigma}'_\tau$, in treatment tr by $\hat{\mathbb{E}}_{tr}[\theta_\tau|\boldsymbol{\sigma}'_\tau]$, where $tr \in \{\text{BASELINE}, \text{DEVELOPER}, \text{RECEIVER}\}$. The corresponding Bayesian benchmark based on the same signal history is written as $\mathop{\mathbb{E}}_{\text{Bayes}}[\theta_\tau|\boldsymbol{\sigma}'_\tau]$.[20]

Next, we lay out our hypotheses in detail.

Hypotheses on Extremeness Our framework in Section 2 posits that the current-state values in agents' subjective best-fit narratives are more *extreme* than their interim beliefs. This yields the following hypothesis regarding the narratives constructed in DEVELOPER.

Hypothesis 1. *On average, the θ_τ^D values in observed narratives in* DEVELOPER *are more* extreme *than the elicited mean beliefs of θ_τ in* BASELINE.

To operationalize this hypothesis, we define our main measure of **extremeness** for a θ_τ value—whether from an elicited narrative or a mean belief—as its absolute distance from 0.5, the prior mean of θ_τ:

$$\text{Ext}(\theta_\tau) \;=\; \big|\theta_\tau - 0.5\big|.$$

We also define an **alternative measure of extremeness** that captures deviations from the prior mean *in the direction of the Bayesian benchmark*:[21]

$$\text{Ext}_{\text{Bayes}}(\theta_\tau) = \begin{cases} \theta_\tau - 0.5, & \text{if } \mathop{\mathbb{E}}_{\text{Bayes}}[\theta_\tau|\boldsymbol{\sigma}'_\tau, r] \geq 0.5, \\ 0.5 - \theta_\tau, & \text{if } \mathop{\mathbb{E}}_{\text{Bayes}}[\theta_\tau|\boldsymbol{\sigma}'_\tau, r] < 0.5. \end{cases}$$

Further, the *extremeness* of subjective best-fit narratives should be partially transmitted to agents' *posterior beliefs*, leading to the next hypothesis:

Hypothesis 2. *On average, the elicited mean beliefs of θ_τ are more* extreme *in* DEVELOPER *and* RECEIVER *than in* BASELINE.

[19] Under this approach, we do not distinguish between narratives that induce the same vector of state values but assign redraw events to different points in time. Such multiplicities can occur because developers are allowed to invoke a redraw event while maintaining that the redraw did not change the state value. These cases were expected—and indeed turn out—to be rare, as, when comparing two such narratives, the one with fewer redraw events dominates in fit. Moreover, this distinction is irrelevant for our hypotheses, which concern only the value θ_τ^D embedded in the narrative.

[20] At evaluation periods $\tau = 15, 20, 25, 30$, this Bayesian benchmark is computed using the most recent ten success/failure realizations *currently shown to participants*, rather than the full historical data available in earlier periods. We verified that this distinction is empirically negligible for our experiment: with $r = 20\%$, the current-data and full-history benchmarks are highly correlated. Results are available upon request.

[21] This alternative measure may not coincide with our main measure because empirical beliefs may update from the prior in the opposite direction of the Bayesian benchmark, in which case the alternative measure is defined to be negative.

Hypotheses on Volatility Our framework further predicts that, as agents progress across evaluation periods, the current state values in their (dynamically evolving) subjective best-fit narratives exhibit greater *volatility* than their interim beliefs. This leads to the following hypothesis:

Hypothesis 3. *On average, the θ_τ^D values in observed narratives in* Developer *are more* volatile *across evaluation periods than the elicited mean beliefs of θ_τ in* Baseline.

To operationalize this hypothesis, we define a measure of **volatility** for θ_τ values as their absolute movement between two consecutive evaluation periods:

$$\text{Vol}(\theta_\tau) = \left|\theta_\tau - \theta_{\tau-5}\right|, \text{ for } \tau \in \{15, 20, 25, 30\}.$$

We also define an **alternative, company-level measure of volatility** as the standard deviation of θ_τ values across all five evaluation periods for the same company:

$$\text{Vol}_{\text{Company}}(\theta_{10}, \theta_{15}, \ldots, \theta_{30}) = \sqrt{\frac{1}{5}\sum_{\tau\in\{10,15,\ldots,30\}}(\theta_\tau - \bar{\theta})^2}, \text{ where } \bar{\theta} = \frac{1}{5}\sum_{\tau\in\{10,15,\ldots,30\}}\theta_\tau.$$

Further, the *volatility* of current-state values in subjective best-fit narratives should likewise be transmitted to agents' *posterior beliefs*, yielding the following hypothesis:

Hypothesis 4. *On average, the elicited mean beliefs of θ_τ are more* volatile *across evaluation periods in* Developer *and* Receiver *than in* Baseline.

Hypotheses on Heterogeneous Effects across Signal Histories Beyond average effects, our framework predicts systematic *heterogeneity* in treatment effects across different signal histories. First, Propositions 1 and 2 suggest that narratives amplify reactions to the data: If the signal history conveys relatively strong information about θ_τ, the shared narratives are predicted to be comparatively more extreme. Therefore, the effect of narratives on extremeness should be disproportionately driven by *informative* signal histories. To test this prediction, we partition the set of signal histories $\mathcal{D}$ into two subsamples based on the magnitude of the deviation of the Bayesian expectation of θ_τ from the prior mean of 0.5. Specifically, we define

$$\mathcal{D}_{\text{informative}} = \left\{\boldsymbol{\sigma}'_\tau \in \mathcal{D} : \left|\mathop{\mathbb{E}}_{\text{Bayes}}[\theta_\tau|\boldsymbol{\sigma}'_\tau, r] - 0.5\right| > \text{median deviation}\right\}.$$

as the subsample of *relatively informative* signal histories.[22] We hypothesize that the extremeness effects predicted by Hypotheses 1 and 2 are stronger for relatively informative signal histories.

Hypothesis 5. *The differences between*

- *the θ_τ^D values in observed narratives in* Developer*, and*

[22] In our sample of signal histories, the median deviation of the Bayesian expectation of θ_τ from 0.5 is approximately 0.181.

- *the elicited mean beliefs of* θ_τ *in* DEVELOPER *and* RECEIVER

relative to the elicited mean beliefs of θ_τ *in* BASELINE*, are larger for* $\mathcal{D}_{informative}$*—the* relatively informative *signal histories—than for* $\mathcal{D}\backslash\mathcal{D}_{informative}$*—the* relatively uninformative *ones.*

An additional implication of this prediction concerns the *volatility* of best-fit narratives and posterior beliefs across evaluation periods. Specifically, as suggested by Proposition 3, such volatility should be disproportionately driven by companies with intrinsically *volatile* signal histories—those for which the Bayesian expectation of θ_τ fluctuates substantially across evaluation periods. To test this implication, we partition the set of companies based on the company-level volatility measure defined earlier. We classify as *relatively volatile* those companies whose

$$\text{Vol}_{\text{Company}}\Big(\mathop{\mathbb{E}}_{\text{Bayes}}[\theta_{10}|\boldsymbol{\sigma}'_{10}, r], \mathop{\mathbb{E}}_{\text{Bayes}}[\theta_{15}|\boldsymbol{\sigma}'_{15}, r], \mathop{\mathbb{E}}_{\text{Bayes}}[\theta_{20}|\boldsymbol{\sigma}'_{20}, r], \mathop{\mathbb{E}}_{\text{Bayes}}[\theta_{25}|\boldsymbol{\sigma}'_{25}, r], \mathop{\mathbb{E}}_{\text{Bayes}}[\theta_{30}|\boldsymbol{\sigma}'_{30}, r]\Big)$$

exceeds the sample median (approximately 0.157). We hypothesize that the volatility effects predicted by Hypotheses 3 and 4 are stronger among relatively volatile companies.

Hypothesis 6. *The differences between*

- *the volatility of* θ^D_τ *values in observed narratives across evaluation periods in* DEVELOPER*, and*
- *the volatility of elicited mean beliefs of* θ_τ *across evaluation periods in* DEVELOPER *and* RECEIVER

relative to the volatility of elicited mean beliefs of θ_τ *across evaluation periods in* BASELINE*, are larger for the* relatively volatile *companies than for the* less volatile *ones.*

3.4 Procedures

We programmed the experiment using oTree (Chen et al., 2016) and conducted it in the TIDE Lab at the University of Alabama, with 80 participants per treatment. The experiment and all hypotheses were preregistered in the AEA RCT Registry under the ID **AEARCTR-0015607**.[23]

Before running the first session, we generated the sample of signal histories shown to participants using the following procedure:

- We first drew ten state-evolution paths $(\theta_1, \theta_2, \ldots, \theta_{30})$, each corresponding to a distinct company. For each company, we then generated its *long* signal history $(\sigma_1, \sigma_2, \ldots, \sigma_{30})$ based on the underlying state-evolution path.

[23]We decided to omit two sets of hypotheses from the main text in order to streamline the text, despite finding strong evidence for both. See Appendix E for the omitted hypotheses and corresponding results.

- For every company, we created a corresponding *mirror company* with an *inverted* state-evolution path and signal history. Specifically, if a company's path and signals are given by $(\theta_1, \theta_2, \ldots, \theta_{30})$ and $(\sigma_1, \sigma_2, \ldots, \sigma_{30})$, respectively, its mirror company's path and signals are $(1-\theta_1, 1-\theta_2, \ldots, 1-\theta_{30})$ and $(1-\sigma_1, 1-\sigma_2, \ldots, 1-\sigma_{30})$. This procedure yielded a total of twenty companies.

- To assign the twenty companies to participants, we divided them into four subsets of five companies each: two subsets consisting of the original companies and two consisting of their mirror counterparts. We then randomly assigned an equal number of participants in each treatment to one of the four subsets. Consequently, each company was evaluated by 20 participants per treatment. Finally, for each participant, we randomized the order in which the five assigned companies were evaluated.

This procedure ensured that: (i) all company data were generated exactly according to the DGP described to participants;[24] (ii) the realized signal histories and their presentation order were balanced across treatments; (iii) the distribution of realized signal histories was symmetric—that is, not systematically skewed toward positive or negative outcomes; and (iv) each realized signal history was evaluated by at least 20 participants per treatment, which ensures that the design is well-powered to control for signal-history fixed effects throughout the analysis.

In the Baseline treatment, participants spent an average of 18.6 minutes on the experiment and earned \$14.9 on average. In the Developer treatment, participants spent an average of 32.3 minutes and earned \$18.9 on average. In the Receiver treatment, participants spent an average of 22.4 minutes and earned \$15.0 on average. All treatments included a \$7.5 base payment.

4 Results

In this section, we present our main experimental findings, which broadly validate the preregistered hypotheses in Section 3.3 on how narrative sharing amplifies belief extremeness and volatility. Throughout the analysis, we employ two samples: the *full sample*, which includes all 80 participants per treatment, and the *highly consistent sample*, which restricts attention to participants who passed our "consistency check" at the end of the experiment.[25]

4.1 The Structure of Constructed Narratives

We begin our analysis by studying the basic structure and properties of the narratives constructed in Developer. First, we focus on a key feature—the number of state transitions embedded in these narratives. Table 1 reports the joint distribution of transition counts in the true best-fit narratives and in the developer-constructed narratives. On average, the true best-fit narratives for our signal histories contain 1.46 state transitions, whereas the

[24] Assigning each participant exclusively to either original or mirror companies prevents dependence across evaluations—no participant evaluates both a company and its mirror.

[25] The *highly consistent* subsample consists of 217 out of 240 participants (90.4%) in the full sample: 70 in Baseline, 73 in Developer, and 74 in Receiver.

developer-constructed narratives contain only 1.14. This difference is highly statistically significant ($p < 0.01$) and is evident throughout the distribution: constructed narratives are more likely than true best-fit narratives to include zero (23.0% vs. 16.0%) or one (49.2% vs. 38.0%) transition, and less likely to include two (21.8% vs. 34.0%) or three or more (6.1% vs. 12.0%) transitions.[26]

Overall, the number of transitions in developer-constructed narratives is positively associated with the number in the true best-fit narratives, although the relationship is far from one-to-one. Developers are more likely to indicate zero transitions when the true best-fit narrative has zero transitions, more likely to indicate one when the true best-fit narrative has one, and so on, with each of these effects being highly statistically significant. At the same time, there is a clear tendency for participants to understate the number of transitions relative to the true best-fit: even when the true best-fit contains two or more transitions, most constructed narratives still contain only zero or one. This suggests a reluctance to attribute "too many" transitions to the data. Indeed, compared with cases in which the true best-fit narrative contains a single transition (typically corresponding to a salient shift in the signals), developers are more likely to indicate *no* transition when the true best-fit contains two or more transitions—cases in which the data likely feature multiple structural shifts.

Table 1: Joint distribution of the numbers of state transitions in *true* best-fit narratives and *developer-constructed* narratives.

Constructed Narrative \ *True* Best-fit Narrative	0	1	2	3	4	Total
0	120 (37.5%)	93 (12.2%)	169 (24.9%)	45 (28.1%)	32 (40.0%)	459 (23.0%)
1	114 (35.6%)	498 (65.5%)	283 (41.6%)	65 (40.6%)	24 (30.0%)	984 (49.2%)
2	61 (19.1%)	117 (15.4%)	202 (29.7%)	39 (24.3%)	17 (21.3%)	436 (21.8%)
3	12 (3.8%)	40 (5.3%)	12 (1.8%)	11 (6.9%)	4 (5.0%)	79 (4.0%)
4	4 (1.3%)	2 (0.3%)	13 (1.9%)	0 (0.0%)	3 (3.8%)	22 (1.1%)
5	3 (0.9%)	10 (1.3%)	1 (0.1%)	0 (0.0%)	0 (0.0%)	14 (0.7%)
6	6 (1.9%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	6 (0.3%)
Total	320	760	680	160	80	2,000

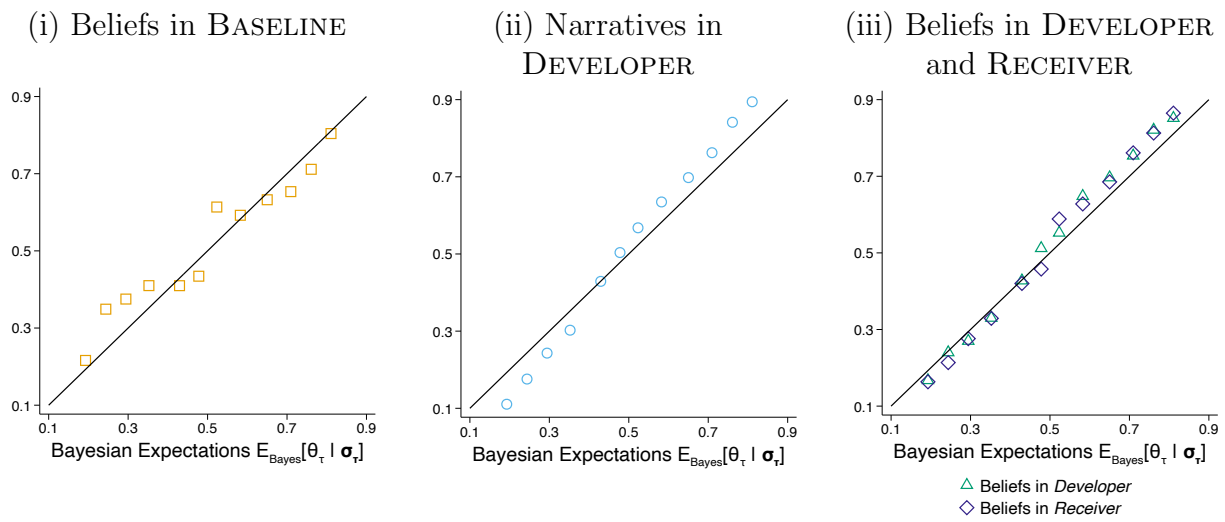
***Notes**: Cell entries are counts, with column percentages in parentheses. Row totals report the marginal distribution of the number of transitions in developer-constructed narratives. The sample consists of all* DEVELOPER *narratives ($N = 2{,}000$) for different signal histories.*

4.2 Narrative Sharing Induces More *Extreme* Beliefs

Next, we present our first main result examining how belief *extremeness* varies systematically across treatments. To visualize belief patterns, we sort signal histories by the implied

[26]The *true best-fit narrative* differs from the *true narrative*. The former refers to the narrative with the highest posterior likelihood given the realized signal history. The latter refers to the actual underlying state-transition path. Given the per-period state-transition probability, a company experiences $9 \times 0.2 \times \frac{100}{101} \approx 1.78$ expected transitions over ten periods; in our sample, the true narratives have an average of 1.88 transitions. Both of these numbers are larger than the average transition numbers of the true best-fit narratives, which is 1.46. Thus, even before considering participants' behavior, the true best-fit narratives already contain fewer transitions than the true underlying paths.

Figure 3: Patterns of beliefs across different treatments.



Notes: *The figure displays binned scatterplots illustrating how elicited narratives and beliefs vary with the Bayesian benchmark across treatments. Panel (i) plots mean beliefs in* BASELINE. *Panel (ii) shows the current-state values in narratives constructed in* DEVELOPER. *Panel (iii) plots mean beliefs in* DEVELOPER *and* RECEIVER.

Bayesian expectation of the current state, $\mathop{\mathbb{E}}_{\text{Bayes}}[\theta_\tau|\boldsymbol{\sigma}'_\tau]$, and plot how reported beliefs covary with this benchmark across treatments in Figure 3.

Panel (i) shows that mean beliefs in BASELINE follow an "inverse-S-shape," consistent with *underinference* (Benjamin, 2019): beliefs are too high (insufficiently pessimistic) when the Bayesian benchmark is low and too low (insufficiently optimistic) when it is high. This pattern indicates that BASELINE participants react too little to information in the data.

Panel (ii) reveals the opposite pattern for the current-state values embedded in narratives constructed in DEVELOPER. For unfavorable signal histories (i.e., those with a low Bayesian expectation), narrative-implied values fall below the Bayesian benchmark, whereas for favorable signal histories they lie above it. Consequently, the DEVELOPER narratives imply more extreme assessments of the company's current state, consistent with Hypothesis 1. As shown in Panel (iii), these narratives also exert a strong influence on participants' beliefs. In both DEVELOPER and RECEIVER, beliefs respond more strongly to information than in BASELINE and systematically overreact relative to the Bayesian benchmark, consistent with Hypothesis 2. The magnitude of overreaction is similar across the two treatments, implying a comparable degree of transmission from narratives to beliefs. Together, these results demonstrate that narrative sharing amplifies reactions to data and generates more extreme beliefs, producing biases that are qualitatively distinct from those observed in BASELINE.

We formalize these findings using the following OLS specification estimated on pooled data across treatments:

$$\text{Ext}(y_{i\boldsymbol{\sigma}t}) = \beta_{\text{DEV}}\mathbb{1}(\text{DEVELOPER}) + \beta_{\text{REC}}\mathbb{1}(\text{RECEIVER}) + \rho_t + \delta_{\boldsymbol{\sigma}} + \varepsilon,$$

where the dependent variable is our main extremeness measure. This measure quantifies the distance between participant i's current state evaluation and the prior mean (0.5) when

observing signal history $\boldsymbol{\sigma}$ in round t of the experiment. In DEVELOPER, we measure two such state evaluations in each round, once by eliciting a narrative (which implies a current state estimate) and a mean belief. In the remaining treatments, we only elicit the mean belief. The variables $\mathbb{1}$(DEVELOPER) and $\mathbb{1}$(RECEIVER) are treatment indicators, ρ_t are evaluation-period fixed effects, and $\delta_{\boldsymbol{\sigma}}$ are signal history fixed effects.

Table 2: Narrative sharing induces more extreme beliefs

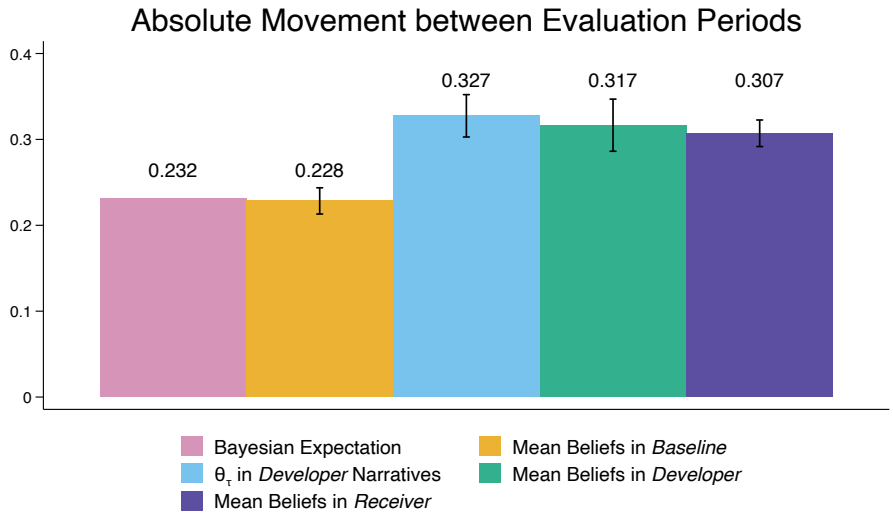
	Main Extremeness Measure: $\lvert\theta_\tau - 0.5\rvert$			
	Narratives vs. BASELINE Beliefs		Beliefs across Treatments	
	(1)	(2)	(3)	(4)
DEVELOPER Narratives	0.063***	0.062***		
	(0.009)	(0.009)		
DEVELOPER Beliefs			0.050***	0.049***
			(0.012)	(0.012)
RECEIVER Beliefs			0.042***	0.042***
			(0.007)	(0.007)
Sample	Full Sample	Highly Consistent	Full Sample	Highly Consistent
Round FE	Yes	Yes	Yes	Yes
Signal History FE	Yes	Yes	Yes	Yes
BASELINE Mean	0.207	0.210	0.207	0.210
Observations	4000	3575	6000	5425

Notes: *Standard errors are clustered by participant. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. The dependent variable is the absolute deviation of the elicited value from 0.5, the prior mean of the state. Columns (1)–(2) compare current-state values in* DEVELOPER *narratives to mean beliefs in* BASELINE. *Columns (3)–(4) compare mean beliefs across all treatments. The omitted category is always mean beliefs in* BASELINE. *The "highly consistent" subsample includes participants who passed the preregistered consistency check.*

Table 2 reports the regression results. Columns (1)–(2) compare the current-state values embedded in DEVELOPER narratives to mean beliefs in BASELINE, corresponding to the comparison between Panels (i) and (ii) of Figure 3. Consistent with Hypothesis 1, narrative-implied state estimates are 6.3 percentage points farther from the prior than BASELINE expectations ($p < 0.01$). Relative to the BASELINE distance of 0.207, this represents an approximately 30% increase in extremeness when moving from BASELINE beliefs to DEVELOPER narratives.

Columns (3)–(4) compare mean beliefs in DEVELOPER and RECEIVER to those in BASELINE, corresponding to the comparison between Panels (i) and (iii) in Figure 3. Consistent with Hypothesis 2, mean beliefs in DEVELOPER and RECEIVER are 24% ($= \frac{0.050}{0.207}$) and 20% ($= \frac{0.042}{0.207}$) more extreme than those in BASELINE, respectively, with both differences significant at the 1% level. Table A1 in the online appendix shows that these findings are robust to using the alternative extremeness measure $\text{Ext}_{\text{Bayes}}$ defined in Section 3.3. In an additional exercise, Table A8 in the online appendix formally examines under- and overinference, indicating that participants significantly underinfer in BASELINE, while they overinfer in both

Figure 4: Narrative sharing induces more volatile beliefs.



Notes: *The figure plots the average of our main volatility measure—the absolute movement between two consecutive evaluation periods—for mean beliefs across treatments and for current-state values in* Developer *narratives. Error bars indicate 95% confidence intervals based on standard errors clustered at the participant level.*

Developer and Receiver.

Taken together, the results indicate that narratives embed extreme assessments of the current state. In turn, these narratives exert a substantial influence on individuals' beliefs, rendering them markedly more extreme than beliefs in Baseline.

4.3 Narrative Sharing Induces More *Volatile* Beliefs

In this section, we further document how belief *volatility* varies systematically across treatments. Figure 4 presents our main measure of volatility—the absolute movement between two consecutive evaluation periods—both for mean beliefs across different treatments and for the current-state values in Developer narratives. For reference, the figure also reports the volatility of the Bayesian expectation as a benchmark.

We find that the volatility of Baseline beliefs across evaluation periods is statistically indistinguishable from that of the Bayesian benchmark ($p = 0.52$), despite Baseline beliefs exhibiting underinference, as shown above. One likely explanation is that elicited beliefs contain idiosyncratic noise, which adds variability across periods and offsets the dampening effect of underinference, resulting in volatility levels close to the Bayesian benchmark.

Importantly, Figure 4 shows that the current-state values embedded in Developer narratives are approximately 43% more volatile than Baseline beliefs and 41% more volatile than the Bayesian benchmark, supporting Hypothesis 3. Moreover, this excess volatility is strongly transmitted to beliefs: mean beliefs in Developer and Receiver are, respectively, 39% and 35% more volatile than Baseline beliefs ($p < 0.01$), consistent with Hypothesis 4.

Taken together, these results show that narrative sharing not only produces more extreme beliefs but also amplifies their movements over time, generating excess volatility. Table A2

in the online appendix reports the formal regression results, while Figure A1 and Table A3 confirm the robustness of these findings to using the alternative company-level volatility measure $\text{Vol}_{\text{Company}}$ defined in Section 3.3.

4.4 Heterogeneous Effects across Signal Histories

In this section, we test additional predictions of our framework concerning heterogeneous effects across observed signal histories. We first examine whether the effect of narratives on belief *extremeness* is stronger for *informative* signal histories—those for which the Bayesian benchmark is more extreme. To do so, we estimate an OLS specification that extends our baseline model by including an indicator for relatively informative signal histories and its interaction with treatment:

$$\begin{aligned}\text{Ext}(y_{i\boldsymbol{\sigma}t}) = \beta_{\text{DEV}}\mathbb{1}(\textsc{Developer}) + \alpha_{\text{DEV}}\mathbb{1}(\textsc{Developer}) \times \mathbb{1}(\boldsymbol{\sigma} \in \mathcal{D}_{\text{informative}}) \\ +\beta_{\text{REC}}\mathbb{1}(\textsc{Receiver}) + \alpha_{\text{REC}}\mathbb{1}(\textsc{Receiver}) \times \mathbb{1}(\boldsymbol{\sigma} \in \mathcal{D}_{\text{informative}}) + \mathbb{1}(\boldsymbol{\sigma} \in \mathcal{D}_{\text{informative}}) + \rho_t + \varepsilon.\end{aligned}$$

In this specification, the dependent variable is our main extremeness measure—the absolute distance of the state evaluation from 0.5. The variables $\mathbb{1}(\textsc{Developer})$ and $\mathbb{1}(\textsc{Receiver})$ are treatment indicators, while $\mathbb{1}(\boldsymbol{\sigma} \in \mathcal{D}_{\text{informative}})$ indicates whether the signal history is classified as *relatively informative*, as defined in Section 3.3. We include round fixed effects, denoted by ρ_t. We do not include signal-history fixed effects, as identification of the interaction coefficients α_{DEV} and α_{REC} relies on variation across signal histories.

Table 3 reports the regression results. As expected, participants' beliefs are, on average, more extreme for relatively informative signal histories. More importantly, Column (1) shows that the interaction between Developer narratives and relatively informative signal histories is positive and statistically significant. This indicates that the extremeness gap between the current-state values embedded in Developer narratives and Baseline beliefs is larger for relatively informative signal histories ($p < 0.01$). Column (2) shows that this interaction pattern extends to beliefs in both Developer and Receiver. Together, these results indicate that the effect of narrative sharing on belief extremeness is substantially larger for relatively informative signal histories. Tables A4 and A5 in the online appendix confirm that these findings are robust to (i) restricting the sample to participants who passed the consistency check and (ii) using the alternative extremeness measure $\text{Ext}_{\text{Bayes}}$ defined in Section 3.3.

Next, we test whether the effect of narratives on belief *volatility* is significantly stronger for *intrinsically volatile* companies—those for which the Bayesian expectation of the current state fluctuates substantially across evaluation periods. To do so, we estimate the following specification:

$$\begin{aligned}\text{Vol}_{i\boldsymbol{\sigma}t} = \beta_{\text{DEV}}\mathbb{1}(\textsc{Developer}) + \alpha_{\text{DEV}}\mathbb{1}(\textsc{Developer}) \times \mathbb{1}(c \text{ is relatively volatile}) \\ +\beta_{\text{REC}}\mathbb{1}(\textsc{Receiver}) + \alpha_{\text{REC}}\mathbb{1}(\textsc{Receiver}) \times \mathbb{1}(c \text{ is relatively volatile}) \\ +\mathbb{1}(c \text{ is relatively volatile}) + \rho_t + \varepsilon,\end{aligned}$$

where the dependent variable is the absolute change in participant i's state evaluation between rounds $t-1$ and t for the same company. The indicators $\mathbb{1}(\textsc{Developer})$ and

Table 3: Stronger effects of narrative sharing for relatively informative signal histories

	Main Extremeness Measure: $\lvert\theta_\tau - 0.5\rvert$	
	(1)	(2)
Relatively Informative	0.066***	0.064***
	(0.009)	(0.009)
Developer Narratives	0.031***	
	(0.011)	
Developer Narratives × Relatively Informative	0.064***	
	(0.012)	
Developer Beliefs		0.032**
		(0.013)
Receiver Beliefs		0.010
		(0.008)
Developer Beliefs × Relatively Informative		0.036***
		(0.012)
Receiver Beliefs × Relatively Informative		0.063***
		(0.012)
Round FE	Yes	Yes
Mean of Dep. Var. for Omitted Category (Baseline Beliefs for Relatively Uninformative Signal Histories)	0.172	0.172
Observations	4000	6000

Notes: *Standard errors are clustered by participant. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. The results in this table use the full sample of 240 participants (80 in each treatment). The dependent variable is the absolute deviation of the elicited value from 0.5, the prior mean of the state. Columns (1) compares current-state values in* Developer *narratives to mean beliefs in* Baseline*. Columns (2) compares mean beliefs across all treatments. The omitted category is always mean beliefs in* Baseline *for the* relatively uninformative *signal histories.* Relatively informative *signal histories are histories with a Bayesian expectation that is more distant from 0.5 than the median in our sample.*

$\mathbb{1}$(Receiver) denote treatment assignment, while $\mathbb{1}(c$ is relatively volatile) indicates whether the signal history corresponds to a company c classified as relatively volatile, as defined in Section 3.3. We include round fixed effects, denoted by ρ_t.

Table 4 presents the results. As expected, participants' evaluations are, on average, more volatile for relatively volatile companies. Column (1) shows that the increase in volatility of the current-state values embedded in Developer narratives—relative to Baseline beliefs—is significantly larger for relatively volatile companies than for less volatile ones ($p < 0.01$). Column (2) shows that similar interaction patterns emerge for beliefs in Developer and Receiver, although the effects are smaller in magnitude and only marginally significant for Developer beliefs ($p = 0.07$). Overall, these results indicate that the effect of narrative sharing on belief volatility is stronger for intrinsically more volatile companies.

Table 4: Stronger effects of narrative sharing for relatively volatile companies

	Main Volatility Measure: $\lvert\theta_\tau - \theta_{\tau-5}\rvert$	
	(1)	(2)
Relatively Volatile	0.076***	0.076***
	(0.008)	(0.008)
Developer Narratives	0.056***	
	(0.013)	
Developer Narratives × Relatively Volatile	0.086***	
	(0.016)	
Developer Beliefs		0.073***
		(0.018)
Receiver Beliefs		0.056***
		(0.011)
Developer Beliefs × Relatively Volatile		0.029*
		(0.016)
Receiver Beliefs × Relatively Volatile		0.045***
		(0.014)
Round FE	Yes	Yes
Mean of Dep. Var. for Omitted Category (Baseline Beliefs for Relatively Involatile Companies)	0.190	0.190
Observations	3200	4800

Notes: *Standard errors are clustered by participant. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. The results in this table use the full sample of 240 participants (80 in each treatment). The dependent variable is the absolute movement of the elicited value across two adjacent evaluation periods. Column (1) compares current-state values in* Developer *narratives to mean beliefs in* Baseline*. Column (2) compares mean beliefs across all treatments. The omitted category is always mean beliefs in* Baseline *for the* relatively involatile *companies.* Relatively volatile *companies are companies with a higher-than-median standard deviation of the Bayesian expectation across evaluation periods.*

Tables A6 and A7 in the online appendix confirm that these findings are robust to (i) restricting the sample to participants who passed the consistency check and (ii) using the alternative company-level volatility measure defined in Section 3.3.

4.5 Additional Evidence on Mechanisms

In this section, we present supplementary results that further clarify the mechanisms through which narrative construction and transmission amplify belief extremeness and volatility. We first show that the extremeness and volatility of constructed narratives arise from two distinct sources: the placement of state-transition (redraw) points and the choice of state values within the resulting segments. Next, we demonstrate that narratives with a higher empirical fit—precisely those that are most extreme and volatile—exert a stronger influence

on receivers' beliefs. Finally, we show that narratives reshape biases in belief formation, attenuating responses to distant signals while amplifying responses to the most recent ones, in some cases generating overreaction to recent information.

4.5.1 Decomposing Narrative Overinference, Extremeness and Volatility

We find that narratives imply systematic overinference from the data relative to the Bayesian benchmark. This overinference operates through two distinct channels. First, a narrative specifies a set of state-transition points (or the absence thereof) over the observed period. When constructed narratives—similar to best-fit narratives—place transitions at locations that maximize contrast across segments, sorting the signal history into periods of sustained success and failure, this *state-transition placement* alone increases extremeness and thus contribute to overinference. Second, conditional on the chosen state-transition points, developers form a *point estimate* of the current state. When this estimate—again resembling the best-fit estimate—gravitates toward the raw share of success signals following the most recent transition, extremeness increases further beyond what is implied by state-transition placement alone. We next decompose overinference, extremeness, and volatility of narratives into these two channels.

To isolate these mechanisms, we define a key counterfactual: the *conditional Bayesian* estimate. Let X denote the number of positive signals and Y the total number of signals observed after the most recent transition in a participant's constructed narrative. The conditional Bayesian estimate is then $\frac{X+1}{Y+2}$, which represents the rational posterior mean conditional on the narrative's state-transition path. Comparing this conditional Bayesian benchmark to the unconditional Bayesian benchmark isolates the bias attributable to state-transition commitment. The remaining overinference reflects deviations of the narrative's point estimate from the conditional Bayesian benchmark, such as a tendency to report the *conditional best-fit* estimate $\frac{X}{Y}$.

Panel A of Table 5 decomposes overinference, extremeness, and excess volatility of DEVELOPER narratives into these two components. Several patterns emerge. First, state-transition placement accounts for between 23% and 38% of overinference, extremeness, and volatility. This indicates that developers systematically place state-transition points at locations that accentuate contrast across segments. The remaining share is driven by point estimates that deviate from the conditional Bayesian benchmark.

Panels B and C provide two counterfactual exercises that help contextualize these results. Panel B isolates the role of point estimation by replacing developers' subjective point estimates of θ_τ with the *conditional best-fit*—the maximum likelihood estimate given the developers' chosen transition points. Comparing Panels A and B shows that all outcome differences would be larger if developers always selected the conditional best-fit, indicating that developers' point estimates are conservative relative to this benchmark. Panel C then replaces developers' chosen transition paths with the *true best-fit* path, effectively presenting the true best-fit narrative and illustrating the extent to which state-transition placement can further increase overinference, extremeness, and volatility.

Taken together, this evidence shows that developers' narratives are biased relative to the Bayesian benchmark both because developers selectively choose state-transition points that maximize contrast and because they engage in non-Bayesian point estimation. At the same

time, these biases exhibit a degree of *conservatism*: developers neither consistently select the true best-fit transition path nor fully match success frequencies in their point estimates. Table A9 in the online appendix shows that these findings are robust to using alternative measures of extremeness and volatility.

Table 5: Overinference, extremeness, and volatility of current-state estimates under different benchmarks.

	Overinference		Extremeness: $\lvert\theta_\tau - 0.5\rvert$		Volatility: $\lvert\theta_\tau - \theta_{\tau-5}\rvert$	
	Coeff.	% of Total	Mean	% of Total Bias	Mean	% of Total Bias
A. Developer Narratives						
(1) Bayesian Benchmark	1.000	–	0.171	–	0.232	–
(2) Empirically Observed Narrative	1.272	100%	0.270	100%	0.327	100%
Decomposition of Total Overinference/Bias:						
– Due to state-transition placement	0.063	23.2%	0.038	38.0%	0.023	24.1%
– Due to point estimation	0.209	76.8%	0.062	62.0%	0.073	75.9%
B. Conditional Best-Fit Narratives						
(3) Narrative w/ Best-Fit Estimate	1.538	100%	0.315	100%	0.396	100%
Decomposition of Total Overinference/Bias:						
– Due to state-transition placement	0.063	11.7%	0.038	26.1%	0.023	14.0%
– Due to point estimation	0.475	88.3%	0.107	73.9%	0.142	86.0%
C. True Best-Fit Narratives						
(4) True Best-Fit Narrative	1.725	100%	0.373	100%	0.537	100%
Decomposition of Total Overinference/Bias:						
– Due to state-transition placement	0.222	30.7%	0.062	30.6%	0.104	34.0%
– Due to point estimation	0.503	69.3%	0.141	69.4%	0.202	66.0%

Notes: *This table decomposes overinference, extremeness, and bias in narratives into two channels: (i) a* ***state-transition placement*** *channel, arising from the selection of a particular state-transition path, and (ii) a* ***point-estimation*** *channel, arising from the choice of state values given that path. Overinference is measured by regressing the corresponding θ_τ estimate on the Bayesian expectation of θ_τ; the table reports the resulting coefficient estimates for each comparison group. Panel A presents this decomposition for empirically observed narratives from the* Developer *treatment. Panel B reports the same decomposition for* conditional best-fit *narratives, estimated using the state-transition path in* Developer *narratives. Panel C reports the decomposition for the theoretical* true best-fit *narratives. "Total overinference/bias" is defined as the difference between the narrative-implied value and the fully Bayesian benchmark. The* ***state-transition placement*** *component is the difference between the "Conditional Bayesian" counterfactual—constructed using the narrative's transition points—and the fully Bayesian benchmark. The* ***point-estimation*** *component is the residual difference.*

4.5.2 Stronger Impact of Better-fit Narratives on Receiver Beliefs

Results in Section 4.2 show that narratives constructed in Developer substantially influence the posterior beliefs of both developers and receivers, amplifying belief extremeness and volatility. We now ask a more nuanced question: Do narratives exert a *stronger* influence when they provide a *better fit* to the observed data? This question is important because, as shown in the conceptual framework in Section 2, higher-fit narratives tend to embed more

extreme and volatile current-state values.[27] If such narratives also receive greater weight in belief formation, they would further magnify belief extremeness and volatility relative to a benchmark in which narratives of different fit affect beliefs equally.

Studying this question is complicated by the fact that beliefs are elicited only once in our design, which prevents a direct estimation of the *weight* participants place on the narrative when forming their posterior beliefs. We therefore take an indirect approach. Specifically, we use the *distance* between a participant's posterior mean belief and the current-state value embedded in the narrative as an inverse proxy for narrative influence: smaller distances indicate greater weight placed on the narrative. We estimate the following specification separately for Developer and Receiver beliefs:

$$\begin{aligned}\text{DistanceBeliefNarrative}_{i\boldsymbol{\sigma}t} \quad &= \beta_1 \text{NarrativeLikelihoodRank}_{i\boldsymbol{\sigma}t} \\ &+\beta_2 \text{DistanceNarrativeBayesian}_{i\boldsymbol{\sigma}t} + \rho_t + \delta_{\boldsymbol{\sigma}} + \varepsilon\end{aligned}$$

The dependent variable is the distance measure described above. The main independent variable, *NarrativeLikelihoodRank*, is defined as the rank of the narrative's likelihood among all narratives constructed for the same signal history, normalized to range from 0 (best fit) to 1 (worst fit).[28] Because we include signal-history fixed effects, the coefficient β_1 captures whether, for the same signal history, worse-fit narratives are associated with larger distances between beliefs and narratives. We also control for *DistanceNarrativeBayesian*, the distance between the narrative's current-state value and the Bayesian posterior mean implied by the same signal history, to account for how far the narrative deviates from the fully Bayesian benchmark.

In Receiver, narratives are randomly assigned to participants. Therefore, this specification identifies the causal effect of narrative fit on receivers' updating toward narratives. In contrast, in Developer, participants observe narratives they constructed themselves. As a result, the coefficient on *NarrativeLikelihoodRank* in Developer reflects correlations rather than causal effects.

Table 6 reports the results. Column (1) shows that, in Developer, narrative likelihood rank is not significantly related to the distance between participants' posterior beliefs and the current-state values embedded in their own narratives. One interpretation is that developers misperceive the quality of their narratives or are overconfident in them, leading them to place consistently high weight on their narratives regardless of fit.

By contrast, Column (2) reveals a statistically significant relationship in Receiver. Receivers place systematically greater weight on better-fit narratives: their posterior beliefs lie closer to the current-state values embedded in narratives with higher likelihood. Quantitatively, moving from the best-fit to the worst-fit narrative for a given signal history is associated with an increase of 0.023 in the distance measure, corresponding to 15% of a standard deviation ($p = 0.050$). This result indicates that narrative fit materially affects the extent to which receivers incorporate narratives into their posterior beliefs.

We provide two additional exercises in the online appendix. Table A10 shows that the pattern is robust to an alternative measure of narrative influence—whether the receiver's

[27] Consistent with this mechanism, we also observe a positive correlation between narrative extremeness and fit in the data; see Figure A4 in the online appendix.

[28] Each of the 100 signal histories in our sample is shown to 20 developers. We rank the resulting 20 narratives by likelihood and normalize the rank to the interval $[0, 1]$.

Table 6: Stronger effects of better-fit narratives on receiver (but not developer) beliefs

	Distance between Belief and Narrative	
	(1) DEVELOPER	(2) RECEIVER
Narrative Likelihood Rank (Normalized)	0.004	0.023**
	(0.016)	(0.012)
Distance between Narrative and Bayesian	0.201***	0.501***
	(0.074)	(0.048)
Round FE	Yes	Yes
Signal History FE	Yes	Yes
Mean of Dependent Variable	0.100	0.116
Observations	2000	2000

Notes: *Standard errors are clustered by participant. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. Column (1) uses the posterior beliefs in* DEVELOPER *and Column (2) uses the posterior beliefs in* RECEIVER*. The dependent variable is the distance between a participant's posterior mean belief and the current state embedded in the observed narrative. The variable "Narrative Likelihood Rank" is defined as the rank of the narrative among all the narratives constructed for the same signal history in our sample, normalized to the interval* $[0, 1]$ *(with 0 indicating the best fit and 1 indicating the worst fit). The variable "Distance between Narrative and Bayesian" is defined as the distance between the current state embedded in the observed narrative and the Bayesian posterior mean of the current state.*

posterior belief moves in the opposite direction from the narrative's current-state value (i.e., the two lie on opposite sides of the prior mean of 0.5). Table A11 further demonstrates that the results are robust to an alternative set of controls.

4.5.3 Narratives Amplify Reaction to Recent Signals and Dampen Reaction to More Distant Ones

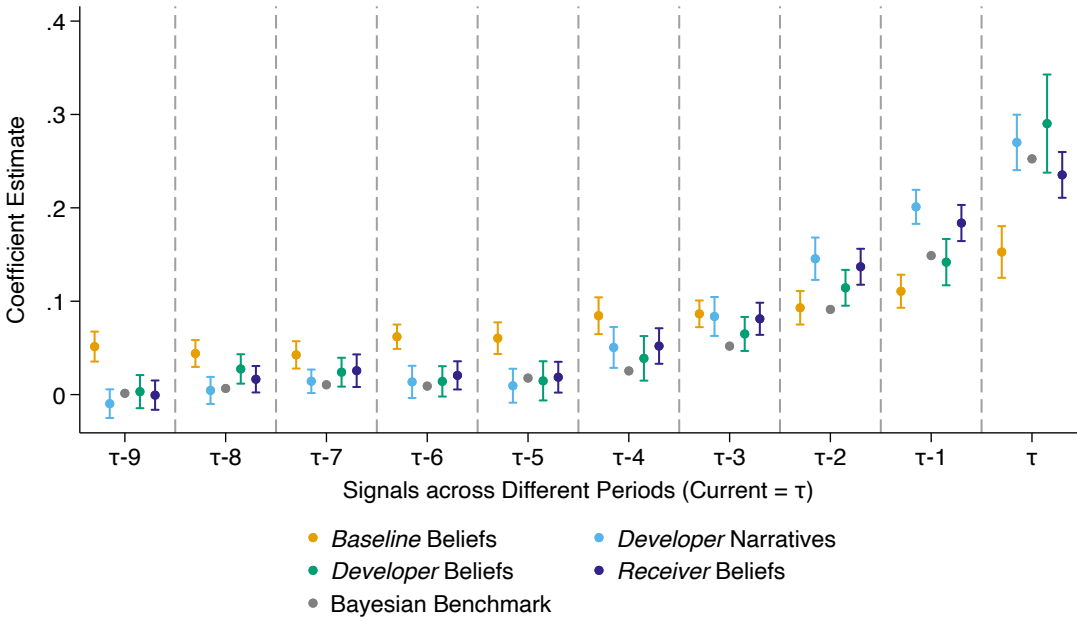
The previous sections documented that narrative sharing leads to beliefs that are more extreme and volatile. In this section, we ask whether narratives simply amplify reactions to all signals uniformly, or whether they alter how agents learn from data in a more nuanced manner. We hypothesize that narratives induce *structural overfitting*: they shift participants' attention toward a "current regime" based on a short window of recent data, effectively downweighting signals that are more distant in the history.

To test this hypothesis, we estimate how sensitive mean beliefs (and current-state values in narratives) are to each of the ten most recent binary signals—the signals readily available to participants at the moment of narrative construction and belief formation—and compare these sensitivity estimates across treatments and to the Bayesian benchmark. We estimate the following specification separately for each treatment:

$$y_{i\boldsymbol{\sigma}\tau} = \alpha + \sum_{t=0}^{9} \beta_t \, \sigma_{\tau-t} + \varepsilon,$$

where $y_{i\boldsymbol{\sigma}\tau}$ is either participant i's mean belief about the current state or, in the case of DEVELOPER, the current-state value in participant i's constructed narrative when observing signal history $\boldsymbol{\sigma}$ in evaluation period τ.

Figure 5: Sensitivities of mean beliefs and narratives to signals across different periods



Notes: *Each panel plots the estimated coefficients β_t on signals from periods τ (most recent) to $\tau-9$ (most distant) from separate regressions estimated within each treatment. "Narratives" refer to the current-state value embedded in the constructed narrative, while "Beliefs" refer to the participants' reported mean beliefs in different treatments. Error bars are 95% confidence intervals derived from standard errors clustered by participant.*

We plot the resulting coefficient series in Figure 5. The results suggest large differences in inference across treatments. First, BASELINE beliefs tend to *smooth* over the signal history: participants underreact to the most recent signals (periods τ and $\tau-1$) while overreacting to more distant signals (periods $\tau-9$ to $\tau-3$), relative to the Bayesian benchmark. This pattern suggests that, without narratives, beliefs are sluggish: participants average over a long history and do not sufficiently account for the obsolescence of old data in a changing environment.

Narrative construction *reverses* this smoothing bias. Current-state values in DEVELOPER narratives exhibit sharply *dampened* sensitivity to distant signals. However, this comes at the cost of an overreaction to the most recent signals (periods $\tau-4$ to τ), where their sensitivity exceeds the Bayesian benchmark. These patterns are transmitted, though somewhat attenuated, to mean beliefs in both DEVELOPER and RECEIVER.

This analysis identifies a key mechanism behind the treatment effects. When moving from BASELINE to DEVELOPER, beliefs shift from *smoothing* over the full signal history to *overreacting* to recent signals. By highlighting a specific state-transition path, narratives concentrate attention on the "now," increasing exposure to the noise in recent realizations and generating the excess volatility documented in Section 4.3.

4.6 Structural Estimation

The reduced-form estimates confirm that narratives induce overreaction and excess volatility. This section develops a structural model to assess whether the framework of Section 2 can quantitatively capture the patterns highlighted in the reduced-form analysis.

4.6.1 Model and Estimation Strategy

Estimating the framework requires estimating a sampling function from which agents draw their narratives; a sample size parameter K, determining the number of narratives each agent considers; and persuasion parameters α_D and α_R denoting the extent of updating towards the shared narrative. We assume that the sampling distribution is a power transformation of the Bayesian posterior: $f(\mathbf{n}_\tau|\boldsymbol{\sigma}_\tau, r) \propto \pi(\mathbf{n}_\tau|\boldsymbol{\sigma}_\tau, r)^\beta$, where $\beta > 0$ is a distortion parameter. This parametric form is relatively flexible and permits Bayesian sampling ($\beta = 1$), sampling from a "flat" posterior ($\beta < 1$) or sampling from a "concentrated" posterior ($\beta > 1$). The parameter β fully pins down the sampling distribution. In addition, we assume that participants report their mean estimates with noise. Denoting the true mean estimate by θ^*_τ, the reported mean estimate is $\theta_\tau = \theta^*_\tau + \varepsilon$, where $\varepsilon \sim N(0, \sigma^2_\varepsilon)$. This noise term allows the model to account for reporting errors and other sources of idiosyncratic variation.

To estimate the model, we use two-step Simulated Method of Moments (SMM). Use $\mu(\xi)$ to denote the moment vector predicted by the framework and $\hat{\mu}$ to denote the vector of empirical moments. The SMM procedure identifies the parameter vector that minimizes $(\mu(\xi) - \hat{\mu})' W (\mu(\xi) - \hat{\mu})$, where W is a weighting matrix. The two step procedure first estimates $\hat{\xi}_1$ while setting W equal to the identity matrix. Based on the first step estimates, the procedure estimates $\hat{W}(\hat{\xi}_1)$. In the second step, it updates the parameter vector by minimizing the objective while setting $W = \hat{W}(\hat{\xi}_1)$. This procedure results in asymptotically efficient estimates, and standard errors for all real-valued parameters can be recovered using the common sandwich formula. Since parameter K is an integer, the sandwich formula does not apply and we do not derive a standard error for its particular parameter.[29]

We estimate this model using history-level moments. Our sample features 48 distinct signal histories,[30] and for each of these histories we aim to match the following five moments: (i)-(iii) the average mean belief in Baseline, Developer, and Receiver, (iv) the average narrative-implied estimate in Developer, (v) the Baseline belief variance. Thus, our structural estimation targets $48 \times 5 = 240$ moments. The moment conditions identify all model parameters, and while they are estimated jointly, we can describe the individual contribution of each movement to the parameter identification. The Baseline mean beliefs identify the sampling distribution parameter β, while the Developer and Receiver beliefs identify the persuasion parameters α_D, α_R. The sample size parameter K is identified both by the average narrative-implied estimate in Developer and by the belief variance in

[29] A standard error for K could be derived via bootstrapping. Because our estimation is computationally very demanding (it requires simulating moments at each iteration step), we did not not implement a bootstrap estimator.

[30] As described in Section 3.4, we perfectly balanced signal histories. While the experiment featured 96 distinct signal histories, half of them (48) were perfect mirrors of the other half. In the structural estimation, we work with a "demirrored" sample that inverts the mirrored signal histories, leaving us with 48 distinct histories.

BASELINE. Intuitively, a larger sample size K implies a smaller belief variance for a given signal history because the mean belief is the average of the sampled narratives, which becomes more concentrated around the sampling distribution mean as K increases. The belief variance in addition also identifies the idiosyncratic noise term σ_ε.[31] Appendix C includes a detailed descriptions of all moments and the estimation procedure.

Allowing for misperceptions of r. So far, we described a procedure to derive a minimal set of parameters to quantify the model. Below, we also present estimation results from a variant relaxing the assumption that participant beliefs accurately reflect the true redraw probability $r = 0.2$. This relaxation is mainly motivated by the findings in Section 4.5.3, which suggest that BASELINE beliefs overweight signals more distant in the past, possibly reflecting an underestimation of the true redraw probability. To allow for misperceptions of r, we assume that agents draw from $f(\mathbf{n}_\tau) \propto \pi(\mathbf{n}_\tau|\boldsymbol{\sigma}_\tau, \hat{r})^\beta$, where $\hat{r}$ is an estimation parameter denoting the perceived redraw probability. Under this specification, agents sample from the Bayesian posterior that emerges when the redraw probability is $\hat{r}$, which is then further distorted by the β parameter. In the estimation, we assume that BASELINE and RECEIVER share a common $\hat{r}$ but allow for a different $\hat{r}$ in DEVELOPER.[32] By relaxing the structural model in this way, we can assess whether changes in the perceived redraw probability drive the treatment differences, rather than persuasion by narratives.

4.6.2 Estimation Results

Table 7: Structural parameter estimates

	Spec. (i): $r = 0.2$		Spec. (ii): Subjective $\hat{r}$	
Parameter	Estimate	(SE)	Estimate	(SE)
β (sampling distortion)	0.761	(0.013)	0.878	(0.018)
K (narratives sampled)	8	—	6	—
σ_ε^2 (reporting noise)	0.014	(0.001)	0.016	(0.001)
α_D (developer persuasion)	0.771	(0.034)	0.575	(0.064)
α_R (receiver persuasion)	0.747	(0.029)	0.741	(0.031)
$\hat{r}_{BL}$ (baseline perceived r)	0.200	—	0.021	(0.002)
$\hat{r}_D$ (developer perceived r)	0.200	—	0.190	(0.003)

Notes: Two-step efficient SMM estimates. Standard errors from the sandwich formula, conditional on $\hat{K}$. In Specification (i), r is fixed at the true value 0.2 for all arms. In Specification (ii), $\hat{r}_{BL}$ and $\hat{r}_D$ are estimated, assuming that$\hat{r}_R = \hat{r}_{BL}$.

[31] The framework predicts $\text{Var}_{f_{\boldsymbol{\sigma}_\tau}(\mathbf{n}_\tau)}(\theta^*_\tau) = \sigma^2_\varepsilon + \frac{\widehat{\text{Var}(\theta_\tau|\boldsymbol{\sigma}_\tau)}}{K}$, where $\widehat{\text{Var}(\theta_\tau|\boldsymbol{\sigma}_\tau)}$ is the empirical belief variance. The idiosyncratic noise term adds a level shift to any signal history variance while K scales the variance in proportion to the sampling variance. Hence, σ^2_ε and K are identified separately.

[32] We impose $r_R = r_{BL}$ for two reasons. First, RECEIVER participants, who do not construct narratives themselves, have no reason to perceive the data-generating environment differently from BASELINE participants. Second, r_R is not separately identified from α_R in RECEIVER mean beliefs, making it difficult to disentangle the two channels from mean moments alone.

Table 7 reports the parameter estimates. Several findings emerge. First, we observe that $\hat{\beta} < 1$, suggesting that belief formation is best described by sampling from a flatter-than-Bayesian prior. This is unsurprising, as we observe underinference in Baseline, which directly depends on this sampling distortion. Second, we observe positive but finite estimates of K, explaining our previous result that, while developers select narratives based on fit, they often fail to identify the true best-fit narrative. Third, we estimate substantial narrative persuasion—in Specification (i), shared narratives enter the posterior beliefs with a weight between 75-77%. While these estimates attenuate somewhat when allowing for misperceived r in Specification (ii), they remain statistically different from zero and economically significant.

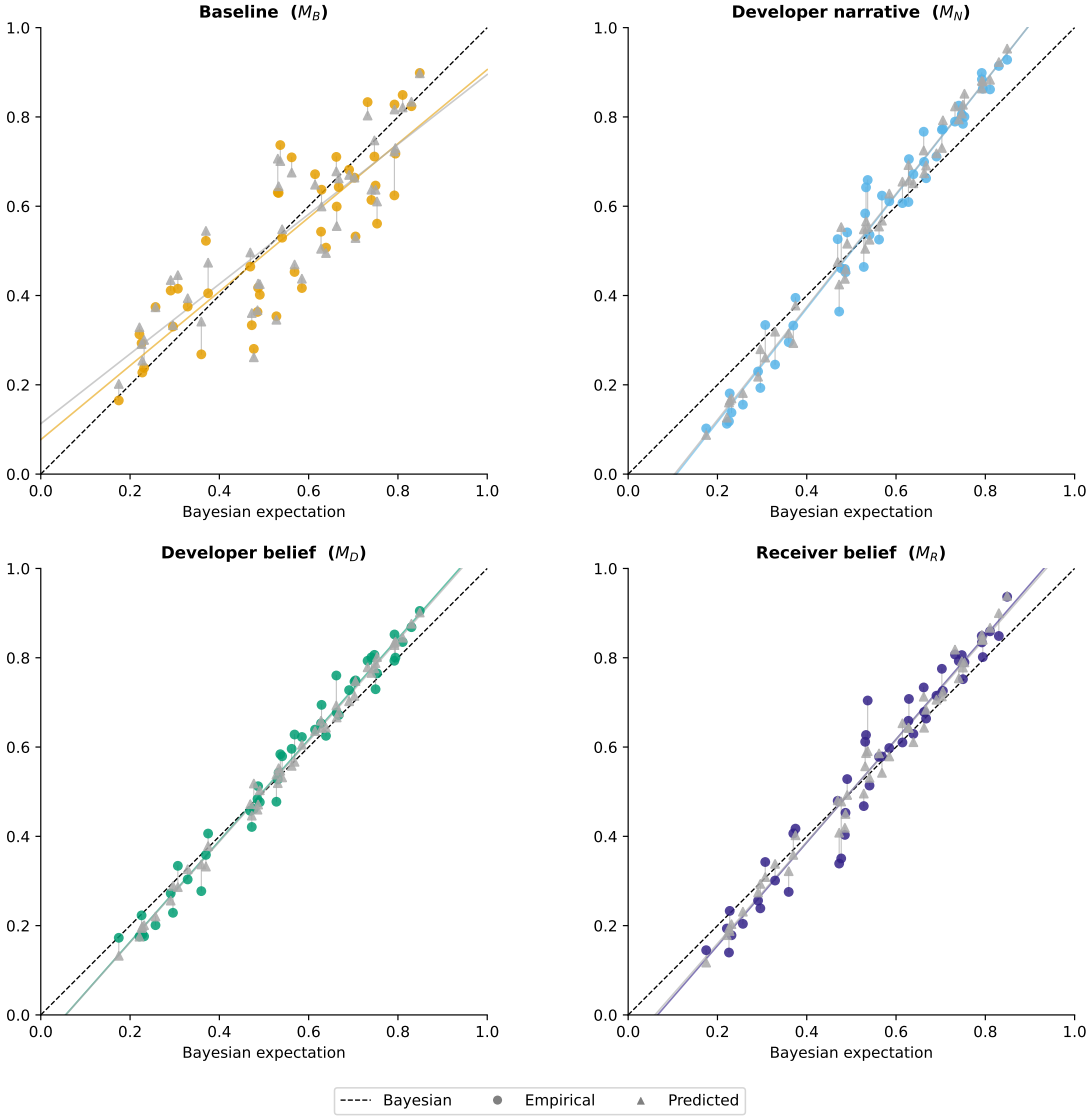
Interestingly, Specification (ii) reveals that Baseline participants perceive the environment as nearly stationary ($\hat{r}_{BL} \approx 0.02$), substantially underestimating the true redraw probability of $r = 0.2$. Narrative construction appears to attenuate this misperception: Developer participants perceive a redraw probability close to the true rate ($\hat{r}_D \approx 0.19$). The gap between both estimates is significant ($p < 0.001$), which provides a structural explanation of why Baseline participants are overly sensitive to signals more distant in the past (Section 4.5.3). While this constitutes a potential confound of our treatment—narrative construction may inadvertently correct a baseline misperception of the environment—we conduct a counterfactual analysis below showing that this channel does not account for substantial parts of the treatment effects.

Assessing the structural model fit. Figure 6 contrasts model predictions of Specification (ii) (Subjective $\hat{r}$) with the empirical results. The panels plots the average mean belief or narrative-implied estimate for every signal history against the history's Bayesian expectation. The individual signal-moments are generally predicted accurately and the model tracks the observed treatment differences well.

In addition, the figure illustrates that beliefs fluctuate much more around the Bayesian expectation in Baseline than in the other treatments. The structural model can replicate these patterns because the subjective posterior accounts for conservative updating ($\beta < 1$) and for misperceptions ($\hat{r}_{BL} < r$). These distortions imply that the interim beliefs for two histories with the same Bayesian expectation may differ drastically—for example, after a reduction in r, the temporal sequence of signals is less important when forming an expectation, which can drive a substantial wedge between two signal histories with the same Bayesian expectations. Through this mechanism, the structural model can track the mean beliefs in Baseline. In comparison, moments in the narrative treatments are less influenced by the distortions to the sampling function, for two reasons: (i) we estimate that $\hat{r}_D \approx r$, implying that the misperception of r has a limited impact on narrative sampling in *Developer*; (ii) the sampling distortion β does not vary the ordinal ranking of narratives in terms of fit for a given history, which is the primary driver of narrative selection. This explains why moments in the narrative treatments fluctuate less around the Bayesian expectation.

In Appendix C, we assess model fit more extensively. We show that the predicted moments highly correlate with the targeted moments and present evidence on out-of-sample performance, where the model manages to match additional, untargeted moments and the observed treatment effect sizes. We also provide additional evidence showing that the Spec-

Figure 6: Structural Model Predictions and Empirical Moments



Notes: *Each panel plots a predicted moment (either a mean belief or the narrative-implied estimate) and its empirical analog, for each of the 48 signal histories observed by participants in the experiment. Predictions and empirical analogs belonging to the same signal history are connected by vertical lines. The solid slope lines are best fit estimates using the model predictions (gray lines) or the empirical analogs (colored lines). These lines often coincide, suggesting that the structural model can account for the extent of over- and underreaction present in the data.*

ification (ii) (subjective $\hat{r}$) tracks the signal sensitivity patterns of Figure 5 closely. Finally, we illustrate that the subjective best-fit narrative is relatively unaffected by the estimated distortions to the sampling distribution.

4.6.3 Counterfactual Exercise: Ruling Out Misperception of Redraw Probability as the Primary Driver of Treatment Effects

The finding that $\hat{r}_D \gg \hat{r}_{BL}$ raises an important concern: could the treatment effects documented in Sections 4.2–4.3 be driven primarily by narrative construction inadvertently correcting participants' misperception of the redraw probability, rather than by narrative persuasion per se? To assess this, we compute counterfactual predictions in which all subjects rely solely on their interim beliefs ($\alpha_i = 0$)—shutting down narrative persuasion entirely—and vary only the perceived redraw probability from $\hat{r}_{BL}$ to $\hat{r}_D$.

Table 8: Counterfactual decomposition: r-shift vs. narrative persuasion

	Extremeness	SD(p)	$\Delta(p_T)$
Data levels			
Baseline	16.35	19.09	22.84
Developer	19.96	23.70	31.66
Model at $\alpha = 0$ (r-shift only)			
At $\hat{r}_{BL} \approx 0.02$	15.42	17.95	21.97
At $\hat{r}_D \approx 0.19$	16.14	19.48	26.93
Treatment effects			
r-shift only ($\alpha = 0$)	+0.72	+1.53	+4.95
Data (Developer − Baseline)	+3.61	+4.60	+8.82
r-shift as % of data	19.9%	33.2%	56.2%

Notes: Based on Specification (ii) parameters ($\hat{\beta} = 0.878$, $K = 6$). The counterfactual sets $\alpha_D = \alpha_R = 0$, so posterior beliefs equal the interim belief plus noise. In this table, beliefs are scaled from 0 to 100 to improve readability. Extremeness $\equiv |\text{belief} - 50|$. SD($p$) is the within-subject standard deviation of beliefs across five evaluation periods, averaged over subjects. $\Delta(p_T) \equiv |p_t - p_{t-5}|$ is the absolute period-to-period belief change, averaged over subjects and periods. Volatility simulated using actual 30-signal experimental histories; 5,000 subjects per base history.

Table 8 presents the results. Shifting the perceived redraw probability from $\hat{r}_{BL}$ to $\hat{r}_D$—while shutting down narrative persuasion entirely—generates only 20% of the observed treatment effect on extremeness. The misperception channel explains a larger share of the volatility effects (33% for SD(p) and 56% for $\Delta(p_T)$), because a higher perceived redraw produces more variable posteriors across evaluation periods, which mechanically amplifies cross-period belief fluctuations. Nevertheless, the misperception channel alone cannot account for the majority of the treatment effects on extremeness or for the changes belief volatility. Narrative persuasion ($\alpha_i > 0$)—the mechanism highlighted by our framework—remains a dominant driver. This provides reassurance that the treatment effects we document are attributable to the narrative channel.

We complement this analysis with additional reduced-form evidence in Appendix D.

There, we employ a k-means algorithm to identify a cluster of "heuristic" BASELINE participants, who essentially ignore the temporal sequence of signals when forming a belief. We then replicate our main results while controlling for cluster membership, showing that the treatment effects remain significant even when focusing on non-heuristic BASELINE participants as a comparison group for the narrative treatments.

5 Conclusion

In this paper, we show that the construction and diffusion of narratives systematically change the nature of the belief biases that agents exhibit—among both developers and receivers of narratives. In a hidden-Markov, regime-switching environment, participants *underinfer* about the current state from the available data when no narrative communication is allowed. However, when participants are prompted to construct singular narratives to make sense of the data—and when these narratives are shared with others—the dominant bias shifts from underreaction to *over*reaction. Moreover, narrative sharing induces excess volatility in participants' beliefs over time as new data arrive and narratives are adapted accordingly. These findings align with the predictions of our framework, which captures the key intuition that narratives that closely fit the data tend to generate beliefs that are both more extreme and more volatile.

Our results have broader implications for both theory and practice. On the theory side, we propose a novel narrative-based mechanism for overreaction in macroeconomic expectations (Bordalo et al., 2020, 2021) and for the excess-volatility puzzle (Shiller, 1981), and we provide experimental evidence in support of this mechanism. Turning to financial markets, our findings imply that popular narratives constructed in response to stock market rallies and crashes—such as the "this time is different" narratives during the dot-com bubble—can causally amplify fluctuations in beliefs and, in turn, market prices. For policymakers, these results suggest that exposing market participants to a plurality of narratives, rather than singular, well-fitting ones, may help temper market frenzies and promote more stable expectations.

Our paper also opens several avenues for future research. First, as discussed above, it is important to examine how agents behave when prompted to consider multiple potential narratives of the same data. Does exposure to a plurality of narratives temper the biases induced by singular, well-fitting narratives, or do agents instead selectively adopt only one narrative? Second, how might agents who face strategic motives—for example, those arising from media competition or reputation building—construct narratives differently from the developers in our experiment? Finally, it is valuable to study, both theoretically and empirically, how narrative-driven belief biases translate into economic actions such as trading and investment, and how such mechanisms may ultimately contribute to broader market phenomena such as asset-price volatility.

References

Afrouzi, H., S. Y. Kwon, A. Landier, Y. Ma, and D. Thesmar (2023): "Overreaction in Expectations: Evidence and Theory," *The Quarterly Journal of Economics*, 138, 1713–1764.

Aina, C. (2025): "Tailored Stories," *Mimeo*.

Aina, C. and F. H. Schneider (2025): "Weighting Competing Models," *Mimeo*.

Ambuehl, S. and H. C. Thysen (2024): "Competing Causal Interpretations: An Experimental Study," *Mimeo*.

Andre, P., I. Haaland, C. Roth, M. Wiederholt, and J. Wohlfart (2025): "Narratives about the Macroeconomy," .

Augenblick, N., M. Backus, D. A. Moore, and A. T. Little (2025a): "Assumptions, Disagreement, and Overprecision: Theory and Evidence," *Mimeo*.

Augenblick, N., E. Lazarus, and M. Thaler (2025b): "Overinference from Weak Signals and Underinference from Strong Signals," *The Quarterly Journal of Economics*, 140, 335–401.

Ba, C., J. A. Bohren, and A. Imas (2025): "Over-and Underreaction to Information: Belief Updating with Cognitive Constraints," .

Bailey, M., R. Cao, T. Kuchler, and J. Stroebel (2018): "The Economic Effects of Social Networks: Evidence from the Housing Market," *Journal of Political Economy*, 126, 2224–2276.

Barron, K. and T. Fries (2025): "Narrative Persuasion," *Mimeo*.

Barron, K., S. Huck, and P. Jehiel (2024): "Everyday Econometricians: Selection Neglect and Overoptimism When Learning from Others," *American Economic Journal: Microeconomics*, 16, 162–198.

Benjamin, D. J. (2019): "Errors in probabilistic reasoning and judgment biases," *Handbook of Behavioral Economics: Applications and Foundations 1*, 2, 69–186.

Blanco, M., D. Engelmann, A. K. Koch, and H.-T. Normann (2010): "Belief Elicitation in Experiments: Is There a Hedging Problem?" *Experimental Economics*, 13, 412–438.

Bordalo, P., J. J. Conlon, N. Gennaioli, S. Y. Kwon, and A. Shleifer (2022a): "Memory and Probability," *The Quarterly Journal of Economics*, 138, 265–311.

Bordalo, P., N. Gennaioli, Y. Ma, and A. Shleifer (2020): "Overreaction in macroeconomic expectations," *American Economic Review*, 110, 2748–2782.

Bordalo, P., N. Gennaioli, and A. Shleifer (2018): "Diagnostic Expectations and Credit Cycles," *The Journal of Finance*, 73, 199–227.

——— (2022b): "Salience," *Annual Review of Economics*, 14, 521–544.

Bordalo, P., N. Gennaioli, A. Shleifer, and S. Terry (2021): "Real Credit Cycles," *NBER WORKING PAPER SERIES*.

Bruner, J. (1986): "Two Modes of Thought," in *Actual Minds, Possible Worlds*, Cambridge, Massachusetts: Harvard University Press.

Burnside, C., M. Eichenbaum, and S. Rebelo (2016): "Understanding Booms and Busts in Housing Markets," *Journal of Political Economy*, 124, 1088–1147.

Charles, C. and C. Kendall (2025): "Causal Narratives," *Mimeo*.

Chen, D. L., M. Schonger, and C. Wickens (2016): "oTree—An Open-Source Platform for Laboratory, Online, and Field Experiments," *Journal of Behavioral and Experimental Finance*, 9, 88–97.

Conlon, J. J. and S. Y. Kwon (2025): "Beliefs from Cues," *Mimeo*.

Cutler, D., J. Poterba, and L. Summers (1988): "What Moves Stock Prices?" Tech. Rep. w2538, National Bureau of Economic Research, Cambridge, MA.

Dasgupta, I., E. Schulz, and S. J. Gershman (2017): "Where Do Hypotheses Come From?" *Cognitive Psychology*, 96, 1–25.

Eliaz, K. and R. Spiegler (2020): "A Model of Competing Narratives," *American Economic Review*, 110, 3786–3816.

Enke, B. (2020): "What You See Is All There Is," *The Quarterly Journal of Economics*, 135, 1363–1398.

Enke, B., T. Graeber, R. Oprea, and J. Yang (2024a): "Behavioral Attenuation," *Mimeo*.

Enke, B., F. Schwerter, and F. Zimmermann (2024b): "Associative Memory, Beliefs and Market Interactions," *Journal of Financial Economics*, 157, 103853.

Falk, A. and F. Zimmermann (2018): "Information processing and commitment," *The Economic Journal*, 128, 1983–2002.

Fan, T. Q. (2024): "Choice-induced Misspecified Mental Models," .

Fan, T. Q., Y. Liang, and C. Peng (2025): "The Inference-Forecast Gap in Belief Updating," *Mimeo*.

Fréchette, G. R., E. Vespa, and S. Yuksel (2025): "Extracting Models From Data Sets: An Experiment," *SSRN Electronic Journal*.

Gallistel, C. R., M. Krishan, Y. Liu, R. Miller, and P. E. Latham (2014): "The perception of probability." *Psychological Review*, 121, 96.

Gennaioli, N. and A. Shleifer (2010): "What Comes to Mind," *The Quarterly Journal of Economics*, 125, 1399–1433.

Graeber, T., S. Noy, and C. Roth (2025): "The Transmission of Reliable and Unreliable Information," *Mimeo*.

Hartzmark, S. M., S. D. Hirshman, and A. Imas (2021): "Ownership, Learning, and Beliefs," *The Quarterly Journal of Economics*, 136, 1665–1717.

Healy, P. J. and G. Leo (2024): "Belief elicitation: a user's guide," *Handbook of Experimental Methodology, ed. by L. Yariv and E. Snowberg, Elsevier*, 1.

Hirshleifer, D. (2015): "Behavioral Finance," *Annual Review of Financial Economics*, 7, 133–159.

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

Hossain, T. and R. Okui (2013): "The Binarized Scoring Rule," *The Review of Economic Studies*, 80, 984–1001.

Kahneman, D. and A. Tversky (1982): "The Simulation Heuristic," in *Judgement under Uncertainty. Heuristics and Biases*, 201–208.

Khaw, M. W., L. Stevens, and M. Woodford (2017): "Discrete adjustment to a changing environment: Experimental evidence," *Journal of Monetary Economics*, 91, 88–103.

Kuchler, T. and J. Stroebel (2021): "Social Finance," *Annual Review of Financial Economics*, 13, 37–55.

Liang, Y. (2025): "Learning from Unknown Information Sources," *Management Science*, 71, 3873–3890.

MacKay, D. J. C. (2003): *Information Theory, Inference, and Learning Algorithms*, Cambridge University Press.

Maturana, G. and J. Nickerson (2019): "Teachers Teaching Teachers: The Role of Workplace Peer Effects in Financial Decisions," *The Review of Financial Studies*, 32, 3920–3957.

Milgrom, P. R. (1981): "Good news and bad news: Representation theorems and applications," *The Bell Journal of Economics*, 380–391.

Musolff, R. and F. Zimmermann (2025): "Model Uncertainty," *Mimeo*.

Phillips, L. D. and W. Edwards (1966): "Conservatism in a Simple Probability Inference Task." *Journal of Experimental Psychology*, 72, 346–354.

Pindyck, R. S. and J. J. Rotemberg (1990): "The Excess Co-Movement of Commodity Prices," *The Economic Journal*, 100, 1173.

Sanborn, A. N. and N. Chater (2016): "Bayesian Brains without Probabilities," *Trends in Cognitive Sciences*, 20, 883–893.

Schwardmann, P., E. Tripodi, and J. J. van der Weele (2022): "Self-Persuasion: Evidence from Field Experiments at International Debating Competitions," *American Economic Review*, 112, 1118–1146.

Schwardmann, P. and J. Van Der Weele (2019): "Deception and Self-Deception," *Nature Human Behaviour*, 3, 1055–1061.

Schwartzstein, J. and A. Sunderam (2021): "Using models to persuade," *American Economic Review*, 111, 276–323.

——— (2024): *Sharing Models to Interpret Data*, Harvard Business School.

Shiller, R. J. (1981): "Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends?" *American Economic Review*, 71, 421–436.

——— (1984): "Stock Prices and Social Dynamics," *Brookings Papers on Economic Activity*, 1984, 457–510.

——— (1987): "The Volatility of Stock Market Prices," *Science*, 235, 33–37.

——— (2017): "Narrative Economics," *American Economic Review*, 107, 967–1004.

——— (2019): *Narrative Economics*, Princeton: Princeton University Press.

Skinner, Q. (1969): "Meaning and Understanding in the History of Ideas," *History and Theory*, 8, 3–53.

Soldà, A., C. Ke, L. Page, and W. Von Hippel (2020): "Strategically Delusional," *Experimental Economics*, 23, 604–631.

Vul, E., N. Goodman, T. L. Griffiths, and J. B. Tenenbaum (2014): "One and Done? Optimal Decisions From Very Few Samples," *Cognitive Science*, 38, 599–637.

White, H. V. (2000): *Metahistory: The Historical Imagination in Nineteenth-Century Europe*, A Johns Hopkins Paperback, Baltimore, Md.: Johns Hopkins Univ. Press, paperback ed., 11. [print.] ed.

A Additional Tables and Figures

A.1 Robustness of Main Results

Table A1: Narrative sharing induces more extreme beliefs.

	Alternative Extremeness Measure			
	Narratives vs. Baseline Beliefs		Beliefs across Treatments	
	(1)	(2)	(3)	(4)
Developer Narratives	0.068***	0.066***		
	(0.010)	(0.009)		
Developer Beliefs			0.047***	0.042***
			(0.012)	(0.011)
Receiver Beliefs			0.055***	0.049***
			(0.009)	(0.009)
Sample	Full Sample	Highly Consistent	Full Sample	Highly Consistent
Round FE	Yes	Yes	Yes	Yes
Signal History FE	Yes	Yes	Yes	Yes
Baseline Mean	0.151	0.162	0.151	0.162
Observations	4000	3575	6000	5425

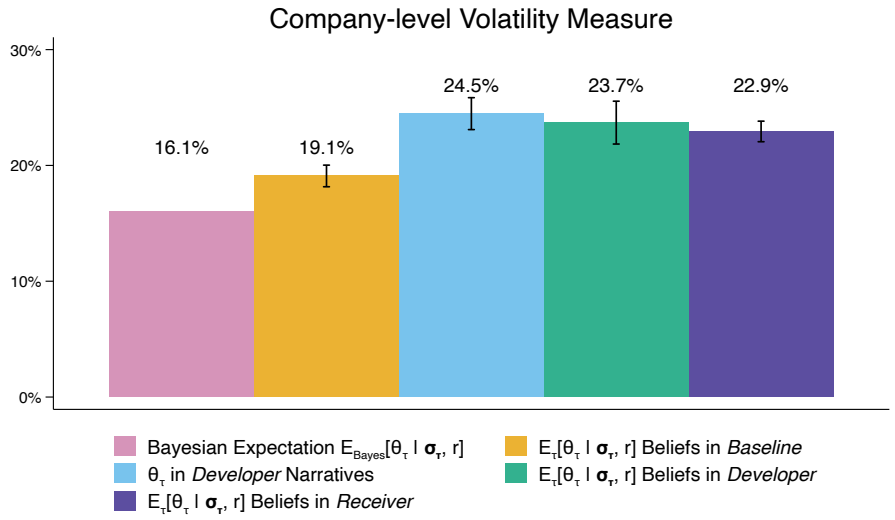
Notes: *Standard errors are clustered by participant. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. The dependent variable is the alternative extremeness measure of the elicited value—its deviation from the 0.5 (the prior mean of the state) in the direction of the Bayesian benchmark. Columns (1)–(2) compare current-state values in* Developer *narratives to mean beliefs in* Baseline*. Columns (3)–(4) compare mean beliefs across all treatments. The omitted category is always mean beliefs in* Baseline*. The "highly consistent" subsample includes participants who passed the preregistered consistency check.*

Table A2: Narrative sharing induces more volatile beliefs.

	Main Volatility Measure: $\lvert\theta_\tau - \theta_{\tau-5}\rvert$			
	Narratives vs. Baseline Beliefs		Beliefs across Treatments	
	(1)	(2)	(3)	(4)
Developer Narratives	0.099***	0.103***		
	(0.014)	(0.015)		
Developer Beliefs			0.088***	0.086***
			(0.017)	(0.018)
Receiver Beliefs			0.079***	0.078***
			(0.011)	(0.011)
Sample	Full Sample	Highly Consistent	Full Sample	Highly Consistent
Round FE	Yes	Yes	Yes	Yes
Signal History FE	Yes	Yes	Yes	Yes
Baseline Mean	0.228	0.231	0.228	0.231
Observations	3200	2860	4800	4340

Notes: *Standard errors are clustered by participant. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. The dependent variable is the absolute movement of the elicited value across two consecutive evaluation periods. Columns (1)–(2) compare current-state values in* Developer *narratives to mean beliefs in* Baseline*. Columns (3)–(4) compare mean beliefs across all treatments. The omitted category is always mean beliefs in* Baseline*. The "highly consistent" subsample includes participants who passed the preregistered consistency check.*

Figure A1: Narrative sharing induces more volatile beliefs.



Notes: *The figure plots the average of our alternative, company-level volatility measure—the standard deviation of the elicited values across all five evaluation periods of the same company—for mean beliefs across different treatments as well as current-state values in* Developer *narratives. Error bars are 95% confidence intervals derived from standard errors clustered by participant.*

Table A3: Narrative sharing induces more volatile beliefs.

	Alternative Company-level Volatility Measure			
	Narratives vs. Baseline Beliefs		Beliefs across Treatments	
	(1)	(2)	(3)	(4)
Developer Narratives	0.054***	0.055***		
	(0.008)	(0.009)		
Developer Beliefs			0.046***	0.045***
			(0.010)	(0.011)
Receiver Beliefs			0.038***	0.037***
			(0.007)	(0.007)
Sample	Full Sample	Highly Consistent	Full Sample	Highly Consistent
Company Order FE	Yes	Yes	Yes	Yes
Signal History FE	Yes	Yes	Yes	Yes
Baseline Mean	0.191	0.193	0.191	0.193
Observations	800	715	1200	1085

Notes: *Standard errors are clustered by participant. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. The dependent variable is the company-level standard deviation of the elicited values across all five evaluation periods. Each observation corresponds to a set of five evaluations of a participant for the same company. Columns (1)–(2) compare current-state values in* Developer *narratives to mean beliefs in* Baseline*. Columns (3)–(4) compare mean beliefs across all treatments. The omitted category is always mean beliefs in* Baseline*. The "highly consistent" subsample includes participants who passed the preregistered consistency check.*

Table A4: Results showing the stronger effects of narrative sharing for relatively informative signal histories, using the *highly consistent* subsample.

	Main Extremeness Measure: $\lvert\theta_\tau - 0.5\rvert$	
	(1)	(2)
Relatively Informative	0.069***	0.068***
	(0.009)	(0.009)
Developer Narratives	0.030***	
	(0.011)	
Developer Narratives × Relatively Informative	0.063***	
	(0.012)	
Developer Beliefs		0.032**
		(0.014)
Receiver Beliefs		0.010
		(0.008)
Developer Beliefs × Relatively Informative		0.034***
		(0.012)
Receiver Beliefs × Relatively Informative		0.061***
		(0.012)
Round FE	Yes	Yes
Mean of Dep. Var. of Omitted Category (Baseline Beliefs for Relatively Uninformative Signal Histories)	0.173	0.173
Observations	3575	5425

Notes: *Standard errors are clustered by participant. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. The results in this table use the "highly consistent" subsample of participants who passed the preregistered consistency check. The dependent variable is the absolute deviation of the elicited value from 0.5, the prior mean of the state. Columns (1) compares current-state values in* Developer *narratives to mean beliefs in* Baseline*. Columns (2) compares mean beliefs across all treatments. The omitted category is always mean beliefs in* Baseline *for the* relatively uninformative *signal histories.* Relatively informative *signal histories are defined as those with a Bayesian expectation that is more distant from 0.5 than the median in our sample.*

Table A5: Results showing the stronger effects of narrative sharing for relatively informative signal histories, using the alternative extremeness measure and the full sample.

	Alternative Extremeness Measure	
	(1)	(2)
Relatively Informative	0.080***	0.083***
	(0.011)	(0.011)
Developer Narratives	0.014	
	(0.009)	
Developer Narratives × Relatively Informative	0.108***	
	(0.014)	
Developer Beliefs		0.007
		(0.010)
Receiver Beliefs		0.019**
		(0.008)
Developer Beliefs × Relatively Informative		0.079***
		(0.015)
Receiver Beliefs × Relatively Informative		0.073***
		(0.015)
Round FE	Yes	Yes
Mean of Dep. Var. for Omitted Category (Baseline Beliefs for Relatively Uninformative Signal Histories)	0.151	0.151
Observations	4000	6000

Notes: *Standard errors are clustered by participant. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. The results in this table use the full sample of 240 participants (80 in each treatment). The dependent variable is the alternative extremeness measure of the elicited value—its deviation from the 0.5 (the prior mean of the state) in the direction of the Bayesian benchmark. Columns (1) compares current-state values in* Developer *narratives to mean beliefs in* Baseline*. Columns (2) compares mean beliefs across all treatments. The omitted category is always mean beliefs in* Baseline *for the* relatively uninformative *signal histories.* Relatively informative *signal histories are defined as those with a Bayesian expectation that is more distant from 0.5 than the median in our sample.*

Table A6: Results showing the stronger effects of narrative sharing for relatively volatile companies, using the *highly consistent* subsample.

	Main Volatility Measure: $\lvert\theta_\tau - \theta_{\tau-5}\rvert$	
	(1)	(2)
Relatively Volatile	0.076***	0.075***
	(0.009)	(0.009)
Developer Narratives	0.056***	
	(0.014)	
Developer Narratives × Relatively Volatile	0.093***	
	(0.017)	
Developer Beliefs		0.068***
		(0.019)
Receiver Beliefs		0.052***
		(0.011)
Developer Beliefs × Relatively Volatile		0.034*
		(0.018)
Receiver Beliefs × Relatively Volatile		0.050***
		(0.014)
Round FE	Yes	Yes
Mean of Dep. Var. for Omitted Category (Baseline Beliefs for Relatively Involatile Companies)	0.193	0.193
Observations	2860	4340

Notes: *Standard errors are clustered by participant. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. The results in this table use the "highly consistent" subsample of participants who passed the preregistered consistency check. The dependent variable is the absolute movement of the elicited value across two adjacent evaluation periods. Column (1) compares current-state values in* Developer *narratives to mean beliefs in* Baseline*. Column (2) compares mean beliefs across all treatments. The omitted category is always mean beliefs in* Baseline *for the* relatively involatile *companies.* Relatively volatile *companies are defined as those with a standard deviation of Bayesian expectations across evaluation periods that is higher than the median in our sample.*

Table A7: Results showing the stronger effects of narrative sharing for relatively volatile companies, using the alternative, company-level volatility measure and the full sample.

	Company-level Volatility Measure of θ_τ	
	(1)	(2)
Relatively Volatile	0.040***	0.040***
	(0.005)	(0.005)
Developer Narratives	0.036***	
	(0.009)	
Developer Narratives × Relatively Volatile	0.035***	
	(0.008)	
Developer Beliefs		0.040***
		(0.011)
Receiver Beliefs		0.031***
		(0.007)
Developer Beliefs × Relatively Volatile		0.012
		(0.009)
Receiver Beliefs × Relatively Volatile		0.015*
		(0.008)
Company Order FE	Yes	Yes
Mean of Dep. Var. for Omitted Category (Baseline Beliefs for Relatively Involatile Companies)	0.171	0.171
Observations	800	1200

Notes: *Standard errors are clustered by participant. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. The results in this table use the full sample of 240 participants (80 in each treatment). The dependent variable is the company-level standard deviation of the elicited values across all five evaluation periods. Each observation corresponds to a set of five evaluations of a participant for the same company. Column (1) compares current-state values in* Developer *narratives to mean beliefs in* Baseline*. Column (2) compares mean beliefs across all treatments. The omitted category is always mean beliefs in* Baseline *for the* relatively involatile *companies.* Relatively volatile *companies are defined as those with a standard deviation of Bayesian expectations across evaluation periods that is higher than the median in our sample.*

A.2 Supplemental Results

A.2.1 Formal tests of over- and underinference

Table A8: Mean beliefs against Bayesian expectations

	Elicited mean belief: $\hat{\mathbb{E}}_{tr}[\theta_\tau \mid \boldsymbol{\sigma}'_\tau]$	
	(1)	(2)
Bayesian exp.	0.823***	0.870***
	(0.0397)	(0.0383)
DEVELOPER × Bayesian expectation	0.315***	0.299***
	(0.0641)	(0.0641)
RECEIVER × Bayesian expectation	0.334***	0.302***
	(0.0514)	(0.0510)
DEVELOPER	-0.153***	-0.145***
	(0.0347)	(0.0357)
RECEIVER	-0.168***	-0.152***
	(0.0284)	(0.0287)
Sample	Full sample	Highly consistent
Round FE	Yes	Yes
$p(H_0 : \text{BASELINE slope} = 1)$	$< .001$	$< .001$
$p(H_0 : \text{DEVELOPER slope} = 1)$	.007	.001
$p(H_0 : \text{RECEIVER slope} = 1)$	$< .001$	$< .001$
Observations	6000	5425

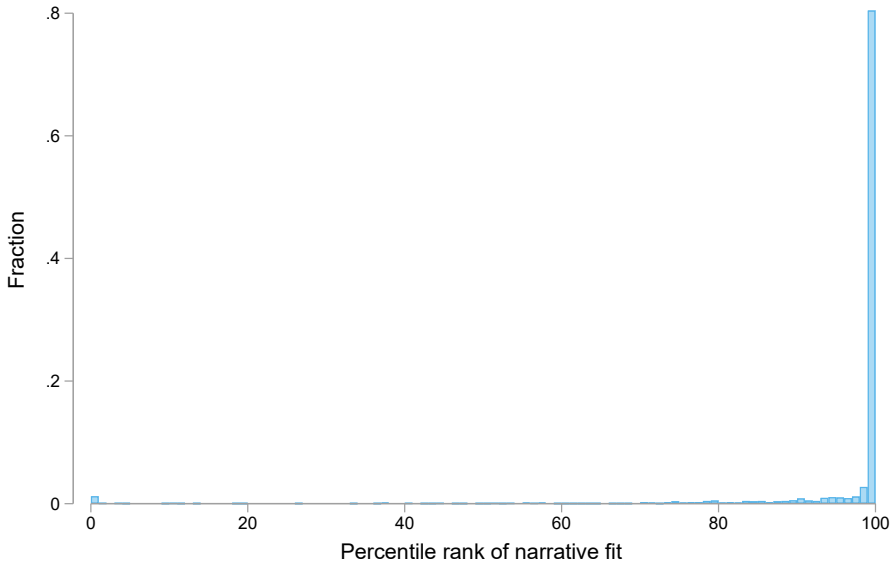
Notes: *Standard errors are clustered by participant. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. The dependent variable is the elicited mean belief.*

A.2.2 Empirical fit of Developer narratives

In this subsection, we provide descriptive statistics about narratives elicited in DEVELOPER, with a focus on fit. Narrative fit is measured by the posterior likelihood. This is useful to rank narratives by fit for a given signal history, but it does not allow us to compare fit across histories: Variation in posterior likelihood across narratives is highly influenced by variation in the underlying signal history. Therefore, we consider a relative fit measure.

Our relative fit measure is the percentile rank achieved by a DEVELOPER narrative's posterior likelihood, relative to the posterior likelihoods that could have been achieved by constructing any other narrative. Figure A2 presents the resulting distribution. This distribution indicates that the vast majority of DEVELOPER narratives (around 90%) rank among the top 10% of narratives that could have been constructed. A substantial majority (80%) rank among the top 1% in terms of fit.

Figure A2: Histogram of the fit achieved by DEVELOPER narratives

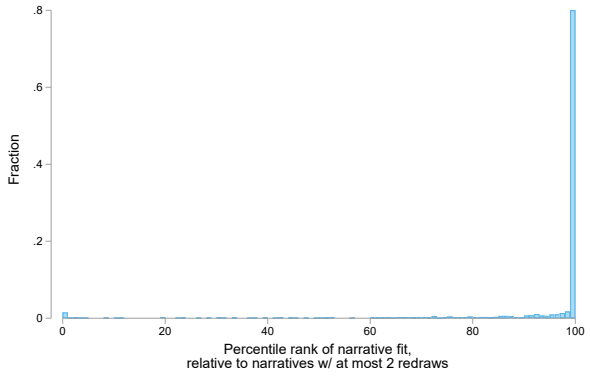


Notes: This figure plots the distribution of the fit of DEVELOPER narratives. The fit is measured in terms of the percentile rank of the narrative's Bayesian posterior likelihood within the Bayesian posterior distribution of the respective signal history the narrative was constructed for. This is achieved as follows: For each signal history encountered by participants in the experiment, we draw 1,000,000 narratives from the Bayesian posterior distribution. We then evaluate each DEVELOPER narrative by ranking it relative to the drawn narratives in terms of the Bayesian posterior likelihood.

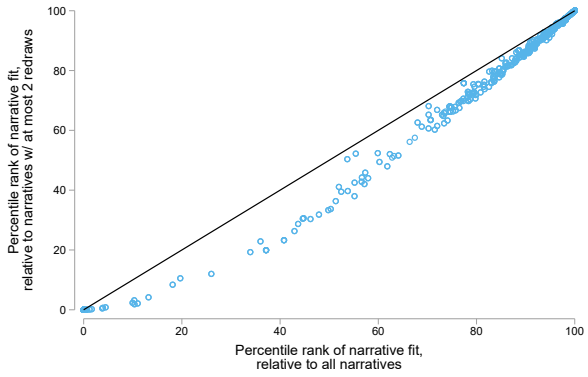
These results are robust to using alternative measures of relative fit. In particular, the result essentially remains the same when we compare DEVELOPER narratives only to narratives which invoke no more than two redraw events. This is important, as many narratives with more than two redraws are mechanically low-fit due to the number of redraws they invoke. In addition, it also compares alike-to-alike: Among DEVELOPER narratives, 92% feature no more than two redraws. Despite the fact that this restriction substantially reduces the comparison group, the resulting fit distribution remains essentially unchanged,

Figure A3: Percentile rank when restricting the comparison to narratives with at most two redraws.

(i) Fit achieved in DEVELOPER



(ii) Comparison of fit rank measures



Notes: The percentile rank in Panel (i) was constructed by ranking each DEVELOPER narrative relative to narratives that with no more than two redraw events. This was achieved by drawing 1,000,000 narratives for each signal history in the experiment and then keeping only those narratives with no more than two redraw events. DEVELOPER narratives were ranked relative to the resultant subset. Panel (ii) is a scatter plot comparing the two relative fit measures. In this plot, the black line is the 45 degree line.

as Panel (i) of Figure A3 shows. Panel (ii) provides an indication of why this adjustment does not change the percentile rank distribution by much: Narratives ranking among the top 1% when compared to all narratives rank among the top 1% when only compared to narratives with no more than two breaks. As a result, under both measures we find that 80% of narratives rank among the top 1% in terms of narrative fit.

A.2.3 Decomposition of Narrative Extremeness and Volatility

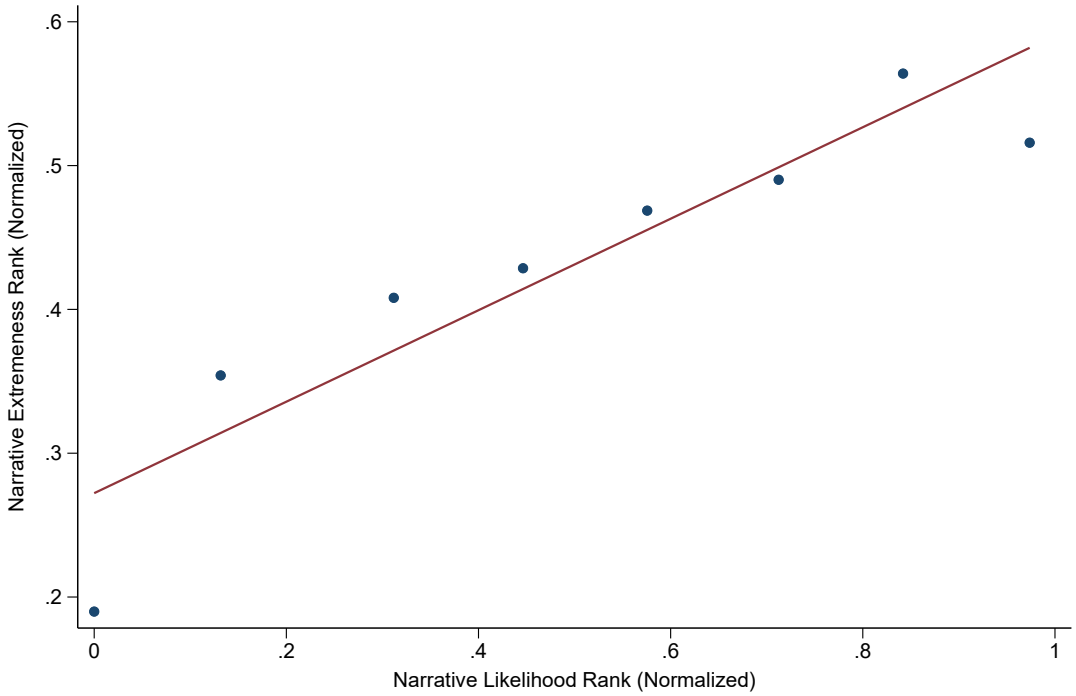
Table A9: Overinference, extremeness, and volatility of current state estimates under different benchmarks

	Overinference		Alternative Extremeness Measure		Company-level Volatility Measure of θ_τ	
	Coeff.	% of Total	Mean	% of Total Bias	Mean	% of Total Bias
A. Developer Narratives						
(1) Bayesian Benchmark	1.000	–	0.171	–	0.161	–
(2) Empirically Observed Narrative	1.272	100%	0.220	100%	0.245	100%
Decomposition of Total Overinference/Bias:						
– Due to state-transition commitment	0.063	23.2%	0.012	23.3%	0.029	33.9%
– Due to point estimation	0.209	76.8%	0.038	76.7%	0.056	66.1%
B. Conditional Best-Fit Narrative						
(3) Narrative w/ Best-Fit Estimate	1.538	100%	0.263	100%	0.290	100%
Decomposition of Total Overinference/Bias:						
– Due to state-transition commitment	0.063	11.7%	0.012	12.5%	0.029	22.0%
– Due to point estimation	0.475	88.3%	0.081	87.5%	0.101	78.0%
C. True Best-Fit Narratives						
(4) True Best-Fit Narrative	1.725	100%	0.273	100%	0.354	100%
Decomposition of Total Overinference/Bias:						
– Due to state-transition commitment	0.222	30.7%	0.029	28.0%	0.057	29.2%
– Due to point estimation	0.503	69.3%	0.074	72.0%	0.137	70.8%

Notes: *This table decomposes overinference, extremeness, and bias in narratives into two channels: (i) a* ***state-transition commitment*** *channel, arising from committing to a single state-transition path, and (ii) a* ***point estimation*** *channel, arising from choosing extreme state values within that path. We measure overinference by regressing the corresponding θ_τ-estimate on the Bayesian expectation of θ_τ. The table plots the coefficient estimate for the different comparison groups. Panel A shows this decomposition for the empirically observed narratives from the* Developer *treatment. Panel B shows the same decomposition for the conditional best-fit narratives, estimated using the subjects' reported probabilities. Panel C shows the decomposition for the theoretical true best-fit narratives. "Total Overinference/Bias" is the difference between the narrative value and the fully Bayesian benchmark. The "Model Selection" component is the difference between the "Conditional Bayesian" counterfactual (using the narrative's transition points) and the fully Bayesian benchmark. The "Point Estimation" component is the remaining difference.*

A.2.4 Narrative Fit and Its Impact on Beliefs

Figure A4: Positive correlation between narrative extremeness and narrative fit



Notes: *The figure displays binned scatter plots illustrating how narrative fit is correlated with narrative extremeness. The Narrative Likelihood Rank is ranks narratives from 0 (best fit) to 1 (worst fit), conditional on the signal history. The Narrative Extremeness Rank ranks narratives analogously from most extreme to least extreme.*

Table A10: Stronger effects of better-fit narratives on receiver (but not developer) beliefs, demonstrated using an alternative outcome measure.

	Dummy: Belief and Narrative Move in Opposite Directions	
	(1)	(2)
	DEVELOPER	RECEIVER
Narrative Likelihood Rank (Normalized)	0.008	0.050**
	(0.022)	(0.020)
Distance between Narrative and Bayesian	0.176**	0.379***
	(0.075)	(0.094)
Round FE	Yes	Yes
Signal History FE	Yes	Yes
Observations	2000	2000

Notes: *Standard errors are clustered by participant. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. Column (1) uses the posterior beliefs in* DEVELOPER *and Column (2) uses the posterior beliefs in* RECEIVER*. The dependent variable is a dummy variable for whether the receiver's posterior belief moves in the opposite direction from the narrative's current-state value (i.e., the two lie on opposite sides of the prior mean of 0.5). The variable "Narrative Likelihood Rank" is defined as the rank of the current narrative among all the narratives constructed for the current signal history in our sample, normalized to the interval* $[0, 1]$ *(with 0 indicating the best fit and 1 indicating the worst fit). The variable "Distance between Narrative and Bayesian" is defined as the distance between the current state embedded in the observed narrative and the Bayesian posterior mean of the current state.*

In Table A11, we present results using the distance between the narrative and the average mean belief in BASELINE (for the same signal history) as a control variable. The average mean belief in BASELINE should provide a more precise counterfactual of what participants would have chosen in the absence of the narrative. However, we estimate this average only with noise. Therefore, in the odd columns of Table A11, we present ORIV regressions that instrument the distance between narrative and BASELINE with the distance between the Bayesian expectation and BASELINE. The results are in line with the results displayed in the main text, with the effect of fit estimated to be stronger in RECEIVER.

Table A11: Stronger effects of better-fit narratives on receiver (but not developer) beliefs

	Distance between Belief and Narrative			
	(1)	(2)	(3)	(4)
Narrative Likelihood Rank (Normalized)	0.00609	0.0132	0.0382***	0.0465***
	(0.0160)	(0.0155)	(0.0119)	(0.0128)
Distance between Narrative and Baseline	0.109	0.239**	0.444***	0.595***
	(0.0764)	(0.0914)	(0.0472)	(0.0568)
Treatment	Developer	Developer	Receiver	Receiver
Estimator	OLS	IV	OLS	IV
Round FE	Yes	Yes	Yes	Yes
Signal History FE	Yes	Yes	Yes	Yes
Observations	2000	2000	2000	2000

Notes: *Standard errors are clustered by participant. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively The dependent variable is the distance between a participant's posterior mean belief and the current state embedded in the observed narrative. The variable "Narrative Likelihood Rank" is defined as the rank of the current narrative among all the narratives constructed for the current signal history in our sample, normalized to the interval* $[0, 1]$ *(with 0 indicating the best fit and 1 indicating the worst fit). The variable "Distance between Narrative and Baseline" is defined as the distance between the current state embedded in the observed narrative and the average mean estimate made for the same signal history in* Baseline. *The odd columns employ IV regressions, instrumenting for the Distance between Narrative and Baseline with the Distance between Narrative and the Bayesian expectation.*

B Mathematical Appendix

This appendix provides details and support to the mathematical statements made throughout Section 2.

B.1 Proof of Lemma 1

The expected posterior mean assessment implied by a shared narrative is equal to

$$m_K^{BF} = \sum_{\mathbf{n}_\tau \in \Omega^\tau} \Pr(\mathbf{n}_i^{BF} = \mathbf{n}_\tau)\theta_\tau(\mathbf{n}_\tau).$$

In this equation,

$$\Pr(\mathbf{n}_i^{BF} = \mathbf{n}_\tau) = \frac{f(\mathbf{n}_\tau)}{p(\mathbf{n}_\tau)}\left[F_1(\mathbf{n}_\tau)^K - (F_1(\mathbf{n}_\tau) - p(\mathbf{n}_\tau))^K\right],$$

where $F_1(\mathbf{n}_\tau) = \sum_{\mathbf{n}'_T \in \Omega^T : f(\mathbf{n}'_\tau) \leq f(\mathbf{n}_\tau)} f(\mathbf{n}'_\tau)$ and $p(\mathbf{n}_\tau) = \sum_{\mathbf{n}'_T \in \Omega^T : f(\mathbf{n}'_\tau) = f(\mathbf{n}_\tau)} f(\mathbf{n}'_\tau)$. Now define

$$F_{K-1}(\mathbf{n}_\tau) \equiv \frac{K}{p(\mathbf{n}_\tau)} \int_{F_1(\mathbf{n}_\tau) - p(\mathbf{n}_\tau)}^{F_1(\mathbf{n}_\tau)} x^{K-1} \mathrm{d}x$$

Solving the integral shows that, equivalently,

$$F_{K-1}(\mathbf{n}_\tau) = \frac{F_1(\mathbf{n}_\tau)^K - (F_1(\mathbf{n}_\tau) - p(\mathbf{n}_\tau))^K}{p(\mathbf{n}_\tau)}$$

Therefore,

$$\Pr(\mathbf{n}_\tau^{BF} = \mathbf{n}_\tau) = f(\mathbf{n}_\tau)F_{K-1}(\mathbf{n}_\tau).$$

Plugging this into the formula for the expected posterior mean assessment and simplifying:

$$\begin{aligned} m_K^{BF} &= \sum_{\mathbf{n}_\tau \in \Omega^\tau} f(\mathbf{n}_\tau)F_{K-1}(\mathbf{n}_\tau)\theta_\tau(\mathbf{n}_\tau) \\ &= \mathbb{E}_{\mathbf{n}_\tau \sim f}\left[F_{K-1}(\mathbf{n}_\tau)\theta_\tau(\mathbf{n}_\tau)\right]. \end{aligned}$$

The covariance between $\theta_\tau(\mathbf{n}_\tau)$ and $F_{K-1}(\mathbf{n}_\tau)$ is defined as:

$$\mathrm{Cov}_{\mathbf{n}_\tau \sim f}(\theta_\tau(\mathbf{n}_\tau), F_{K-1}(\mathbf{n}_\tau)) = \mathbb{E}_{\mathbf{n}_\tau \sim f}\left[F_{K-1}(\mathbf{n}_\tau)\theta_\tau(\mathbf{n}_\tau)\right] - \mathbb{E}_{\mathbf{n}_\tau \sim f}\left[\theta_\tau(\mathbf{n}_\tau)\right]\mathbb{E}_{\mathbf{n}_\tau \sim f}\left[F_{K-1}(\mathbf{n}_\tau)\right].$$

Note that $\mathbb{E}_{\mathbf{n}_\tau \sim f}\left[\theta_\tau(\mathbf{n}_\tau)\right] = m^I$ and that

$$\mathbb{E}_{\mathbf{n}_\tau \sim f}\left[F_{K-1}(\mathbf{n}_\tau)\right] = \sum_{\mathbf{n}_\tau \in \Omega^\tau} \frac{f(\mathbf{n}_\tau)}{p(\mathbf{n}_\tau)}\left[F_1(\mathbf{n}_\tau)^K - (F_1(\mathbf{n}_\tau) - p(\mathbf{n}_\tau))^K\right]$$

Index narratives by j, which is increasing in fit: $f(\mathbf{n}_{\tau,1}) < f(\mathbf{n}_{\tau,2}) < f(\mathbf{n}_{\tau,3}) < \dots f(\mathbf{n}_{\tau,J})$. Then, the sum term in the equation above is equal to

$$\sum_{j=1}^{J} \frac{f(\mathbf{n}_{\tau,j})}{p(\mathbf{n}_{\tau,j})} \left[F_1(\mathbf{n}_{\tau,j})^K - (F_1(\mathbf{n}_{\tau,j}) - p(\mathbf{n}_{\tau,j}))^K\right]$$
$$= \Pr(\text{draw narrative w/ fit } f(\mathbf{n}_{\tau,1}) \text{ or lower}) - \Pr(\text{draw narrative w/ strictly lower fit than } f(\mathbf{n}_{\tau,1})) + \dots$$
$$+ \Pr(\text{draw narrative w/ fit } f(\mathbf{n}_{\tau,J}) \text{ or lower}) - \Pr(\text{draw narrative w/ strictly lower fit than } f(\mathbf{n}_{\tau,J}))$$
$$= 1.$$

Therefore, $\mathbb{E}_{\mathbf{n}_\tau \sim f}\left[F_{K-1}(\mathbf{n}_\tau)\right] = 1$ and the covariance becomes

$$\begin{aligned} \text{Cov}_{\mathbf{n}_\tau \sim f}(\theta_\tau(\mathbf{n}_\tau), F_{K-1}(\mathbf{n}_\tau)) &= \mathbb{E}_{\mathbf{n}_\tau \sim f}\left[F_{K-1}(\mathbf{n}_\tau)\theta_\tau(\mathbf{n}_\tau)\right] - m^I \\ \Rightarrow \mathbb{E}_{\mathbf{n}_\tau \sim f}\left[F_{K-1}(\mathbf{n}_\tau)\theta_\tau(\mathbf{n}_\tau)\right] &= \text{Cov}_{\mathbf{n}_\tau \sim f}(\theta_\tau(\mathbf{n}_\tau), F_{K-1}(\mathbf{n}_\tau)) + m^I. \end{aligned}$$

Plugging this into for the expected posterior mean assessment and simplifying gives the following equivalence provided in the main text (Fit-state covariance):

$$m_K^{BF} - m^I = \psi_K \equiv \text{Cov}_{\mathbf{n}_\tau \sim f}(\theta_\tau(\mathbf{n}_\tau), F_{K-1}(\mathbf{n}_\tau)).$$

B.2 Proof of Proposition 1

Consider the signal history that generates the most pessimistic interim mean belief, $(0, \dots, 0)$. Clearly, if p^* exists, then $p^* > m^I(\boldsymbol{\sigma}_\tau = (0, \dots, 0), r)$. We will now demonstrate that for this particular history, $m_K^{BF} < m^I$ and $m_{i,K}^P < m^I$. Since $m^{BF} - m^I = \psi_K$ and $m_{i,K}^P - m^I = \alpha_{\chi_i}\psi_K$ it is sufficient for us to show that $\psi_K < 0$ for any $K > 1$.

Claim 1: F_{K-1} is increasing in $\pi(\mathbf{n}_\tau)$, with strict increase if $K > 1$. Take any two narratives $\mathbf{n}''_\tau, \mathbf{n}'_\tau \in \Omega^\tau$ with $f(\mathbf{n}''_\tau) > f(\mathbf{n}'_\tau)$ that are adjacent in fit. That is, there is no narrative with a fit in $(f(\mathbf{n}'_\tau), f(\mathbf{n}''_\tau))$. Then, we aim to show that

$$F_{K-1}(\mathbf{n}''_\tau) - F_{K-1}(\mathbf{n}'_\tau) \geq 0.$$

This is equivalent to

$$\frac{F_1(\mathbf{n}''_\tau)^K - (F_1(\mathbf{n}''_\tau) - p(\mathbf{n}''_\tau))^K}{p(\mathbf{n}''_\tau)} - \frac{F_1(\mathbf{n}'_\tau)^K - (F_1(\mathbf{n}'_\tau) - p(\mathbf{n}'_\tau))^K}{p(\mathbf{n}'_\tau)} \geq 0.$$

Note that, because of the adjacency, $F_1(\mathbf{n}''_\tau) - p(\mathbf{n}''_\tau) = F_1(\mathbf{n}'_\tau)$. Hence, the condition simplifies to

$$\frac{F_1(\mathbf{n}''_\tau)^K}{p(\mathbf{n}''_\tau)} + \frac{(F_1(\mathbf{n}'_\tau) - p(\mathbf{n}'_\tau))^K}{p(\mathbf{n}'_\tau)} \geq \frac{F_1(\mathbf{n}'_\tau)^K}{p(\mathbf{n}''_\tau)} + \frac{F_1(\mathbf{n}'_\tau)^K}{p(\mathbf{n}'_\tau)}.$$

Multiplying everything by $\frac{p(\mathbf{n}'_\tau)p(\mathbf{n}''_\tau)}{p(\mathbf{n}'_\tau)+p(\mathbf{n}''_\tau)}$ brings:

$$\frac{p(\mathbf{n}'_\tau)}{p(\mathbf{n}'_\tau) + p(\mathbf{n}''_\tau)} F_1(\mathbf{n}''_\tau)^K + \frac{p(\mathbf{n}''_\tau)}{p(\mathbf{n}'_\tau) + p(\mathbf{n}''_\tau)} (F_1(\mathbf{n}'_\tau) - p(\mathbf{n}'_\tau))^K \geq F_1(\mathbf{n}'_\tau)^K.$$

Note that we can replace $F_1(\mathbf{n}'_\tau)$ with $\frac{p(\mathbf{n}'_\tau)}{p(\mathbf{n}'_\tau)+p(\mathbf{n}''_\tau)}F_1(\mathbf{n}''_\tau)+\frac{p(\mathbf{n}''_\tau)}{p(\mathbf{n}'_\tau)+p(\mathbf{n}''_\tau)}(F_1(\mathbf{n}'_\tau)-p(\mathbf{n}'_\tau))$, as they are equivalent:

$$\begin{aligned}
F_1(\mathbf{n}'_\tau) &= \frac{p(\mathbf{n}'_\tau)}{p(\mathbf{n}'_\tau)+p(\mathbf{n}''_\tau)}F_1(\mathbf{n}''_\tau)+\frac{p(\mathbf{n}''_\tau)}{p(\mathbf{n}'_\tau)+p(\mathbf{n}''_\tau)}(F_1(\mathbf{n}'_\tau)-p(\mathbf{n}'_\tau)) \\
\Rightarrow (p(\mathbf{n}'_\tau)+p(\mathbf{n}''_\tau))F_1(\mathbf{n}'_\tau) &= p(\mathbf{n}'_\tau)F_1(\mathbf{n}''_\tau)+p(\mathbf{n}''_\tau)(F_1(\mathbf{n}'_\tau)-p(\mathbf{n}'_\tau)) \\
\Rightarrow p(\mathbf{n}'_\tau)\underbrace{F_1(\mathbf{n}'_\tau)}_{=F_1(\mathbf{n}''_\tau)-p(\mathbf{n}''_\tau)} &= p(\mathbf{n}'_\tau)F_1(\mathbf{n}''_\tau)-p(\mathbf{n}'_\tau)p(\mathbf{n}''_\tau) \\
&\Rightarrow 0=0.
\end{aligned}$$

As a consequence, the condition becomes

$$\beta\cdot g(F_1(\mathbf{n}''_\tau))+(1-\beta)\cdot g(F_1(\mathbf{n}'_\tau)-p(\mathbf{n}'_\tau))\geq g(\beta\cdot F_1(\mathbf{n}''_\tau)+(1-\beta)\cdot[F_1(\mathbf{n}'_\tau)-p(\mathbf{n}'_\tau)]),$$

where $g(x)=x^K$ and $\beta=\frac{p(\mathbf{n}'_\tau)}{p(\mathbf{n}'_\tau)+p(\mathbf{n}''_\tau)}$. This condition holds by convexity of $g(x)$. If $K>1$, $g(x)$ is strictly convex, which concludes the proof of this claim.

Claim 2: For $\boldsymbol{\sigma}_\tau=(0,\dots,0)$, π is strictly decreasing in $\theta_\tau(\mathbf{n}_\tau)$. The Bayesian posterior likelihood of a narrative $\mathbf{n}_\tau$ can be written as:

$$\pi(\mathbf{n}_\tau)=\sum_{l=1}^{\tau}\Pr(\text{last redraw before } l|\mathbf{n}_\tau)\Pr(\theta_\tau=\theta_\tau(\mathbf{n}_\tau)|\text{last redraw before } l).$$

Crucially, $\Pr(\theta_\tau=\theta_\tau(\mathbf{n}_\tau)|\text{last redraw before } l)$ only depends on $\theta_\tau(\mathbf{n}_\tau)$ but not on values of θ_t before the most recent redraw. For the signal history $(0,\dots,0)$, this probability strictly decreases as θ_τ increases. In addition, $\Pr(\text{last redraw before } l|\mathbf{n}_\tau)$ is independent of $\theta_\tau(\mathbf{n}_\tau)$. The claim directly follows.

Claim 3: For $\boldsymbol{\sigma}_\tau=(0,\dots,0)$, F_K is decreasing in $\theta_\tau(\mathbf{n}_\tau)$ and strictly decreasing if $K>1$. This directly follows from combining the two previous claims.

Claim 4: For $\boldsymbol{\sigma}_\tau=(0,\dots,0)$, $\psi_K<0$ if $K>1$. This is a direct consequence of Claim 3.

This establishes that, for signal history $(0,\dots,0)$, $m_K^{BF},m_{i,K}^P<m^I$. By symmetry, this also implies that, for signal history $(1,\dots,1)$, $m_K^{BF},m_{i,K}^P>m^I$. Since these are the most extreme signal histories, the statement in the proposition is true if we choose p^* to be equal to the expected mean estimate under the second-most pessimistic history (which is $(1,0,\dots,0)$). This concludes the proof.

B.3 Proof of Proposition 2

We show that $\mathrm{Var}_{\boldsymbol{\sigma}_T\sim\mathbb{P}_T}(m_{i,K}^P)\geq\mathrm{Var}_{\boldsymbol{\sigma}_T\sim\mathbb{P}_T}(m^I)$ under this condition. From the definition of the variance:

$$\begin{aligned}
\mathrm{Var}_{\boldsymbol{\sigma}_T\sim\mathbb{P}_T}(m_{i,K}^P) &= \mathrm{Var}_{\boldsymbol{\sigma}_T\sim\mathbb{P}_T}(m^I+\alpha_{\chi_i}\psi_K) \\
&= \mathrm{Var}_{\boldsymbol{\sigma}_T\sim\mathbb{P}_T}(m^I)+\mathrm{Var}_{\boldsymbol{\sigma}_T\sim\mathbb{P}_T}(\alpha_{\chi_i}\psi_K)+2\alpha_{\chi_i}\mathrm{Cov}_{\boldsymbol{\sigma}_T\sim\mathbb{P}_T}(\psi_K,m^I).
\end{aligned}$$

Hence,

$$\begin{aligned}
&\text{Var}_{\boldsymbol{\sigma}_T\sim\mathbb{P}_T}(m^P_{i,K}) \geq \text{Var}_{\boldsymbol{\sigma}_T\sim\mathbb{P}_T}(m^I)\\
\Rightarrow \text{Var}_{\boldsymbol{\sigma}_T\sim\mathbb{P}_T}(m^I) + \text{Var}_{\boldsymbol{\sigma}_T\sim\mathbb{P}_T}(\alpha_{\chi_i}\psi_K) + \alpha_{\chi_i}\text{Cov}_{\boldsymbol{\sigma}_T\sim\mathbb{P}_T}(\psi_K, m^I) &\geq \text{Var}_{\boldsymbol{\sigma}_T\sim\mathbb{P}_T}(m^I)\\
\Rightarrow \text{Var}_{\boldsymbol{\sigma}_T\sim\mathbb{P}_T}(\alpha_{\chi_i}\psi_K) + 2\alpha_{\chi_i}\text{Cov}_{\boldsymbol{\sigma}_T\sim\mathbb{P}_T}(\psi_K, m^I) &\geq 0.
\end{aligned}$$

By defintion, $\text{Var}_{\boldsymbol{\sigma}_T\sim\mathbb{P}_T}(\alpha_{\chi_i}\psi_K) \geq 0$. It directly follows that $\text{Cov}_{\boldsymbol{\sigma}_T\sim\mathbb{P}_T}(\psi_K, m^I) \geq 0$ is a sufficient condition for the variance of the posterior expectation to exceed the variance of the prior expectation.

B.4 Proof of Proposition 3

We defined:

$$\text{Vol}_{\boldsymbol{\sigma}_T\sim\mathbb{P}_T}(m_{\tau'}, m_{\tau''}) \equiv \mathbb{E}_{\boldsymbol{\sigma}_T\sim\mathbb{P}_T}\left[(m_{\tau''} - m_{\tau'})^2\right] = \text{Var}_{\boldsymbol{\sigma}_T\sim\mathbb{P}_T}(m_{\tau''} - m_{\tau'}).$$

Consider the volatility of the posterior mean:

$$\begin{aligned}
\text{Vol}_{\boldsymbol{\sigma}_T\sim\mathbb{P}_T}(m^P_{\tau'}, m^P_{\tau''}) &= \text{Var}_{\boldsymbol{\sigma}_T\sim\mathbb{P}_T}(m^I_{\tau''} + \alpha_{\chi_i}\psi_{K,\tau''} - m^I_{\tau'} - \alpha_{\chi_i}\psi_{K,\tau'})\\
&= \text{Var}_{\boldsymbol{\sigma}_T\sim\mathbb{P}_T}(\Delta m^I + \alpha_{\chi_i}\Delta\psi_K)\\
= \text{Var}_{\boldsymbol{\sigma}_T\sim\mathbb{P}_T}(\Delta m^I) + \text{Var}_{\boldsymbol{\sigma}_T\sim\mathbb{P}_T}(\alpha_{\chi_i}\Delta\psi_K) &+ 2\alpha_{\chi_i}\text{Cov}_{\boldsymbol{\sigma}_T\sim\mathbb{P}_T}(\Delta m^I, \Delta\psi_K).
\end{aligned}$$

Hence,

$$\begin{aligned}
&\text{Vol}_{\boldsymbol{\sigma}_T\sim\mathbb{P}_T}(m^P_{\tau'}, m^P_{\tau''}) > \text{Vol}_{\boldsymbol{\sigma}_T\sim\mathbb{P}_T}(m^I_{\tau'}, m^I_{\tau''})\\
\Rightarrow \text{Var}_{\boldsymbol{\sigma}_T\sim\mathbb{P}_T}(\alpha_{\chi_i}\Delta\psi_K) &+ 2\alpha_{\chi_i}\text{Cov}_{\boldsymbol{\sigma}_T\sim\mathbb{P}_T}(\Delta m^I, \Delta\psi_K) \geq 0.
\end{aligned}$$

By definition, $\text{Var}_{\boldsymbol{\sigma}_T\sim\mathbb{P}_T}(\alpha_{\chi_i}\Delta\psi_K) \geq 0$. It directly follows that $\text{Cov}_{\boldsymbol{\sigma}_T\sim\mathbb{P}_T}(\Delta m^I, \Delta\psi_K) \geq 0$ is a sufficient condition for the posterior volatility to exceed the interim volatility.

B.5 Additional simulation results

The tables below presents simulation results for different parameterizations of the information environment. In each of them, the sufficient conditions stated in Propositions 2 and 3 hold.

B.5.1 Extremeness

The results below quantify the variance (extremeness) increase between interim mean belief and subjective best-fit estimate and provide computational evidence that the fit-state covariance co-moves with the interim mean (i.e., that the condition of Proposition 2 holds).

Tables B1 and B2 display the ratio of the variance of the expected best fit estimate, m^{BF}, and the variance of the expected interim mean estimate, m^I, i.e., they display $Var(m^{BF})/Var(m^I)$ under the assumption that agents sample from the Bayesian posterior. A value > 1 suggests that the variance of the best fit is larger.

Table B1: Variance Ratio $Var(m^{BF})/Var(m^{I})$ for $\Omega = \{0.3, 0.7\}$

τ	$r = 0.2$			$r = 0.4$		
	$K = 2$	$K = 10$	$K = 100$	$K = 2$	$K = 10$	$K = 100$
2	2.122	3.373	3.387	2.039	2.754	2.764
5	1.786	3.355	3.389	1.915	4.364	4.659
10	1.554	2.806	3.018	1.648	3.269	4.535

Notes: *The table displays the ratio of the variances of the subjective best fit and the interim mean. These variances were derived in simulations that, for each possible signal realization of a given information environment, simulate the Bayesian posterior through a FFBS sampler. The FFBS sampler draws 2,000 possible data generating processes for each signal history, with the draw probability equal to the Bayesian posterior likelihood. We calculate both the expected subjective best-fit and the expected interim mean based on the resultant simulated Bayesian posterior.*

Table B2: Variance Ratio $Var(m^{BF})/Var(m^{I})$ for $\Omega = \{0, 0.01, \ldots, 1\}$

τ	$r = 0.2$			$r = 0.4$		
	$K = 2$	$K = 10$	$K = 100$	$K = 2$	$K = 10$	$K = 100$
2	1.922	3.696	4.238	2.208	5.119	6.913
5	1.455	2.458	3.462	1.721	3.375	5.165
10	1.319	1.926	2.894	1.505	2.577	3.801

Notes: *The table displays the ratio of the variances of the subjective best fit and the interim mean. These variances were derived in simulations that, for each possible signal realization of a given information environment, simulate the Bayesian posterior through a FFBS sampler. The FFBS sampler draws 2,000 possible data generating processes for each signal history, with the draw probability equal to the Bayesian posterior likelihood. We calculate both the expected subjective best-fit and the expected interim mean based on the resultant simulated Bayesian posterior.*

Tables B3 and B4 display the correlation $Corr(\psi_K, m^I)$ for different parameterizations of the information environment under the assumption that agents sample from the Bayesian posterior. These correlations are positive and relatively large. Since the correlation is a sign-preserving transformation of the covariance, a positive correlation is a sufficient condition for the posterior mean to be more extreme than the interim mean (Proposition 2).

Table B3: Correlation $Corr(\psi_K, m^I)$ for $\Omega = \{0.3, 0.7\}$

τ	$r = 0.2$			$r = 0.4$		
	$K = 2$	$K = 10$	$K = 100$	$K = 2$	$K = 10$	$K = 100$
2	0.688	0.449	0.447	0.792	0.772	0.742
5	0.665	0.477	0.473	0.774	0.505	0.486
10	0.700	0.430	0.411	0.879	0.610	0.485

Notes: *The table displays the correlation between the fit-state covariance and the interim mean, for different parameterizations. These correlations were derived in simulations that, for each possible signal realization of a given information environment, simulate the Bayesian posterior through a FFBS sampler. The FFBS sampler draws 2,000 possible data generating processes for each signal history, with the draw probability equal to the Bayesian posterior likelihood. We calculate both the expected subjective best-fit and the expected interim mean based on the resultant simulated Bayesian posterior.*

Table B4: Correlation $Corr(\psi_K, m^I)$ for $\Omega = \{0, 0.01, \ldots, 1\}$

τ	$r = 0.2$			$r = 0.4$		
	$K = 2$	$K = 10$	$K = 100$	$K = 2$	$K = 10$	$K = 100$
2	0.971	0.954	0.961	0.979	0.907	0.860
5	0.974	0.779	0.654	0.965	0.895	0.854
10	0.961	0.824	0.637	0.969	0.918	0.880

Notes: *The table displays the correlation between the fit-state covariance and the interim mean, for different parameterizations. These correlations were derived in simulations that, for each possible signal realization of a given information environment, simulate the Bayesian posterior through a FFBS sampler. The FFBS sampler draws 2,000 possible data generating processes for each signal history, with the draw probability equal to the Bayesian posterior likelihood. We calculate both the expected subjective best-fit and the expected interim mean based on the resultant simulated Bayesian posterior.*

B.5.2 Volatility

The results below quantify the volatility increase between interim mean and subjective best-fit and provide computational evidence that changes in the fit-state covariance co-move with changes in the interim mean (the sufficient condition in Proposition 3).

Tables B5 and B6 display the ratio of the volatility of the expected best fit estimate, m^{BF}, and the variance of the expected interim mean estimate, m^I, i.e., they display $Vol(m^{BF})/Vol(m^I)$. A value > 1 suggests that the volatility of the best fit is larger.

Table B5: Volatility Ratio $Vol(m^{BF})/Vol(m^I)$ for $\Omega = \{0.3, 0.7\}$

$\Delta\tau$ (from $\tau = 10$)	$r = 0.2$			$r = 0.4$		
	$K = 2$	$K = 10$	$K = 100$	$K = 2$	$K = 10$	$K = 100$
1	1.430	4.153	6.253	1.968	5.035	9.551
3	1.510	3.127	4.295	1.842	3.687	5.492
5	1.537	2.819	3.681	1.789	3.312	4.613

Notes: *The table displays the ratio of the volatility of the subjective best fit and the interim mean. These volatilities were derived in simulations that first randomly generate 10,000 signal histories for each information environment with $T = 15$. For each simulated signal history, we simulate the Bayesian posterior at points $\tau = 10, 11, 13, 15$ using an FFBS sampler. The FFBS sampler draws 2,000 possible data generating processes for each signal history, with the draw probability equal to the Bayesian posterior likelihood. We calculate the volatilities based on the resultant simulated Bayesian posterior in each evaluation period. The table displays volatility between periods $\tau = 10$ and $\tau = 11$ in Row 1, between periods $\tau = 10$ and $\tau = 13$ in Row 2, between periods $\tau = 10$ and $\tau = 15$ in Row 3.*

Table B6: Volatility Ratio $Vol(m^{BF})/Vol(m^I)$ for $\Omega = \{0, 0.01, \ldots, 1\}$

$\Delta\tau$ (from $\tau = 10$)	$r = 0.2$			$r = 0.4$		
	$K = 2$	$K = 10$	$K = 100$	$K = 2$	$K = 10$	$K = 100$
1	1.430	4.153	6.253	1.968	5.035	9.551
3	1.510	3.127	4.295	1.842	3.687	5.492
5	1.537	2.819	3.681	1.789	3.312	4.613

Notes: *The table displays the ratio of the volatility of the subjective best fit and the interim mean. These volatilities were derived in simulations that first randomly generate 10,000 signal histories for each information environment with $T = 15$. For each simulated signal history, we simulate the Bayesian posterior at points $\tau = 10, 11, 13, 15$ using an FFBS sampler. The FFBS sampler draws 2,000 possible data generating processes for each signal history, with the draw probability equal to the Bayesian posterior likelihood. We calculate the volatilities based on the resultant simulated Bayesian posterior in each evaluation period. The table displays volatility ratios between periods $\tau = 10$ and $\tau = 11$ in Row 1, between periods $\tau = 10$ and $\tau = 13$ in Row 2, between periods $\tau = 10$ and $\tau = 15$ in Row 3.*

Tables B7 and B8 display the correlation $Corr(\Delta\psi_K, \Delta m^I)$ for different parameterizations of the information environment. These correlations are positive and relatively large. Since the correlation is a sign-preserving transformation of the covariance, a positive correlation is a sufficient condition for the posterior mean to be more volatile than the interim mean (Proposition 3).

Table B7: Correlation $Corr(\Delta\psi_K, \Delta m^I)$ for $\Omega = \{0.3, 0.7\}$

$\Delta\tau$ (from $\tau = 10$)	$r = 0.2$			$r = 0.4$		
	$K = 2$	$K = 10$	$K = 100$	$K = 2$	$K = 10$	$K = 100$
1	0.428	0.189	0.196	0.951	0.861	0.810
3	0.664	0.431	0.423	0.956	0.839	0.678
5	0.739	0.509	0.481	0.958	0.841	0.654

Notes: *The table displays the correlation between the change in the fit-state covariance and the change in the interim mean. These correlations were derived in simulations that first randomly generate 10,000 signal histories for each information environment with $T = 15$. For each simulated signal history, we simulate the Bayesian posterior at points $\tau = 10, 11, 13, 15$ using an FFBS sampler. The FFBS sampler draws 2,000 possible data generating processes for each signal history, with the draw probability equal to the Bayesian posterior likelihood. We calculate the correlation based on the resultant simulated Bayesian posteriors. The table displays correlation between periods $\tau = 10$ and $\tau = 11$ in Row 1, between periods $\tau = 10$ and $\tau = 13$ in Row 2, between periods $\tau = 10$ and $\tau = 15$ in Row 3.*

Table B8: Correlation $Corr(\Delta\psi_K, \Delta m^I)$ for $\Omega = \{0, 0.01, \ldots, 1\}$

$\Delta\tau$ (from $\tau = 10$)	$r = 0.2$			$r = 0.4$		
	$K = 2$	$K = 10$	$K = 100$	$K = 2$	$K = 10$	$K = 100$
1	0.890	0.749	0.695	0.976	0.952	0.929
3	0.924	0.803	0.703	0.964	0.922	0.880
5	0.933	0.813	0.675	0.961	0.916	0.874

Notes: *The table displays the correlation between the change in the fit-state covariance and the change in the interim mean. These correlations were derived in simulations that first randomly generate 10,000 signal histories for each information environment with $T = 15$. For each simulated signal history, we simulate the Bayesian posterior at points $\tau = 10, 11, 13, 15$ using an FFBS sampler. The FFBS sampler draws 2,000 possible data generating processes for each signal history, with the draw probability equal to the Bayesian posterior likelihood. We calculate the correlation based on the resultant simulated Bayesian posteriors. The table displays correlation between periods $\tau = 10$ and $\tau = 11$ in Row 1, between periods $\tau = 10$ and $\tau = 13$ in Row 2, between periods $\tau = 10$ and $\tau = 15$ in Row 3.*

B.6 Extension: Sharing Top-M Narratives

We extend the derivation to the case where the developer shares the set of their top-M narratives, $\mathcal{S}_M = \{\mathbf{n}_{(1)}, \ldots, \mathbf{n}_{(M)}\}$, where narratives are ordered by fit such that $f(\mathbf{n}_{(1)}) \geq \cdots \geq f(\mathbf{n}_{(K)})$. The developer's estimate is the average of these narratives:

$$m^{BF}_{K,M} = \frac{1}{M} \sum_{j=1}^{M} \theta_\tau(\mathbf{n}_{(j)}).$$

We show that

$$m^{BF}_{K,M} - m^I = \psi_{K,M} \equiv \mathrm{Cov}_{\mathbf{n}_\tau \sim f}(\theta_\tau(\mathbf{n}_\tau), F_{K,M}(\mathbf{n}_\tau)),$$

where $F_{K,M}(\mathbf{n}_\tau)$ is a weighting function that generalizes $F_{K-1}(\mathbf{n}_\tau)$.

The expected estimate is:

$$\mathbb{E}[m^{BF}_{K,M}] = \frac{1}{M}\sum_{j=1}^{M}\mathbb{E}[\theta_\tau(\mathbf{n}_{(j)})].$$

Let $G_{K,M}(x)$ be the probability that a narrative with fit quantile x ranks among the top M narratives in a sample of size K. This occurs if, among the other $K-1$ narratives, at most $M-1$ have a higher fit. The probability of a higher fit is $1-x$. Thus:

$$G_{K,M}(x) = \sum_{k=0}^{M-1}\binom{K-1}{k}(1-x)^k x^{K-1-k}.$$

Following the logic of Section B.1, the probability that a specific narrative $\mathbf{n}_\tau$ is selected into the top M set is:

$$\Pr(\mathbf{n}_\tau \in \mathcal{S}_M) = f(\mathbf{n}_\tau)\frac{K}{p(\mathbf{n}_\tau)}\int_{F_1(\mathbf{n}_\tau)-p(\mathbf{n}_\tau)}^{F_1(\mathbf{n}_\tau)} G_{K,M}(x)dx.$$

Let $F_{K,M}(\mathbf{n}_\tau) \equiv \frac{K}{Mp(\mathbf{n}_\tau)}\int_{F_1(\mathbf{n}_\tau)-p(\mathbf{n}_\tau)}^{F_1(\mathbf{n}_\tau)} G_{K,M}(x)dx$. Then the expected average estimate is:

$$\begin{aligned} m^{BF}_{K,M} &= \sum_{\mathbf{n}_\tau\in\Omega^\tau}\theta_\tau(\mathbf{n}_\tau)\frac{1}{M}\Pr(\mathbf{n}_\tau\in\mathcal{S}_M) \\ &= \sum_{\mathbf{n}_\tau\in\Omega^\tau}\theta_\tau(\mathbf{n}_\tau)f(\mathbf{n}_\tau)F_{K,M}(\mathbf{n}_\tau) \\ &= \mathbb{E}_{\mathbf{n}_\tau\sim f}[\theta_\tau(\mathbf{n}_\tau)F_{K,M}(\mathbf{n}_\tau)]. \end{aligned}$$

We verify that $\mathbb{E}_{\mathbf{n}_\tau\sim f}[F_{K,M}(\mathbf{n}_\tau)] = 1$.

$$\begin{aligned} \mathbb{E}_{\mathbf{n}_\tau\sim f}[F_{K,M}(\mathbf{n}_\tau)] &= \sum_{\mathbf{n}_\tau} f(\mathbf{n}_\tau)\frac{K}{Mp(\mathbf{n}_\tau)}\int_{F_1(\mathbf{n}_\tau)-p(\mathbf{n}_\tau)}^{F_1(\mathbf{n}_\tau)} G_{K,M}(x)dx \\ &= \frac{K}{M}\int_0^1 G_{K,M}(x)dx. \end{aligned}$$

The integral $\int_0^1 G_{K,M}(x)dx$ is the probability that a random draw from $U[0,1]$ is in the top M of K draws, which is M/K. Thus the expectation is 1. Using the definition of covariance:

$$\begin{aligned} \mathrm{Cov}_{\mathbf{n}_\tau\sim f}(\theta_\tau(\mathbf{n}_\tau), F_{K,M}(\mathbf{n}_\tau)) &= \mathbb{E}_{\mathbf{n}_\tau\sim f}[\theta_\tau(\mathbf{n}_\tau)F_{K,M}(\mathbf{n}_\tau)] - \mathbb{E}_{\mathbf{n}_\tau\sim f}[\theta_\tau(\mathbf{n}_\tau)]\mathbb{E}_{\mathbf{n}_\tau\sim f}[F_{K,M}(\mathbf{n}_\tau)] \\ &= m^{BF}_{K,M} - m^I. \end{aligned}$$

This confirms the result. Note that for $M=1$, $G_{K,1}(x) = x^{K-1}$, recovering the previous result. Conversely, $G_{K,K}(x) = 1$ and $m^{BF}_{K,K} = m^I$: If the developer sends every narrative, their estimate is unbiased. For $1 < M < K$, the function $G_{K,M}(x)$ is increasing in x, i.e, it puts higher weight on larger quantiles. This implies that $F_{K,M}(\mathbf{n}_\tau)$ assigns larger weight to high-fit narratives and $\mathrm{Cov}_{\mathbf{n}_\tau\sim f}(\theta_\tau(\mathbf{n}_\tau), F_{K,M}(\mathbf{n}_\tau))$ remains a measure of the correlation between the estimate $\theta_\tau(\mathbf{n}_\tau)$ and the fit of the narrative: Whenever $M < K$, the estimate is biased towards high-fit.

B.7 Optimal Combination of Estimates and Selection Neglect

In this section, we formalize the boundedly rational benchmark for optimally combining estimates and define the specific error committed by agents in our framework.

Consider two agents, i (Receiver) and j (Developer), who both form noisy interim estimates of the mean state, m_i^I and m_j^I, based on independent sampling procedures. If they were to exchange these mean estimates directly, the optimal posterior belief for agent i would be a precision-weighted average:

$$m_i^P = \beta m_i^I + (1-\beta) m_j^I,$$

where $\beta \in (0,1)$ reflects the relative precision of the two estimates. For example, if both agents sample K narratives from the same distribution, their estimates have equal precision, implying $\beta = 1/2$.

In our framework, however, the developer does not share their mean estimate m_j^I. Instead, they share their subjective best-fit narrative, which implies an estimate $m_{j,K}^{BF}$. As shown in Equation (Fit-state covariance), this estimate is systematically biased relative to the mean:

$$m_{j,K}^{BF} = m_j^I + \psi_K.$$

A rational receiver aware of this selection process should first "de-bias" the shared narrative to recover the developer's mean estimate:

$$\hat{m}_j^I = m_{j,K}^{BF} - \psi_K.$$

The receiver should then combine this estimated mean with their own:

$$m_i^{P,\text{rational}} = \beta m_i^I + (1-\beta)(m_{j,K}^{BF} - \psi_K).$$

Comparing this to the behavioral updating rule used in our framework, $m_i^P = (1-\alpha)m_i^I + \alpha m_{j,K}^{BF}$, we see that the behavioral agent effectively sets $\beta = 1-\alpha$ but fails to subtract the bias term ψ_K. This constitutes *selection neglect*: the agent treats the selected best-fit narrative as if it were a representative mean estimate, thereby inheriting the bias of the selection process.

C Structural Estimation Details

This section provides the full specification, estimation details, and additional results for the structural estimation reported in Section 4.6.

C.1 Signal Histories and De-Mirroring

As described in Section 3, the experiment generates binary signal sequences $\boldsymbol{\sigma} = (s_1, \ldots, s_{10})$, where each $s_t \in \{0, 1\}$ is drawn with probability equal to the hidden state θ_τ. Beliefs are elicited at five evaluation periods ($\tau \in \{10, 15, 20, 25, 30\}$). At each evaluation period, subjects observe the 10 most recent binary signals. Because consecutive evaluation periods overlap by 5 signals, each evaluation period produces a distinct 10-signal history.

We exploit the symmetry of the signal structure to reduce the data. *De-mirroring* maps each history to its bitwise complement (replacing s_t with $1 - s_t$ and transforming beliefs via $p \to 100 - p$), and *folding* retains only the favorable member of each mirror pair (Bayesian posterior > 50). The 10 base signal sequences $\times$ 5 evaluation periods $\times$ 2 (original and mirror) yield 100 distinct 10-signal histories; de-mirroring reduces these to 50 distinct 10-signal histories, of which 48 are unique after accounting for coincidental overlaps. This yields 48 signal histories for estimation, indexed $h = 0, \ldots, 47$. Our estimation matches the cross-sectional mean and BASELINE variance of beliefs for each such history.

C.2 Perceived Redraw Probability

The hidden state evolves as a Markov chain: at each period, the state either remains unchanged with probability $1 - r$ or redraws uniformly from the grid $\{0, 0.01, \ldots, 1.00\}$ with probability r, where $r = 0.2$ is the true redraw probability. The perceived redraw probability r enters the model through the Bayesian posterior, computed via Forward-Filtering Backward-Sampling (FFBS).

We consider two specifications:

- **Specification (i): $r = 0.2$.** All treatment arms use the true redraw probability. This assumes participants correctly perceive the rate of regime change.
- **Specification (ii): treatment-specific r.** We allow participants to form beliefs based on a misperception of r. BASELINE and RECEIVER participants share a common perceived r_{BL}; DEVELOPER participants use a separate r_D. We impose $r_R = r_{BL}$ on the grounds that RECEIVER participants, who do not construct narratives, share the BASELINE perception. This specification addresses the concern that the narrative construction task may inadvertently change participants' perception of the data-generating process (see Section 4.6).

C.3 Narrative Sampling and Interim Beliefs

We model narrative sampling as draws from a distorted posterior. Let $\pi(\mathbf{n}_\tau \mid \boldsymbol{\sigma}, r)$ denote the Bayesian posterior probability of narrative $\mathbf{n}_\tau$ given signal history $\boldsymbol{\sigma}$ and perceived redraw probability r, computed via FFBS. Because r governs the regime-transition matrix

used in the FFBS recursion, it shapes the posterior distribution and therefore the sampling distribution. The subjective sampling distribution is:

$$f(\mathbf{n}_\tau \mid \boldsymbol{\sigma}, r) \propto \pi(\mathbf{n}_\tau \mid \boldsymbol{\sigma}, r)^\beta, \tag{1}$$

where $\beta > 0$ is a distortion parameter. When $\beta = 1$, sampling is Bayesian. When $\beta < 1$, the sampling distribution is flatter than the Bayesian posterior—agents are more likely to sample narratives that the Bayesian posterior considers unlikely, a form of underreaction.

Each subject samples K narratives independently from $f(\cdot \mid \boldsymbol{\sigma}, r)$ and forms an *interim belief* equal to the average current-period state θ_τ across the sampled narratives:

$$\bar{\theta}_{\tau,i} = \frac{1}{K} \sum_{k=1}^{K} \theta_\tau^{(k)}, \qquad \theta_\tau^{(k)} \sim f(\cdot \mid \boldsymbol{\sigma}, r). \tag{2}$$

Among the K sampled narratives, the subject identifies a *best-fit narrative*—the one with the highest posterior probability $f(\mathbf{n}_\tau \mid \boldsymbol{\sigma}, r)$—and its associated current-period state $\theta_\tau^{(\text{best})}$.

C.4 Posterior Beliefs by Treatment Arm

We model the elicited beliefs in each treatment arm as follows, suppressing the conditioning on signal history h for readability.

Baseline beliefs. In the BASELINE treatment, subjects report their interim belief plus idiosyncratic noise:

$$p_{BL,i} = \bar{\theta}_{\tau,i} + \varepsilon_{BL,i}, \qquad \varepsilon_{BL,i} \sim N(0, \sigma_\varepsilon^2). \tag{3}$$

The noise σ_ε^2 captures reporting noise, rounding, and other sources of idiosyncratic variation. The same noise variance is used for all treatment arms.

Developer beliefs. In the DEVELOPER treatment, subjects sample K narratives, identify a best-fit narrative, and report a belief. We model the reported belief as a weighted average of the subject's interim belief and the best-fit narrative's prediction, plus noise:

$$p_{D,i} = (1 - \alpha_D) \cdot \bar{\theta}_{\tau,i} + \alpha_D \cdot \theta_{\tau,i}^{(\text{best})} + \varepsilon_{D,i}, \qquad \varepsilon_{D,i} \sim N(0, \sigma_\varepsilon^2). \tag{4}$$

The parameter $\alpha_D \in [0, 1]$ measures developer *self-persuasion*: the degree to which constructing a best-fit narrative shifts the developer's belief away from the interim expectation. When $\alpha_D = 0$, the developer's belief equals the interim belief; when $\alpha_D = 1$, it equals the best-fit narrative's prediction. Crucially, the developer's interim belief $\bar{\theta}_{\tau,i}$ and best-fit narrative $\theta_{\tau,i}^{(\text{best})}$ are derived from the *same* set of K draws, creating a positive covariance between them.

Receiver beliefs. In the RECEIVER treatment, subjects observe a narrative constructed by a developer and report a belief:

$$p_{R,i} = (1 - \alpha_R) \cdot \bar{\theta}_{\tau,i}^{R} + \alpha_R \cdot \theta_{\tau,j}^{(\text{best},D)} + \varepsilon_{R,i}, \qquad \varepsilon_{R,i} \sim N(0, \sigma_\varepsilon^2), \tag{5}$$

where $\bar{\theta}^R_{\tau,i}$ is the receiver's own interim belief (from the receiver's independent draw of K narratives at perceived r_R) and $\theta^{(\text{best},D)}_{\tau,j}$ is the developer's best-fit narrative (from the developer's draws at r_D). The parameter α_R measures *narrative persuasion*. The receiver's interim belief and the developer's best-fit narrative come from *different* individuals' draws, so they are statistically independent.

All simulated beliefs are clipped to $[0, 1]$.

C.5 Model-implied Moments

For each signal history h, we derive model-implied means and variances from the belief equations above.

Mean moments. Since $\mathbb{E}[\bar{\theta}_{\tau,i}] = \mathbb{E}_g[\theta_\tau \mid h]$, the mean equations are:

$$M_B(h) = \mathbb{E}_g[\theta_\tau \mid h], \tag{6}$$

$$M_D(h) = (1-\alpha_D) \cdot \mathbb{E}_g[\theta_\tau \mid h] + \alpha_D \cdot \mathbb{E}[\theta^{(\text{best})}_\tau \mid h], \tag{7}$$

$$M_R(h) = (1-\alpha_R) \cdot \mathbb{E}_g[\theta_\tau \mid h] + \alpha_R \cdot \mathbb{E}[\theta^{(\text{best})}_\tau \mid h]. \tag{8}$$

In Specification (ii), M_D uses the posteriors at r_D and M_R uses r_R for the interim belief but r_D for the best-fit narrative (since receivers observe a developer's narrative).

Variance moments. The variance decompositions differ between developers and receivers because of the covariance structure.

Baseline. Since $\text{Var}(\bar{\theta}_{\tau,i} \mid h) = \text{Var}_g(\theta_\tau \mid h)/K$:

$$V_B(h) = \frac{\text{Var}_g(\theta_\tau \mid h)}{K} + \sigma^2_\varepsilon. \tag{9}$$

Developer. The developer's interim belief and best-fit narrative come from the same K draws, producing a positive covariance:

$$V_D(h) = (1-\alpha_D)^2 \cdot \frac{\text{Var}_g(\theta_\tau \mid h)}{K} + \alpha_D^2 \cdot V_{\text{bf}}(h) + 2(1-\alpha_D)\alpha_D \cdot C_{\text{bf}}(h) + \sigma^2_\varepsilon, \tag{10}$$

where $V_{\text{bf}}(h) \equiv \text{Var}(\theta^{(\text{best})}_\tau \mid h)$ is the variance of the best-fit narrative's state estimate across independent sets of K draws, and $C_{\text{bf}}(h) \equiv \text{Cov}(\bar{\theta}_{\tau,i},\, \theta^{(\text{best})}_{\tau,i} \mid h)$ is the covariance between the interim belief and the best-fit prediction within the same set of draws. Both are computed by Monte Carlo simulation.

Receiver. The receiver's interim belief (from their own draws) and the developer's best-fit narrative (from the developer's draws) are independent:

$$V_R(h) = (1-\alpha_R)^2 \cdot \frac{\text{Var}_g(\theta_\tau \mid h, r_R)}{K} + \alpha_R^2 \cdot V_{\text{bf}}(h) + \sigma^2_\varepsilon. \tag{11}$$

There is no covariance term because the receiver's interim belief and the developer's best-fit narrative come from different individuals' independent draws.

Moment conditions. We match 5 moment types for each of the 48 histories, yielding $5 \times 48 = 240$ targeted moment conditions:

1. $M_B(h)$: mean baseline belief (eq. 6)
2. $V_B(h)$: within-cell variance of baseline beliefs (eq. 9)
3. $M_N(h)$: mean developer narrative prediction ($\mathbb{E}[\theta_\tau^{(\text{best})}]$)
4. $M_D(h)$: mean developer belief (eq. 7)
5. $M_R(h)$: mean receiver belief (eq. 8)

The variance moments $V_D(h)$ and $V_R(h)$ are not targeted in estimation; they are computed as untargeted (out-of-sample) moments using the common noise variance σ_ε^2 estimated from baseline variance residuals.

C.6 Parameters, Profiling, and Estimation

Parameter vectors. Specification (i) estimates $\theta = (\beta, K, \sigma_\varepsilon^2, \alpha_D, \alpha_R)$—5 parameters (4 continuous + K discrete). Specification (ii) adds r_{BL} and r_D, giving 7 parameters (6 continuous + K discrete).

Profiling. For given outer parameters (β, r) (or (β, r_{BL}, r_D)), the inner parameters $(K, \sigma_\varepsilon^2, \alpha_D, \alpha_R)$ are profiled out in closed form or by one-dimensional grid search, avoiding high-dimensional numerical optimization:

- K is searched over a discrete grid $\{1, 2, \ldots, 50\}$.
- For each K, the noise variance σ_ε^2 is solved from baseline variance residuals:

$$\hat{\sigma}_\varepsilon^2 = \frac{\sum_h w_h^{(v)} \left[\widehat{V}_B^{\text{data}}(h) - \text{Var}_g(\theta_\tau \mid h)/K\right]}{\sum_h w_h^{(v)}},$$

  truncated at ≥ 0, where $w_h^{(v)}$ are the optimal weights for variance moments.
- The persuasion parameters α_D and α_R are estimated by weighted least squares, regressing treatment belief residuals on narrative predictions.

The same noise variance σ_ε^2 enters the variance equations for all treatment arms (eqs. 9–11). The outer search over (β, r) proceeds by grid search followed by Nelder-Mead refinement.

Weighting. The estimation uses two-step efficient SMM. Step 1 applies diagonal normalization (equal total weight per moment type). Step 2 uses optimal diagonal weights: $w_h = n_h / s_h^2$ for mean moments and $w_h = (n_h - 1)/(2s_h^4)$ for variance moments, where n_h is the cell size and s_h^2 is the within-cell sample variance.

Standard errors. Standard errors are computed from the sandwich formula $V(\hat{\theta}) = (G'\Sigma^{-1}G)^{-1}$, where G is the $240 \times p$ Jacobian of model moments with respect to parameters (using analytical derivatives where available and numerical central differences for β and r) and Σ is the diagonal matrix of moment sampling variances (with entries s_h^2/n_h for means and $2s_h^4/(n_h - 1)$ for variances). Standard errors are conditional on $\hat{K}$.

C.7 Assessing Model Fit

C.7.1 Fit to Targeted and Untargeted Moments

The model fits the targeted mean moments well. Cross-history correlations between data and model-implied mean beliefs exceed 0.93 for all treatment arms in both specifications. In Specification (ii), the baseline mean correlation rises from 0.55 to 0.96, as the low $\hat{r}_{BL}$ allows the model to track signal histories where baseline beliefs deviate substantially from the $r = 0.2$ Bayesian benchmark.

A key diagnostic is the model's ability to predict untargeted moments—quantities that are never used in estimation but emerge from the structural assumptions. The model uses a single noise parameter $\hat{\sigma}_\varepsilon^2$, estimated from cross-sectional belief variance in Baseline. As a result, the model's predictions for cross-sectional belief variances in Developer and Receiver are *untargeted*—as Panel B of Table C1 shows, these untargeted variance predictions track the data reasonably well, with cross-history correlations of 0.83 for Developer beliefs and 0.60–0.61 for Receiver beliefs.[33]

The model also matches belief volatility across evaluation periods, another dimension not used in estimation. We simulate subjects through the actual 30-signal experimental histories, computing new posteriors, narrative samples, and beliefs at each of the five evaluation periods. Panel C of Table C1 reports two key volatility statistics: extremeness and absolute period-to-period belief change ($\Delta(p_T)$). The model closely tracks both: Developer $\Delta(p_T)$ is 31.3 vs. data 31.7 in Specification (ii), and Receiver $\Delta(p_T)$ is 31.1 vs. data 30.7. The model slightly underpredicts Baseline extremeness (15.4 vs. 16.4), but accurately captures the treatment effects.

C.7.2 Fit at the Signal-history Level

The structural model is estimated by matching moments for each of 48 distinct signal histories, so the relevant question for model fit is not merely whether aggregate statistics are well matched, but whether the model tracks the cross-history variation in beliefs. Tables C2 and C3 present data and model-implied mean beliefs for a purposefully selected subset of signal histories under Specifications (i) and (ii), respectively. The final column reports the Bayesian benchmark (computed at $r = 0.2$).

The selected histories are chosen to span the range of Bayesian posterior beliefs (from near 50 to above 85) and, importantly, to include cases that are most informative about

[33] The variance patterns reveal that narrative sharing amplifies cross-sectional disagreement. In the data, Developer belief variance (403) is 52% larger than Baseline variance (266), while Receiver variance (288) is 8% larger. Signal histories where the Bayesian posterior is near 50—i.e., the most ambiguous histories—produce the greatest narrative diversity and hence the highest Developer belief variance.

Table C1: Aggregate model fit: targeted, untargeted, and out-of-sample moments

		Spec. (i): $r = 0.2$		Spec. (ii): $r_R = r_{BL}$	
Moment	Data	Model	Corr	Model	Corr
Panel A: Targeted moments					
Baseline mean belief	65.15	64.83	0.553	64.23	0.960
Baseline belief variance	265.6	204.0	0.273	206.7	0.334
Developer narrative mean	71.99	71.46	0.941	71.51	0.940
Developer belief mean	69.82	70.01	0.964	69.23	0.958
Receiver belief mean	70.70	69.86	0.937	69.63	0.942
Panel B: Untargeted variance moments					
Developer narrative variance	314.4	310.0	0.903	302.0	0.907
Developer belief variance	402.8	343.9	0.826	295.8	0.826
Receiver belief variance	287.8	318.9	0.604	325.3	0.613
Panel C: Untargeted aggregate statistics					
Extremeness, Baseline	16.35	14.83	—	15.42	—
Extremeness, Developer	19.96	19.98	—	19.31	—
Extremeness, Receiver	20.94	19.82	—	19.66	—
$\Delta(p_T)$, Baseline	22.84	25.14	—	21.97	—
$\Delta(p_T)$, Developer	31.66	33.09	—	31.26	—
$\Delta(p_T)$, Receiver	30.71	32.51	—	31.10	—

Notes: Panel A moments are matched in estimation. Panel B variance moments are computed post-estimation using the common noise $\hat{\sigma}_\varepsilon^2$ identified from Baseline belief variance, combined with the structural variance decomposition (Appendix C). Panel C reports aggregate statistics not used in estimation. "Data" and "Model" report observation-weighted averages across all 48 signal histories. "Corr" is the observation-weighted cross-history correlation between data and model moments (not applicable for Panel C aggregate statistics). Extremeness $\equiv |\text{belief} - 50|$. $\Delta(p_T) \equiv |p_t - p_{t-5}|$ is the absolute period-to-period belief change, computed per subject across five evaluation periods and then averaged. Panel C model predictions are simulated using actual 30-signal experimental histories with period-specific posteriors; 5,000 simulated subjects per base history. See Appendix C for full aggregate statistics including mean beliefs and SD(p).

the model's strengths and limitations. Three types of histories are highlighted: (1) histories where Baseline beliefs in the data closely track the Bayesian benchmark, where the model fits well under both specifications (e.g., $h = 9, 20, 27$); (2) histories where Baseline beliefs in the data deviate substantially from the $r = 0.2$ Bayesian benchmark, which are problematic for Specification (i) but well fit by Specification (ii) (e.g., $h = 0, 46, 47$); and (3) histories where the model faces persistent challenges even under Specification (ii), often reflecting idiosyncratic sample-level features such as small cell sizes or unusual compositions of signal orderings (e.g., $h = 35$).

Specification (i). When r is fixed at 0.2 for all arms, the model's Baseline predictions are anchored to the Bayesian posterior. This produces good fit for histories where data Baseline means happen to be close to the Bayesian benchmark (e.g., $h = 20$: data 68.2,

model 67.2, Bayes 69.1), but large misses for histories where data BASELINE means diverge. For instance, history $h = 0$ has a data BASELINE mean of 35.4 but the model predicts 52.9 (close to the Bayesian 52.8), a gap of 17.5 percentage points. Similarly, $h = 46$ (data 72.0, model 52.8) and $h = 35$ (data 73.7, model 52.6) exhibit large BASELINE misfits. These gaps propagate into the treatment predictions because the treatment belief equations combine the interim belief (anchored to the Bayesian posterior) with the best-fit narrative prediction; when the BASELINE level is wrong, the treatment-arm predictions are pulled in the wrong direction as well. The aggregate cross-history correlation for BASELINE means is only 0.55 under this specification (Table C1).

Specification (ii). Allowing treatment-specific r dramatically improves the per-history BASELINE fit. The estimated $\hat{r}_{BL} \approx 0.02$ generates posteriors that are flatter and more sensitive to the overall signal composition rather than recent signals, closely tracking the empirical BASELINE levels. Histories that were badly fit under Specification (i) are now well matched: $h = 0$ (data 35.4, model 34.4), $h = 46$ (data 72.0, model 73.8), and $h = 47$ (data 89.9, model 89.8). The BASELINE cross-history correlation rises from 0.55 to 0.96. Treatment-arm predictions also improve because the treatment belief equations for DEVELOPER and RECEIVER participants use $r_D \approx 0.19$ (close to the true value), maintaining the steep recency gradient needed to match treatment patterns.

Table C2: Data vs. model moments for selected histories (Specification (i): $r = 0.2$)

	BASELINE		DEVELOPER Narr.		DEVELOPER Belief		RECEIVER Belief		
h	Data	Model	Data	Model	Data	Model	Data	Model	Bayes
0	35.4	52.9	46.4	54.7	47.8	50.3	46.8	49.8	52.8
3	62.4	74.6	89.9	87.3	79.3	81.6	84.9	81.0	79.2
9	71.8	75.3	86.3	87.7	80.0	84.1	80.2	83.7	79.4
14	83.5	77.4	89.8	90.7	82.7	89.0	85.5	88.8	82.5
20	68.2	67.2	71.1	72.1	72.8	71.2	71.5	71.1	69.1
27	82.4	78.3	91.4	91.6	86.9	89.5	84.9	89.2	83.0
35	73.7	52.6	65.9	53.5	58.4	58.1	70.4	58.6	53.7
43	59.8	51.9	45.9	49.4	52.4	51.8	47.2	52.0	50.9
46	72.0	52.8	53.9	46.7	53.9	52.5	65.0	53.1	52.3
47	89.9	80.0	92.8	94.4	90.5	93.4	93.7	93.2	84.9

Notes: Selected histories span the range of Bayesian posteriors and include cases where data baselines closely track and deviate from the Bayesian benchmark. Full results for all 48 histories available upon request. BASELINE misfits are largest for histories where data beliefs diverge from the $r = 0.2$ Bayesian posterior (e.g., $h = 0$, 35, 46), where Specification (i) is constrained by the assumption that all participants perceive $r = 0.2$.

Table C3: Data vs. model moments for selected histories (Specification (ii): Subj. $\hat{r}$

	Baseline		Developer Narr.		Developer Belief		Receiver Belief		
h	Data	Model	Data	Model	Data	Model	Data	Model	Bayes
0	35.4	34.4	46.4	54.3	47.8	53.5	46.8	49.2	52.8
3	62.4	72.1	89.9	87.8	79.3	83.3	84.9	83.8	79.2
9	71.8	73.0	86.3	88.2	80.0	83.8	80.2	84.3	79.4
14	83.5	79.7	89.8	91.5	82.7	87.0	85.5	88.5	82.5
20	68.2	67.1	71.1	72.0	72.8	70.4	71.5	70.7	69.1
27	82.4	83.5	91.4	92.3	86.9	87.7	84.9	90.1	83.0
35	73.7	70.3	65.9	55.6	58.4	54.8	70.4	59.4	53.7
43	59.8	57.5	45.9	49.2	52.4	50.2	47.2	51.4	50.9
46	72.0	73.8	53.9	45.2	53.9	48.5	65.0	52.6	52.3
47	89.9	89.8	92.8	95.1	90.5	90.0	93.7	93.7	84.9

Notes: Specification (ii) achieves substantially better Baseline fit than Specification (i). The low $\hat{r}_{BL} \approx 0.02$ allows the model to match Baseline levels that deviate from the $r = 0.2$ Bayesian benchmark (e.g., $h = 0$: model 34.4 vs. data 35.4; $h = 47$: model 89.8 vs. data 89.9). Treatment-arm fit is broadly similar across specifications because Developer and Receiver participants use $\hat{r}_D \approx 0.19$, close to the true value.

C.7.3 Fit to Belief Sensitivities to Different Signals

A central finding of Section 4.5.3 is that Baseline beliefs weight signals relatively uniformly across positions, while treatment beliefs exhibit steep recency gradients—placing disproportionate weight on the most recent signals. This pattern is the primary empirical motivation for Specification (ii), which allows treatment-specific perceived r. In this appendix, we examine whether the structural model reproduces these signal sensitivity profiles and, in particular, whether the estimated $\hat{r}_{BL}$ and $\hat{r}_D$ values can account for the stark difference between Baseline and treatment sensitivity patterns.

Methodology. We regress both data and model-implied beliefs on the 10 binary signal indicators:

$$\text{belief}_h = \gamma_0 + \sum_{t=1}^{10} \gamma_t \cdot s_{h,t} + u_h,$$

where $s_{h,t} \in \{0, 1\}$ is the t-th signal in history h, using n-weighted OLS (with n_h equal to the number of subjects who observed history h). The coefficients γ_t measure how much an additional favorable signal at position t shifts beliefs, in percentage points. A Bayesian agent who believes the environment is nearly stationary (low r) should weight all signals roughly equally, because past signals remain informative about the current state. Conversely, an agent who perceives frequent regime change (high r) should discount early signals heavily and place most weight on recent ones, since older signals may reflect a now-obsolete regime. The slope of the coefficient profile—from γ_1 to γ_{10}—therefore provides a diagnostic of how participants perceive the persistence of the data-generating process.

Figures C1–C4 plot the resulting coefficient series for each treatment arm. Each figure compares the data coefficients to model-implied coefficients evaluated at two r values: $\hat{r}_D \approx 0.19$ (close to the true redraw probability) and $\hat{r}_{BL} \approx 0.02$ (the estimated BASELINE perception). This comparison reveals a fundamental tension that no single r can resolve.

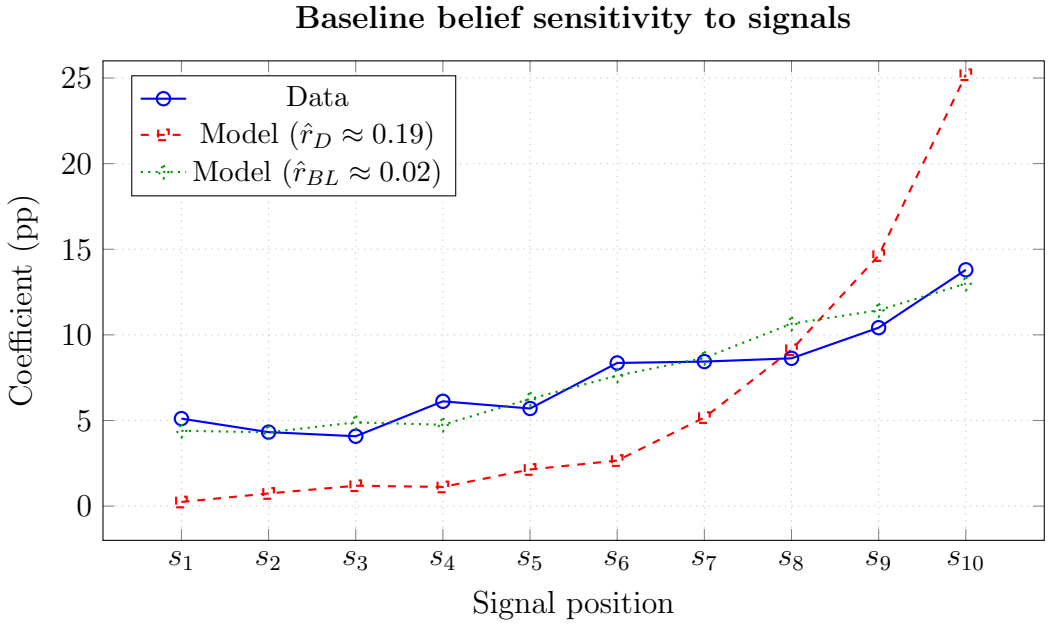


Figure C1: Signal sensitivity for BASELINE beliefs. The data profile is relatively flat, with coefficients ranging from roughly 4–14 pp. At $\hat{r}_D \approx 0.19$, the model generates a steep recency gradient that far overshoots the data (especially for early signals s_1–s_5, which the model predicts to be nearly uninformative). At $\hat{r}_{BL} \approx 0.02$, the flat profile matches the data closely, confirming that BASELINE participants behave as if the environment were nearly stationary.

Baseline vs. treatment sensitivity. Figure C1 illustrates the core problem with assuming a common $r = 0.2$ for all arms. The data show that BASELINE participants place substantial weight on even the earliest signals ($\gamma_1 \approx 5$ pp), a pattern consistent with perceiving the environment as highly persistent. At $r = 0.2$, the model predicts that s_1 should contribute only $\sim$0.2 pp—essentially zero—because under that rate of regime change, a signal from 10 periods ago carries little information about the current state. The low $\hat{r}_{BL} \approx 0.02$ resolves this by making all 10 signals comparably informative.

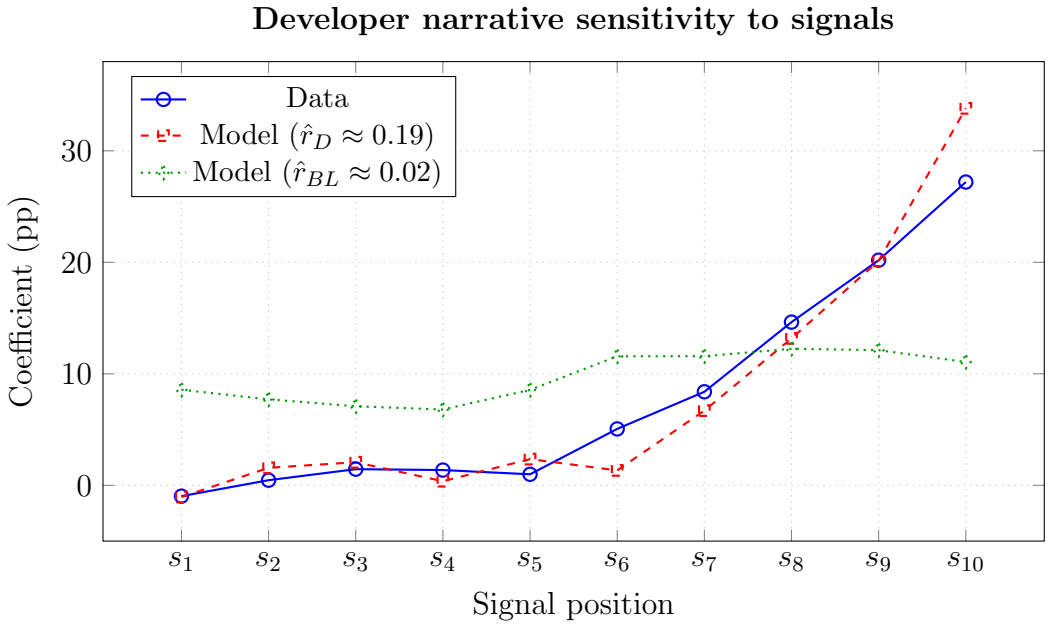


Figure C2: Signal sensitivity for DEVELOPER narratives. The data profile is steep: s_{10} contributes ∼27 pp while s_1 contributes ∼−1 pp. At $\hat{r}_D \approx 0.19$, the model reproduces this steep gradient closely. At $\hat{r}_{BL} \approx 0.02$, the model generates a nearly flat profile ($\gamma_t \approx 7$–13 pp for all t), failing to capture the dominance of recent signals.

Treatment sensitivity profiles. Figures C2–C4 show that the treatment arms exhibit steep, monotonically increasing coefficient profiles—the hallmark of beliefs driven by an r close to the true value. For DEVELOPER narratives (Figure C2), the most recent signal (s_{10}) contributes roughly 27 pp to beliefs, while the earliest signal (s_1) contributes approximately zero. The model at $\hat{r}_D \approx 0.19$ closely replicates this pattern (correlation 0.98, RMSE 2.6 pp), whereas the model at $\hat{r}_{BL} \approx 0.02$ predicts a flat profile that is qualitatively wrong (correlation 0.72, RMSE 8.0 pp). This demonstrates that the perceived r is the key determinant of the signal sensitivity shape: a high r generates steep recency, while a low r generates flatness.

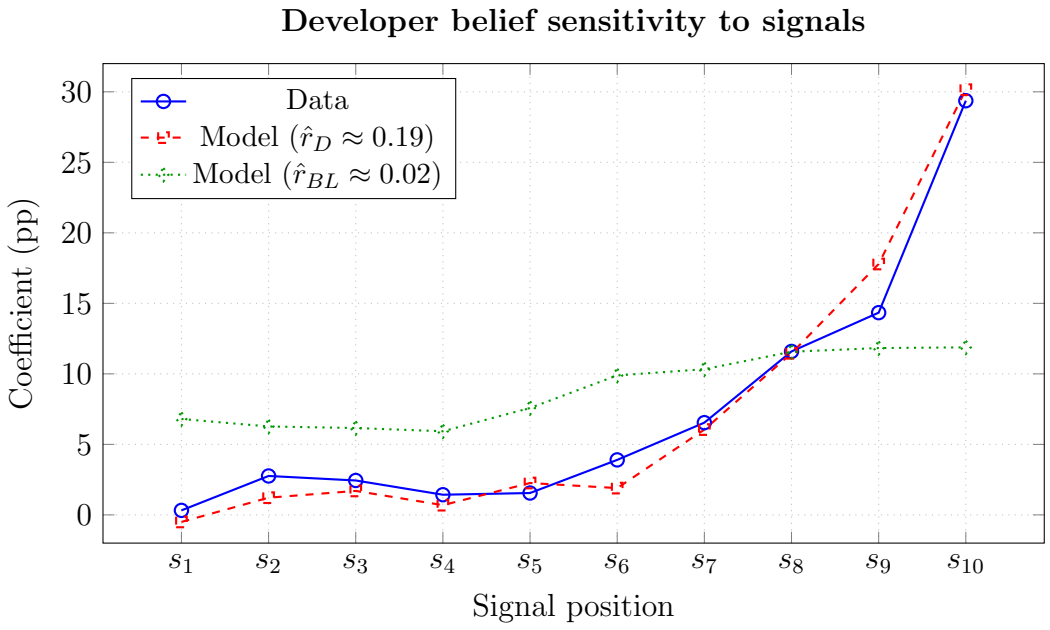


Figure C3: Signal sensitivity for DEVELOPER beliefs. DEVELOPER beliefs inherit the steep recency gradient from the narrative: s_{10} contributes ∼29 pp, while early signals contribute near zero. The model at $\hat{r}_D \approx 0.19$ fits closely (correlation 0.99, RMSE 1.5 pp). At $\hat{r}_{BL} \approx 0.02$, the flat profile misses the data entirely.

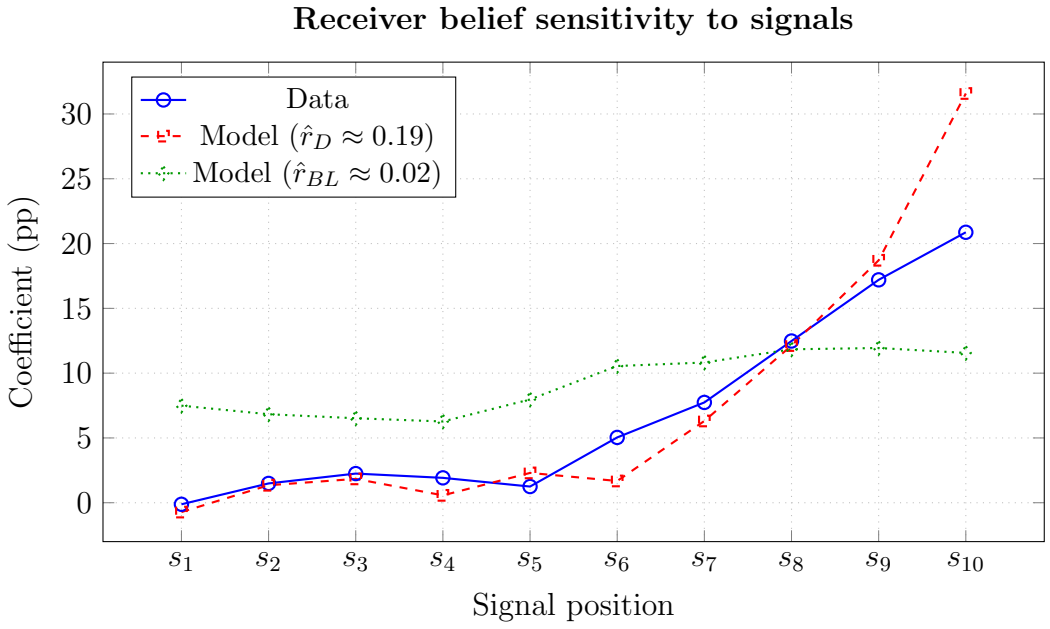


Figure C4: Signal sensitivity for RECEIVER beliefs. The steep recency gradient mirrors DEVELOPER beliefs, with s_{10} contributing ∼21 pp. The model at $\hat{r}_D \approx 0.19$ captures the shape (correlation 0.96), though it slightly overpredicts the s_{10} coefficient. At $\hat{r}_{BL} \approx 0.02$, the flat profile again fails to reproduce the steep data gradient.

Summary of fit. Table C4 summarizes the fit of the signal sensitivity profiles across all treatment arms. Two patterns stand out. First, within each r value, the correlations and RMSE values reveal which treatment arms are well explained. At $\hat{r}_D \approx 0.19$, all three treatment arms (DEVELOPER narratives, DEVELOPER beliefs, RECEIVER beliefs) are fit with correlations above 0.96 and RMSE values of 1.5–3.7 pp, but the BASELINE fit is poor (RMSE 5.2 pp) because the model generates too steep a recency gradient for BASELINE beliefs. At $\hat{r}_{BL} \approx 0.02$, the BASELINE fit is excellent (RMSE 1.0 pp), but all three treatment arms are poorly fit (RMSE 5.7–8.0 pp) because the model generates nearly flat profiles that miss the steep data gradients.

Second, no single r value can simultaneously fit BASELINE and treatment sensitivity profiles. This is the fundamental tension that motivates Specification (ii): BASELINE participants behave as if the environment is nearly stationary ($r \approx 0.02$), while treatment participants—who have constructed or observed narratives—behave as if they correctly perceive the true rate of regime change ($r \approx 0.19$). The structural model accommodates this by allowing r to differ across treatment arms, with the narrative construction task appearing to attenuate a BASELINE tendency toward perceiving excessive stationarity.

Table C4: Signal sensitivity fit: data vs. model coefficients

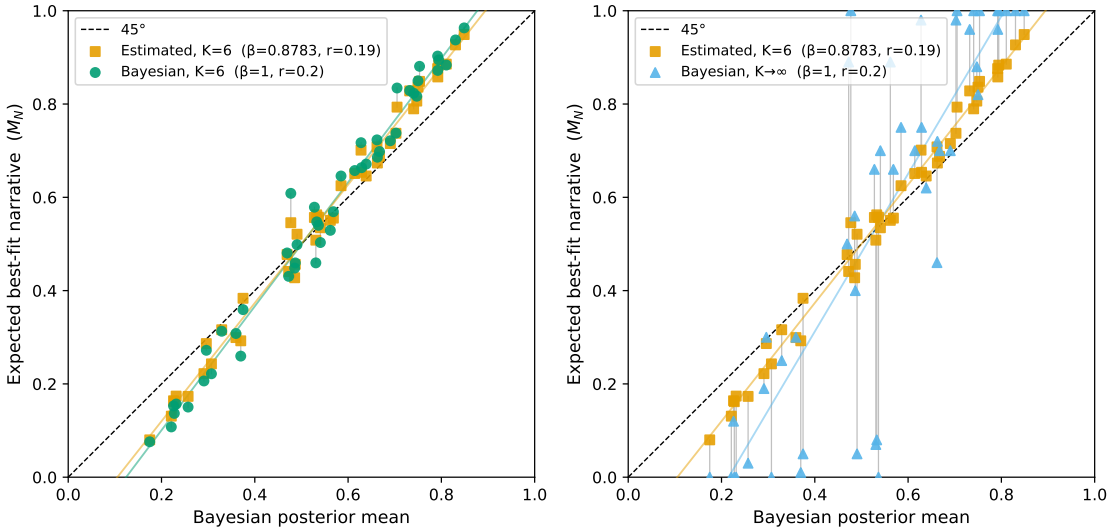
	$\hat{r}_D \approx 0.19$		$\hat{r}_{BL} \approx 0.02$	
Treatment	Corr.	RMSE	Corr.	RMSE
BASELINE belief	0.932	5.23	0.946	0.99
DEVELOPER narrative	0.977	2.59	0.721	8.04
DEVELOPER belief	0.991	1.47	0.778	6.98
RECEIVER belief	0.964	3.66	0.849	5.68

Notes: Each column reports the correlation and RMSE (in percentage points) between data and model-implied signal sensitivity coefficients ($\gamma_1, \ldots, \gamma_{10}$) at the indicated r value. At $\hat{r}_D \approx 0.19$, treatment arms are well fit but BASELINE is not. At $\hat{r}_{BL} \approx 0.02$, BASELINE is well fit but treatment arms are not. This fundamental tension—BASELINE beliefs require a low r (flat profile) while treatment beliefs require a high r (steep recency gradient)—confirms the need for treatment-specific perceived r in Specification (ii).

C.8 Sensitivity of Subjective Best-fit Estimates to Sampling Parameters

Figure C5 displays two scatterplots illustrating the sensitivity of the subjective best-fit estimates to distortions to the sampling distribution (β and r) and to changes in the number of sampled narratives (K). Each dot is a unique signal history in the experiment. Predictions made for the same signal history are connected by vertical lines. The solid lines are linear fit estimates. As the figure illustrates, the finite K has a large impact on the nature of developer narratives, while the distortions to the sampling distribution have little influence.

Figure C5: Sensitivity of Subjective Best-fit Estimates to Sampling Distortions



D Belief Sensitivity to Different Signals: Additional Analysis

Both our reduced form and structural estimates suggest that BASELINE participants underestimate the redraw probability r. In this appendix section, we explore whether this is driven by a heuristic behavior of some BASELINE participants who ignore the temporal sequence of signals, essentially treating the data-generating process as if it features no break.

To detect heterogeneity among BASELINE participants, we perform a cluster analysis, sorting participants based on their individual signal sensitivity. We obtain an individual sensitivity measure by estimating the following regression specification for every participant in the sample, using their data over 25 rounds:

$$\theta_\tau = \alpha + \beta_1 \left(\sum_{t=0}^{4} \sigma_{\tau-t} \right) + \beta_2 \left(\sum_{t=5}^{9} \sigma_{\tau-t} \right) + \varepsilon.$$

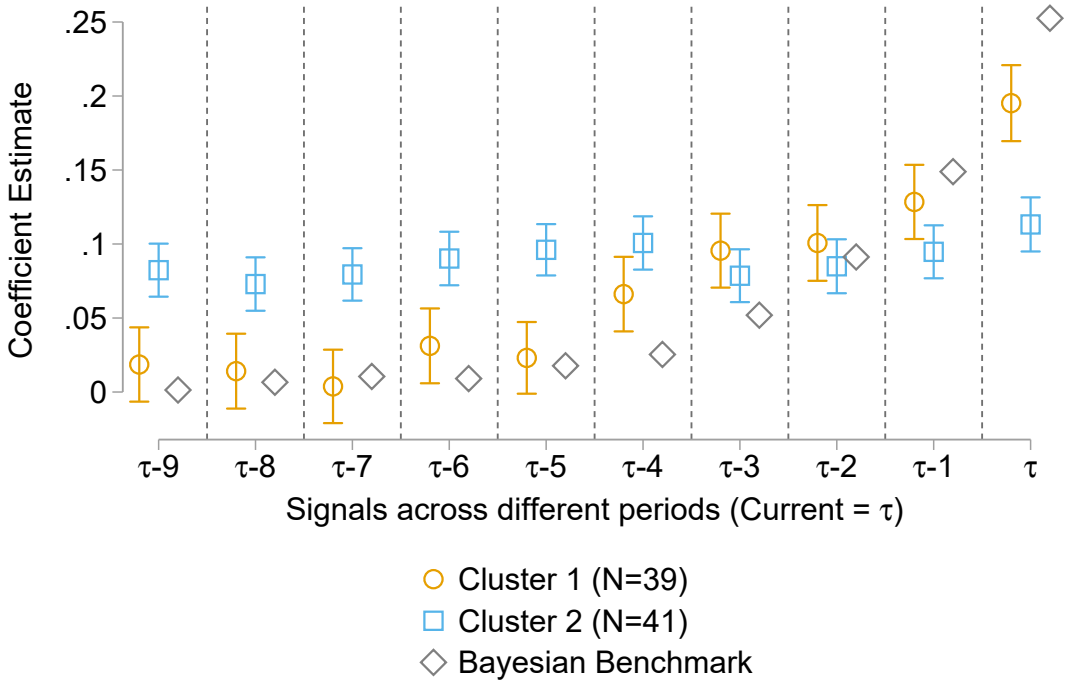
The resulting coefficients $(\hat{\beta}_1, \hat{\beta}_2)$ denote a participant's sensitivity to the five most recent or the five most distant signals, respectively. This is a more aggregated version of the specification employed in Section 4.5.3 but it is in the same spirit: It provides us with a measure of how sensitive a participant is to distant signals, relative to more recent ones. We use this aggregate measure to improve coefficient precision, as we only have 25 observations per participant.

After obtaining a vector $(\hat{\beta}_1, \hat{\beta}_2)$ for each BASELINE participant, we employ a k-means clustering algorithm to sort participants into two clusters based on the similarity of their vectors. The algorithm identifies two participant clusters of roughly equal size. We use the identified clusters to estimate the same regression as in the main text,

$$\theta_\tau = \alpha + \sum_{t=0}^{9} \beta_t \sigma_{\tau-t} + \varepsilon,$$

separately for each cluster. Figure D1 presents results indicating that the algorithm indeed identified heterogeneous sensitivity patterns among participants. The coefficient estimates for Cluster 2 (41/80 participants) are essentially independent across different time periods, suggesting that Cluster 2 consists of "heuristic" participants who do not account for the temporal sequence of signals when forming their beliefs. Instead, the reactions of Cluster 1 (39/80 participants) to the temporal signal sequence are quite pronounced.

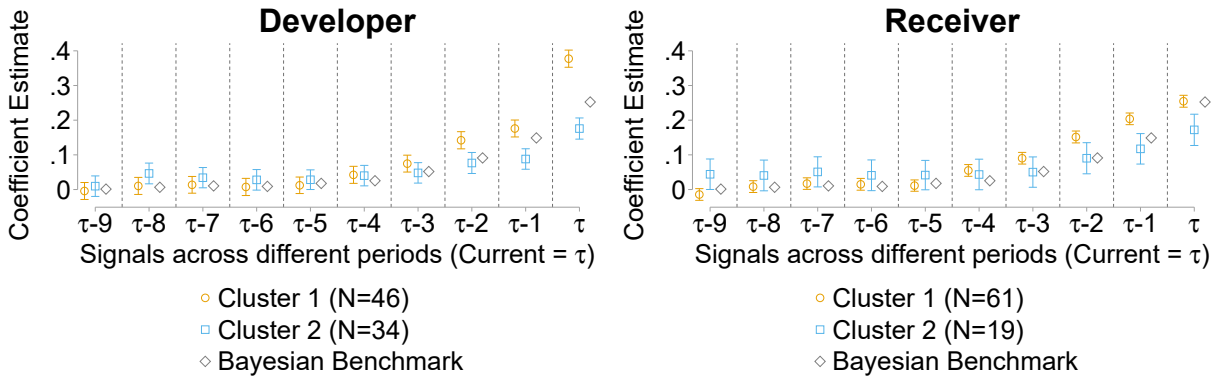
Figure D1: Sensitivity of mean beliefs to signal realizations, by Baseline cluster



Notes: *Each panel plots the estimated coefficients β_t on signals from periods τ (most recent) to $\tau - 9$ (most distant) from separate regressions estimated within each cluster. Error bars are 95% confidence intervals derived from standard errors clustered by participant.*

We also performed the same clustering exercise in the narrative sharing treatments to investigate whether heuristic participants also are present in these treatments. Figure D2 displays the estimated signal sensitivity for every cluster, which is always sloping upwards. While there is also some heterogeneity in sensitivity, it is less pronounced and the fully flat pattern observed among Cluster 2 in Baseline cannot be replicated. Therefore, the heuristic tendency to treat signals as independent of one another appears to be a special feature of Baseline.

Figure D2: Sensitivity of mean beliefs to signal realizations, by DEVELOPER and RECEIVER cluster



Notes: *Each panel plots the estimated coefficients β_t on signals from periods τ (most recent) to $\tau - 9$ (most distant) from separate regressions estimated within each cluster. Error bars are 95% confidence intervals derived from standard errors clustered by participant.*

In the two tables below, we replicate our main extremeness and volatility results while controlling for whether BASELINE participants belong to the heuristic cluster (Cluster 2). This analysis provides several insights. First, relative to Cluster 1 BASELINE participants, heuristic cluster members have less extreme and less volatile beliefs. This is expected, as these cluster members effectively regard a larger sample size when forming their beliefs. Second, DEVELOPER and RECEIVER treatments still generate more volatile and extreme beliefs, relative to the non-heuristic Cluster 1 BASELINE participants. Therefore, our main results remain robust when using the more signal-sensitive BASELINE participants as a comparison group.

Table D1: Main extremeness results while controlling for BASELINE cluster membership

	Main Extremeness Measure: $\lvert\theta_\tau - 0.5\rvert$			
	Narratives vs. BASELINE Beliefs		Beliefs across Treatments	
	(1)	(2)	(3)	(4)
Heuristic Cluster	-0.023**	-0.032***	-0.023**	-0.032***
	(0.010)	(0.010)	(0.010)	(0.010)
DEVELOPER Narratives	0.051***	0.045***		
	(0.011)	(0.010)		
DEVELOPER Beliefs			0.038***	0.032**
			(0.013)	(0.013)
RECEIVER Beliefs			0.030***	0.025***
			(0.009)	(0.008)
Sample	Full Sample	Highly Consistent	Full Sample	Highly Consistent
Round FE	Yes	Yes	Yes	Yes
Signal History FE	Yes	Yes	Yes	Yes
BASELINE Cluster 1 Mean	0.219	0.225	0.219	0.225
Observations	4000	3575	6000	5425

Notes: *Standard errors are clustered by participant. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. The dependent variable is the absolute deviation of the elicited value from 0.5, the prior mean of the state. Columns (1)–(2) compare current-state values in* DEVELOPER *narratives to mean beliefs in* BASELINE. *Columns (3)–(4) compare mean beliefs across all treatments. The omitted category is always mean beliefs in* BASELINE. *The subsample includes participants who passed the preregistered consistency check. These specifications replicate our main extremeness results while including a Heuristic Cluster indicator specifying whether a* BASELINE *participant belongs to Cluster 2. The reference group are Cluster 1* BASELINE *participants.*

Table D2: Main volatility results while controlling for Baseline cluster membership

	Main Volatility Measure: $\lvert\theta_\tau - \theta_{\tau-5}\rvert$			
	Narratives vs. Baseline Beliefs		Beliefs across Treatments	
	(1)	(2)	(3)	(4)
Heuristic Cluster	-0.072***	-0.086***	-0.073***	-0.086***
	(0.014)	(0.014)	(0.013)	(0.013)
Developer Narratives	0.062***	0.058***		
	(0.016)	(0.016)		
Developer Beliefs			0.051***	0.041**
			(0.018)	(0.019)
Receiver Beliefs			0.042***	0.033***
			(0.013)	(0.012)
Sample	Full Sample	Highly Consistent	Full Sample	Highly Consistent
Round FE	Yes	Yes	Yes	Yes
Signal History FE	Yes	Yes	Yes	Yes
Baseline Cluster 1 Mean	0.265	0.274	0.265	0.274
Observations	3200	2860	4800	4340

Notes: *Standard errors are clustered by participant. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. The dependent variable is the absolute movement of the elicited value across two adjacent evaluation periods. Columns (1)–(2) compare current-state values in* Developer *narratives to mean beliefs in* Baseline*. Columns (3)–(4) compare mean beliefs across all treatments. The omitted category is always mean beliefs in* Baseline*. The subsample includes participants who passed the preregistered consistency check. These specifications replicate our main extremeness results while including a Heuristic Cluster indicator specifying whether a* Baseline *participant belongs to Cluster 2. The reference group are Cluster 1* Baseline *participants.*

E Evidence for Additional Preregistered Hypotheses

We preregistered two sets of additional hypotheses not reported in the main text. First, we preregistered hypotheses on the relation between the mean beliefs and the Bayesian benchmark, where we anticipate the over- and underinference results. Second, we preregistered a hypothesis on mean beliefs being more likely to match the global best-fit estimate in the narrative sharing treatments. Our empirical results are in line with every omitted hypothesis. Here, we provide the detailed evidence.

Comparison with the Bayesian benchmark

Hypothesis 7. *In all treatments, mean beliefs are positively associated with the Bayesian expectation.*

Hypothesis 8a. *Mean beliefs in* Developer *are more sensitive to the Bayesian benchmark than those in* Baseline.

Hypothesis 8b. *Mean beliefs in* Receiver *are more sensitive to the Bayesian benchmark than those in* Baseline.

Table E1 presents preregistered tests of the above hypotheses. We find that mean beliefs are positively correlated with the Bayesian benchmark, supporting Hypothesis 7; we also find that mean beliefs are more strongly correlated with the Bayesian benchmark in Developer and Receiver than in Baseline, as evidenced by the significantly positive treatment interaction terms. These results are supportive of Hypotheses 8a and 8b.

Transmission between narratives and mean beliefs

Hypothesis 9a. *Mean beliefs in* Developer *are more likely to coincide with the best-fit estimate than those in* Baseline.

Hypothesis 9b. *Mean beliefs in* Receiver *are more likely to coincide with the best-fit estimate than those in* Baseline.

Table E2 presents preregistered tests of the above hypotheses. We perform regressions with an indicator for whether the mean estimate is within one percentage point of the best fit estimate as the dependent variable and treatment indicators as the independent variables. We find significantly positive Developer and Receiver indicators, supporting both hypotheses.

Table E1: Preregistered test of Hyp. 8

	Elicited mean belief: $\hat{\mathbb{E}}_{tr}[\theta_\tau \mid \boldsymbol{\sigma}'_\tau]$	
	(1)	(2)
Bayesian exp.	0.823***	0.870***
	(0.0397)	(0.0383)
DEVELOPER × Bayesian expectation	0.315***	0.299***
	(0.0641)	(0.0641)
RECEIVER × Bayesian expectation	0.334***	0.302***
	(0.0514)	(0.0510)
DEVELOPER	-0.153***	-0.145***
	(0.0347)	(0.0357)
RECEIVER	-0.168***	-0.152***
	(0.0284)	(0.0287)
Sample	Full sample	Highly consistent
Round FE	Yes	Yes
Observations	6000	5425

Notes: *Standard errors are clustered by participant. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. The dependent variable is the elicited mean belief.*

Table E2: Preregistered test of Hyp. 9

	(1) $\mathrm{I}(p_T^e - p_T^* \in (-1,1))$	(2) $\mathrm{I}(p_T^e - p_T^* \in (-1,1))$
DEVELOPER	0.0770***	0.0799***
	(0.0202)	(0.0222)
RECEIVER	0.0465***	0.0415***
	(0.0140)	(0.0147)
Sample	Whole	Highly consistent
Round FE	Yes	Yes
Data × Company FE	Yes	Yes
Observations	6000	5425

Notes: *Standard errors are clustered by participant. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.*

F Experimental Instructions

F.1 Baseline Treatment

Introduction to the study

In this study, you will be asked to evaluate the state of 5 hypothetical companies labelled A, B, C, D, and E.

Each of the companies that you will evaluate produces a fictitious tech product called a Widget. As the built-in technology advances very quickly, the product is outdated after about a year. At the start of each year, the company, therefore, discontinues the previous year's model and releases a new model of the Widget. In every year, the company's new Widget model is either "a success" or "a failure."

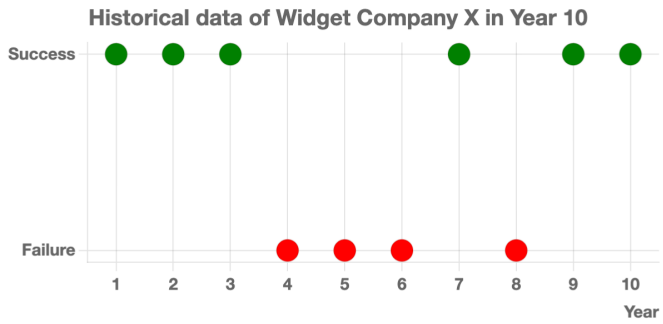
To help you evaluate each company, you will receive several pieces of information. These are explained below.

THE PIECES OF INFORMATION ARE:

1. *Historical performance data*

Whenever you make an evaluation, you will read some historical data of the company. The historical data shows whether the company generated "a success" or "a failure" with the models it produced in each of the past 10 years (e.g., from Year 1 to Year 10).

As a hypothetical example, the data could look like the following:



In a graph like this, a green ball indicates success, and a red ball indicates failure. In this example, you can see that Widget Company X was successful with the model it produced in 6 out of the 10 years.

2. *A Widget company's probability of success depends on its CEO*

In any given year, a company's probability of success is fully determined by its current CEO. During the tenure of a specific CEO, its probability of success will be the same in every year. For any given CEO, this probability is a random number between 0% and 100%. You can think of a good CEO as one whose probability of success is higher than 50% and of a bad CEO as one whose probability of success is lower than 50%. Similarly, you can think of a very good CEO as one whose probability of success is closer to 100% and of a very bad CEO as one whose probability of success is closer to 0%.

3. *The CEO may change over time*

The CEO of a company may change over time due to random factors. Specifically, from one year to the next, there is a fixed **1 in 5 (20%)** chance that the company will switch to a new CEO. This means that, **on average,** there is a CEO change **every 5 years.** However, the tenure of individual CEOs could vary greatly, since CEO changes occur at random points in time. Whenever there is a new CEO, their probability of success will also be a random number between 0% and 100%.

Please note: Whether there is a **CEO change** in any given year is completely random and **does not depend on the performance of the current CEO.**

Although all the Widget companies you will evaluate share the basic characteristics described here, they are completely independent of one another. This means that **the performance of one company does not affect the performance of another company.** Thus, you should evaluate them individually.

Next

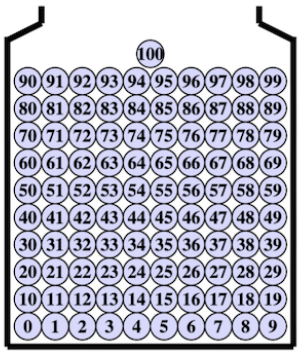
More details on how the CEO affects company outcome in every year

For each Widget company, You can think of the computer as going through the following process.

---

1. *CEO in Year 1*

When the company starts in Year 1, the computer first determines the type of its first CEO. To do so, the computer draws a ball at random from the following urn (and puts it back afterwards).



The drawn number will become the first CEO's probability of success.

For example, **if the computer draws** 70, then the first CEO's probability of success will be 70%. Consequently, the company's probability of success is 70% for each year in which the first CEO is in place.

---

2. *Performance outcome in Year 1*

After drawing the first CEO's probability of success, the computer determines whether the company has a success or failure in Year 1. To do so, the computer draws from an urn with 100 balls, which are either green or red. The number of green balls is equal to the first CEO's probability of success. For example, **if the first CEO's probability of success is 70%**, then there will be 70 green balls and 30 red balls in that urn, as shown below. (For a different type of CEO, the number of green balls in the urn will be different.)

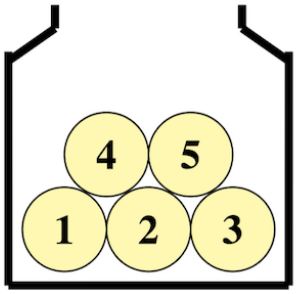


If the ball drawn from the urn is green, then the company is successful in Year 1. If the ball drawn from the urn is red, then the company fails in Year 1.

---

3. *CEO and performance outcome in later years*

In each of the later years, the computer first determines whether the CEO of the company will change in that year. To do so, the computer draws a ball at random from the following urn.



If the computer draws any number between (1) to (4), the CEO from the previous year will remain in place for the current year, and the company will have the same probability of success as before.

If the computer draws (5), the CEO will change and there will be a new CEO. **If there is a new CEO, the computer will determine the probability of success of the new CEO by drawing another ball at random from the first urn above** (see "1. *CEO in Year 1*").

The computer will then determine the success or failure of the company in that year in the same way as in Year 1 (see "2. *Performance outcome in Year 1*"). Specifically, the computer will draw from an urn with 100 balls, which are either green or red, and the number of green balls will be set to the **current** CEO's probability of success. A green ball drawn leads to success in that year, and a red ball drawn leads to failure.

Summary

These details about the process are meant to help you understand the following key points:

- There is a fixed 1 in 5 (20%) chance that the company will switch to a new CEO in any given year.
- When there is a new CEO, their probability of success has nothing to do with the previous CEOs.
- The company's probability of success in any year depends on the type of the CEO active in that year. It is not directly affected by its success or failure in earlier years.

Previous Next

Your task: evaluating each company

You will evaluate each Widget company by providing an estimate of the company's **current probability of success.**

> Later, we will just use **"PoS%"** as a short hand for "probability of success." Remember: The PoS% of a CEO can be any number between 0% and 100%, and it determines the PoS% of the company during all the years in which the CEO is in place.

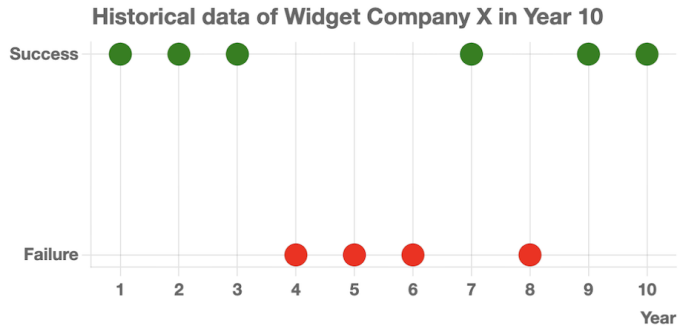
Your bonus will depend on the accuracy of your evaluations (we will provide you with the details on the next page).

---

1. *Making the evaluation*

While evaluating each Widget company, you will have access to its historical performance data spanning **the past 10 years**.

For example, you will first evaluate each company in **Year 10** of its history. You will receive historical data which could look like the following chart.



As your evaluation, we will ask you:

> On average, what is your estimate of the company's **Current PoS% in Year 10**?

Recall that the company's Current PoS% in Year 10 is the PoS% of the CEO currently in place in Year 10. In other words, it will be the company's PoS% in Year 11 **if** the CEO currently active in Year 10 remains in place in Year 11.

Example of a possible evaluation:

> On average, my estimate of the company's Current PoS% in Year 10 is **65%**.

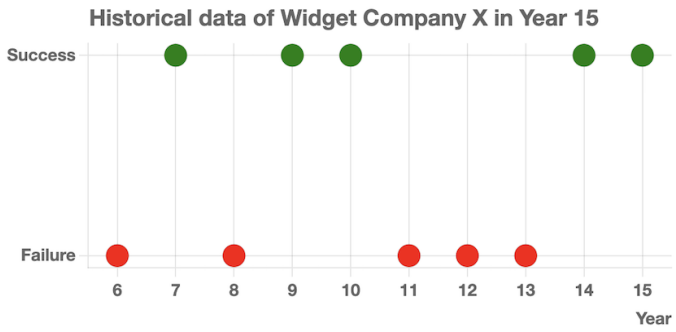
IMPORTANT: As you make the evaluation, you will **not** be told whether and when the company's CEO has changed in the past 10 years. However, you know that there is always a **1 in 5 (20%) chance** that the CEO changes from any year to the next. Thus, there might have been no, one, or multiple CEO changes during the 10-year period.

---

2. *Making new evaluations over time with new information*

You will evaluate each Widget company at 5 points in time. From one point to the next, 5 years will elapse for the company, and new data of company performance in the elapsed years will become available. This implies that you will make separate evaluations for each company in Year 10, 15, 20, 25, and 30 of its history.

When you evaluate a company, you will always have access to its performance data spanning **the past 10 years**. For example, this is how the available data in Year 15 may look like:



Your later evaluations of a company will follow the same structure as your first evaluation. For example, in Year 15, we will ask you:

> On average, what is your estimate of the company's **Current PoS% in Year 15**?

Note: The company's CEO **may or may not** change between different years of evaluation (e.g., between Year 10 and Year 15).

Previous Next

Details about how your bonus is calculated

Your goal is to provide an accurate evaluation for each company. Your payment will, therefore, depend on **the accuracy of your evaluations**.

Specifically, at the end of the study, the computer will randomly choose **one** of the evaluations you make during the study to be **the evaluation that counts** towards your bonus. We will use a "PoS% bonus formula" to determine whether you win a **$8** bonus based on the accuracy of your PoS% estimate in **the evaluation that counts**. This formula is explained below.

Explanation of the PoS% bonus formula: We will use a fixed formula to assess your estimate of a company's PoS%. The formula determines your chance of winning a bonus based on how close your estimate is to the truth. The basic idea of the formula is: The closer an estimate is to the truth, the more likely you are to win a bonus. Therefore, it is in your best interest to make all your estimates as accurately as possible.

If you would like to see further details of the PoS% bonus formula, you can click on the following button:

Click here to see the PoS% bonus formula

In short, you can win a bonus up to **$8**. To maximize your bonus, you should try to provide accurate evaluations.

Previous Next

Overview of the procedures for each Widget company

The experiment will consist of 5 rounds. In each round, you will be assigned a new Widget company.

As described previously, each round will consist of the same procedures:

1. You will see the historical performance data of the company from Year 1 to Year 10. You will then estimate its Current PoS% in Year 10.
2. You will see the historical performance data of the company from Year 6 to Year 15. You will then estimate its Current PoS% in Year 15.
3. ... (repeat)
4. You will see the historical performance data of the company from Year 21 to Year 30. You will then estimate its Current PoS% in Year 30.

Previous Next

Comprehension questions

Please answer the following questions to make sure that you understand the experimental instructions.

You can use the navigation bar above to quickly access specific screens of the instructions.

How many different companies will you evaluate?

How many times will you evaluate each company?

How many years of historical data will be available to you when you make each evaluation?

What is the chance that a company's CEO changes in any given year?

In this study, which of the following statements about CEO changes is correct?

CEO change between two evaluations of the same company

Between your two evaluations of the same company (e.g., between Year 25 and Year 30), will there be a change in its CEO?

In this study, which of the following statements about a company's success is correct?

Under a given CEO (i.e., if there is no CEO change), how does a company's probability of success change every year?

What do we mean by the "probability of success" of a company in Year 10?

What should you do to maximize your bonus?

You can use the navigation bar above to access the previous instructions.

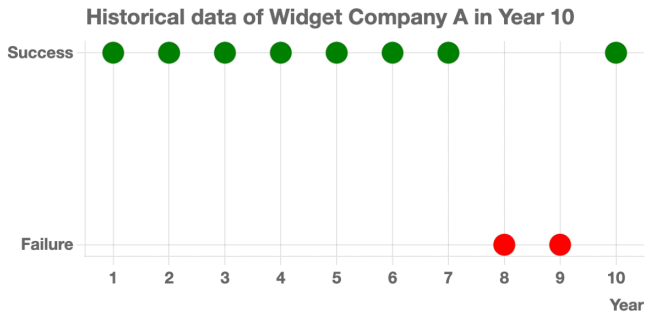
Your 1st evaluation of Company A

This is your **1st** evaluation of **Company A**. You will estimate Company A's Current PoS% in **Year 10**. In other words, this will be the company's PoS% in Year 11 **if** the CEO currently active in Year 10 remains in place in Year 11.

Recall that there is always a **1 in 5 (20%) chance** that the company switches to a new CEO from one year to the next.

Please read the historical data of Company A in the past 10 years:

Historical Data



On average, what is your estimate of Company A's Current PoS% in **Year 10**?

____ %

Figure F1: A typical round (elicitation of mean belief)

Final task

You are approaching the end of the study.

In order to better understand your decisions, we will ask you to evaluate two more Widget companies: Company S and Company T. To save time, we will only ask you to evaluate each of them **once**, that is, in **Year 10** of the company.

Note again that all the Widget companies are independent of each other, but share the basic features (CEO changes, etc.) described at the start of the study. There is always a **1 in 5 (20%) chance** that a company switches to a new CEO from one year to the next.

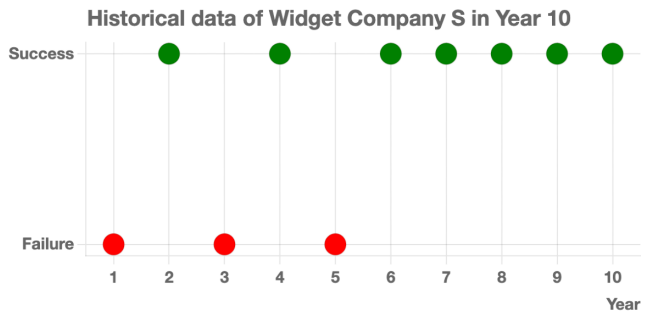
Your evaluations of Company S and Company T will not affect your bonus payment.

Evaluation of Company S

Now, we will ask you to estimate Company S's Current PoS% in **Year 10**. In other words, this will be the company's PoS% in Year 11 **if** the CEO currently active in Year 10 remains in place in Year 11.

Please read the historical data of Company S from Year 1 to Year 10:

Historical Data



On average, what is your estimate of Company S's Current PoS% in **Year 10**?

____ %

Final task

You are approaching the end of the study.

In order to better understand your decisions, we will ask you to evaluate two more Widget companies: Company S and Company T. To save time, we will only ask you to evaluate each of them **once**, that is, in **Year 10** of the company.

Note again that all the Widget companies are independent of each other, but share the basic features (CEO changes, etc.) described at the start of the study. There is always a **1 in 5 (20%) chance** that a company switches to a new CEO from one year to the next.

Your evaluations of Company S and Company T will not affect your bonus payment.

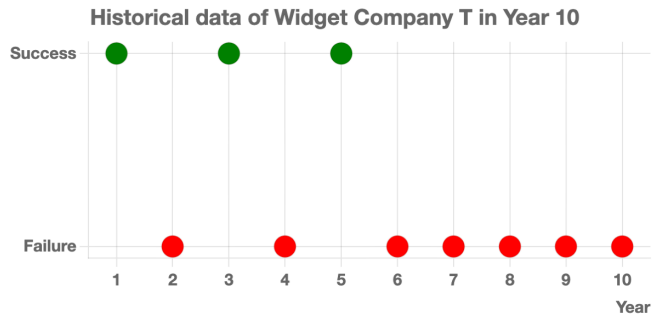
Evaluation of Company T

Now, we will ask you to estimate Company T's Current PoS% in **Year 10**. In other words, this will be the company's PoS% in Year 11 **if** the CEO currently active in Year 10 remains in place in Year 11.

Note that Company T is a different company from Company S.

Please read the historical data of Company T from Year 1 to Year 10:

Historical Data



On average, what is your estimate of Company T's Current PoS% in **Year 10**?

%

F.2 Developer Treatment

Your task: evaluating each company

You will evaluate each Widget company **in two steps**.

- First, you will provide your best guess of **whether and how** the company's probability of success has changed over time.
- Second, you will provide an estimate of the company's **current probability of success.**

> Later, we will just use **"PoS%"** as a short hand for "probability of success." Remember: The PoS% of a CEO can be any number between 0% and 100%, and it determines the PoS% of the company during all the years in which the CEO is in place.

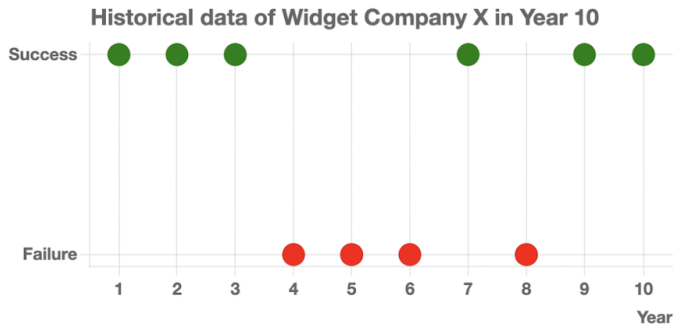
Your bonus will depend on the accuracy of your evaluations (we will provide you with the details on the next page).

---

1. *First step of evaluation*

While evaluating each Widget company, you will have access to its historical performance data spanning **the past 10 years**.

For example, you will first evaluate each company in **Year 10** of its history. You will receive historical data which could look like the following chart.



As the first step of your evaluation, we will ask you:

> What is your best guess on **whether and how** the company's PoS% has changed **from Year 1 to Year 10** (as a result of potential CEO changes)?

As the first step of your evaluation, we will ask you:

> What is your best guess on **whether and how** the company's PoS% has changed **from Year 1 to Year 10** (as a result of potential CEO changes)?

One example of a possible first-step evaluation:

> My best guess is that: The company did not have any CEO change from Year 1 to Year 10. The company had a PoS% of 55% from Year 1 to Year 10.

Another example of a possible first-step evaluation:

> My best guess is that: The company had CEO changes in Year 4 and Year 7. The company had a PoS% of 90% from Year 1 to Year 3, a PoS% of 20% from Year 4 to Year 6, and a PoS% of 75% from Year 7 to Year 10.

---

2. *Second step of evaluation*

As the second step of your evaluation, we will ask you:

> On average, what is your estimate of the company's **Current PoS% in Year 10**?

Recall that the company's Current PoS% in Year 10 is the PoS% of the CEO currently in place in Year 10. In other words, it will be the company's PoS% in Year 11 **if** the CEO currently active in Year 10 remains in place in Year 11.

Example of a possible second-step evaluation:

> On average, my estimate of the company's Current PoS% in Year 10 is 65%.

IMPORTANT: As you make the evaluation, you will **not** be told whether and when the company's CEO has changed in the past 10 years. However, you know that there is always a **1 in 5 (20%) chance** that the CEO changes from any year to the next. Thus, there might have been no, one, or multiple CEO changes during the 10-year period.

Details about how your bonus is calculated

Your goal is to provide an accurate evaluation for each company. Your payment will, therefore, depend on **the accuracy of your evaluations**.

Specifically, at the end of the study, the computer will randomly choose **two** of the evaluations that you made during the study to be **the two evaluations that count** towards your bonus. Let's call them **Evaluation 1** and **Evaluation 2**.

- For **Evaluation 1**, your bonus will depend on **the first step** (i.e., your guess of whether and how the company's PoS% has changed in the past 10 years). Specifically, you will win a **$4** bonus if you correctly identify the exact CEO change years (or the lack thereof) in the data. Moreover, we will randomly pick one of the 10 years, and use a "PoS% bonus formula" to determine whether you win another **$4** bonus based on the accuracy of your guess of the PoS% in that year. This formula is explained below.
- For **Evaluation 2**, your bonus will depend on **the second step** (i.e., your estimate of the company's Current PoS%). We will again use the "PoS% bonus formula" to determine whether you win another **$8** bonus based on how accurate your estimate is.

Explanation of the PoS% bonus formula: We will use a fixed formula to assess your estimate of a company's PoS%. The formula determines your chance of winning a bonus based on how close your estimate is to the truth. The basic idea of the formula is: The closer an estimate is to the truth, the more likely you are to win a bonus. Therefore, it is in your best interest to make all your estimates as accurately as possible.

If you would like to see further details of the PoS% bonus formula, you can click on the following button:

Click here to see the PoS% bonus formula

In short, you can win a bonus up to **$16** in total. To maximize your bonus, you should try to provide accurate evaluations.

Previous Next

Overview of the procedures for each Widget company

The experiment will consist of 5 rounds. In each round, you will be assigned a new Widget company.

As described previously, each round will consist of the same procedures:

1. You will see the historical performance data of the company from Year 1 to Year 10. You will then make your guess of how its PoS% has evolved from Year 1 to Year 10 and estimate its current PoS% in Year 10.
2. You will see the historical performance data of the company from Year 6 to Year 15. You will then make your guess of how its PoS% has evolved from Year 6 to Year 15 and estimate its current PoS% in Year 15.
3. ... (repeat)
4. You will see the historical performance data of the company from Year 21 to Year 30. You will then make your guess of how its PoS% has evolved from Year 21 to Year 30 and estimate its current PoS% in Year 30.

Previous Next

You can use the navigation bar above to access the previous instructions.

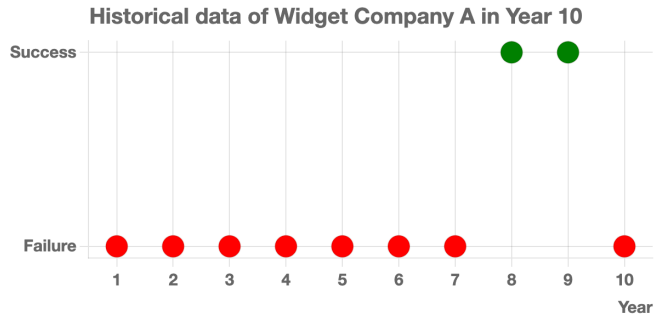
Your 1st evaluation of Company A — the first step

This is your **1st** evaluation of **Company A**. You will assess Company A in **Year 10**.

Recall that there is always a **1 in 5 (20%) chance** that the company switches to a new CEO from one year to the next.

Please read the historical data of Company A in the past 10 years:

Historical Data



What is your best guess on **whether and when** Company A's CEO has changed **from Year 1 to Year 10**?

Please select all the years in which you think there was a CEO change (i.e., a new CEO replaced the CEO from in the previous year).
If your best guess is that there was no CEO change from Year 1 to Year 10, please select the last option.

Year 1 | Year 2 | Year 3 | Year 4 | Year 5 | Year 6 | Year 7 | Year 8 | Year 9 | Year 10

My best guess is that there was no CEO change from Year 1 to Year 10.

You can select or unselect each option at any time.

Based on your guess above, what is your best guess on **how** Company A's PoS% has evolved **from Year 1 to Year 10**?

Please fill in each blank below. Note: As you complete each blank, the chart above will update to show your answers.

The CEO active from Year 1 to Year 10 has a PoS% of _____ %.

Figure F2: A typical round — Step 1 (narrative development)

You can use the navigation bar above to access the previous instructions.

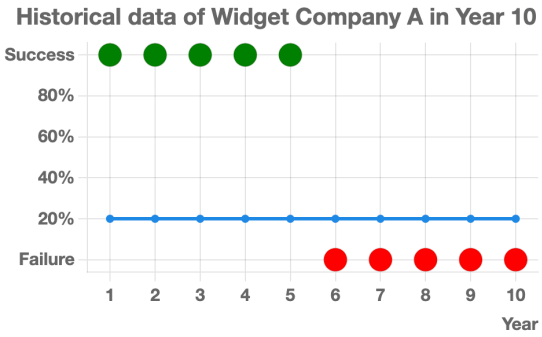
Your 1st evaluation of Company A — the second step

You will now estimate Company A's Current PoS% in **Year 10**. In other words, this will be the company's PoS% in Year 11 **if** the CEO currently active in Year 10 remains in place in Year 11.

Recall that there is always a **1 in 5 (20%) chance** that the company switches to a new CEO from one year to the next.

Please read the historical data of Company A in the past 10 years:

Historical Data



First-step Evaluation

*There was no CEO change from Year 1 to Year 10. The CEO active **from Year 1 to Year 10** has a PoS% of **20%**.*

Your first-step evaluation is displayed as a **blue line** in the chart to the left.

On average, what is your estimate of Company A's Current PoS% in **Year 10**?

____ %

Figure F3: A typical round — Step 2 (elicitation of mean belief)

F.3 Receiver Treatment

Your task: evaluating each company

You will evaluate each Widget company by providing an estimate of the company's **current probability of success.**

> Later, we will just use **"PoS%"** as a short hand for "probability of success." Remember: The PoS% of a CEO can be any number between 0% and 100%, and it determines the PoS% of the company during all the years in which the CEO is in place.

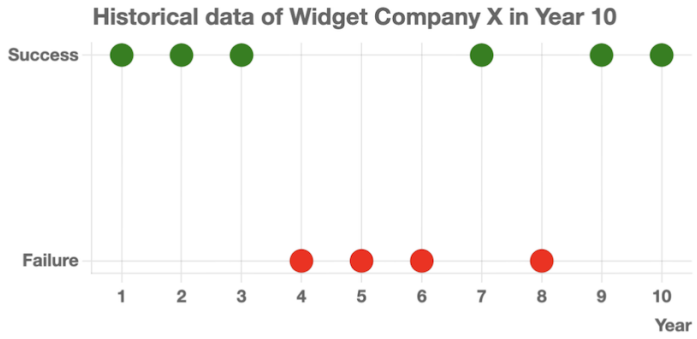
Your bonus will depend on the accuracy of your evaluations (we will provide you with the details on the next page).

---

1. *Making the evaluation*

While evaluating each Widget company, you will have access to its historical performance data spanning **the past 10 years**.

For example, you will first evaluate each company in **Year 10** of its history. You will receive historical data which could look like the following chart.



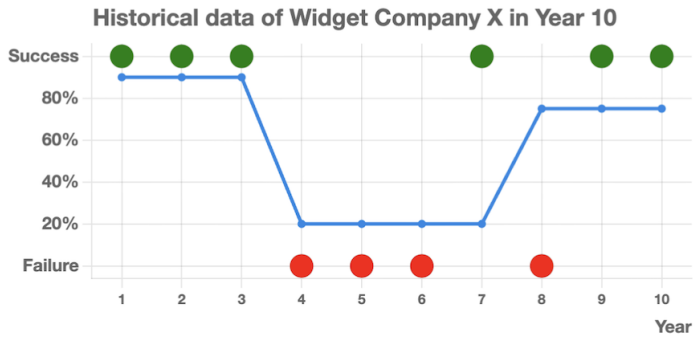
You will receive this dataset together with the evaluation of **a randomly matched participant** based on the same historical data. The evaluation of this participant will consist of:

1. Their best guess of **whether and when** the company's CEO has changed in the past 10 years.
2. (Based on the guess above) their estimate of the **PoS%** of each of the CEOs in the past 10 years.

As an example, suppose the matched participant's best guess is as follows:

> The company had CEO changes in Year 4 and Year 7. The company had a PoS% of 90% from Year 1 to Year 3, a PoS% of 20% from Year 4 to Year 6, and a PoS% of 75% from Year 7 to Year 10.

You would receive this information in the format of the following chart, where the blue line shows the matched participant's evaluation:



Then, as your own evaluation, we will ask you:

> On average, what is your estimate of the company's **Current PoS% in Year 10**?

Recall that the company's Current PoS% in Year 10 is the PoS% of the CEO currently in place in Year 10. In other words, it will be the company's PoS% in Year 11 **if** the CEO currently active in Year 10 remains in place in Year 11.

Example of a possible evaluation:

> On average, my estimate of the company's Current PoS% in Year 10 is **65%**.

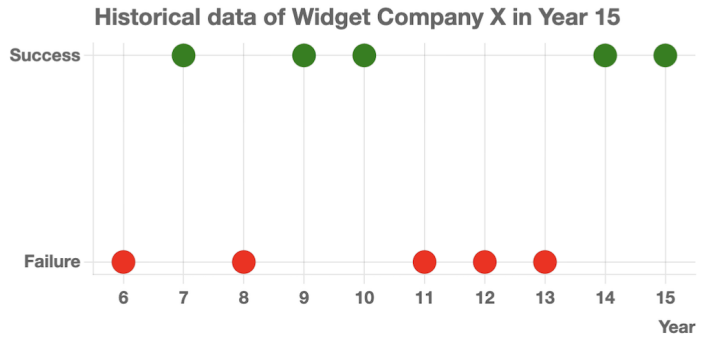
IMPORTANT: No participant is told whether and when the company's CEO has changed in the past 10 years. The matched participant just made their own guess based on the historical data, knowing that there is always a **1 in 5 (20%) chance** that the CEO changes from any year to the next, and that there might have been no, one, or multiple CEO changes during the 10-year period.

---

2. *Making new evaluations over time with new information*

You will evaluate each Widget company at 5 points in time. From one point to the next, 5 years will elapse for the company, and new data of company performance in the elapsed years will become available. This implies that you will make separate evaluations for each company in Year 10, 15, 20, 25, and 30 of its history.

When you evaluate a company, you will always have access to its performance data spanning **the past 10 years**. For example, this is how the available data in Year 15 may look like:



You will again see the evaluation of **another randomly matched participant** based on the same historical data (the evaluation is not shown in the example above).

Your later evaluations of a company will follow the same structure as your first evaluation. For example, in Year 15, we will ask you:

> On average, what is your estimate of the company's **Current PoS% in Year 15**?

Note: The company's CEO **may or may not** change between different years of evaluation (e.g., between Year 10 and Year 15). Also, you will always be matched with a new participant when you make a new evaluation of the same company. That is, the guesses you see at different points for the same company are made by **different participants**.

Previous Next

Details about how your bonus is calculated

Your goal is to provide an accurate evaluation for each company. Your payment will, therefore, depend on **the accuracy of your evaluations**.

Specifically, at the end of the study, the computer will randomly choose **one** of the evaluations you make during the study to be **the evaluation that counts** towards your bonus. We will use a "PoS% bonus formula" to determine whether you win a **$8** bonus based on the accuracy of your PoS% estimate in **the evaluation that counts**. This formula is explained below.

Explanation of the PoS% bonus formula: We will use a fixed formula to assess your estimate of a company's PoS%. The formula determines your chance of winning a bonus based on how close your estimate is to the truth. The basic idea of the formula is: The closer an estimate is to the truth, the more likely you are to win a bonus. Therefore, it is in your best interest to make all your estimates as accurately as possible.

If you would like to see further details of the PoS% bonus formula, you can click on the following button:

Click here to see the PoS% bonus formula

In short, you can win a bonus up to **$8**. To maximize your bonus, you should try to provide accurate evaluations.

Previous Next

Overview of the procedures for each Widget company

The experiment will consist of 5 rounds. In each round, you will be assigned a new Widget company.

As described previously, each round will consist of the same procedures:

1. You will see the historical performance data of the company from Year 1 to Year 10. In addition, you will see a matched participant's best guess of how its PoS% has evolved from Year 1 to Year 10. You will then estimate its Current PoS% in Year 10.
2. You will see the historical performance data of the company from Year 6 to Year 15. In addition, you will see another matched participant's best guess of how its PoS% has evolved from Year 6 to Year 15. You will then estimate its Current PoS% in Year 15.
3. ... (repeat)
4. You will see the historical performance data of the company from Year 21 to Year 30. In addition, you will see another matched participant's best guess of how its PoS% has evolved from Year 21 to Year 30. You will then estimate its Current PoS% in Year 30.

Previous Next

You can use the navigation bar above to access the previous instructions.

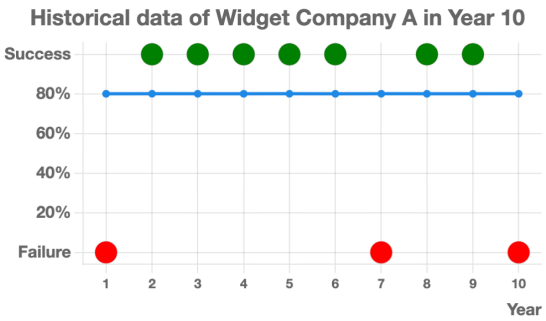
Your 1st evaluation of Company A

This is your **1st** evaluation of **Company A**. You will estimate Company A's Current PoS% in **Year 10**. In other words, this will be the company's PoS% in Year 11 **if** the CEO currently active in Year 10 remains in place in Year 11.

You have been randomly matched with a participant who had exactly the same information as you and made their own evaluation of the company. You are always matched with different participants across sequential evaluations of the same company.

Please read the historical data of Company A in the past 10 years, as well as the evaluation of the matched participant:

Historical Data



Evaluation of the matched participant

*There was no CEO change from Year 1 to Year 10. The CEO from **Year 1 to Year 10** has a **PoS% of 80%**.*

The matched participant's evaluation is displayed as a **blue line** in the chart to the left.

On average, what is your estimate of Company A's Current PoS% in **Year 10**?

%

Figure F4: A typical round (elicitation of mean belief)