# The Transmission of Reliable and Unreliable Information*

Thomas Graeber Shakked Noy Christopher Roth

July 10, 2026

**Abstract**

Information often spreads and influences beliefs regardless of its reliability. We show that this occurs in part because indicators of reliability tend to be lost in the process of word-of-mouth transmission. We conduct controlled experiments where participants listen to economic forecasts and pass them on through voice messages. Other participants listen either to original or transmitted audio recordings and report incentivized beliefs. Across various transmitter incentive schemes, a claim's reliability is lost in transmission much more than the claim itself. Reliable and unreliable information, once filtered through transmission, impact listener beliefs similarly. Mechanism experiments show that reliability is lost not because it is perceived as less relevant or harder to transmit, but because it is less likely to *come to mind* during transmission. Evidence from our experiments, a large corpus of everyday conversations, and economic TV news shows that contextual cues can bring reliability to mind and induce its transmission, but situations in which people share information seldom contain such cues.

*Keywords:* Information Transmission, Word-of-mouth, Reliability, Memory.

* An earlier version of this paper was titled "Lost in Transmission". We thank the editors and four anonymous referees for helpful comments and suggestions. We also thank Ned Augenblick, John Conlon, Stefano DellaVigna, Jeanne Hagenbach, David Hirshleifer, Ryan Oprea, Gautam Rao, Frank Schilbach, Andrew Schotter, Michael Thaler, Sevgi Yuksel and Florian Zimmermann for helpful comments and suggestions. We thank Simon Cordes, Pietro Ducco, Maximilian Fell, Milena Jessen, Frida Kueper, Malte Kornemann, Nicolas Roever, Constantin Schesch, Georg Schneider and Till Stange for outstanding research assistance. All of the data collections were preregistered on the AEA RCT registry (12119, 17479, 17533, 17540). This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (ERC Starting Grant: 101160770 - VIRAL). Roth acknowledges funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2126/2-390838866, the Hans-Kelsen Preis, and the Research Council of Norway through its Centre of Excellence Scheme (FAIR project No 262675). Human subjects approval was given by Harvard IRB and the University of Cologne ethics committee, and an exemption was granted by MIT COUHES. Graeber: University of Zurich, thomas.graeber@gmail.com. Noy: Massachusetts Institute of Technology, snoy@mit.edu. Roth: University of Cologne, ECONtribute, CEPR, and MPI for Behavioral Economics Bonn, roth@wiso.uni-koeln.de.

# 1 Introduction

The extent to which information spreads and influences beliefs is often unrelated to its reliability. Consider three examples. The "4% rule" for retirement savings, a standard piece of financial advice, originated as a descriptive finding in a 1994 analysis of historical data that the author warned were unrepresentative and dangerous to extrapolate from. The author later expressed surprise at how quickly his caveats dropped away as the finding passed from mouth to mouth and solidified into a "rule" (Bengen, 2024). One of the most viral health tips of the 2000s claimed that eating six small meals a day promoted weight loss. The tip was repeated on broadcast television shows, in bestselling books, and throughout the multi-billion-dollar weight loss industry (ABC News, 2005; Gower, 2002). Those repeating the claim rarely said anything about its provenance or evidentiary basis; it appears to have originated either in a 1967 book by a physician who claimed sensationally to have helped 10,000 patients lose weight, or in a 1991 randomized controlled trial with a sample size of 7 (Stillman and Baker, 1967; Tai et al., 1991). A 2014 RAND retrospective about the 2002 National Intelligence Estimate on weapons of mass destruction in Iraq concluded that the report "contained several qualifiers that were dropped" when it was transformed into an executive summary, and that "as the draft NIE went up the intelligence chain of command, the conclusions were treated increasingly definitively" (Gompert et al., 2014).

Economic theory predicts that more reliable information should influence beliefs more. The preceding anecdotes, however, suggest that unreliable claims frequently influence behavior too much, and reliable claims too little, in part because the claims themselves get passed from person to person while information about their credibility or evidentiary basis is lost in transmission.

This paper studies whether, and why, this is true. Borrowing standard terminology, we distinguish between the realization of a signal (the "level" of a claim) and the precision of that signal (the "reliability" of the claim). This distinction gives us a natural benchmark: we ask whether the precision of signals (reliability) tends to be lost in word-of-mouth transmission *more* than the signal realizations (level). We then ask whether the information people transmit is influenced by cost-benefit calculations about the payoffs to transmitting different kinds of information, or by cognitive constraints such as selective attention or memory. To study these questions, we use controlled online experiments and observational data from everyday conversations and TV news.

**Design.** Our experiments involve more than 9,000 participants. In a *transmitter* experiment, participants listen to a one-minute message giving a qualitative forecast about an economic variable, and are incentivized to record themselves faithfully passing on the information they heard. In a subsequent *listener* experiment, participants listen to either the original forecasts or transmitted versions of those forecasts before stating incentivized beliefs.

We base the original forecasts on actual news coverage, and cross-randomize their level and reliability. We vary the *level*—high or low—by switching whether the original message argues for an increase or a decrease in the relevant variable. We vary the *reliability*—reliable or unreliable—using manipulations that weave certainty- or uncertainty-denoting words into an otherwise-identical text, and manipulations that change multiple implicit and explicit signals

of reliability, including the speaker's confidence, credentials, stated sources of evidence, fluency, and vocabulary. Both level and reliability are communicated in qualitative terms only, mimicking how people naturally communicate. Transmitters are incentivized based on how close the belief updates induced by their voice messages are to the average belief updates induced by the original messages. Such incentives motivate a faithful transmission of information, which is ubiquitous in the real world: sales employees relay customer feedback to developer teams, analysts brief executives, and friends share financial advice sourced from media.

**Main results.** Our main finding is that information about the reliability of a prediction is lost in transmission more than twice as much as information about the prediction's level. We refer to this finding as *differential information loss,* and document it using three sets of analyses.

In our first analyses, we examine the transcripts of transmitted messages. Nearly all of the messages include some statement about the level of the original prediction, but only a third mention the original prediction's reliability or include other markers of reliability, such as uncertainty prefixes. Transmitted messages, containing an average of 114 words (8-10 sentences), tend to be only half as long as the original messages. Yet even the longest decile of transmitted messages, which are about as long as the original messages, mention reliability less than 30% of the time. Many messages go on at length, and in great detail, about the level of the original forecast without mentioning its reliability.

What ultimately matters is the preservation of the information content of the original messages, which our transcript analysis does not fully capture: the original information could be preserved using fewer or different words. We therefore analyze two sets of beliefs: listeners' beliefs about the level and reliability of the original messages ("message beliefs") and their belief updates about the underlying economic state discussed in the recordings ("state belief updates").

We begin with message beliefs. Among listeners who directly hear the original messages, a shift from a low-level to a high-level message changes beliefs about the prediction's level by 1.37 standard deviations (SDs). Among listeners who hear transmitted versions of those messages, beliefs shift by only 0.88 SDs. This indicates a $100\times[(1.37-0.88)/1.37] \approx 36\%$ loss of sensitivity to variation in the level of the original prediction. By contrast, loss of reliability information is more than two times as large. Among listeners who hear the original messages, switching from a weak-reliability to a strong-reliability message shifts beliefs about the message's reliability by 1.18 SDs. The corresponding shift for listeners who hear transmitted recordings is 0.12 SDs, meaning 90% of the variation in information about a message's reliability is lost in transmission.

In both cases, the loss of sensitivity to our manipulations is driven by a *symmetric* compression of beliefs towards an intermediate value. After transmission, high forecasts are perceived as predicting a less high level, and low forecasts as predicting a less low level. Similarly, reliable messages are perceived as less reliable, but unreliable messages are perceived as more reliable.

We then turn to state belief updates, that is, listeners' belief updates about the economic variables discussed in the recordings. State belief updates are the object of ultimate economic relevance and the object on which transmitters are incentivized. Listeners who directly hear the original messages update their beliefs in a qualitatively Bayesian way: they update in the di-

rection of the message's prediction, and those who hear strong-reliability versions of a message update twice as strongly, on average, as those who hear weak-reliability versions. By contrast, listeners who hear transmitted versions of the messages update about the same amount on average from weak-reliability and strong-reliability messages—the distinction between weak- and strong-reliability messages is almost completely lost in transmission. Level and reliability information loss interact: for highly reliable original messages, loss of level and reliability information push in the same direction, strongly attenuating belief updates towards zero. By contrast, for unreliable original messages, the two types of loss counteract each other, resulting in smaller effects on belief updates. This suggests that the primary consequence of the phenomenon we document is to drag down the influence of reliable information relative to unreliable information.

**Mechanisms.** We next ask what drives the differential information loss we document. On the one hand, reliability information could be lost more than level information as the result of a deliberate tradeoff, either because the perceived benefits of transmitting reliability information are lower or the perceived cognitive costs are higher. On the other hand, differential loss could result from some non-deliberate mechanism. For example, reliability information might not *come to mind* at the moment of recording the voice message. In a series of mechanism experiments, we reject the first two explanations and find support for the third.

We begin by studying beliefs about the benefits of passing on level versus reliability information. When asked, transmitters in our main experiment on average deem both equally important. We also conduct an additional experiment that explicitly and equally incentivizes transmitters to pass on level and reliability information, effectively holding beliefs about the relative benefits of transmitting the two dimensions constant. Even under these conservative incentives, we find pronounced differential information loss (36% for level and 65% for reliability).

Next, we ask whether the perceived cognitive costs of transmitting reliability information are higher. We conduct an experiment where transmitters are allowed to decide whether their bonus payment will depend on their transmission of level information or reliability information. *Ex ante*, a majority choose to be incentivized based on their transmission of *reliability* information and expect it to be easier to communicate, a view that does not change after having experienced the task, suggesting that higher perceived costs of transmitting reliability cannot explain our results.

Finally, we extend our analysis of mechanisms beyond perceived benefits and costs to examine the possibility that reliability information may simply fail to come to mind during the cognitively taxing task of information transmission. Our previous results hint at this possibility: more than 60% of transmitters do not mention reliability information at all in their messages, even when *ex post* remembering this information, saying that it is equally important as level information, and believing it is easier to transmit.

We focus on the hypothesis that reliability is less likely to come to mind because contexts prompting transmission typically do not contain "cues" associated with reliability (drawing on standard models of cue-based retrieval of memories; Bordalo et al., 2025b; Kahana, 2012). Several pieces of evidence show that reliability is much more likely to be passed on when brought to the transmitter's attention via contextual cues. This is true for a range of cue types, both for

cues randomly varied in our experiments and, more suggestively, for naturally occurring cues in observational field data. It is true for indirect, payoff-irrelevant cues of reliability incidentally varied in our baseline experiments. It is true for explicit and highly salient cues in a supplementary experiment where we replicate our baseline design but confront transmitters with a prominent reminder to transmit reliability on the screen where they complete their recordings. Finally, it is true for naturally occurring cues in everyday language, where requests for information that contain reliability cues are much more likely to be succeeded by answers containing reliability information; and on economic TV news coverage, where discussion of economic uncertainty tracks a benchmark measure of true uncertainty more closely during periods with more extraneous reliability cues. Despite the importance of cues, we show using a large dataset of spoken conversations that requests for information virtually never explicitly mention its reliability, suggesting that generic requests for information transmission may not typically be associated with (and hence cue) reliability. Cues of reliability are not, however, costless: in our supplementary experiment that reminds transmitters to pass on reliability, reliability loss falls sharply, but the loss of level information rises significantly—indeed, level information is now lost at a higher rate than reliability—indicating that the two compete for shared cognitive resources.

In summary, our mechanism analyses suggest that failures to pass on reliability information do not stem from transmitters representing the transmission problem correctly but consciously choosing to place lower weight on the transmission of reliability information; rather, reliability simply tends not to come to mind, because the situations that prompt people to pass on information rarely draw attention to it. In the language of Bordalo et al. (2025b), reliability is missing from the default way transmitters represent the transmission problem.

**Related literature.** Our results show that reliable and unreliable information, once filtered through transmission, become hard to distinguish. This effect of transmission may operate alongside and compound a distinct *updating* bias: Griffin and Tversky (1992), Massey and Wu (2005), and Augenblick et al. (2025), among others, argue that even conditional on knowing the precision or diagnosticity of a signal, people overinfer from unreliable signals and underinfer from reliable ones.[1] In our experiments, the effect of such an updating bias is held constant across listeners to original and transmitted messages by design, allowing us to identify the distinct effect of transmission. Jointly, transmission-induced loss of reliability indicators and people's insufficient sensitivity to received reliability indicators may reduce the influence of reliable and trustworthy information relative to unreliable information. The transmission loss we document also has a distinct consequence. While individual belief updating biases are privately costly, they do not necessarily affect others if they do not influence the way signal realizations are repeated to peers. By contrast, failures to pass on reliability information impose externalities, including on sophisticated peers who *would* respond appropriately to reliability information if it reached them. We demonstrate these consequences in simulation exercises calibrated to match results from our experiments, which we also use to explore the consequences of transmission-induced

[1]Note that signal "strength" in the model of Griffin and Tversky (1992) is related to level information in our framework, and signal "weight" to reliability information.

information loss for patterns of belief convergence and polarization.

Our focus on the transmission of qualitative stories about economic variables relates to a growing literature on the diffusion of narratives (Hirshleifer, 2020; Shiller, 2017). The loss of reliability we document in oral transmission may be part of a broader hypothesized phenomenon (e.g., Hirshleifer, 2020): as stories get told and re-told, they are simplified in the specific sense that nuance is lost. Recent contributions in this literature have focused on the role of narratives for belief formation (Andre et al., 2025; Barron and Fries, 2024; Graeber et al., 2024a,b; Kendall and Charles, 2025). Other papers that focus on the incentives governing the supply of information include Thaler (2025), Thaler et al. (2025), and Serra-Garcia (2025).

We also relate to a literature on how belief formation is shaped by selective attention (Ba et al., 2024; Graeber, 2023; Hartzmark et al., 2021), memory (Bordalo et al., 2023, 2021, 2025b), and cues (Bordalo et al., 2025a, 2026b, 2025b; Conlon and Kwon, 2025; Gennaioli and Shleifer, 2010). We differ from this literature in our focus on how selective memory and cues affect the *supply* of information. Our findings are consistent with work in psychology demonstrating that reliability is rarely scrutinized during habitual information processing, with "epistemic vigilance" being triggered only by specific cues such as inconsistency or explicit uncertainty (Kahneman, 2011; Sperber et al., 2010). They also connect to Stanovich (2018) and others' framework of "miserly" processing, according to which people substitute easy answers for hard ones to save cognitive effort. Overcoming these cognitive shortcuts requires two things: recognizing when the easy answer is insufficient and the problem demands more careful thought, and having the mental tools to produce a more sophisticated answer in that case. In our case, the bottleneck appears to lie in the former step: reliability information is lost not because transmitters lack the capacity to incorporate it, but because it fails to come to mind unless explicitly cued.

Our paper builds on a large literature on social learning (Banerjee, 1992; Bikhchandani et al., 1992; Bursztyn et al., 2014; Conlon et al., 2025; Golub and Sadler, 2016; Golub and Jackson, 2010; Mobius and Rosenblat, 2014; Weizsäcker, 2010), information diffusion (Akçay and Hirshleifer, forthcoming; Banerjee et al., 2013, 2019; Han et al., forthcoming), face-to-face interactions (Atkin et al., 2022; Battiston et al., 2021), and verbalization (Batista et al., 2024). We differ from this literature in our focus on the transmission of qualitative information in the form of spoken narratives and the investigation of cognitive mechanisms that shape transmission of different kinds of information.[2]

# 2 Baseline Experimental Design

Our baseline design comprises two experiments. In the transmitter experiment, respondents listen to a recording and are incentivized to pass on the information contained in the recording. In the listener experiment, different respondents listen to either the original recordings or trans-

[2]Transmission experiments are also studied outside economics (Carlson, 2018, 2019; Mesoudi and Whiten, 2008), reaching at least back to Bartlett's 1932 studies on *serial reproduction* of stories from memory (Bartlett, 1995).

mitted versions before forming their beliefs.[3]

Our baseline study design is guided by the following objectives: (i) an experimental setting in which we can quantify the transmission rates of different kinds of information in natural-language spoken messages, (ii) well-defined incentives for transmission, (iii) systematic variation in different types of information in the original recordings, and (iv) an incentive-compatible belief elicitation in the listener experiment to quantify information loss due to transmission.

## 2.1 Transmitter Experiment

**Structure of the experiment.** In the transmitter experiment, respondents listen to one recording containing two separate opinions about two economic variables, in a random order: home price growth in an anonymous U.S. city and revenue growth of an anonymous U.S. retail company. The city and retailer are New York City and Walmart, respectively, which is not revealed to participants so that they lack strong priors and cannot search for additional information. This ensures that belief formation is, as much as possible, based only on the information we provide in the original recordings. The opinions are written and recorded by us; respondents are informed that these opinions are based on real media commentary on these topics, and are told at the end of the survey that other participants heard recordings arguing for the opposite conclusions. The recording containing both opinions lasts for 2-3 minutes, with each opinion lasting 1-1.5 minutes.[4] Respondents are then asked to separately record their own verbalizations of the two opinions they listened to, and finally answer several belief questions about each topic. Appendix Figure A1 shows the structure of the transmitter survey.

**Speech recordings.** We collect audio recordings, which have several advantages over written text for our purposes. First, oral information transmission is natural: it is the dominant form of communication in daily life, and an important source of information through conversations as well as consumption of television, radio, or podcasts. Second, unlike written communication, the spontaneity of oral communication provides a testing ground for analyzing how cognitive constraints affect information transmission and social learning. A vast literature has examined differences between written and spoken text production (e.g., Akinnaso, 1982; Berger and Iyengar, 2013; Chafe and Tannen, 1987). Written text tends to be more formal, structured, pre-meditated, and involves higher cognitive effort (e.g., Bourdin and Fayol, 2002). Third, speech data allow us to capture critical features of natural language that are mostly absent from written texts and may be essential to the communication of reliability, including tone, emphasis, and disfluencies such as pauses, repetitions, revisions, hesitations, or filler words.

**Transmitter incentives.** The design of our baseline transmitter incentives directly follows our conceptualization of a message's information content as *the average belief movement induced*

[3]The full set of experimental instructions for all experiments can be found at the following link: `https://raw.githubusercontent.com/cproth/papers/master/LiT_instructions.pdf`.

[4]We provide people with the two forecasts consecutively in the same recording, rather than separately playing each forecast before the respondent records their verbalization of it, because this mimics an aspect of transmission in the real world: people are, over time, exposed to multiple pieces of information on various topics, before eventually relaying some information to others.

*by that message.* For each topic, transmitters are tasked with recording a message that induces belief changes that are as close as possible to the average belief changes induced by the original message they listened to. Specifically, one in ten transmitters is selected to be eligible for a $20 bonus payment. Their probability of receiving the payment (conditional on eligibility) is a quadratic function of the distance between the average belief change induced by their message and the average belief change induced by the original message, among two sets of listeners who will hear either their message or the original message. We explain to respondents that in order to maximize their chances of receiving the bonus, they should pass on anything from the original message that they think would be relevant for how people change their beliefs.

This incentive scheme is motivated by our conceptualization of information content and is thus the natural starting point for our experiments. However, there are many alternative possible schemes, some of which may seem less complicated and/or more explicit. Three remarks are in order. First, the optimal strategy under this incentive scheme is to perfectly replicate the original message. Insofar as cognitive constraints prevent this, the optimal strategy is to equate the marginal returns to transmitting level and reliability information by balancing the relative marginal costs of transmitting each type against the relative marginal impact of each type on listeners' belief updates. We study the actual impacts of each type of information on belief updates in Section 3 and transmitters' beliefs about the costs and benefits of transmitting each type in Section 4. Transmitters' beliefs may be biased, which would be a source of transmission distortions that we would want to capture.

Second, incentives based on listeners' *belief changes* (rather than *posteriors*) incentivize transmission of all relevant pieces of information in the original message. If transmitters were incentivized by the accuracy of listeners' posteriors, the optimal strategy might be to "do the updating for the listener:" form a Bayesian posterior after listening to the original recording and simply report this quantitative posterior in the transmitted message. Because transmitters do not know listeners' priors or how their beliefs might react to different pieces of information, incentives based on belief changes encourage them to pass on all information in the original message.[5] We consider this a naturalistic feature of our scheme: in practice, people most often transmit information without knowing which aspects of the original information the audience wants to learn about and what their priors are, motivating transmission of the substantive information content.

Third, although the quantitative formula underlying the incentive scheme is complicated, we explain the scheme in intuitive terms ("you should pass on all information you think is relevant to how people change their beliefs"). To ensure high levels of understanding, only participants who pass a comprehension question on transmitter incentives are allowed to take part in our study. In Section 4, we explore alternative transmission incentive schemes.

**Structure of original recordings.** The original recordings have the following general structure. First, they introduce the variable of interest, i.e., home price growth or revenue growth

[5]Even under our incentive scheme, rational transmitters might, instead of passing on the original information, communicate the degree of belief movement they think should occur given their assumed distribution of prior beliefs, updating rules, etc. In practice, this behavior appears negligible: we observe no transmitter recordings suggesting attempts to convey predicted belief movements.

of a retailer. They then put forward some arguments justifying why the variable of interest will increase or decrease. For example, the speaker mentions that as consumers' disposable incomes decrease due to inflation, they often switch towards lower-price retailers, such as the U.S. retailer in question; or that issuance of new residential construction permits in the U.S. city being discussed has slowed down recently, meaning housing supply will increasingly fall behind growing demand. Towards the end of the message, the speaker states explicitly whether they believe the variable will increase or decrease over the coming year. Throughout the recording, the reliability of the prediction is explicitly or implicitly communicated using techniques we discuss below. Full transcripts of the messages and the audio recordings are available in Appendix F.

The design of these messages is motivated by the nature of real-world commentary on economic topics such as house price or company revenue growth. Such commentary usually justifies predictions with substantive arguments about the variables of interest, e.g., relating to market conditions or broader trends in the economy. The arguments in our messages are drawn from real media reporting on these topics. Moreover, such messages communicate reliability with both explicit and implicit markers.

**Experimental variation: original recording contents.** The design of our original recordings is guided by our distinction between the *level* and *reliability* of a prediction about a variable. We make the following observations about this distinction. First, this distinction is parsimonious, theoretically appealing, and general. To perform a belief update from any piece of information, a Bayesian agent always requires both a signal value and a signal precision. Moreover, level and reliability are always—implicitly or explicitly—conveyed by any forecast. For example, even the absence of explicit confidence or reliability statements could itself be an indicator of the forecast's reliability. Second, our distinction connects with previous belief formation research: for example, some research suggests that people pay insufficient attention to the *weight* or precision of evidence when forming their beliefs in abstract and quantitative updating tasks (Griffin and Tversky, 1992; Massey and Wu, 2005). Third, note that our taxonomy is different from the distinction between the first and second moment of the forecast state. Specifically, reliability is an attribute of a signal structure rather than a property of the distribution of the forecast state.

To leverage the level-reliability distinction in our experiments, we randomize these two features of the original message recordings: whether the message argues for an increase or a decrease in the level of the variable and whether the message is reliable or unreliable.

We randomly assign respondents to two kinds of reliability manipulations. Respondents in the *naturalistic condition* hear recordings that vary reliability using a combination of explicit statements about confidence, evidence quality, and speaker competence, as well as implicit markers of reliability such as verbal fluency and vocabulary. For example, a high-reliability message sounds highly fluent with a sophisticated vocabulary, cites respectable sources of evidence, and mentions relevant credentials. A low-reliability message is full of disfluencies, expresses low confidence, cites obviously unreliable sources, and admits a lack of relevant credentials.

Meanwhile, respondents in the *modular condition* receive recordings that are identical except for a set of explicit markers indicating either high or low reliability (e.g., "definitely" vs.

"possibly", "will" vs. "might", etc.) and explicit confidence statements ("I am highly confident" vs. "I am not at all confident"). Respondents in the *modular condition* are assigned to one of the following three conditions: (i) Strong reliability; (ii) Weak reliability; and (iii) Neutral reliability (where the markers and confidence statements are simply omitted).[6]

These two types of manipulations serve different purposes: the *naturalistic condition* embraces the full range of linguistic tools through which reliability of a statement may be expressed in practice, at the cost of a loss of control about which precise component drives perceptions of reliability. The *modular condition*, by contrast, provides this control by allowing us to trace the loss of specific reliability words or phrases, at the cost of focusing attention on just these modular elements. While interpretation of verbal uncertainty prefixes can vary from person to person (Vogel et al., 2022), this variability is constant between listeners to original and transmitted messages, meaning tracking the loss of these words should suffice to decompose information loss. Because both manipulations end up producing very similar results, we report our main results pooling both conditions, and show disaggregated results in Appendix Figure A7.

Our reliability manipulations most closely approximate real-world situations where a person is learning from a stranger, about whose reliability they have no strong prior. In these cases, people infer a speaker's reliability from the way the speaker talks, the claims the speaker makes, and what the speaker says about their background. All of the participants in our experiment are strangers to each other and must infer reliability only from the contents of voice recordings. Situations like this abound in everyday life, in contexts such as social media, television, conferences, public venues, social gatherings or professional settings.

Finally, we randomize whether the recording is made by a male or female actor. This is not a focus of analysis and we randomize simply for symmetry, and so that each topic a transmitter listens to is discussed by a different voice. We find no evidence that the effects of any of our manipulations, or the effects of transmission, vary with the original voice's gender.

The different margins of randomization are stratified: each transmitter hears two recordings, one with an "increase" and one with a "decrease," one with "strong reliability" and one with "weak reliability," and one with a male voice and one with a female voice.[7]

**Beliefs.** After recording themselves, transmitters answer the same beliefs questions that listeners do, so we defer discussion of those questions to the following subsection.

## 2.2 Listener Experiment

**Structure and treatments.** This experiment draws on the speech recordings collected in the transmitter experiment. It lets us quantify transmission-induced information distortions by measuring and comparing the information content of the original messages and transmitted versions of those messages.

[6]As pre-specified, our main analysis focuses on comparisons between weak and strong reliability for simplicity. Appendix Figure A4 shows belief updates including the neutral-reliability condition.

[7]Then, if exactly one of the two topics is in the modular condition, that topic has a 33% chance of getting switched to "neutral reliability". If both topics are in the modular condition, there is a 66% chance that one of the two topics is randomly switched to "neutral reliability."

Recall that our experiments involve forecasts about two topics: (i) the change in home price growth in a U.S. city and (ii) the change in revenue growth of a U.S. retailer, both for the upcoming year. For each of the two topics, participants in the listener survey first state their prior belief about the outcome variable of interest and then listen to a recording about the variable before answering a set of beliefs questions. The order of the topics is randomized. For each topic, respondents are randomly matched to a transmitter and listen either to the same original recording as the transmitter heard, or that transmitter's message. We implement a 30% chance of hearing the original and a 70% chance of hearing a transmitted recording. We oversample transmitted recordings as they are by construction more heterogeneous compared to original recordings. Appendix Figure A2 shows the survey structure.

Listeners are told whether they are listening to the original message or another participant's transmitted version. They could take this information into account when updating their beliefs about the message content, e.g., by discounting the reliability of *any* transmitted message relative to a corresponding original message. However, as discussed in Section 3, we find no evidence that transmission on average affects the perceived level or reliability of the original messages.

**Beliefs.** After listening to a recording, respondents are incentivized to guess the realization of the target variable—change in house price growth or change in revenue growth over the next 12 months—as well as the level and reliability of the prediction in the original message.

We separately elicit beliefs about the state of the variable under discussion, referred to as *state beliefs* henceforth, as well as beliefs about the original message's contents, called *message beliefs*, for two reasons. A listener's state beliefs are the most economically relevant objects. However, belief movements about the state are also affected by respondents' priors and prior confidence, making it difficult to back out respondents' perceptions of the level and reliability of the original prediction. Directly eliciting beliefs about the message's level and reliability circumvents this issue and brings us closer to the objects of interest in our guiding distinction and our treatment manipulations. Moreover, belief updates about the state are *simultaneously determined* by a message's level and reliability. This means that loss of level information affects respondents' sensitivity to reliability information and vice versa, preventing us from cleanly distinguishing level and reliability information loss based solely on state belief updates. The same is not true for message beliefs, which separate out the original message's level and reliability.

For each topic, we hence elicit three outcome variables: state belief movements (posteriors about the economic variable minus priors); and two message beliefs: the respondent's belief about the level of the original message's prediction and the respondent's belief about its reliability. We elicit respondents' priors about the state, and use belief movements as our outcome rather than posteriors. This reduces noise resulting from idiosyncratic respondent heterogeneity in usage of the belief scales. Because respondents do not know anything about the context the message concerns, their priors simply capture tendencies in their response behavior.

To measure respondents' state beliefs we ask them about the change of the variables of interest in the next 12 months. For home price growth, this question reads as:

> How will house price growth in this city change over the next 12 months?

Our two unknown states are *changes in growth rates* because this permits a natural prior of zero and reasonably symmetric possibilities around that prior. This lets us shift beliefs symmetrically up or down with our high- or low-level messages, creating clean variation in the information content of the recordings. To elicit respondents' corresponding message beliefs about the level of the prediction, we ask the following question:

> How do you think the person [whose opinion you just heard/whose opinion was summarized in the recording] predicts house price growth in this city will change over the next 12 months?

To measure respondents' message beliefs about the reliability of the prediction, we ask:

> How reliable do you think the prediction given by the person [whose opinion you just heard/whose opinion was summarized in the recording] is? Specifically, what do you think is the probability that this person's forecasts about changes in house price growth in this city are roughly correct? Concretely, assuming that the true change in house price growth is a number called X, what do you think is the likelihood that this person's prediction will fall within 1% of X, i.e., between X-1% and X+1%?

**Incentives for accuracy.** Respondents are told that one in ten respondents will be randomly chosen to be eligible for a $20 bonus payment, which will be based on one of the incentivized items in the survey. State beliefs are always directly incentivized based on the true development of the variable over the next year.[8] Message beliefs are unincentivized for a randomly selected 50% of respondents. For the other half of respondents, the question is phrased as a second-order question ("your job is to predict what people who heard the same recording as you would on average respond to the direct question") and responses are incentivized based on the accuracy of their guess about other participants' average guess.[9] Results based on incentivized versus unincentivized message beliefs are virtually identical, as shown in Appendix Figure A8.

## 2.3 Sample and Procedures

We conducted our experiments on Prolific, a widely used online platform to conduct social science experiments (Eyal et al., 2021). The transmitter experiment and listener experiment were run with 540 and 1,510 U.S. respondents, respectively, in November 2023. Table A4 records summary statistics for all our experimental samples. All of the data collections were preregistered on the AEA RCT registry (#12119, #17479, #17533, #17540). As preregistered, we drop recordings below the 5th percentile of recording length or transcript word length (as a proxy for empty or

[8] State beliefs are incentivized with the following formula: Probability of winning $20 [in %] $= 100 - 10(\text{Estimate [in \%]} - \text{True state of the world in 12 months [in \%]})^2$.

[9] Responses are incentivized with the following formula for beliefs about the originator's prediction and reliability, respectively: Probability of winning $20 [in %] $= 100 - \alpha(\text{Response [in \%]} - \text{Average response to direct question [in \%]})^2$, where $\alpha = 10$ for level and $\alpha = 2$ for reliability. This approach allows us to incentivize these beliefs in the absence of a "true state", since the original recordings were provided by us and there is no corresponding originator belief.

contentless recordings). Following this restriction, our baseline transmitter experiment yields a total of 1,010 valid speech recordings.

# 3 What Is Lost in Transmission?

Our main finding in this paper is *differential information loss*: information about the reliability of a forecast is lost in transmission much more strongly than information about its level. In this section, we demonstrate differential information loss in three distinct and complementary ways: first, by examining the transcripts of transmitted messages for mentions of level and reliability; second, by analyzing listeners' *message beliefs* about the level and reliability of the original forecast, separately for listeners who hear original messages versus listeners who hear transmitted versions; and third, by analyzing listeners' *state beliefs* about the economic variables discussed in the original forecasts. Each set of analyses has its own advantages and drawbacks.

## 3.1 Transcript Analysis

We begin by examining the transcripts of transmitted messages to see whether they contain statements about the level or reliability of the original forecasts.

Panel (a) of Figure 1 displays the share of transmitter transcripts classified as containing statements about the level of the original prediction or about its reliability. For reliability, we adopt a maximally broad notion of what counts as communicating reliability, incorporating all of the components we use to vary reliability in the original recordings. This includes explicit statements about reliability or confidence as well as the use of certainty or uncertainty markers like "might" or "definitely." We separately show results of human coding and of automated coding using the large language model GPT-4 (Asirvatham et al., 2026). The figure illustrates that the different coders and the large language model come to similar conclusions.[10] See Appendix Table A2 for some examples of transmitted messages and their hand-codings.

The key finding of Panel (a) is that while most transmitted scripts contain statements about the level (between 87 and 95 percent), a far smaller fraction of transmitted scripts contain statements indicating the reliability of the original message (between 30 and 45 percent). Panel (b) shows that this is true independent of the length of the transmitted message: even among transmitted messages that are 200-300 words long (longer than the original messages), only 20% are unanimously agreed upon by our coders to contain statements about reliability. Longer messages tend to differ from shorter ones primarily in providing a much higher level of detail about the original message's arguments for its level prediction. Appendix Figure A10 also shows that the fraction of scripts containing statements about level or reliability is fairly stable across our four level × reliability conditions (high versus low level, and weak versus strong reliability).

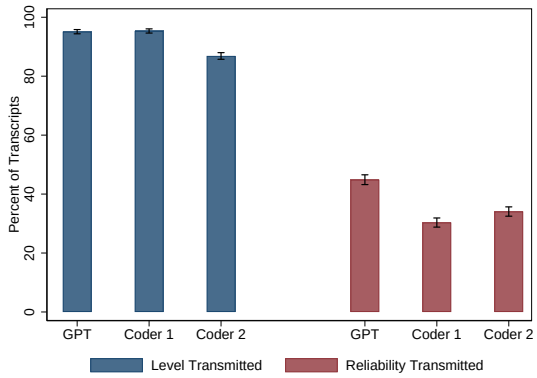
Transmitted recordings also include many disfluencies—hesitations, "um" statements, self-corrections, and so on—which could influence how listeners perceive the reliability of the orig-

[10]For level, if one human coder identifies level as being passed on, the other does with 91% probability, and GPT does with 98% probability. For reliability, the corresponding numbers are 60% and 75%. In our analysis of beliefs data where we split according to handcoded classifications, we restrict to transcripts where our coders agree unanimously.

inal forecast. However, Appendix Figure A11 shows that the average number of disfluencies in transmitted scripts does not vary with whether the original forecast was high- or low-reliability, and the presence of disfluencies in transmitted scripts does not affect listener beliefs about the reliability of the original forecast (analyzed in the next section). The reliability of the original forecasts does not appear to be effectively communicated through disfluencies.

**Result 1.** *Transmitted messages are more than two times as likely to mention the level of the original prediction than to mention its reliability.*

**a)** Percent of Scripts Containing Statements about Level and Reliability




**b)** Percent Unanimously Containing Level or Reliability, by Script Length

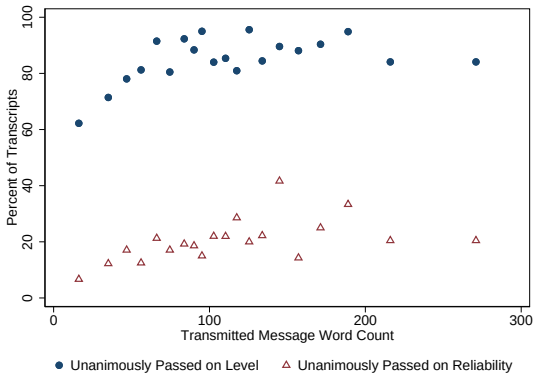



Figure 1: This figure presents data from our baseline experiment (Belief Movement Incentives). Panel (a) shows the percent of transcripts that are coded as conveying any information about the level and reliability of the original forecast, separately by two human coders and GPT-4, with standard error bars. Panel (b) shows binned scatterplots of the percentage of scripts unanimously classified by our coders and GPT as containing statements about level or reliability, respectively, against the word count of the script. $N = 540$ transmitters, 1,010 transcripts.

These results, while transparent and strongly suggestive, do not conclusively establish that reliability information is lost in transmission more than level information. For example, perhaps listeners only need to hear 10% of the reliability cues in the original messages to grasp their true reliability, so that 90% of the original cues can be dropped without loss; or perhaps reliability is passed on in a way not picked up by our codings. To address this possibility, we turn to an analysis of listeners' *message beliefs*, their beliefs about the level and reliability of the original messages. We will confirm our finding of differential reliability loss, but also find that our binary hand-codings from this section do not tell the full story: there is substantial loss of reliability information even among messages we coded in this section as passing on at least some reliability.

## 3.2 Message Beliefs

To provide independent measures of level and reliability information, we separately elicit listeners' message beliefs about the level and reliability of the original prediction, using questions described in Section 2.2. Figure 2 presents results on message beliefs. We z-score message beliefs

within each topic × reliability manipulation quadrant to make the aggregation across experimental conditions more comparable; results with raw beliefs are available in Appendix Figure A3.

Panel (a) examines message beliefs about the level of the original prediction. The blue dots show the average beliefs of listeners who directly hear original recordings. Listeners who hear a *low*-level original recording believe the level of the prediction is 1.37 SDs lower on average than listeners who hear a *high*-level original recording. Meanwhile, the orange dots show the beliefs of listeners who hear transmitted versions of the original recordings. Here, the difference between the beliefs of listeners who hear transmitted versions of *low*-level recordings and those who hear transmitted versions of *high*-level recordings is only 0.88 SDs, indicating $100 \times [(1.37 - 0.88)/1.37] \approx 36\%$ loss of sensitivity to level information. In other words, listeners who hear transmitted recordings are 36% less sensitive to variations in the level of the original predictions, compared to listeners who directly hear the original predictions. Formally, the *change in slope* statistic printed in the plot is calculated from a regression of the form

$$\text{LevelBelief}_i = \beta_0 + \beta_1 \text{HighLevel}_i + \beta_2 \text{Transmitted}_i + \beta_3 (\text{HighLevel}_i \times \text{Transmitted}_i) + \varepsilon_i, \quad (1)$$

where $\text{LevelBelief}_i$ is the listener's belief about the level of the original prediction (z-scored at the topic by reliability manipulation type level); $\text{HighLevel}_i$ is a dummy for the original forecast having a high level; and $\text{Transmitted}_i$ is a dummy for the participant listening to a transmitted version of the original forecast. Standard errors are two-way clustered at the voice recording and listener level.[11] The change in slope statistic is simply $-100 \times (\beta_3/\beta_1)$.

Panel (b) examines listeners' message beliefs about the reliability of the original predictions. Here, the sensitivity loss is nearly three times as strong. Listeners hearing the original messages believe the strong-reliability messages are 1.18 SDs more reliable than the weak-reliability messages on average. Among listeners hearing transmitted versions of the original messages, this difference is only 0.12 SDs, indicating roughly 90% loss of sensitivity. A formal test of equality of the two information loss statistics rejects the null at $p < 0.001$, $\chi^2 = 74.5$.

Figure 2 further illustrates that, in both cases above, transmission weakens the distinction between high- and low-level messages (or weak- and strong-reliability messages) by *symmetrically compressing listeners' beliefs towards an intermediate value*. This is compatible with the following dynamic: listeners hold an average prior about level or reliability that is located halfway between our two manipulations; they update away from this prior when hearing a message; and the strength of this update is weakened by noise introduced during transmission. If more noise is introduced for the reliability than the level, this compression will be stronger for reliability.

The finding of nearly symmetrical compression also shows that, contrary to an intuitive hypothesis, the fact that a message is transmitted does *not* reduce its perceived reliability on average: instead, transmission causes strong-reliability messages to be perceived as less reliable, and weak-reliability messages to be perceived as more reliable.

**Interpretation.** The finding that transmission symmetrically compresses level and reliability beliefs admits a cognitive-imprecision interpretation according to which transmission adds sym-

[11] Standard errors are virtually identical for different ways of clustering.

metric, zero-mean noise to the level and reliability of the original message, attenuating listeners' perceptions of those parameters towards a default belief as the true parameters become difficult to discern. This interpretation connects our findings to noisy processing models popular in the recent literature (e.g., Ba et al., 2024; Enke and Graeber, 2023). Alternatively, transmission-induced compression towards an intermediate belief could reflect a form of ignorance or feeling of "I don't know" (Fischhoff and Bruine De Bruin, 1999), or a process of anchoring-and-adjustment (Tversky and Kahneman, 1974) caused by listening to a transmitted message. This could be true if listeners find it difficult to decode or interpret the contents of transmitted messages, give up, and retreat to a default belief. Under this interpretation, the strength of attenuation reflects the difficulty of decoding each type of information. We remain agnostic about which exact interpretation is the most accurate.

**Result 2.** *Verbal transmission induces substantial information loss. This information loss differs for different types of information: Loss of reliability information is about three times as large as loss of level information.*

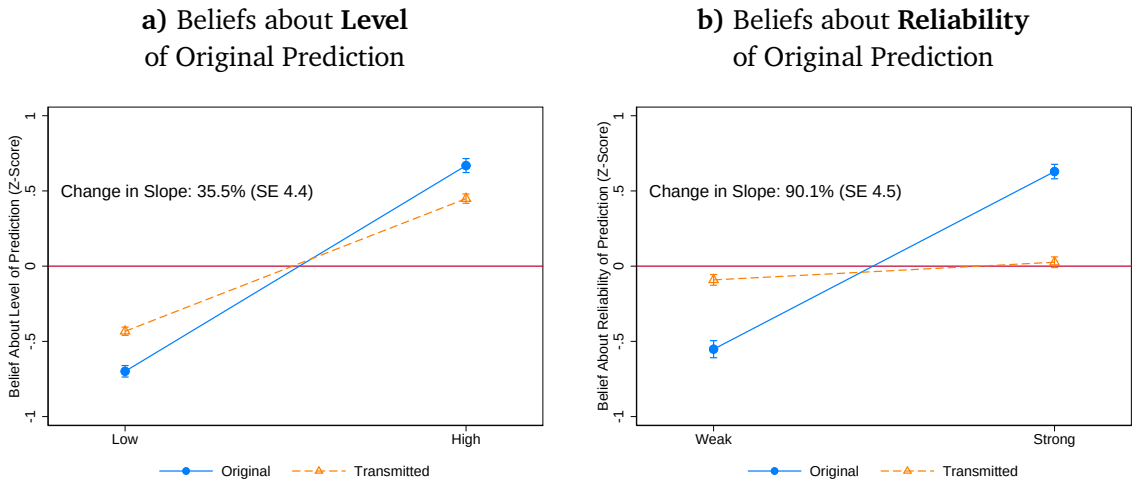



Figure 2: Slopes in the two panels differ at $p < 0.001, \chi^2 = 74.5$. This figure presents data from our baseline experiment (Belief Movement Incentives). It shows listeners' beliefs about the level and reliability of the prediction in the original message, separately by whether the original message is low- vs high-level or weak- vs strong-reliability, and separately by whether the listener hears the original message or a transmitted version of it. Dots are mean beliefs (z-scored at the topic by reliability manipulation level) and bars are standard error bars (1 SE each direction). See Appendix Figure A3 for results with raw (non-z-scored) beliefs and Appendix Tables A6 and A7 for regression tables. $N = 1,510$ listeners and 540 transmitters.

**Relation to transcript analysis.** How much of the differential information loss in Figure 2 simply reflects the complete omission of reliability indicators from two-thirds of transcripts documented in the previous section? Appendix Section D decomposes differential information loss into a component driven by the absence of reliability indicators (extensive margin) and a component that remains even in messages with at least one reliability indicator (intensive

margin). We find that only about a quarter of the differential loss in Figure 2 is driven by the extensive margin; as we show, even messages that contain at least one reliability cue drop many of the cues seeded in the original messages, and this matters for listeners' beliefs.

## 3.3 State Belief Updates

Message beliefs, analyzed above, enable us to separately track level and reliability information through the transmission process. However, they are artificial objects. The objects of immediate economic relevance are listeners' *state belief updates*, their belief updates about the economic variables discussed in the recordings (home price growth and revenue growth).

**Differential information loss in state belief updates.** We can adapt the specification in Equation 1 to measure level and reliability information loss using state belief updates instead of message beliefs. The results of this analysis are printed in Panel (c) of Figure 3. We calculate that transmission reduces the sensitivity of listeners' belief updates to our level manipulations by 30%, and reduces the sensitivity to reliability manipulations by 90%, strikingly similar to the numbers calculated using message beliefs.

Informally, Panel (c) of Figure 3 shows that listeners of the original messages update, on average, twice as much from strong-reliability messages compared to weak-reliability messages. Listeners to transmitted versions, meanwhile, update almost the same amount from weak- and strong-reliability messages. This is what underlies our finding of 90% reliability information loss.

**Interpreting the effects of transmission loss on state belief updates.** In addition to showing strong differential loss of reliability information, Figure 3 displays a rich set of patterns that we now discuss. Panel (a) displays the average state belief updates of listeners who directly hear original recordings, across the four categories of our level/reliability cross-randomization. We pool data from both topics, home price and revenue growth, and separately z-score belief movements for comparability. The panel shows that state belief updates are sensitive to both our level and reliability manipulations. In particular, listeners adjust their beliefs in a qualitatively Bayesian manner: they move in the direction of the forecast they receive, with the strength of the update moderated by the reliability of the forecast.

Panel (b) of Figure 3 shows predictions for the effects of transmission loss on state belief updates. Panel (c) shows the actual effects of transmission, which match the predicted effects.

To understand the predictions and results, observe that the loss of *level* information should uniformly shrink listeners' belief updates towards zero (the mean belief update, given z-scoring). This is because, as Panel (a) of Figure 2 shows, transmission symmetrically compresses beliefs about the level of the original prediction towards the mean value. This should in turn compress belief updates towards the mean belief update, given that average priors are the same across experimental conditions. Hence, across all four conditions, we predict that level information loss should attenuate belief updates towards zero (the green arrows in Panel (b) of Figure 3).

Meanwhile, the loss of *reliability* information should have different effects in the strong versus weak reliability conditions. Loss of reliability information symmetrically compresses listeners' beliefs about the reliability of the original messages towards the mean (Panel (b) of Figure 2).

This means that transmission causes *strong-reliability* messages to be perceived as *less reliable*. This, in turn, should shrink belief updates from strong-reliability messages, since the size of a listener's belief update should be smaller the lower the perceived reliability of the signal. Hence we predict that in the strong-reliability conditions, reliability information loss should attenuate belief updates towards zero (the purple arrows in the leftmost and rightmost conditions in Panel (b) of Figure 3). Conversely, transmission causes *weak-reliability* messages to be perceived as *more reliable*. This means reliability information loss should strengthen belief updates *away* from zero in the weak-reliability conditions (the purple arrows in the two middle conditions).

Overall, we obtain an unambiguous prediction that in the strong-reliability conditions—where both level and reliability information loss push in the same direction—transmission should cause belief updates to shrink strongly towards zero. Meanwhile, in the weak-reliability conditions, level information loss pushes towards zero and reliability loss pushes away from zero; without knowing which effect dominates, we have an ambiguous prediction for the effect of transmission belief updates in these conditions.

Panel (c) of Figure 3 shows empirical results that exactly bear out these predictions. In the strong-reliability conditions, transmission causes average belief updates to shrink in size by about 50%. Meanwhile, in the weak-reliability conditions, the opposing effects of level and reliability information loss seem to roughly cancel out, and average belief updates barely change.

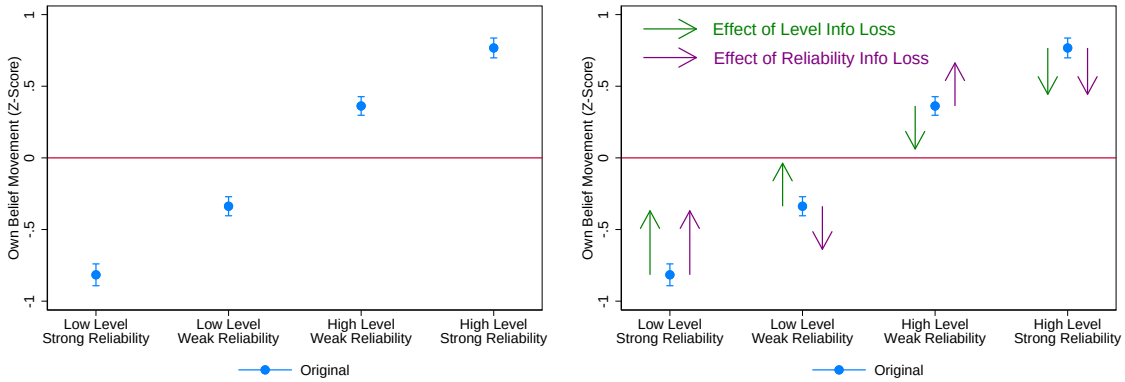
Appendix Figure A5 validates our comparative-static explanation of the empirical results by splitting transmitters according to whether they are coded as passing on reliability (see Section 3.1). Consistent with our story, transmitters who fail to pass on reliability information induce overreactions among listeners in the weak-reliability buckets and more severe underreactions among listeners in the strong-reliability buckets.

**Implications of transmission loss.** Summing up, Figure 3 shows that transmission-induced information loss has two impacts on downstream belief updates. First, listeners to original messages update about twice as much from strong-reliability messages as from weak-reliability messages; by contrast, listeners to transmitted versions update to the same degree from weak- and strong-reliability messages. That is, transmission *reduces the relative influence of strong-reliability messages.* This happens mainly through a reduction in the absolute influence of strong-reliability messages; the absolute influence of low-reliability messages does not change much on average because reliability loss is offset by level information loss pushing in the opposite direction. This suggests that the main impact of transmission loss is to reduce the impact of information from trustworthy sources. Second, averaging across all four conditions, listeners' absolute belief updates are 30% smaller when listening to transmitted messages, an effect driven by the strong-reliability conditions.[12] This means that transmission *reduces the average impact of new information on beliefs*. Whether transmission-induced information loss leads to more convergence or more polarization of beliefs in response to new information is more subtle; it depends on the cueing process underlying information loss. We demonstrate and discuss this in Appendix C.

[12]Technically, the figure shows that *z-scored* belief updates are smaller, but this is also true for mean raw belief updates; the mean raw belief update is $\approx 0$.

**a)** Belief Updates among Listeners to Original Messages

**b)** Predicted Effects of Transmission Loss on Belief Updates




**c)** Belief Updates Among Listeners to Original versus Transmitted Messages

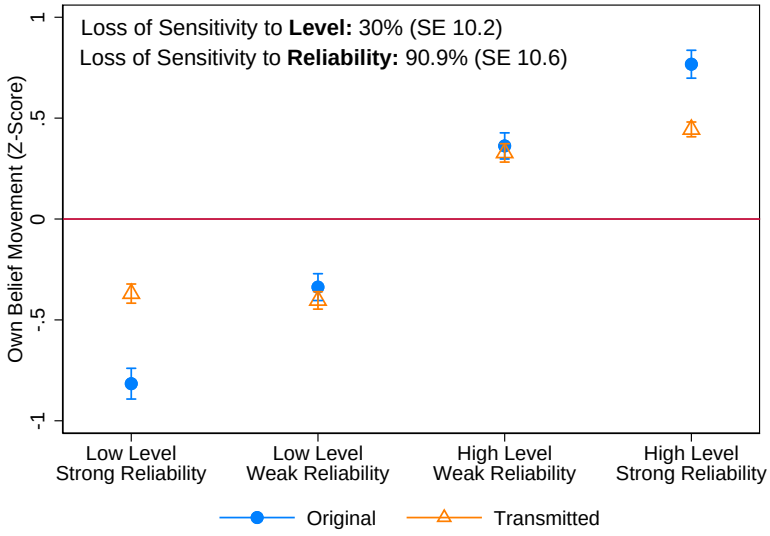



Figure 3: This figure shows average belief movements (posterior minus prior) about the economic variable from our baseline experiment (Belief Movement Incentives). Panel (a) shows average belief movements about the economic variable across the four different level-reliability conditions, only for listeners who directly hear the original messages. Dots are mean beliefs and bars are standard error bars (1 SE each direction). Panel (b) adds illustrative arrows. Panel (c) adds the corresponding beliefs of listeners hearing transmitted versions of the messages. $N = 1,510$ listeners and 540 transmitters. The loss of sensitivity to level information is calculated from a regression of the form: $\text{BeliefUpdate}_i = \alpha_0 + \alpha_1 \text{HighLevel}_i + \alpha_2 \text{StrongReliability}_i + \alpha_3 \text{Transmitted}_i + \alpha_4(\text{HighLevel}_i \times \text{Transmitted}_i) + \xi_i$. The loss of sensitivity to reliability information is calculated from a regression of the form $\text{BeliefUpdate}_i \times (2 \times \text{HighLevel}_i - 1) = \gamma_0 + \gamma_1 \text{HighLevel}_i + \gamma_2 \text{StrongReliability}_i + \gamma_3 \text{Transmitted}_i + \gamma_4(\text{StrongReliability}_i \times \text{Transmitted}_i) + \zeta_i$, where we flip the sign of low-level belief updates to make the effects of StrongReliability comparable across low- and high-level messages. Appendix Table A5 gives regression versions of these results. Appendix Figure A4 shows these results restricting to the Modular manipulation and including the neutral-reliability condition. Appendix Figure A5 shows empirical results validating the arrows in Panel (b). Appendix Figure A3 shows raw (non-z-scored) beliefs. Appendix Figure A6 shows Wasserstein distances between the belief update distributions.

**Result 3.** *Verbal transmission weakens the average effect of new information on beliefs. It also*

*reduces the relative influence of strong-reliability, compared to weak-reliability, information.*

# 4 Mechanisms Underlying Differential Loss

What drives the differential loss of reliability and level information? In this section, we test different potential mechanisms. To structure this analysis, we distinguish between mechanisms that involve a deliberate decision by the transmitters to prioritize passing on level information, and mechanisms that involve transmitters subconsciously or non-deliberately failing to pass on reliability information. If differential loss results from transmitters' deliberate decisions, it arises either because (i) the *perceived benefits* of transmitting reliability information are lower or (ii) the *perceived costs* of transmitting reliability information are higher. Alternatively, differential loss may not result from a deliberate cost-benefit tradeoff; for example, (iii) reliability information may simply *fail to come to mind* at the moment of recording the voice message, e.g., due to some kind of attention or memory constraint. We examine each of these three possibilities in turn.

This distinction is not merely mechanical. If transmitters are weighing costs and benefits, the natural way to improve how reliability diffuses is to change their incentives. If reliability instead fails to come to mind, incentives have limited bite, and what matters is whether the situation draws attention to reliability in the first place. We return to this contrast, and its practical implications, in Section 4.3.4.

## 4.1 Perceived Benefits of Transmitting Level and Reliability

We first consider the perceived benefits of, or incentives for, communicating level versus reliability information. Perceived incentives are a natural starting point: in practice, people pass on information in a variety of different circumstances, and the objective of such information transmission can vary widely, from informing to persuading to entertaining the recipient. It is likely that people (at least partly) tailor the contents they transmit to the specific requirements of the situation. The differential loss observed in our data might be an artifact of our setup that induces specific (perceived) transmission incentives, or it may be a more fundamental property of transmission that is likely to occur robustly across different transmission settings.

### 4.1.1 Evidence from Baseline Experiment

We begin by examining three additional pieces of evidence from our baseline experiment. After recording their messages, participants in our transmitter survey are randomized between three sets of supplementary questions. First, we elicit beliefs about the importance of passing on reliability versus level information, to test whether participants (mistakenly) believe that passing on reliability information would not affect listeners' belief updates and hence their probability of receiving the bonus payment. To the contrary, we find that respondents believe that passing on reliability information increases their probability of receiving the bonus payment by about the same amount as passing on level information: the average response is 71% for level and 68% for reliability. This is true even among respondents whom we classify as *not* passing on reliability information in their recordings (averages of 73% versus 67%).

Second, to test whether respondents are aware that they are omitting specific information, we ask another one-third of respondents whether they included level information and reliability information in their recordings. In line with our analyses of transcripts, we find that 64% of respondents admit to not passing on reliability information (compared to 31% for level).

Third, to examine whether people forget or do not pay attention to the incentive scheme, we examine whether, at the end of the survey, the final one-third of respondents still pass the initial comprehension checks about their incentives. We find that 90% of respondents correctly answer both questions about the incentives.

Taken together, these results suggest that — when asked — people remember the incentive scheme and believe it incentivizes them to pass on reliability, yet they admit to not passing on reliability in their actual recordings. This provides a first sign that the loss of reliability information is due to the failure of reliability to come to mind during the transmission process.

### 4.1.2 Additional Evidence: Incentives for Content Transmission

In the baseline experiment, transmitter bonuses were based on the induced belief movements of listeners, leaving transmitters free to pick and choose which dimensions of the original content they believe will be relevant for listeners' belief updates. In a supplementary experiment, we instead explicitly fix the relative incentives to transmit level and reliability information, making beliefs about the relative importance of level and reliability information for belief updates irrelevant for transmission decisions.

This experiment and its results are described in detail in Appendix Section A.1.3. Half of the participants are told that their message will be scored by a group of "evaluators" on a scale of 0 to 10 according to the fidelity with which it preserves the "content and meaning" of the original message. Compared to our baseline incentives, this scheme should incentivize transmitters to pass on reliability information regardless of their beliefs about its importance for listeners' belief updates, because the instructions encompass *all* the contents of the original message. The other 50% of transmitters are explicitly told that evaluators will separately score the fidelity of their message's transmission of the *level* and the *reliability* information in the original message, and that both scores receive equal weight in their probability of receiving the bonus payment. This scheme has two distinctive features. First, it ensures that transmission of reliability information is *explicitly* just as payoff-relevant as transmission of level information. Second, unlike any of our other incentive schemes, it introduces transmitters to the distinction between level and reliability information. It therefore produces a very conservative test of differential information loss.

Results of the experiment are described in Appendix Section A.1.3 and reported in Appendix Figures A16 and A17. We find large differential information loss in both new incentive schemes, though differential loss is slightly smaller than in our baseline experiments. We conclude that differences in perceived importance of transmitting level versus reliability information cannot fully explain our main result.

## 4.2 Perceived Costs of Transmitting Level and Reliability

Next, we turn our focus to the second possible driver of differential loss and examine the subjectively perceived *costs* or *difficulty* of transmitting level versus reliability information. Here we can distinguish between the *(ex-ante) anticipated* and *(ex-post) experienced* costs of transmitting each type of information. Transmitters might deliberately omit reliability information because they *expect* it to be more costly or difficult to transmit; alternatively, they might try to transmit reliability information but then *experience* it as being very difficult to properly transmit. Our analysis in Section 3.1, which found that 60% of transmitted transcripts do not include anything about the reliability of the original message, implies that transmitters are not even *trying* to transmit reliability, suggesting that anticipated costs are more likely to be relevant than experienced ones.

As a direct test of the initially *anticipated* costs of transmitting level versus reliability information, we conduct an additional transmitter experiment in which we study whether transmitters prefer to be paid for their performance in transmitting information about (i) the level of the original prediction or (ii) the reliability of the original prediction. By "performance," we mean an external evaluator's assessment of how well the transmitter's message passed on the level or reliability information, respectively, following the second incentive scheme from Section 4.1.2. We also elicit transmitters' expectations about how difficult transmitting level or reliability information will be. To test whether experienced costs deviate from anticipated ones, we study whether transmitters' beliefs about the difficulty of transmitting level versus reliability information change after experiencing the transmission process.

Panel (a) of Appendix Figure A18 shows that 62 percent of respondents choose to transmit information about the reliability of the prediction, and the average perceived difficulty of transmitting reliability information is slightly lower than for level. Differences in the perceived difficulty of communicating level and reliability information are relatively small, both measured before (Panel (b), $t = 0.64, p = 0.53$) and after the recording (Panel (c), $t = 2.4, p = 0.02$). This suggests that transmitting reliability information is, if anything, *easier*, and makes it hard to see how higher anticipated or experienced costs of transmitting reliability information could play a role in driving our results. Virtually all respondents pass on the characteristic they chose.

There is no heterogeneity in perceived costs that could generate the pattern of differential information loss we observe. For example, suppose that the 62% of people choosing to transmit reliability are capable of transmitting both types of information in the main experiment, but the 38% choosing to transmit level information find transmitting reliability to be prohibitively costly. This could generate differential information loss even if transmitting level is perceived as harder on average. But we find no such heterogeneity: the groups choosing to transmit level versus reliability information give similar average difficulty ratings and have similar average difficulty gaps between the parameter they choose to transmit and the other.

**Result 4.** *Mechanism experiments suggest that differential transmission loss of reliability information in our experiment is not the result of a deliberate decision: it is not driven by the subjectively perceived benefits or costs of transmission.*

## 4.3 What Comes to Mind

Having established that differential loss does not appear to be the result of a deliberate prioritization of level information, we examine the possibility that transmitters subconsciously or non-deliberately neglect to include reliability information due to attentional or memory constraints. Our previous results—which show that, when asked, transmitters believe passing on reliability is important, believe it is easier than passing on level, yet mostly do not even try to pass on reliability—hint at this kind of attentional neglect of reliability.

### 4.3.1 Conceptual Framework

To structure our investigation, we build on simple stylized facts from the science of memory. When a person is prompted to transmit information, what "comes to mind" depends on the situation or request prompting transmission (the "cue") and the probabilities that different pieces of information stored in memory will be activated by that cue (e.g., Bordalo et al., 2025b; Kahana, 2012). Specifically, when someone encounters an original piece of information, various aspects of that information (facts, arguments, the confidence level and characteristics of the speaker, and so on) are stored in their memory as separate traces. When that person later encounters a situation prompting transmission of the information, the probability with which each trace comes to mind depends on two things. The first is the *similarity* or association between the trace and the cue. For example, an explicit request for reliability information, or a request that otherwise reminds the transmitter about reliability or uncertainty, is more likely to trigger retrieval of traces pertaining to reliability. The second is *interference* between different memory traces: retrieval of a specific trace is less likely when many other similar traces are stored in memory and could be confused with the focal trace. For example, a hiring manager who has recently read several hundred CVs may struggle to remember which one a specific qualification came from.

This distinction between cue-based similarity and interference naturally suggests two reasons why reliability information may be less likely to come to mind.

**Cues.** The "cue" in our transmission experiments is the generic request to recall and pass on the information from the original message. Whether this cue triggers retrieval of level or reliability information with higher probability depends on the perceived similarity or association between the cue and level or reliability information, respectively.

Anecdotal evidence suggests that real-life requests for transmission typically cue level information but rarely reliability information. People often ask about substantive economic facts, beliefs, or projections, but rarely directly request confidence levels or reliability assessments from their interlocutors. This could create an association between transmission and level information that leads generic transmission contexts (like our experimental context, which carefully avoids explicitly cueing either level or reliability) to naturally cue level information, but not reliability.

**Interference.** A second factor, whose consequences for transmission we do not test, is that reliability may be harder to retrieve because it is easily confused across sources. Predictions about levels usually arrive wrapped in rich, context-specific arguments and examples, which

makes them distinctive and easy to keep apart. Reliability is typically expressed in generic terms ("maybe," "I'm fairly sure"), so reliability statements from different conversations resemble one another and can blur together. Because we cannot directly measure whether this similarity generates retrieval interference, we set this channel aside and focus on cues.[13]

### 4.3.2 Cues and Reliability Transmission: Experimental Evidence

We now provide evidence on the role of cues in reliability transmission. We begin by showing that transmitters remember the reliability of the original message when asked, so reliability information is available for retrieval rather than being irretrievably forgotten. We then provide four pieces of evidence that cues of reliability increase the rate of reliability transmission: two experimental pieces of evidence in this section, and then two observational pieces in Section 4.3.3. Finally, we show that explicit cues of reliability are rare in practice.

**Transmitters remember reliability when cued.** Once specifically cued, transmitters in our baseline experiment are able to accurately recall both the level and reliability of the original predictions. Specifically, after they complete their recordings, we present transmitters with the same set of beliefs questions we pose to listeners (see Section 2). Appendix Figure A12 demonstrates that there is virtually no memory loss about the original message's reliability: several minutes after hearing the original recording and after performing the cognitively demanding task of recording their own voice message in the interim, transmitters are *just as sensitive to variations in reliability* as listeners whose beliefs are elicited immediately after hearing the original recordings. Transmitters hence do not *forget* the reliability information, meaning that, if cued to attend to reliability, they should be able to retrieve and transmit it.

**Irrelevant cues in our baseline experiment.** An incidental feature of our baseline transmitter experiments is that they seeded random variation in exposure to subtle, payoff-irrelevant reliability cues. One of our reliability manipulations — the *modular* manipulation — varied whether an otherwise-identical text contained certainty-denoting prefixes, uncertainty-denoting prefixes, or *no* reliability prefixes.[14] Consider two respondents, A and B, who both heard the same low-reliability Retail recording. Suppose that respondent A heard a neutral-reliability Housing recording, while B heard a high-reliability Housing recording. Respondents A and B both heard exactly the same Retail recording, but respondent B may be more likely to pass on its reliability for three reasons. First, B encountered more reliability language in total (hearing both a low- and a high-reliability recording, rather than a low- and a neutral-reliability recording). Second, B heard reliability language more recently relative to the moment of transmission. Third, the difference in reliability between the two recordings heard by B could create a contrast effect raising the salience of reliability. Note that the cue here is subtle — consisting of reliability words seeded in a different recording than the one under consideration — and payoff-irrelevant, having noth-

[13]Psychologists call the difficulty of attributing a memory to its correct origin a source monitoring error (Hovland and Weiss, 1951; Johnson, 1997); several studies tie it to interference between similar memories (Henkel and Franklin, 1998; Johnson et al., 1993; Lindsay and Johnson, 1989).

[14]Appendix A.1.5 provides additional details.

ing to do with the payoff for transmitting the focal recording. It is also delivered much closer to the moment of transmission than the incentives — only a minute or two beforehand — ensuring that anything it cues is more likely to be top of mind during transmission.

Column (1) of Table 1 regresses an indicator for whether a transmitter's message for a particular recording is unanimously classified by our LLM and human coders as transmitting reliability markers on an indicator for whether the *other* recording the transmitter heard contained reliability indicators. (The independent variable is 0 for transmitters whose other recording was a no-prefix *modular* recording, and 1 for all other transmitters.) The results show that hearing such payoff-irrelevant cues in the *other* recording increases the probability of transmitting reliability information for *this* recording by 10.7 percentage points. Columns (2) and (3) confirm, using more granular LLM codings, that the effect is entirely driven by increased transmission of reliability indicators consistent with the original recording (high indicators for high-reliability recordings, and low indicators for low-reliability recordings). This pattern is informative about *how* the cue operates. Rather than importing the reliability content of the *other* recording into the focal message, the cue increases faithful transmission of the focal recording's *own* reliability. This is what we would expect if the cue makes reliability salient as a dimension, prompting transmitters to attend to and reproduce the focal recording's actual reliability, rather than causing reliability content to spill across recordings. Unlike our next set of results, we find no evidence of crowding out of level transmission. We return to Columns (4) and (5) in section 4.3.3.

Table 1: Reliability Cues Increase Transmission of Reliability

| | Experimental Results: Reliability Indicators Passed On | | | British National Corpus: Answer Contains Reliability Indicator | |
|---|---|---|---|---|---|
| | (1) All Words (Unanimous) | (2) Aligned Words (LLM) | (3) Misaligned Words (LLM) | (4) All Questions | (5) Secondhand Info Questions |
| Other Recording Contained Reliability Prefixes | 0.107*** | 0.100** | -0.014 | | |
| | (0.035) | (0.043) | (0.030) | | |
| Question Contained Reliability Prefixes | | | | 0.207*** | 0.224*** |
| | | | | (0.003) | (0.003) |
| Constant | 0.130*** | 0.274*** | 0.123*** | 0.151*** | 0.166*** |
| | (0.028) | (0.037) | (0.027) | (0.002) | (0.002) |
| Observations | 702 | 702 | 702 | 90057 | 70751 |

Standard errors in parentheses
* p<0.1, ** p<0.05, *** p<0.01

*Note*: Columns (1)-(3) report results from our baseline transmitter experiment. An observation is a transmitter-by-topic, restricted to transmitter-by-topics where the transmitter heard a high- or low-reliability recording about that topic. The independent variable in all three columns is an indicator for whether the recording the transmitter heard for the *other* topic was *not* a neutral-reliability modular recording. The dependent variable in Column (1) is an indicator for whether our LLM and human coders unanimously classify the transmitter's transmitted message as containing reliability indicators. In Column (2), it is an indicator for whether the LLM classified the message as containing a reliability indicator aligned with the original message (high-reliability indicators for high-reliability messages, and vice versa; our human coders did not code at this level of granularity). In Column (3), it is an indicator for whether the LLM classified the message as containing a reliability indicator misaligned with the original message. Columns (4) and (5) use data from the British National Corpus. An observation is a line in one of the conversations in the corpus. Column (4) restricts to conversational lines that our LLM identifies as being questions, while Column (5) additionally restricts to questions that are about secondhand information the questioner's interlocutor would not have direct access to (e.g., not questions about the interlocutor's name or experience at a restaurant). The dependent variable in both columns is an indicator for whether the response to the question in the next conversational line contains any reliability indicators and the independent variable is an indicator for whether the question itself contains any reliability indicators.

**Salient cues of reliability during transmission.** As another test of the importance of cues in determining the transmission of reliability information, we conduct an additional *high salience* experiment that increases the *during-transmission* salience of the distinction between the level and reliability of the original message. This experiment tests the hypothesis that differential information loss decreases when transmitters are directly reminded about the level–reliability distinction during the recording process, effectively equally cueing level and reliability. The design also allows us to assess whether emphasizing one dimension of information crowds out the transmission of the other.

The design closely follows the *explicit incentives* treatment described in Section 4.1.2, in which transmitters were explicitly incentivized to transmit both the level and reliability of the original message's prediction. It adds three features to increase the salience of the level-reliability incentives at the time of recording: First, we add additional, more heavy-handed comprehension questions in which respondents need to correctly answer which types of information they need to transmit in the experiment. Second, just prior to each recording we ask respondents: "What do you have to pass on well to maximize your chances of receiving a bonus? Tick all that apply" with the following response options: (i) level of the speaker's prediction; (ii) reliability of the speaker's prediction. Respondents can only proceed once they correctly answer this question by selecting both. Third, on the actual recording page we add the following reminder: "Remember: Your

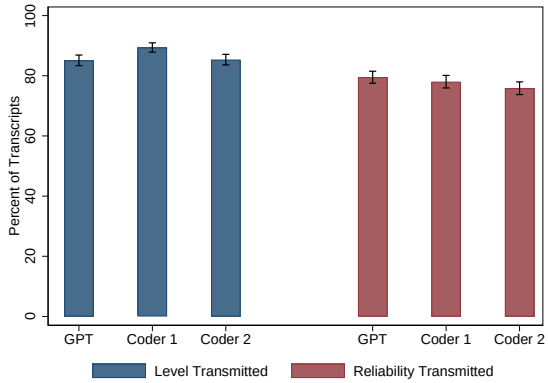
bonus payment is based equally on how well you pass on both of the following: (i) The level of the speaker's prediction. (ii) The reliability of the speaker's prediction." This cue, delivered *at* the moment of transmission, is again delivered much closer temporally to transmission compared to the presentation of the incentives. This reminder is presented in large, red font. The experiment was conducted on Prolific in November 2023 with 244 transmitters and 1,010 listeners.

Figure 4 summarizes the results. Consistent with the hypothesis that reliability information is more likely to come to mind when it is explicitly cued during transmission, Panel (a) shows that transmitters in this design talk much more about reliability, with nearly 80% of transmitted transcripts containing at least some information about the original prediction's reliability, compared to just 30-40% in our previous experiments. The share of transcripts containing level information decreases, from 90-95% to 80-90%. Commensurately, Panels (b) and (c) document a large reduction in reliability information loss relative to our baseline experiment (from 90% to 40%), which is accompanied by a statistically significant increase in level information loss (from 36% to 53%). This pattern of results plausibly reflects crowding-out of level information as transmitters talk more about reliability, suggesting that the transmission of level and reliability information draw on common cognitive resources, such that increasing the transmission of one can come at the expense of the other.

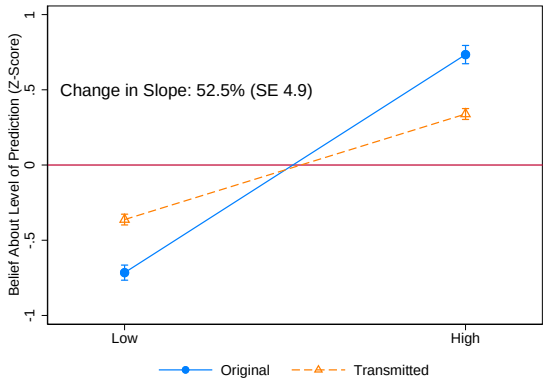
Panel (c) shows that distortions of reliability information disappear entirely for weak-reliability messages but remain for strong-reliability messages. On the one hand, this may suggest that indicators of weak reliability are more salient or easier to transmit once transmitters have reliability in mind. On the other hand, this pattern may reflect a symmetric loss akin to the one documented before, coupled with an overall downward shift of perceived reliability that equally applies to all transmitted messages. Two pieces of evidence favor the latter explanation: our LLM and human handcoders unanimously find that reliability is equally likely to be transmitted in the high- versus low-reliability conditions, and a handcoding-based adjustment to the beliefs figure described in Appendix Section A.1.2 eliminates the apparent asymmetric loss.

Panel (d) documents the consequences for the overall pattern of listeners' state belief updates. Transmission strongly attenuates belief movements towards zero on average. This is driven by the level information loss; moreover, the offsetting force of reliability information loss for weak-reliability messages, which pushed belief updates for those messages away from zero, is now absent (see the detailed discussion of forces in Section 3.3). As a result, transmission mostly preserves the distinction between weak- and strong-reliability messages: listeners update less than half as much from weak-reliability compared to strong-reliability messages, regardless of whether they hear original or transmitted recordings. However, this also means that average belief updates from transmitted messages are shrunk even further than in our baseline experiment.

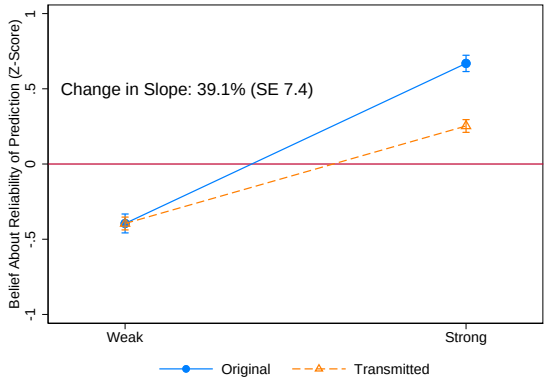
**a)** Fraction of Scripts Containing Level/Reliability Statements




**b)** Salience Experiment: Level Info Loss




**c)** Salience Experiment: Reliability Info Loss




**d)** Belief Updates About the Economic State

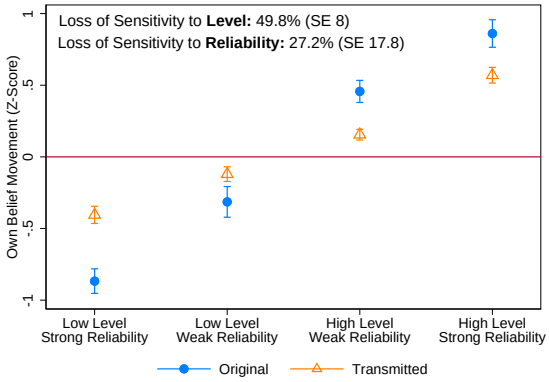



Figure 4: This figure presents data from the *high salience* experiment. Panel (a) replicates Panel (a) of Figure 1, showing which fraction of transmitted scripts contain statements about the level or reliability of the original prediction. Panels (b) and (c) replicate Figure 2, showing beliefs about the original message's level and reliability, separately by whether level is low/high, reliability is weak/strong, and the listener is hearing an original or transmitted message. Panel (d) replicates Panel (c) of Figure 3, showing listeners' average belief updates about the economic variable. Bars are standard error bars. $N = 1{,}010$ listeners and 244 transmitters.

Taken together, this experiment highlights a fundamental tradeoff. Making reliability salient at the moment of transmission restores the differential influence of weak- versus strong-reliability messages, but it does so by reallocating limited cognitive resources, thereby increasing level information loss and further attenuating overall belief updating. As a result, it aggravates the fact that in a population with heterogeneous priors, transmission loss slows down belief convergence on the basis of new information: belief updates in the intermediate quadrants of Figure 4 Panel (d) are more severely attenuated towards zero. Of course, such a slowdown may be desirable if this convergence would otherwise happen on the basis of unreliable information.

### 4.3.3 Cues and Reliability Transmission: Field Evidence

The previous section used two sources of experimental variation to show that cues increase the rate of reliability transmission. We now show more suggestively that, in real-world transmission

contexts, cues of reliability are associated with increased rates of reliability transmission.

**Cues in economic TV news.** We first examine data on economic cable news broadcasts. For the sake of brevity, we here briefly describe the main ideas behind our methodology and the main results, deferring a detailed exposition to Appendix E.

Our measure of the fidelity of uncertainty transmission on cable news is the correlation between the number of uncertainty-denoting words and phrases used in economic segments and a benchmark measure of "true" economic uncertainty calculated from newspapers (Baker et al., 2016). The use of newspapers as a benchmark fits our previously-discussed distinction between written text—which is more premeditated and less subject to cognitive and memory constraints—and spontaneous word-of-mouth discussion or punditry, which is more vulnerable to the cognitive constraints we study. We test whether higher quantities of uncertainty language on a channel's news segments *yesterday*—which we treat as cues that incidentally bring reliability to mind—are associated with a greater responsiveness of uncertainty language in economic segments to (our benchmark of) true economic uncertainty *today*. We find that this is in fact the case, and that this is robust to a range of specifications and placebo tests.

In other words, when pundits on a certain channel discussed uncertainty more yesterday, their discussion of uncertainty *today* is more sensitive to true uncertainty. This is consistent with the idea that recent experiences discussing uncertainty cue people to attend to it more, leading them to transmit it with higher fidelity. The fact that our outcome of interest is a slope means that serial correlation in uncertainty discussion would not by itself generate our results, and our use of day fixed effects (isolating between-channel variation in uncertainty discussion on preceding days) rules out effects of true economic uncertainty on attention to uncertainty. However, we emphasize that this analysis is only suggestive: for example, one alternative interpretation is that the act of communicating uncertainty on preceding days shifts commentators' beliefs about the importance of transmitting uncertainty, rather than acting as an unconscious cue.

**Cues in everyday conversations.** We next turn to the British National Corpus, a dataset consisting of 1,251 recordings of everyday conversations between individuals in the UK, recorded at home or in other settings using participants' smartphones (Love et al., 2017). We use a large language model to annotate each of the 1.2 million lines of usable conversational text in the corpus. We ask the model to identify lines containing questions; we further distinguish between questions concerning information that the speaker's interlocutor would have direct access to (such as what happened to them that morning) and secondhand information (such as a weather forecast or piece of economic or political news). Such questions, by definition, contain a request for level information of some kind; we ask the model to additionally classify whether the question contains an explicit request for reliability information, such as the interlocutor's certainty level or the source of the information. As we discuss below, such explicit requests are almost completely nonexistent. We therefore also code whether the question contains reliability markers or other references to reliability that could act as indirect cues. Column (4) of Table 1 shows that a question containing such a marker is 21 percentage points more likely to be followed by an answer containing a reliability indicator. This remains true in Column (5), where we restrict to

questions requesting second-hand information. Indirect cues of reliability in questions therefore raise the probability that the answer to the question mentions reliability.

This analysis is only suggestive: it could also reflect omitted factors shifting reliability expressions in both questions and answers, or reliability cues shifting interlocutors' beliefs about the importance of transmitting reliability information. However, the results of this analysis are consistent with the evidence from our experiments and TV news: even subtle cues appear to help trigger reliability transmission.

**Everyday requests for information rarely cue reliability.** The preceding analysis suggests that cues are consequential. However, evidence from the British National Corpus suggests they are rare in practice. Of the 70,751 questions about secondhand information that the LLM identifies, only 2,911 (4.1%) contain an explicit request for information about the certainty or reliability of the interlocutor's answer. Explicit requests for reliability information are hence very rare in everyday spoken conversation. We also ask the LLM to code whether each question contains any *subtler* cues of reliability, according to a maximally broad notion that we define as any language relating to certainty or possibility. Even according to this definition, less than half (46%) of questions contain such cues, compared to 19% of all conversational lines.

### 4.3.4 Summary and Discussion

**Summary.** Across four pieces of evidence, cues that cause transmitters to attend to reliability—ranging from indirect, payoff-irrelevant cues to highly salient, payoff-relevant ones—robustly increase rates of reliability transmission. However, indirect cues of reliability are uncommon in spoken conversations, and explicit cues are very rare. This may lead generic transmission contexts to become more associated with level than reliability information, meaning that reliability tends not to be cued in such contexts, explaining failures to transmit it.

In Appendix C, we use our simulation exercises to explore the implications of cue-based reliability transmission for patterns of belief updating and polarization: cue-induced transmission of reliability consistently improves belief accuracy but can either increase or decrease polarization, depending on the nature of the cue and the axis along which polarization is measured.

**Caveats and Discussion.** This section has argued that the results of our experiments are best understood not as reflecting deliberate cost-benefit analyses supporting prioritization of level rather than reliability information, but involuntary cognitive mechanisms reflecting "what comes to mind." Importantly, we do not claim that cost–benefit considerations are always absent in the field. To the contrary, such considerations may well amplify or attenuate the patterns we document, depending on context. Our claim is narrower: even abstracting from strategic motives, there is a systematic cognitive asymmetry that makes reliability less likely to be retrieved and verbalized during transmission. The experimental results establish this mechanism in isolation; the observational data show that the preconditions for this mechanism—infrequent cues of reliability—are pervasive in naturally occurring communication.

As previously mentioned, our finding that people neglect to transmit reliability information complements previous evidence that people neglect reliability when updating their beliefs (Grif-

fin and Tversky, 1992). Cognitive psychologists distinguish between a "representational" stage of decision-making in which an agent constructs a mental representation of the decision problem they face, including relevant considerations, and a "computational" stage in which those considerations are processed and aggregated into a decision (see, e.g., Ba et al., 2024). Current interpretations of reliability neglect in belief updating focus on the *computational* stage: people attend to reliability information but either struggle to assess its quantitative magnitude (Augenblick et al., 2025) or imprecisely code and are uncertain how to incorporate it into belief-updating calculations (Enke et al., 2025). Our findings for the case of transmission suggest instead that reliability neglect here originates in the *representational* stage, with reliability failing to come to mind in the first place. This potential difference in mechanisms may not be surprising. Failures to bring reliability to mind when recalling a message from memory during transmission seem plausible, while such failures seem less likely to occur in controlled belief-updating problems where numeric measures of signal precision are explicitly presented to decision-makers. Conversely, computational difficulties seem more likely to arise in updating—when one must assess the quantitative magnitude of reliability and incorporate it into a Bayesian calculation—than in transmission—when one can simply repeat indicators of reliability.

A complementary reading draws on Bordalo et al. (2026a), who show that people tend to represent an inference problem as either a sampling or base-rate problem or as a simple inference problem that neglects the prior, foregrounding one statistic rather than integrating both as Bayes' rule requires. Our transmission setting exhibits an analogous representational competition on the supply side: the default representation of the transmission problem foregrounds a forecast's level and backgrounds its reliability, and our High Salience experiment shows that forcing reliability into the representation suppresses the transmission of level (Figure 4). This view rationalizes the limited bite of our incentive variations, since incentives that operate conditional on a representation do little when reliability is absent from the default one, just as payoffs do not easily induce people to combine base-rate and likelihood information.

One implication of the apparent origin of failures to transmit reliability in the representational stage of decision-making is that reliability transmission will be relatively insensitive to incentives or stakes, and more sensitive to contextual factors that shift a transmitter's representation of the transmission problem at the moment of transmission. Even a payoff-irrelevant cue—one that has no claim on a cost-benefit calculation—substantially increases reliability transmission in our experiments, while explicit incentives to transmit reliability (in our Content Transmission Incentives experiment, where the incentives are introduced before the original recording is heard and several minutes before transmission) do not prevent the emergence of large differential loss. This pattern fits naturally with the Bordalo et al. (2025b) framework, in which incentives operate conditional on a representation: if reliability is absent from the default representation of the transmission problem, incentives delivered in advance of transmission may have limited purchase, while cues delivered at the moment of transmission can change the representation itself and hence be more impactful. A natural prediction, which our design does not directly test, and which we flag as a direction for future work, is that incentives delivered *at* the moment of

transmission should behave more like contextual cues and hence be more effective.

# 5 Economic Consequences and Robustness

We close the paper by studying the economic consequences of the transmission loss we document and assessing the robustness of our experimental findings.

## 5.1 Consequences

### 5.1.1 Financial Investing Experiment

To assess whether transmission-induced losses of reliability information have economically meaningful consequences, we study a financial investment task. Decision making under risk, particularly investment, is a canonical economic setting where both level and reliability beliefs translate directly into consequential choices. Moreover, investment decisions are a context where the transmission of qualitative narratives has been argued to be particularly influential (Hirshleifer, 2020; Shiller, 2017), especially given the recent rise of social-media-driven stock and crypto investing.

This setting also lets us address a concern about interpretation. Our main outcomes so far are beliefs about the level and reliability of the original messages as well as incentivized state belief updates. While the latter are incentivized based on forecast accuracy, one might still worry that listeners' reported assessments of reliability reflect a judgment of how reliable a message *sounds* rather than how willing they are to act on it. To address this concern, we use a setting in which listeners' decisions translate directly into payoffs, so that perceived reliability is revealed through behavior with real stakes.

**Design.** This experiment departs from our baseline in three ways. First, the original narratives are about imminent earnings announcements for two real-world companies. We cross-randomize whether the original message argues that earnings will be higher or lower than the current consensus forecast. Second, transmitters are paid based on the realized performance of an incentivized investment decision made by the listener who hears their message. Specifically, listeners are tasked with allocating $100 across assets that pay off depending on whether the relevant company's earnings announcement ends up higher, lower, or within a small range around the current consensus forecast. One percent of listeners have their decisions implemented and the transmitters matched to those listeners receive the same payoffs as the listeners. Third, we vary the *uncertainty* (rather than reliability) expressed in the original messages by manipulating the speaker's description of the distribution of possible outcomes. In the low-uncertainty case, the speaker says repeatedly that their prediction will certainly turn out to be true; in the high-uncertainty case, the speaker says repeatedly that the opposite could be true as well, and it's possible to imagine things going either way. This allows us to test the robustness of our results to a new way of communicating reliability or uncertainty which is particularly natural in the investment setting.

**Logistics.** This preregistered experiment was conducted on Prolific in December 2025 with 299 transmitters and 1,287 listeners.

**Results.** The results, summarized in Panels (a) and (b) of Figure A14, and discussed in more detail in Appendix Section A.1.2, replicate our core finding: level information is lost at a rate of 14.5% and certainty information is lost at a rate of 77.4%. A formal test of equality of the two information loss statistics rejects the null at $p < 0.001, \chi^2 = 20.82$. Panel (c) shows that about 90% of transmitted messages contain level information, while only 50% contain reliability information. Panel (d) shows that listeners' investment decisions react to information in a risk-averse way: they invest in the direction of the message they receive, shifting their investment more strongly in response to high-certainty messages than low-certainty messages. Transmission causes the sensitivity of the investment decision to level information to decrease by 20.6% and its sensitivity to certainty information to decrease by 59.3%, though these estimates are less precise due to the noisiness of the investment outcome.

**Economic implications: back-of-the-envelope calculation.** Using data on the beliefs and investment decisions of listeners who directly hear the original messages, we estimate listeners' risk aversion assuming homogeneous CRRA preferences, estimating $\hat{\gamma} = 2.09$ (see Appendix A.1.2 for details; our estimate falls into the 1-3 range commonly found in the literature; see Elminejad et al., 2025). We use our estimated utility function to calculate the welfare distortions implied by the level and certainty information loss we document. Eliminating both distortions would improve investment surplus by 19.5% in dollar terms and 6.5% in utility terms. Eliminating certainty information loss would improve surplus by 7.6% in dollar terms and 4.3% in utility terms (since the importance of reliability information is increasing in risk aversion).[15]

These magnitudes should be taken with a grain of salt, since they are sensitive to a range of required assumptions. Nevertheless, they provide a simple proof of concept that the information loss we document can matter nontrivially for welfare.

### 5.1.2 Diffusion of Information

The efficiency reductions documented above reflect a transmission-induced decline in the sensitivity of investment decisions to the certainty or reliability of information. Investment decisions could also be insufficiently sensitive to the reliability of new information due to private belief updating biases like those documented by Griffin and Tversky (1992) and Augenblick et al. (2025), among others. Appendix Section C describes the results of simple simulation exercises demonstrating that in the presence of such biases, transmission-induced loss of reliability information both compounds the insensitivity of decisions to the reliability of new information and—distinctively—partially shifts the distribution of efficiency losses onto sophisticated agents who *would* react accurately to reliability information had it reached them.

## 5.2 Robustness

We now summarize the design and results of three preregistered experiments probing the robustness of our main results to alternative design choices. Table 2 provides an overview of our results across all experiments. Full descriptions of experiments and results are available in Appendix B.

[15] By surplus, we mean the gains of eliminating information loss as a percentage of the gains of moving from the default no-information equal-allocation portfolio to the optimal portfolio.

**Extensive-margin transmission decisions.** A valid concern is that our baseline results may partly be driven by the fact that transmitters *have to* pass on all messages. In the real world, individuals might simply choose not to pass on information they perceive as less credible, preventing unreliable information from diffusing and mistakenly being treated as reliable. To evaluate this possibility, we adapted the investment-transmission experiment described in Section 5.1 to allow transmitters the option of not recording a voice message.

Transmitters are given the same incentives as in the main investment design: their payoff is higher the greater the realized payoff of the investment made by their listener. As before, they are told that the listener will receive no information about the company other than their voice message. We explicitly tell respondents they have the option of not making a recording if they believe that the original message does not contain information that would be useful to pass on.

Appendix Figure A19 plots the share of transmitters choosing to record a message in each of the four level × certainty quadrants for the original message. Overall, transmitters choose to record a message 81% of the time, and this frequency barely varies across quadrants; when the original message is high-certainty, transmitters are more likely to record a message by a statistically insignificant 2.7 percentage points. When transmitters do choose to pass on a message, they mention reliability 42% of the time according to our GPT classification, somewhat lower than in the main investment experiment in Section 5.1.

The insensitivity of transmitters' sharing decisions to the certainty level of the original message reflects two factors. First, the certainty or reliability of the original information is not top-of-mind for transmitters deciding whether to record a message (see Appendix Table A8). This finding is consistent with evidence on the sharing of fake news on social media collected by Pennycook et al. (2021). Second, both in our experimental context and in the real world, choosing not to pass on a message is an ineffective remedy for the phenomenon we document. Declining to pass on a message avoids the potential downside of causing low-reliability information to be mistakenly interpreted as high-reliability (or vice versa), but it also means completely depriving the listener of any of the original information. This is reflected in the strong majority view among transmitters that passing on a message would be helpful to the listener.

**Quantitative communication.** Our baseline experiment used purely qualitative scripts because this imitates the majority of real-world communication. However, many important situations do involve the transmission of quantitative predictions or statements of numerical subjective probabilities. We therefore examine the robustness of our results to the addition of numerical statements about level and reliability to our original scripts. The experimental design is virtually identical to our baseline but adds quantitative information about both level and reliability to the original scripts. Quantitative information about the level is conveyed by providing a point estimate of the change in revenue growth. Quantitative reliability information is communicated via a probabilistic confidence statement. This experiment has the added benefit of alleviating potential concerns that our baseline results are driven by people perceiving our level manipulations as "more binary" or "more qualitative" than our reliability manipulations, and finding it easier to pass on binary or qualitative information. By communicating both level and reliability in ex-

actly the same way at one point in the transcripts (through a single numerical percentage, e.g., a 10% increase in house price growth and a 90% confidence level), this experiment minimizes extraneous differences in the way level and reliability information are communicated.

Panels (a) and (b) of Appendix Figure A20 show that differential loss *strengthens* under this design: level information loss falls by nearly two-thirds, to 12.5%, while reliability loss remains 89.0%. A formal test of equality of the two information loss statistics rejects the null at $p < 0.001, \chi^2 = 22.29$.

**Transmission incentives based on the movement of belief distributions.** Transmitters in our baseline experiment were incentivized to transmit messages that matched, as closely as possible, the mean belief updates induced by listening to the original messages. A potential concern is that this focus on *mean* belief updates might bring level information to mind more strongly than reliability information, whereas a broader focus on the *distribution* of beliefs would bring reliability to mind more easily.

To address this possibility, we replicate our main results with an alternative incentive scheme. Rather than eliciting listeners' mean beliefs, we elicit their full distribution of beliefs using a graphical interface (Crosetto and De Haan, 2023). Transmitters are shown this interface, told how listeners will use it, and told that their probability of receiving the bonus payment depends on a measure of the distance between the average belief distribution induced by their message and the average belief distribution induced by the original message.

Panels (a) and (b) of Appendix Figure A21 show that differential information loss is unchanged under this design, at 33% for level and 88% for reliability, underscoring the robustness of differential loss of reliability information. A formal test of equality of the two information loss statistics rejects the null at $p < 0.01, \chi^2 = 6.99$.

# 6 Conclusion

Our economic decisions often rely on information sourced from others through verbal communication. Does the process of verbal transmission systematically distort economic information? We conduct a series of tightly controlled experiments to answer this question. Our experiments show that different types of information are subject to different degrees of transmission loss: the reliability of a prediction dissipates much more in the transmission process than the prediction's level, reducing the influence of reliable information on beliefs, both in absolute terms and relative to unreliable information. Mechanism experiments demonstrate that reliability information is lost in transmission largely because it fails to come to mind during the transmission process, not because of gaps in perceived benefits or costs of transmitting level versus reliability information. A simple associative memory framework suggests that reliability is less likely to come to mind because it is less likely to be cued in everyday contexts. Examinations of everyday conversations and economic TV news yield support for this possibility.

A striking feature of our evidence on cues is how little it takes to induce reliability transmission. Payoff-irrelevant prefixes in an unrelated recording, indirect linguistic cues in a conversation partner's question, and even yesterday's cable news coverage all shift reliability transmission

Table 2: Comparison of Results across Incentive Schemes and Experimental Designs

| | | % Scripts | | Loss (SE) | | | |
|---|---|---|---|---|---|---|---|
| **Experiment** | **Incentive Scheme** | **L** | **R** | **L** | **R** | $p$ | **Key Feature** |
| ***Panel A.*** *Baseline* | | | | | | | |
| Belief Movement | Mean belief change | 95 | 45 | 35.5 (4.4) | 90.1 (4.5) | <.001 | Core design |
| ***Panel B.*** *Alternative incentive schemes* | | | | | | | |
| Implicit Content | Holistic fidelity | 96 | 44 | 31.0 (7.8) | 73.7 (5.9) | <.001 | Content transm. incentivized |
| Explicit Content | Level & reliab. ratings | 94 | 48 | 35.6 (5.2) | 65.3 (9.0) | <.001 | Equal L&R incentive |
| Belief Distribution | Distribution shift | 85 | 29 | 32.9 (18.3) | 87.6 (7.5) | <.01 | Beyond mean beliefs |
| Investment Payoffs | Shared listener payoff | 89 | 52 | 14.5 (8.8) | 77.4$^{\dagger}$ (10.3) | <.001 | Aligned incentives |
| ***Panel C.*** *Design variations* | | | | | | | |
| Quantitative Scripts | Baseline scheme | 91 | 53 | 12.5 (15.5) | 89.0 (5.0) | <.001 | Numerical statements |
| ***Panel D.*** *Mechanism experiment* | | | | | | | |
| High Salience | Explicit content + cue | 85 | 80 | 52.5 (4.9) | 39.1 (7.4) | <.05 | Salient cues during transmission |

*Notes.* L = level, R = reliability. "% Scripts" = share of transmitted messages classified by GPT as containing level/reliability statements. Losses (%, SE in parentheses) from message beliefs via change-in-slope (Eq. 1); $p$ tests equality of L and R losses. All designs are 2 × 2 in level (high/low) × reliability (strong/weak). $^{\dagger}$Certainty/uncertainty in the investment experiment.

today. This sensitivity to subtle, incidental contextual factors — combined with the relative insensitivity to incentives delivered in advance — is the signature of a representational mechanism in the sense of Bordalo et al. (2025b): reliability is absent from the default representation of the transmission problem but easily brought back in by context. The corresponding implication is that interventions to improve how reliable information diffuses need not strengthen incentives so much as change the representation, though our results show this can come at the cost of level transmission.

Our findings speak to a variety of real-world phenomena, including viral fake news, persistent belief polarization, and failures of expert communication. The scope of our experimental results is, by necessity, limited. Longer delays between information consumption and transmission, multiple stages of transmission, and the richer set of incentives facing communicators in the real world may cause *even more extreme* differential information loss than what we document. Participants in our experiment are incentivized to faithfully convey the original information; by contrast, real-world communicators often face incentives to entertain, persuade, or impress, all of which may militate in favor of dropping caveats, admissions of uncertainty, or other reliability indicators. While incentives to persuade or entertain might result in even stronger reliability loss in real-world contexts, this may be counterbalanced by other factors, like the ability of listeners to ask follow-up questions probing the transmitter's confidence.

# References


**ABC News**, "The Three Hour Diet," 2005.

**Akinnaso, F Niyi**, "On the differences between spoken and written language," *Language and speech*, 1982, *25* (2), 97–125.

**Akçay, Erol and David Hirshleifer**, "Social Finance as Cultural Evolution, Transmission Bias and Market Dynamics," *Proceedings of the National Academy of Sciences*, forthcoming, *7* (43).

**Andre, Peter, Ingar Haaland, Christopher Roth, Mirko Wiederholt, and Johannes Wohlfart**, "Narratives about the Macroeconomy," *Review of Economic Studies*, 2025.

**Asirvatham, Hemanth, Elliott Mokski, and Andrei Shleifer**, "GPT as a Measurement Tool," NBER Working Paper 34834, National Bureau of Economic Research February 2026.

**Atkin, David, M Keith Chen, and Anton Popov**, "The returns to face-to-face interactions: Knowledge spillovers in Silicon Valley," Technical Report, National Bureau of Economic Research 2022.

**Augenblick, Ned, Eben Lazarus, and Michael Thaler**, "Overinference from weak signals and underinference from strong signals," *Quarterly Journal of Economics*, 2025.

**Ba, Cuimin, J Aislinn Bohren, and Alex Imas**, "Over-and underreaction to information," *Available at SSRN*, 2024.

**Baker, Scott R, Nicholas Bloom, and Steven J Davis**, "Measuring economic policy uncertainty," *The quarterly journal of economics*, 2016, *131* (4), 1593–1636.

**Banerjee, Abhijit, Arun G Chandrasekhar, Esther Duflo, and Matthew O Jackson**, "The Diffusion of Microfinance," *Science*, 2013, *341* (6144), 1236498.

**_ , _ , _ , and _** , "Using gossips to spread information: Theory and evidence from two randomized controlled trials," *The Review of Economic Studies*, 2019, *86* (6), 2453–2490.

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

**Barron, Kai and Tilman Fries**, "Narrative Persuasion," Technical Report, WZB Discussion Paper 2024.

**Bartlett, Frederic Charles**, *Remembering: A study in experimental and social psychology*, Cambridge university press, 1995.

**Batista, Rafael M., Juliana Schroeder, Aastha Mittal, and Sendhil Mullainathan**, "Misarticulation: Why We Sometimes Feel Our Words Don't Match Our Thoughts," *Working Paper*, 2024.

**Battiston, Diego, Jordi Blanes i Vidal, and Tom Kirchmaier**, "Face-to-face communication in organizations," *The Review of Economic Studies*, 2021, *88* (2), 574–609.

**Bengen, William P.**, "The 4% Rule," `https://www.bengenfs.com/the-4-percent-rule/` 2024. Accessed July 26, 2025.

**Berger, Jonah and Raghuram Iyengar**, "Communication channels and word of mouth: How the medium shapes the message," *Journal of consumer research*, 2013, *40* (3), 567–579.

**Bikhchandani, Sushil, David Hirshleifer, and Ivo Welch**, "A theory of fads, fashion, custom, and cultural change as informational cascades," *Journal of Political Economy*, 1992, *100* (5), 992–1026.

**Bordalo, Pedro, Giovanni Burro, Nicola Gennaioli, Gad Nacamulli, and Andrei Shleifer**, "Ads as Cues," Working Paper 34387, National Bureau of Economic Research October 2025.

__ , **John J Conlon, Nicola Gennaioli, Spencer Y Kwon, and Andrei Shleifer**, "Memory and probability," *The Quarterly Journal of Economics*, 2023, *138* (1), 265–311.

__ , **John J. Conlon, Nicola Gennaioli, Spencer Yongwook Kwon, and Andrei Shleifer**, "How People Use Statistics," *Review of Economic Studies*, 2026, *93* (1), 250–285.

__ , **Katherine Coffman, Nicola Gennaioli, Frederik Schwerter, and Andrei Shleifer**, "Memory and Representativeness.," *Psychological Review*, 2021, *128* (1), 71.

__ , **Nicola Gennaioli, Florencio Lopez de Silanes, Simon G. Schröder, Andrei Shleifer, and Maarten van Rooij**, "The Psychology of Macroeconomic Expectations," 2026. Unpublished manuscript.

__ , __ , **Giacomo Lanzani, and Andrei Shleifer**, "A cognitive theory of reasoning and choice," Technical Report, National Bureau of Economic Research 2025.

**Bourdin, Beatrice and Michel Fayol**, "Even in adults, written production is still more costly than oral production," *International journal of Psychology*, 2002, *37* (4), 219–227.

**Bursztyn, Leonardo, Florian Ederer, Bruno Ferman, and Noam Yuchtman**, "Understanding mechanisms underlying peer effects: Evidence from a field experiment on financial decisions," *Econometrica*, 2014, *82* (4), 1273–1301.

**Carlson, Taylor N**, "Modeling political information transmission as a game of telephone," *The Journal of Politics*, 2018, *80* (1), 348–352.

__ , "Through the grapevine: Informational consequences of interpersonal political communication," *American Political Science Review*, 2019, *113* (2), 325–339.

**Chafe, Wallace and Deborah Tannen**, "The relation between written and spoken language," *Annual review of anthropology*, 1987, *16* (1), 383–407.

**Conlon, John J and Spencer Y Kwon**, "Beliefs from Cues," 2025.

__ , **Malavika Mani, Gautam Rao, Matthew W Ridley, and Frank Schilbach**, "Not Learning from Others," Technical Report, National Bureau of Economic Research 2025.

**Crosetto, Paolo and Thomas De Haan**, "Comparing input interfaces to elicit belief distributions," *Judgment and Decision making*, 2023, *18*, e27.

**Elminejad, Ali, Tomas Havranek, and Zuzana Irsova**, "Relative risk aversion: A meta-analysis," *Journal of Economic Surveys*, 2025, *39* (5), 2315–2333.

**Enke, Benjamin and Thomas Graeber**, "Cognitive Uncertainty," *Quarterly Journal of Economics*, 2023.

__ , __ , **Ryan Oprea, and Jeffrey Yang**, "Behavioral attenuation," Technical Report, National Bureau of Economic Research 2025.

**Eyal, Peer, Rothschild David, Gordon Andrew, Evernden Zak, and Damer Ekaterina**, "Data quality of platforms and panels for online behavioral research," *Behavior Research Methods*, 2021, pp. 1–20.

**Fischhoff, Baruch and Wändi Bruine De Bruin**, "Fifty–fifty= 50%?," *Journal of Behavioral Decision Making*, 1999, *12* (2), 149–163.

**Gennaioli, Nicola and Andrei Shleifer**, "What Comes to Mind," *Quarterly Journal of Economics*, 2010, *125* (4), 1399–1433.

**Golub, B and ED Sadler**, "Learning in Social Networks," in "The Oxford Handbook of the Economics of Networks," Oxford University Press, 2016.

**Golub, Benjamin and Matthew O Jackson**, "Naive learning in social networks and the wisdom of crowds," *American Economic Journal: Microeconomics*, 2010, *2* (1), 112–149.

**Gompert, David C, Hans Binnendijk, and Bonny Lin**, "The US invasion of Iraq, 2003," *Blinders, Blunders, and Wars*, 2014, pp. 161–74.

**Gower, Timothy**, "Going Strong," 2002.

**Graeber, Thomas**, "Inattentive Inference," *Journal of the European Economic Association*, 2023, *21* (2), 560–592.

**_ , Christopher Roth, and Constantin Schesch**, "Explanations," 2024.

**_ , _ , and Florian Zimmermann**, "Stories, Statistics, and Memory," *Quarterly Journal of Economics*, 2024, *139* (4), 2181–2225.

**Griffin, Dale and Amos Tversky**, "The Weighing of Evidence and the Determinants of Confidence," *Cognitive psychology*, 1992, *24* (3), 411–435.

**Han, Bing, David Hirshleifer, and Johan Walden**, "Social Transmission Bias and Investor Behavior," *Journal of Financial and Quantitative Analysis*, forthcoming.

**Hartzmark, Samuel M, Samuel D Hirshman, and Alex Imas**, "Ownership, learning, and beliefs," *The Quarterly journal of economics*, 2021, *136* (3), 1665–1717.

**Henkel, Linda A and Nancy Franklin**, "Reality monitoring of physically similar and conceptually related objects," *Memory & Cognition*, 1998, *26* (4), 659–673.

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

**Hovland, Carl I and Walter Weiss**, "The influence of source credibility on communication effectiveness," *Public opinion quarterly*, 1951, *15* (4), 635–650.

**Johnson, Marcia K**, "Source monitoring and memory distortion," *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 1997, *352* (1362), 1733–1745.

**_ , Shahin Hashtroudi, and D Stephen Lindsay**, "Source monitoring.," *Psychological bulletin*, 1993, *114* (1), 3.

**Kahana, Michael Jacob**, *Foundations of human memory*, OUP USA, 2012.

**Kahneman, Daniel**, *Thinking, Fast and Slow*, New York: Farrar, Straus and Giroux, 2011.

**Kendall, Chad W and Constantin Charles**, "Causal Narratives," Technical Report, National Bureau of Economic Research 2025.

**Lindsay, D Stephen and Marcia K Johnson**, "The eyewitness suggestibility effect and memory for source," *Memory & Cognition*, 1989, *17* (3), 349–358.

**Love, Robbie, Claire Dembry, Andrew Hardie, Vaclav Brezina, and Tony McEnery**, "The Spoken BNC2014: Designing and building a spoken corpus of everyday conversations," *International Journal of Corpus Linguistics*, 2017, *22* (3), 319–344.

**Massey, Cade and George Wu**, "Detecting regime shifts: The causes of under-and overreaction," *Management Science*, 2005, *51* (6), 932–947.

**Mesoudi, Alex and Andrew Whiten**, "The multiple roles of cultural transmission experiments in understanding human cultural evolution," *Philosophical Transactions of the Royal Society B: Biological Sciences*, 2008, *363* (1509), 3489–3501.

**Mobius, Markus and Tanya Rosenblat**, "Social learning in economics," *Annu. Rev. Econ.*, 2014, *6* (1), 827–847.

**Pennycook, Gordon, Ziv Epstein, Mohsen Mosleh, Antonio A Arechar, Dean Eckles, and David G Rand**, "Shifting attention to accuracy can reduce misinformation online," *Nature*, 2021, *592* (7855), 590–595.

**Serra-Garcia, Marta**, "The Attention–Information Tradeoff," 2025.

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

**Sperber, Dan, Fabrice Clément, Christophe Heintz, Olivier Mascaro, Hugo Mercier, Gloria Origgi, and Deirdre Wilson**, "Epistemic vigilance," *Mind & Language*, 2010, *25* (4), 359–393.

**Stanovich, Keith E**, "Miserliness in human cognition: The interaction of detection, override and mindware," *Thinking & Reasoning*, 2018, *24* (4), 423–444.

**Stillman, Irwin Maxwell and Samm Sinclair Baker**, *The Doctor's Quick Weight Loss Diet*, Prentice-Hall, 1967.

**Tai, Mary M., Peter Castillo, and F. Xavier Pi-Sunyer**, "Meal size and frequency: effect on the thermic effect of food," *The American Journal of Clinical Nutrition*, 1991, *54* (5), 783–787.

**Thaler, Michael**, "The supply of motivated beliefs," *arXiv preprint arXiv:2111.06062*, 2025.

**__ , Mattie Toma, and Victor Yaneng Wang**, "Numbers Tell, Words Sell," *Working Paper*, 2025.

**Tversky, Amos and Daniel Kahneman**, "Judgment under Uncertainty: Heuristics and Biases: Biases in judgments reveal some heuristics of thinking under uncertainty.," *science*, 1974, *185* (4157), 1124–1131.

**Vogel, Hannah, Sebastian Appelbaum, Heidemarie Haller, and Thomas Ostermann**, "The interpretation of verbal probabilities: A systematic literature review and meta-analysis," *German Medical Data Sciences 2022–Future Medicine: More Precise, More Integrative, More Sustainable!*, 2022, pp. 9–16.

**Weizsäcker, Georg**, "Do we follow others when we should? A simple test of rational expectations," *American Economic Review*, 2010, *100* (5), 2340–2360.

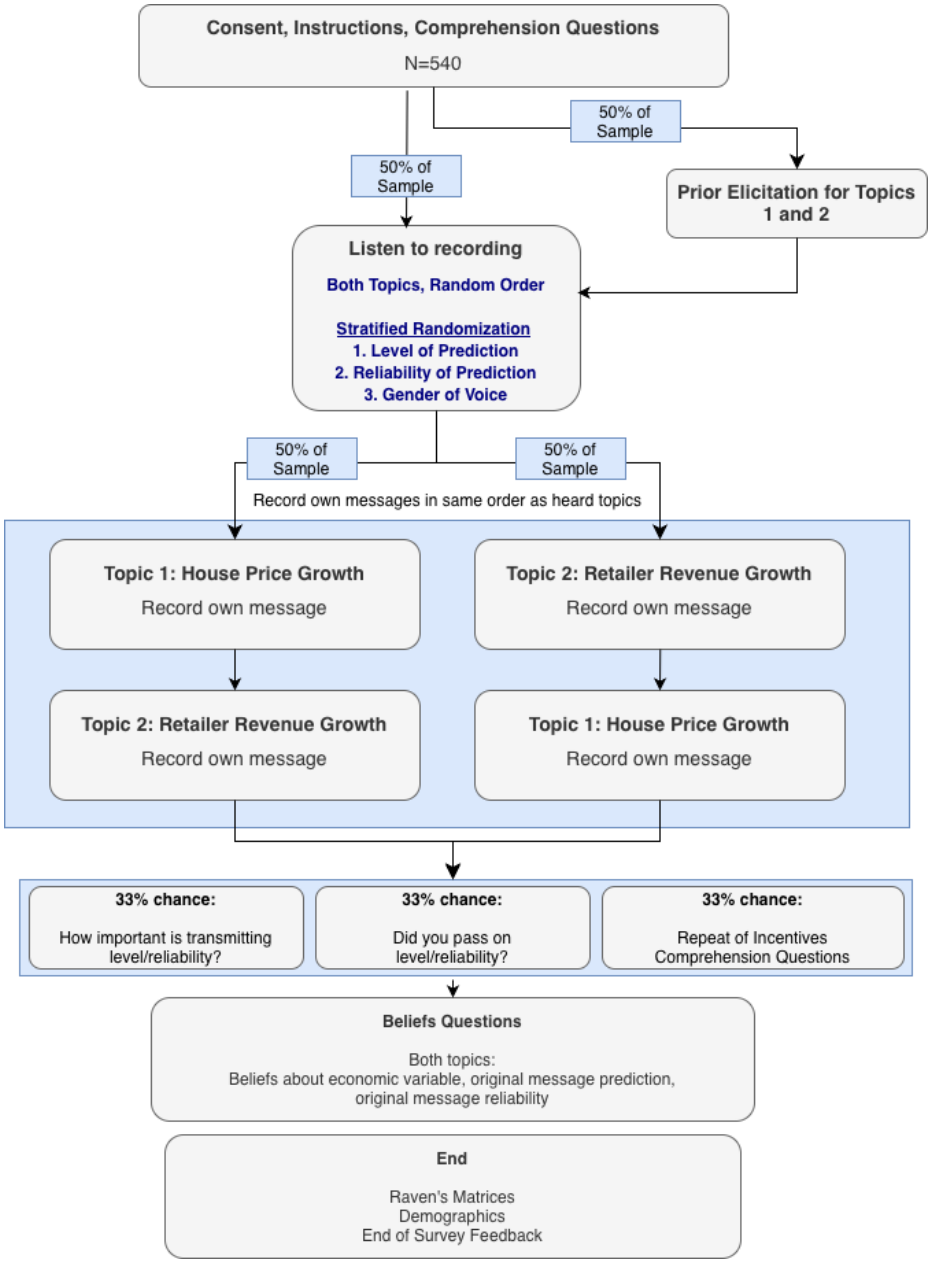
# Online Appendix

The Transmission of Reliable and Unreliable Information

Thomas Graeber, Shakked Noy, and Christopher Roth

## A Additional Exhibits

**a)** Transmitter Experimental Design




Appendix Figure A1: This figure shows the design of our baseline transmitter experiment.

**a)** Listener Experimental Design

Appendix Figure A2: This figure shows the design of our baseline listener experiment.

Appendix Table A1: Overview of Main Data Collections

| Role | Incentive scheme | Sample | Additional design features | Main outcomes | AEA RCT # |
|---|---|---|---|---|---|
| **Baseline experiments** | | | | | |
| Transmitter | Belief movement | Prolific (540) | None | Speech content; beliefs about originator level and reliability | 12119 |
| Listener | Belief movement | Prolific (1,510) | Original vs. transmitted recording | Beliefs about state; beliefs about originator level and reliability | 12119 |
| **Mechanism experiments** | | | | | |
| **Content transmission incentives** | | | | | |
| Transmitter | Content transmission | Prolific (501) | Explicit incentives for content transmission | Speech content; beliefs about state; beliefs about originator | 12119 |
| Listener | Content transmission | Prolific (1,509) | Original vs. transmitted recording | Beliefs about state; beliefs about originator level and reliability | 12119 |
| **Choice of incentives** | | | | | |
| Transmitter | Choice of incentives | Prolific (97) | Transmitters choose which information to emphasize | Incentive choice; perceived difficulty of transmitting level vs. reliability | 12119 |
| **High salience of reliability** | | | | | |
| Transmitter | High salience | Prolific (244) | Salient reminders of transmission incentives | Speech content; beliefs about originator level and reliability | 12119 |
| Listener | High salience | Prolific (1,010) | Original vs. transmitted recording | Beliefs about state; beliefs about originator level and reliability | 12119 |
| **Robustness experiments** | | | | | |
| **Quantitative scripts** | | | | | |
| Transmitter | Average belief movement | Prolific (181) | Scripts contain quantitative information on level and reliability | Speech content; beliefs about originator level and reliability | 12119 |
| Listener | Average belief movement | Prolific (834) | Original vs. transmitted recording | Beliefs about state; beliefs about originator level and reliability | 12119 |
| **Belief distribution incentives** | | | | | |
| Transmitter | Belief distribution | Prolific (225) | Incentives target belief distributions | Speech content; beliefs about originator level and reliability | 17479 |
| Listener | Belief distribution | Prolific (892) | Original vs. transmitted recording | Belief distributions over state; beliefs about originator level and reliability | 17479 |
| **Investment incentives** | | | | | |
| Transmitter | Investment incentives | Prolific (299) | Incentives tied to listener investment payoff | Speech content; beliefs about originator level and reliability | 17533 |
| Listener | Investment incentives | Prolific (1,287) | Original vs. transmitted recording | Investment choice; beliefs about state; beliefs about originator level and reliability | 17533 |
| **Extensive margin experiment** | | | | | |
| Transmitter | Investment incentives | Prolific (296) | Optional transmission decision | Transmission choice; speech content (if transmitted) | 17540 |

*Notes.* All experiments implement a $2 \times 2$ design that varies the level of the economic variable (high vs. low) and the reliability of the original message (high vs. low). Sample sizes refer to respondents who completed the survey and satisfied pre-specified inclusion criteria.

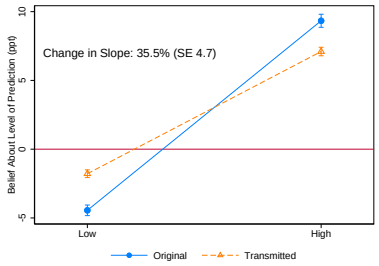
Appendix Table A2: Example Transmitted Messages

| Handcoded Classification | Message Text |
|---|---|
| Passed on level and reliability | In the second recording. Um it was stated very confidently that prices of houses is going to go down and that there's very good um scientific evidence for this. And also that right now, there is a huge difference between um mortgage rates and house prices. |
| Passed on level and reliability | The retail company in question sells things at a lower price than its competitors. And because of the current climate, that's something that appeals to most people at this time. However, this sort of thing is not that easily predicted. So though my prediction is that the companies growth, they will grow, they will be positive for them. It's not guaranteed. |
| Passed on level but not reliability | The price of a home will continue to rise throughout the next year. Not only due to rising interest rates in order to obtain a mortgage, but for the cost to build a new home and obtain permits for building the home as well as the materials required. |
| Passed on level but not reliability | Ok. This prediction is on the change in revenue growth of a large U.S. retail company and specifically this U.S. retail company operates in the budget friendly market is affordable to consumers. And with that in mind, we have to consider that interest rates are the driving force in this economy. Interest rates affect the consumers in the, it affects their debt and with a higher interest rates, their interest costs are often increased and increase their overall debt. And that means the discretionary income is reduced. And when consumers have less discretion, discretionary income, they look toward, uh, they look toward retailers that of affordable and price friendly merchandise. And that means this particular U.S. retail company that operates with a niche in budget friendly prices will lead to a higher revenue growth in the upcoming year. |
| Passed on reliability but not level | Oh, I love you. The change in uh revenue growth of retail companies um was a little difficult to understand in the second message. She didn't sound really confident and kind of jumped around a bit and then even gave her own kind of confirmation bias by what she was hearing up a bar by random guys, but things that she didn't even really understand. Um So something about, you know, as banks print money, there's more money available which takes the value of the dollar, meaning prices go up because it's not as valuable anymore. Um That's kind of the general gist I got of it is over, flood of money means the drive up of prices because it's just not valuable. |

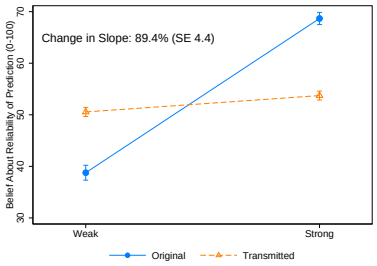
## A.1 Additional Figures

### A.1.1 Baseline Experiment: Belief Movement Incentives

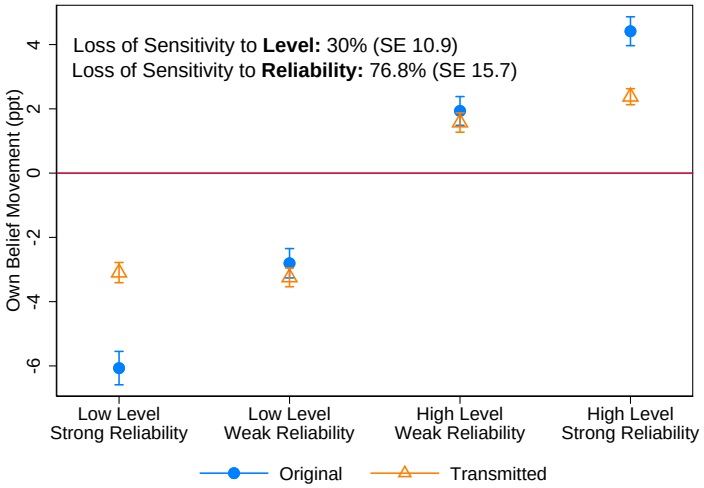
**a)** Level Message Beliefs: Not Z-Scored




**b)** Reliability Message Beliefs: Not Z-Scored

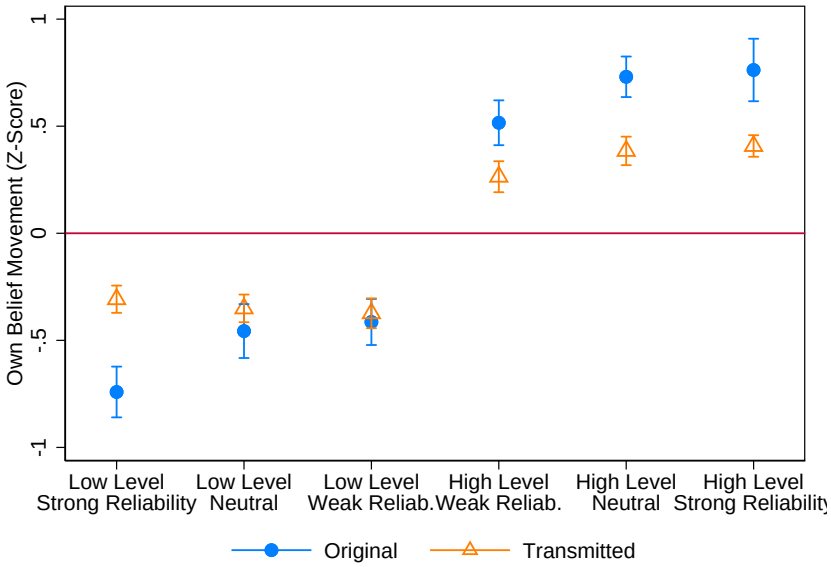



**c)** State Belief Updates: Not Z-Scored



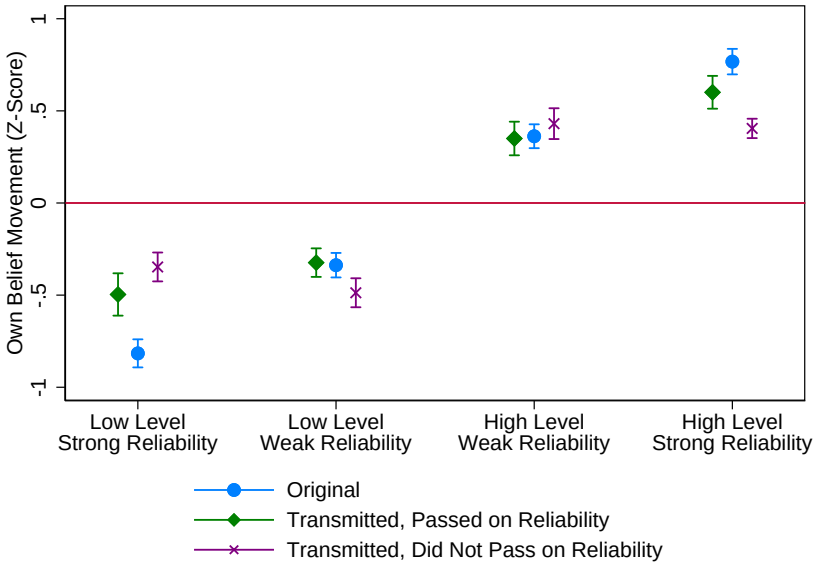

Appendix Figure A3: This figure presents data from our baseline experiment (Belief Movement Incentives). It replicates Figures 2 and 3 using raw (rather than z-scored) beliefs.

**a)** Belief Updates About Economic Variable:
Modular Manipulation Only, Including Neutral Reliability Conditions



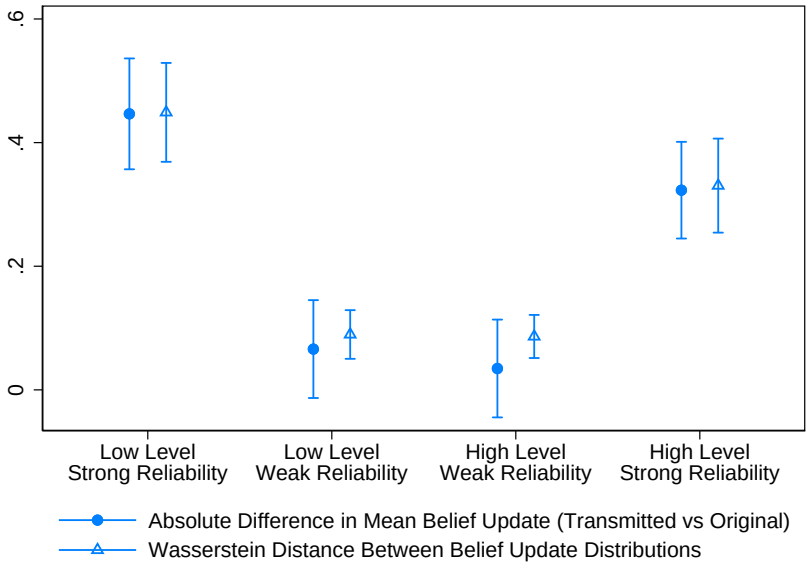

Appendix Figure A4: This figure presents data from our baseline experiment (Belief Movement Incentives). It is an alternative version of Panel (c) of Figure 3. It shows the average belief updates of listeners, restricting to the Modular reliability manipulation, which has a weak-reliability, strong-reliability, and neutral-reliability condition (the last of which simply omits the uncertainty- or certainty- denoting prefixes and statements that constitute the first two manipulations).

**a)** Belief Updates About Economic Variable, by Reliability Transmission



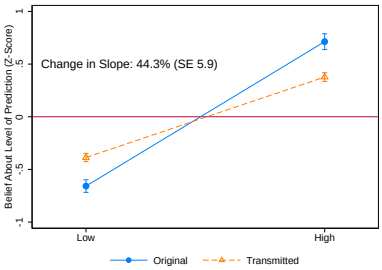

Appendix Figure A5: This figure presents data from our baseline experiment (Belief Movement Incentives). It is an alternative version of Panel (c) of Figure 3. It shows the average belief updates of listeners, splitting listeners who hear transmitted recordings by whether the transmitted recording is unanimously considered by our handcoders to have passed on reliability (green diamonds) or is unanimously considered to have not passed on reliability but passed on the level (purple X's).

**a)** Belief Updates About Economic Variable: Mean Difference vs Wasserstein Distance

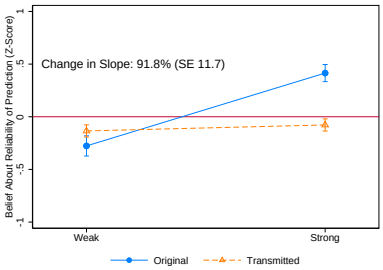



Appendix Figure A6: This figure presents data from our baseline experiment (Belief Movement Incentives). It is an alternative version of Panel (c) of Figure 3. It shows the absolute mean difference in belief updates between listeners who directly heard the original messages versus those who heard transmitted versions, as well as the Wasserstein distance between the two distributions of belief updates, in each quadrant of our level × reliability manipulation. Standard errors for the Wasserstein distance are bootstrapped.

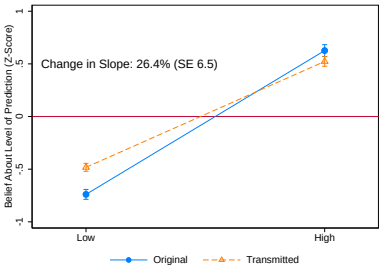
**a)** Level Info Loss: Modular Manipulation



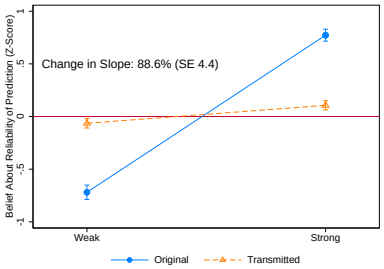

**b)** Reliability Info Loss: Modular Manipulation




**c)** Level Info Loss: Naturalistic Manipulation

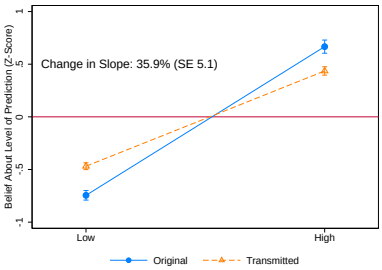



**d)** Reliability Info Loss: Naturalistic Manipulation

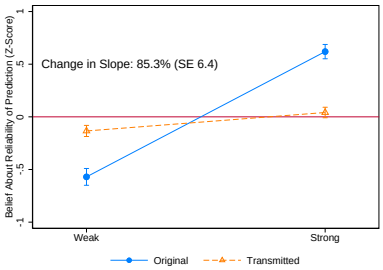



Appendix Figure A7: This figure presents data from our baseline experiment (Belief Movement Incentives). It replicates Figure 2, showing beliefs about the level and reliability of the original prediction, separately by respondents in our *modular* versus *naturalistic* reliability manipulations.

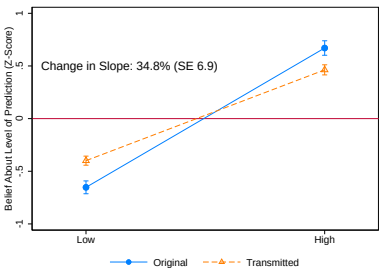
**a)** Level Info Loss: No Incentives



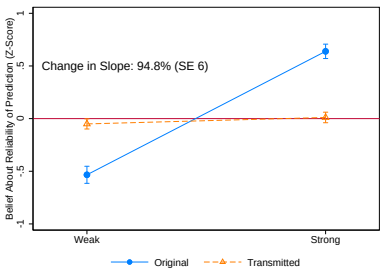

**b)** Reliability Info Loss: No Incentives




**c)** Level Info Loss: Second-Order Incentives

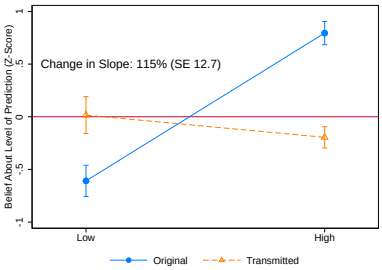



**d)** Reliability Info Loss: Second-Order Incentives

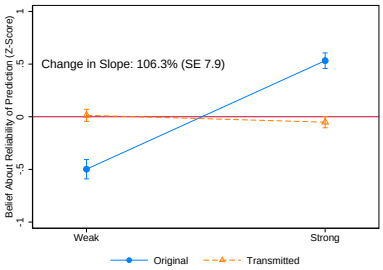



Appendix Figure A8: This figure presents data from our baseline experiment (Belief Movement Incentives). It replicates Figure 2, showing beliefs about the level and reliability of the original prediction, separately by respondents who are asked these questions directly and not incentivized, compared to respondents who are asked these as second-order belief questions and incentivized according to how closely they match the average beliefs of the unincentivized respondents.

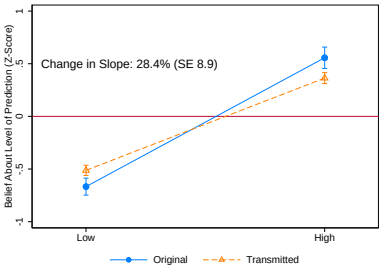
**a)** Level Info Loss: Not Passed On



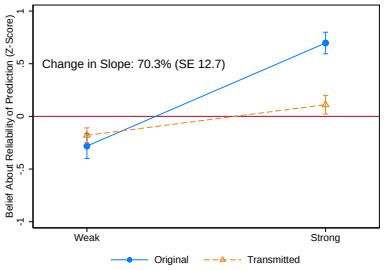

**b)** Reliability Info Loss: Not Passed On




**c)** Level Info Loss: Passed On

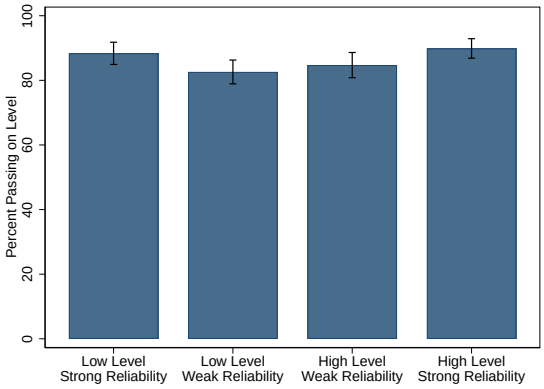



**d)** Reliability Info Loss: Passed On

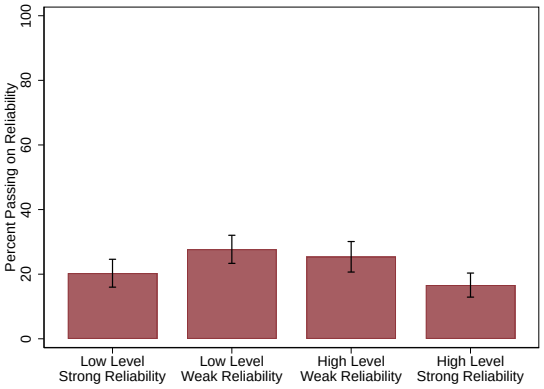



Appendix Figure A9: This figure presents data from our baseline experiment (Belief Movement Incentives). It replicates Figure 2, showing beliefs about the level and reliability of the original prediction. Panels (a) and (b) restrict to recordings that both human coders and GPT-4 unanimously agree do *not* contain information about the level (Panel (a)) or reliability (Panel (b)). Panels (c) and (d) restrict to recordings that are unanimously agreed to contain information about the level or reliability.

**a)** Fraction of Scripts Containing Statements about **Level**



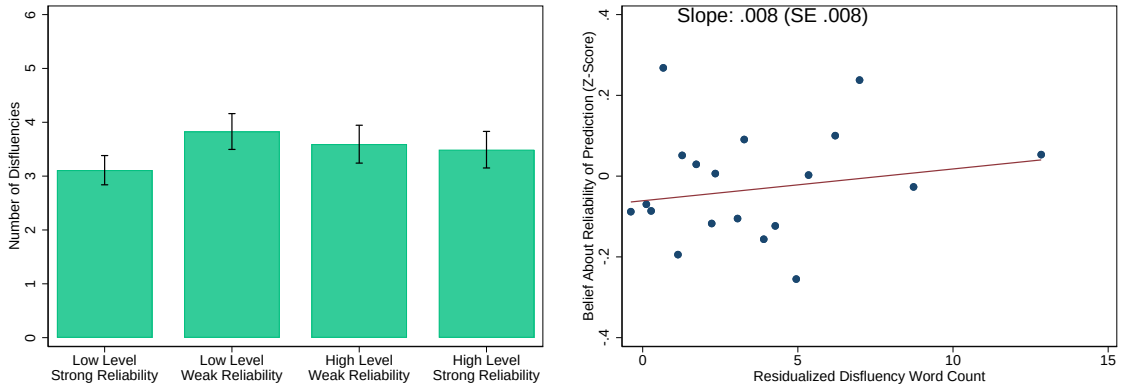

**b)** Fraction of Scripts Containing Statements about **Reliability**




Appendix Figure A10: This figure disaggregates Figure 1 by the four conditions in our level × reliability manipulation. It shows the percent of transmitted messages that are unanimously classified by our coders as containing statements about level or reliability.

**a)** Average Number of Disfluencies in Transmitted Scripts

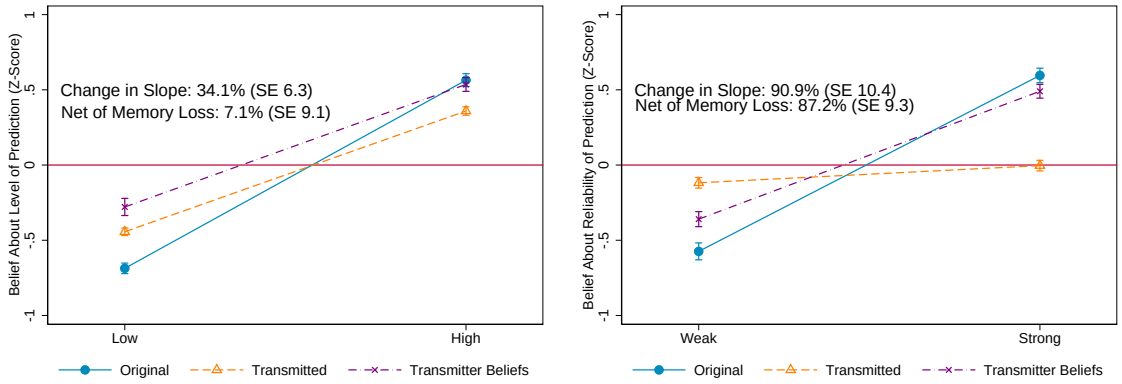
**b)** Disfluencies Do Not Predict Listener Beliefs




Appendix Figure A11: This figure looks at disfluencies (hesitations, "um" statements), automatically counted by GPT-4 and encompassing various kinds of disruptions in the flow of the transmitted transcripts. Panel (a) plots the average number of disfluencies per transmitted script by the four conditions in our level × reliability manipulation. Panel (b) shows a binned scatter plot of listener beliefs about the reliability of the original message on the number of disfluencies in the transmitted recording's transcript, controlling for the transmitted message's overall word count and topic fixed effects.

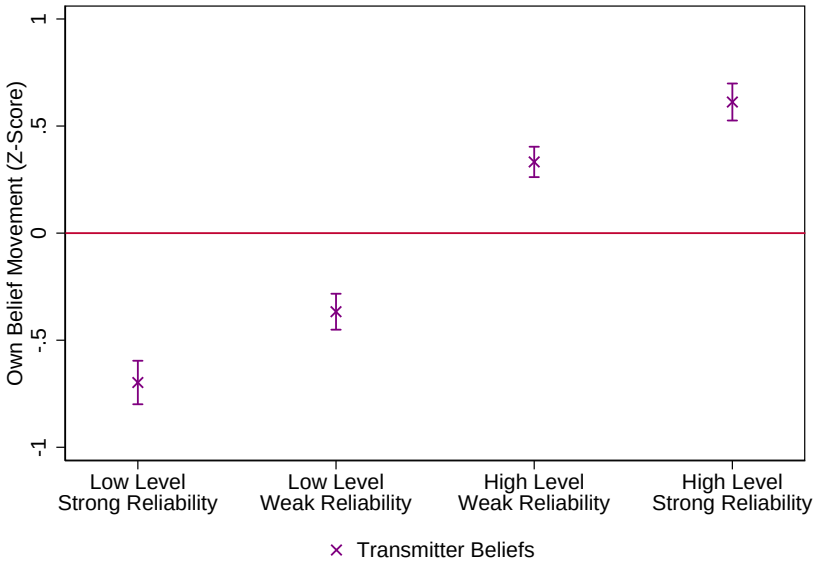
**a)** Transmission of **Level**

**b)** Transmission of **Reliability**




Appendix Figure A12: This figure presents data from our baseline experiment (Belief Movement Incentives). It replicates Figure 2 but adds a line representing the beliefs of the *transmitters* who create the recordings. The "net of memory loss" statistics compare the orange line to the purple line instead of the blue line. Beliefs in this case are z-scored *after* pooling transmitters' beliefs into the sample.

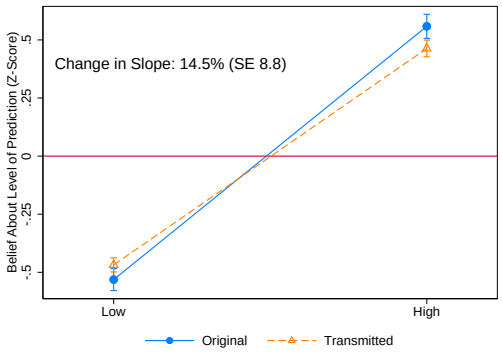
**a)** Belief Updates About Economic Variable: Transmitters



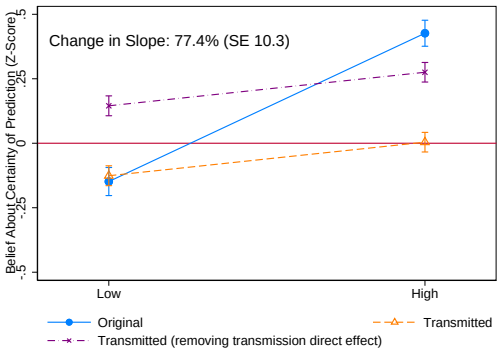

Appendix Figure A13: This figure replicates Figure 3 using state-beliefs data from the transmitters in our baseline experiment instead of listeners.

### A.1.2 Supplementary Experiment: Investment Task

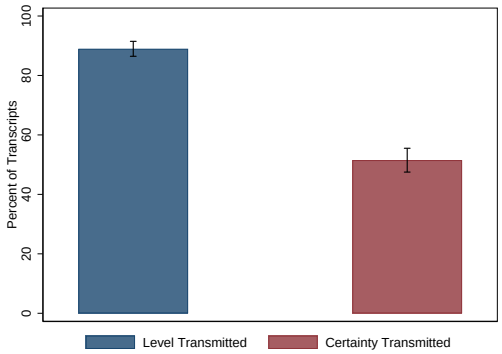
**a)** Beliefs About **Level** of Original Prediction



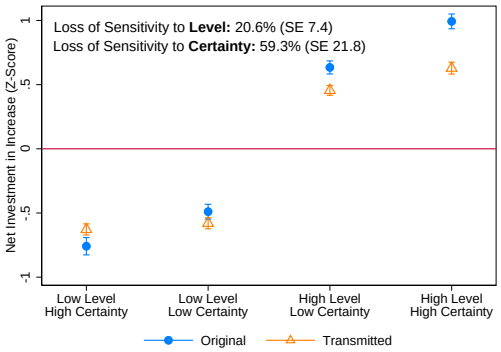

**b)** Beliefs About **Reliability** of Original Prediction




**c)** Fraction of Scripts Containing Level/Certainty Statements (GPT Coding)




**d)** Mean Investment (% Invested in "Increase" Asset Minus % Invested in "Decrease" Asset)




Appendix Figure A14: Slopes in Panels (a) and (b) differ at $p < 0.001, \chi^2 = 20.82$. This figure presents data from our version of the baseline experiment that uses earnings-announcement narratives and listener-investment-based incentives. Panels (a) and (b) replicate Figure 2, showing listeners' beliefs about the level and uncertainty of the prediction in the original message, separately by whether the original message is low- vs high-level or low- versus high-certainty, and separately by whether the listener hears the original message or a transmitted version of it. The purple line in Panel (b) adjusts transmitted reliability beliefs upwards by the gap between the average reliability beliefs of listeners hearing original messages and the average reliability beliefs of listeners hearing transmitted messages that are coded by GPT as not containing any reliability information, controlling for the reliability of the original message. Dots are mean beliefs and bars are standard error bars (1 SE each direction). Panel (c) replicates Figure 1. It shows the fraction of transmitted messages classified by GPT-4 as containing statements about the level or certainty of the original forecast. Panel (d) replicates Figure 3, but using our investment outcome instead of beliefs about the state variable. In the investment task, listeners allocate \$100 between an asset that pays off if the state is higher than expected, one that pays off if the state is roughly as expected, and one that pays off if the state is lower than expected; the outcome we use is the percent invested in the first bucket minus percent invested in the third. $N = 299$ transmitters and 1,287 listeners.

**Interpretation of findings.** Certainty beliefs are distorted less symmetrically in this design than in our baseline experiment, with perceptions of high-certainty messages being attenuated much more strongly. There are two potential explanations for this. One is that the combination of the investment setting and new certainty manipulation raised the salience of uncertainty, which disproportionately drew attention to low-certainty markers. Consistent with this explanation, our *high salience* mechanism experiment in Section 4.3, which directly ramps up the salience of reliability during the transmission task, finds that the elimination of reliability information loss is larger in the low-reliability condition (see Figure 4 Panel (c)). An alternative explanation is that, in this experiment, transmission both (i) symmetrically compressed reliability beliefs towards zero and (ii) shifted those beliefs downwards on average. Below, we find support for this latter explanation by examining LLM handcodings of reliability transmission.

These explanations cannot be separately identified using belief data alone. However, we can try to tease them apart using our LLM codings of the content of transmitted messages. In our baseline experiment, 48% of transmitted messages in the low-reliability condition and 41% in the high-reliability condition are coded by the LLM as passing on reliability information; in this experiment, the corresponding figures are 59% and 45%—just a mild increase in the asymmetry between low- and high-reliability messages.

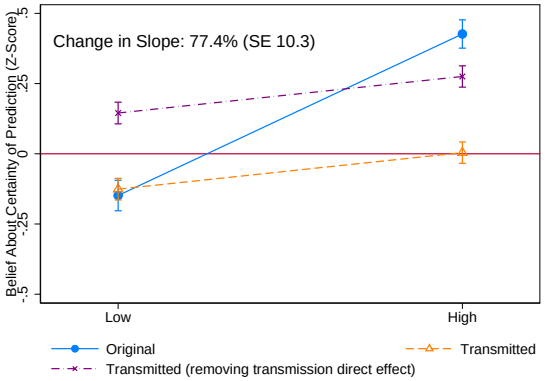
Moreover, a natural benchmark for isolating the direct effects of transmission on reliability beliefs is the subset of transmitted messages that GPT coded as containing neither certainty nor uncertainty markers. Messages in this subset plausibly represent a neutral middle ground, lacking any specific reliability language. If they are nevertheless perceived as less reliable on average than the original messages, this suggests a direct negative effect of transmission on reliability beliefs. (For example, one possibility is that because transmitted messages contain more disfluencies, listeners trust them less on average.)

In our baseline experiment, we find no evidence of such an effect: the reliability beliefs of listeners hearing original messages are the same, on average, as those of listeners hearing transmitted messages with no reliability indicators, suggesting no direct average effect of transmission on reliability beliefs. By contrast, in the investment experiment, the reliability beliefs of listeners hearing original messages are 0.27 standard deviations *higher*, on average, than the reliability beliefs of listeners hearing transmitted messages coded by GPT as not containing reliability indicators.[1] This indicates that transmission reduced the average perceived reliability of messages, perhaps because transmitters in this experiment felt less confident given the more complex finance-related content of the original messages. The purple line in Panel (b) of Figure A14 adjusts transmitted relia-

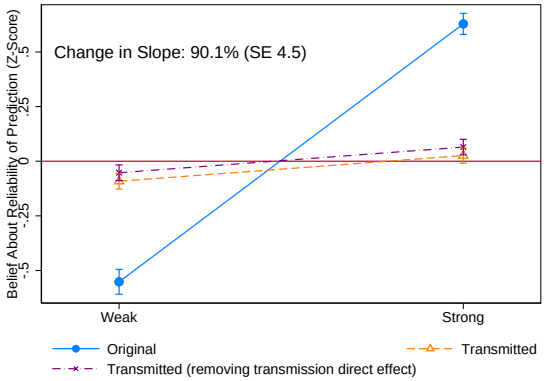
[1]When making these comparisons, we control for whether the original message was high or low reliability to account for the fact that the latter group can be selected on reliability.

bility beliefs upwards by 0.27 SDs to account for this; after this adjustment, reliability loss no longer appears asymmetric. Appendix Figure A15 shows that this adjustment makes a big difference in this experiment, but not in our baseline experiment. Panel (c) shows that it also makes a difference in our high salience experiment, the other place where we find apparently asymmetric reliability loss.

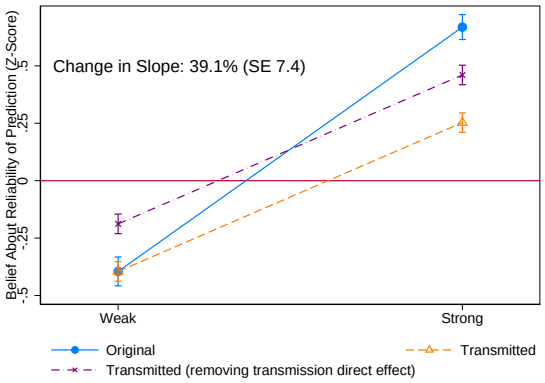
**a)** Adjusting for Transmission-Induced Reductions in Average Reliability in *Investment* Experiment




**b)** Adjusting for Transmission-Induced Reductions in Average Reliability in *Baseline* Experiment




**c)** Adjusting for Transmission-Induced Reductions in Average Reliability in *High Salience* Experiment




Appendix Figure A15: Panel (a) directly replicates Panel (b) of Figure A14. Panel (b) replicates this figure using data from our baseline experiment, and Panel (c) using data from our *high salience* experiment (replicating Panel (c) of Figure 4). In all three panels, the purple line is the orange line adjusted upwards by the gap in average reliability beliefs between (1) listeners directly hearing the original messages and (2) listeners hearing transmitted messages that were coded by GPT as *not* containing any reliability indicators. We calculate this gap controlling for the reliability of the original messages in each case.

**Back-of-the-envelope welfare calculation.** We proceed as follows. We assume listeners have homogeneous CRRA utility with coefficient $\gamma$ and zero outside wealth. The state takes values $\{-10, 0, 10\}$ (these are the values we use in the quantitative-communication

experiment) and listeners have a uniform prior over these possibilities; this is the prior our design tries to induce, where all three possibilities are equally reasonable *ex ante*.

Because we want to compute counterfactuals in which we separately eliminate level and reliability information loss, we map listeners' message beliefs (rather than state beliefs) into investment choices via an implied posterior distribution over $\{-10, 0, 10\}$. We interpret the listener's level message belief $b$ as the expected value of the state assuming the listener fully trusts the original message, and convert this into a probability distribution by assuming it reflects a mixture between 0 and 10 (if $b > 0$) or 0 and $-10$ (if $b < 0$). We treat the listener's certainty message belief $c \in [0, 100]$ as a trust weight $w(c) = c/100$. If $b > 0$, the listener's posterior probabilities on the three possible states are $\{(1-w(c))\frac{1}{3}, (1-w(c))\frac{1}{3} + w(c)(1-\frac{b}{10}), (1-w(c))\frac{1}{3} + w(c)\frac{b}{10}\}$.

We estimate $\hat{\gamma}$ by minimum distance, matching the observed portfolio shares of listeners directly hearing the original messages to the implied optimal shares given the posteriors we derive from their message beliefs. This produces $\hat{\gamma} = 2.09$.

To conduct welfare calculations for listeners hearing transmitted messages, we assume in each level × reliability quadrant that the optimal investment is the investment implied by $\hat{\gamma}$ and the posterior implied by the mean message beliefs of listeners in that quadrant who directly heard original messages.

To compute efficiency losses under belief attenuation, we compare optimal portfolios under the ground-truth correct beliefs to the optimal portfolios chosen given beliefs attenuated by the amounts estimated in our experiment: either attenuating beliefs with respect to variation in level (by shrinking the gap between the low-level and high-level quadrants) or attenuating beliefs with respect to variation in certainty (by shrinking the gap between the low-certainty and high-certainty quadrants). We calculate the raw gap in utility between these investments (given that the ground-truth is true) and express it as a percentage of the raw utility gain from moving from the default portfolio that splits investments equally across the three states to the optimal portfolio given our ground-truth probabilities.

### A.1.3 Supplementary Experiment: Content Transmission Incentives

**Design.** This experiment is virtually identical to the baseline experiment, except that half of respondents are generically incentivized to pass on *all* of the information in the original messages (*implicit incentives*), while half are explicitly and equally incentivized to pass on both the level and reliability of the original forecast (*explicit incentives*).

In particular, respondents are informed that, if selected for bonus payment eligibility (a 10% chance), a different group of participants will score transcripts of their recordings on a scale of 0 to 10, where 0 corresponds to "Nothing conveyed in meaning" and 10

corresponds to "Everything conveyed in meaning". This group, which we refer to as the *evaluators*, is distinct from the listeners. If the average score a transmitter's recordings receive is at least an 8, the transmitter will receive a $20 bonus payment. Between subjects, we randomly assign transmitters to two variants of the incentive scheme. In *implicit incentives*, participants are given the following instructions:

> The other participants will answer the following question about your voice message: How accurately did the voice message convey the content and meaning of what the speaker said?

Compared to the original transmitter incentives, this incentive scheme should incentivize transmitters to pass on reliability information regardless of their beliefs about its importance for listeners' belief updates, because the instructions encompass *all* the contents of the original message.

In the *explicit incentives* condition, we go one step further by informing respondents that the evaluators will answer two questions about the message, one about the level of the prediction and one about the reliability of the prediction:

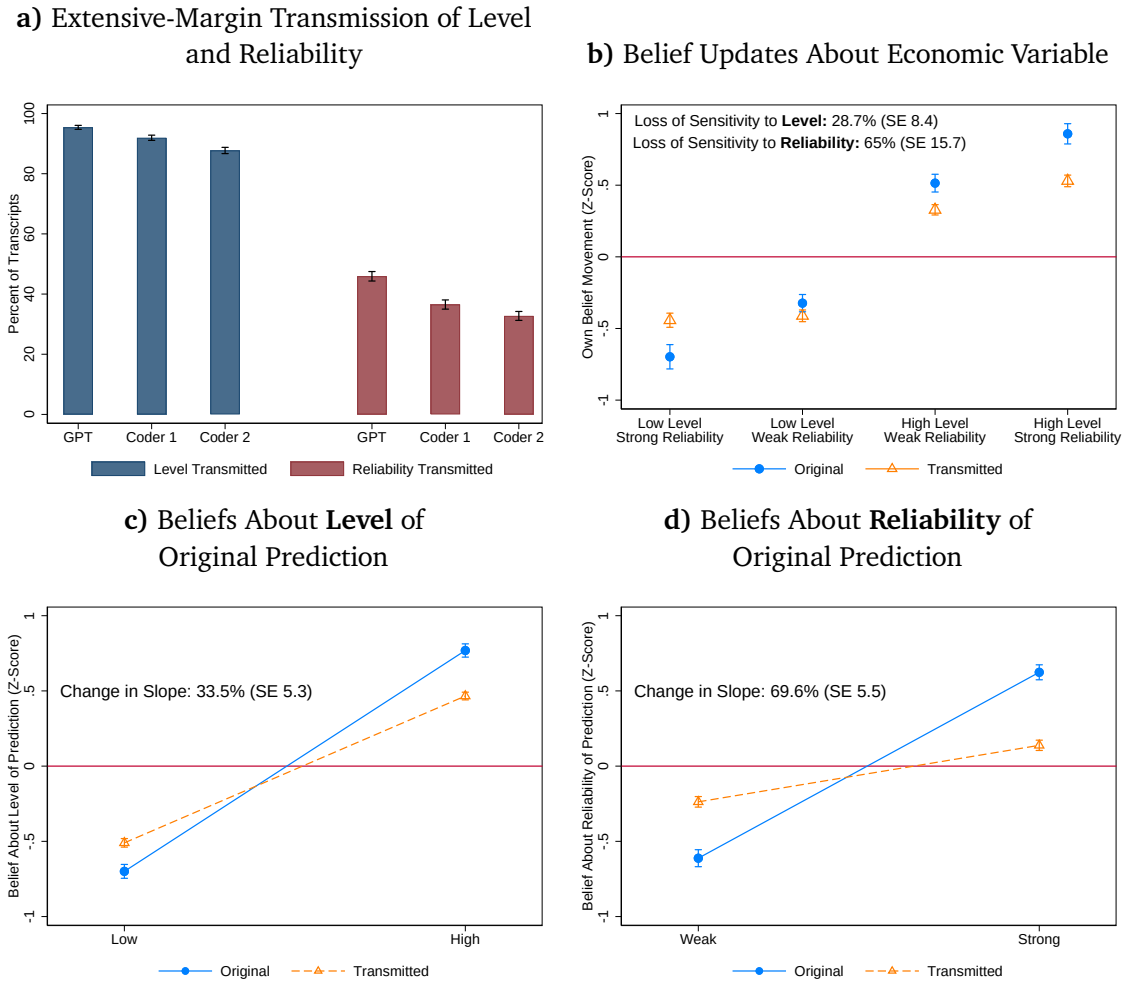
> The other participants will answer two questions about your voice message: How accurately was the speaker's prediction about the level of the economic variable conveyed in the voice message? How accurately was the speaker's assessment of the reliability of their forecast conveyed in the voice message?

The explicit incentive scheme has two main features. First, it ensures that transmission of the reliability of the prediction is, by design and explicitly, equally as payoff-relevant as the transmission of the prediction's level. Second, unlike both the baseline and implicit schemes, it introduces transmitters explicitly to the level-reliability distinction.

**Logistics.** The additional transmission and listener experiments were run with 501 and 1,509 U.S. respondents from Prolific, respectively, in September 2023.

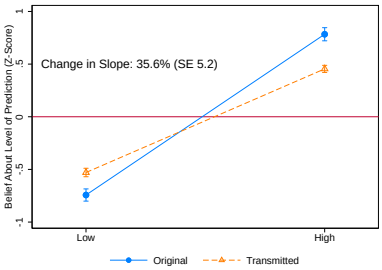
**Results.** Appendix Figure A16 shows results, pooling the implicit and explicit schemes. Panel (a) shows substantial differential loss in our analysis of scripts: only 30-40% of scripts mention reliability, compared to 85-95% mentioning level. Panels (c) and (d) analyze message beliefs, showing 33.5% loss of level information and 69.6% loss of reliability. This is similar to, albeit slightly smaller than, the 35% vs. 90% differential information loss in our baseline experiment. A formal test of equality of the two information loss statistics rejects the null at $p < 0.001, \chi^2 = 27.1$. Appendix Figure A17 shows that results are fairly similar across the implicit and explicit incentive schemes. The loss of sensitivity to level is 35.6% for explicit incentives and 31% for implicit incentives. The loss of sensitivity to reliability is 65.3% for explicit incentives and 73.7% for implicit

incentives. Taken together, increasing the payoff relevance and salience of reliability information through implicit and explicit incentive schemes reduces reliability information loss only modestly. Thus, differences in perceived importance of transmitting level versus reliability information do not fully explain our main result.

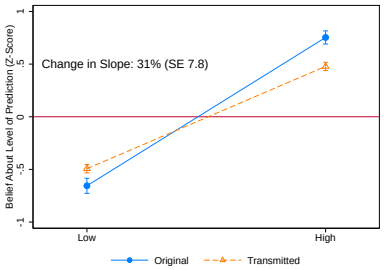



Appendix Figure A16: This figure presents data from the Content Transmission Incentives experiment, pooling the implicit and explicit incentive conditions. Panel (a) replicates Figure 1. It shows the fraction of transmitted messages classified by GPT-4 and our two human coders as containing statements about the level or reliability of the original forecast. Panel (b) replicates Figure 3, showing belief updates about the economic variable by level × reliability quadrant and by whether the listener heard a transmitted or original recording. Panels (c) and (d) replicate Figure 2, showing listeners' beliefs about the level and reliability of the prediction in the original message, separately by whether the original message is low- vs high-level or weak- vs strong-reliability, and separately by whether the listener hears the original message or a transmitted version of it. Dots are mean beliefs and bars are standard error bars (1 SE each direction). $N = 1,509$ listeners and 501 transmitters. Appendix Figure A17 splits these graphs by implicit versus explicit incentives.

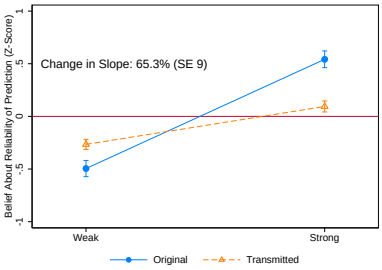
**a)** Level: Explicit Incentives



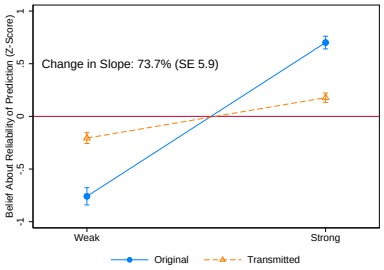

**b)** Level: Implicit Incentives




**c)** Reliability: Explicit Incentives

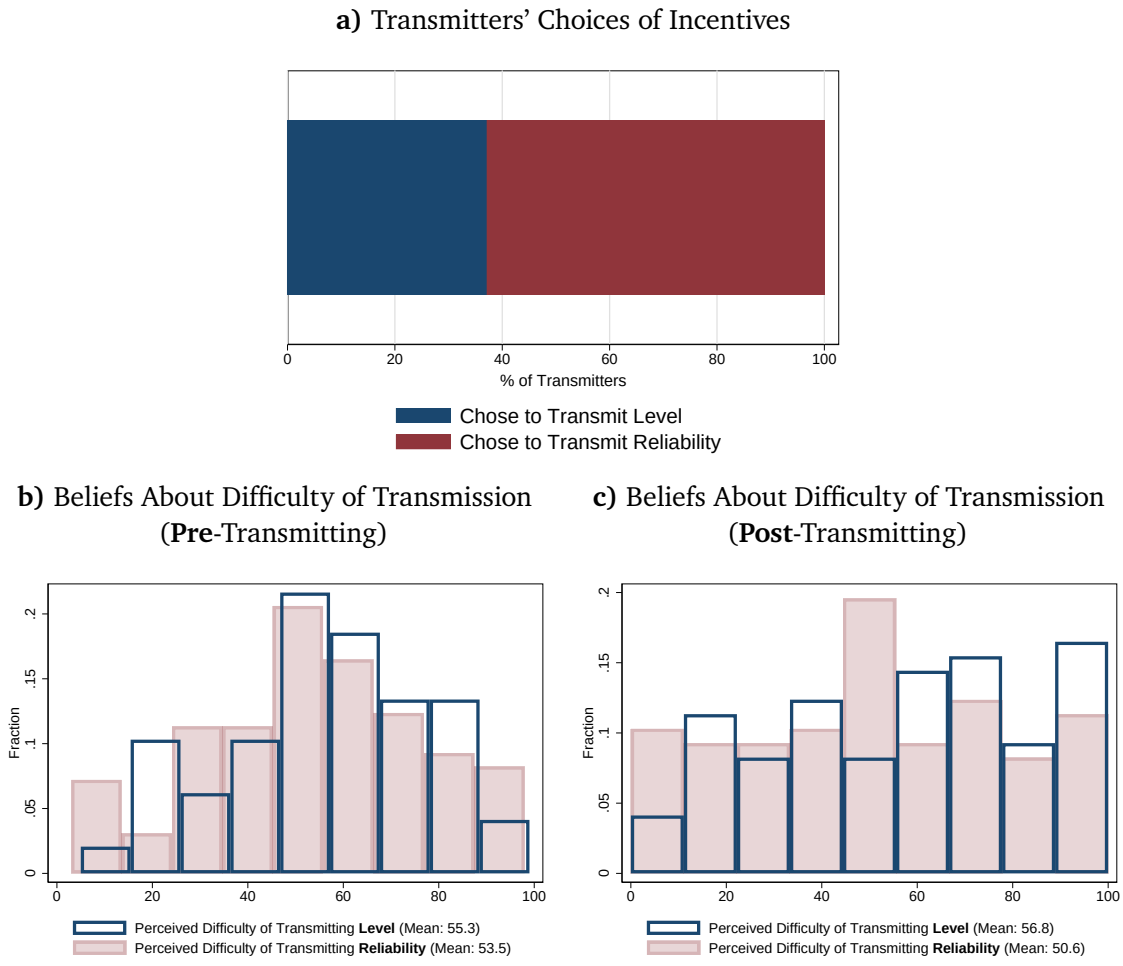



**d)** Reliability: Implicit Incentives




Appendix Figure A17: This figure replicates Figure 2 in the Content Transmission Incentives data, separately by respondents randomized into the explicit and implicit transmission incentives.

### A.1.4 Supplementary Experiment: Choice of Incentives




Appendix Figure A18: This figure presents data from the *choice of incentive* experiment. Panel (a) shows the share of people choosing to be incentivized based on their transmission of level information versus reliability information. Panel (b) shows the distribution of respondents' beliefs about the difficulty of transmitting level and reliability, before they complete the transmission task. Panel (c) shows respondents' beliefs about the difficulty of transmission, after completing the transmission task. $N = 97$ transmitters.

### A.1.5 Identifying Variation in the Incidental-Cue Analysis

This appendix details the variation underlying Columns (1)-(3) of Table 1, which show that exposure to payoff-irrelevant reliability cues in one recording raises the probability that a transmitter conveys reliability when verbalizing the other recording. Because this variation was generated incidentally by our baseline randomization rather than designed for this analysis, we describe its source and document balance.

**Source of variation.** Recall from Section 2 that each transmitter hears two recordings, one about home prices and one about retail revenue, before recording verbalizations of

both. The reliability of each recording is randomized. In the modular manipulation, a recording is assigned to a strong-reliability version (certainty prefixes and confidence statements), a weak-reliability version (uncertainty prefixes and low-confidence statements), or a neutral version (prefixes and confidence statements omitted). Assignment proceeds in two steps. First, the two recordings a transmitter hears are stratified so that one is strong-reliability and one is weak-reliability, one argues for an increase and one for a decrease, and one uses a male and one a female voice; the modular-versus-naturalistic manipulation is assigned independently for each recording. Second, modular recordings are switched to neutral as follows: if exactly one of the two recordings is modular, that recording is switched to neutral with probability 33%; if both recordings are modular, exactly one of the two, chosen at random, is switched to neutral with probability 66%.

For Columns (1)-(3), an observation is a transmitter-by-recording, restricted to focal recordings that are strong- or weak-reliability (not neutral). The independent variable indicates whether the *other* recording the transmitter heard was non-neutral, that is, carried reliability cues. Identification rests on the fact that, holding the focal recording's content fixed, whether the other recording carries reliability cues is determined by the random assignment of the neutral condition and of the modular-versus-naturalistic manipulation.

**Why the cue is incidental and payoff-irrelevant.** A transmitter's bonus for the focal recording depends only on how closely the belief movements induced by their focal verbalization track those induced by the focal original recording (Section 2.1). The reliability content of the other recording is irrelevant to this payoff. A transmitter optimizing their focal-recording payoff therefore has no reason to attend to the reliability of the other recording. The fact that this variation was not engineered for the present analysis reinforces this interpretation: nothing in the design directed transmitters' attention to the reliability of the non-focal recording, so any resulting increase in reliability transmission reflects an incidental cognitive cue rather than a response to incentives.

**Balance.** Appendix Table A3 regresses the cue indicator on characteristics of the focal recording: its level, its reliability, the voice gender, and the order in which it was heard. Consistent with random assignment, these characteristics do not jointly predict the cue indicator ($F(4) = 0.98$, $p = 0.45$). By construction, the cue indicator is mechanically related to whether the focal recording is itself modular or naturalistic, because the neutral condition exists only within the modular manipulation; we therefore control for the focal manipulation type throughout.

Appendix Table A3: Balance in Incidental-Cue Assignment

| | (1)<br>Other Recording Contained Reliability Prefixes |
|---|---|
| High-Level Focal Recording | 0.040<br>(0.027) |
| Strong-Reliability Focal Recording | 0.003<br>(0.027) |
| Female Voice | -0.015<br>(0.027) |
| Transmitter Female | -0.032<br>(0.034) |
| Transmitter Age | -0.000<br>(0.001) |
| Transmitter Employed | 0.003<br>(0.042) |
| Transmitter BA+ | 0.014<br>(0.036) |
| Transmitter White | 0.005<br>(0.052) |
| Transmitter Black | -0.108<br>(0.068) |
| Constant | 0.819***<br>(0.082) |
| Observations | 702 |
| Transmitters | 424 |
| Joint F-statistic | 0.984 |
| Joint p-value | 0.453 |

Standard errors in parentheses
* p<0.1, ** p<0.05, *** p<0.01

*Note*: This table presents balance checks for the incidental-cue analysis in Columns (1)–(3) of Table 1. An observation is a transmitter-by-recording, restricted to focal recordings that are high- or low-reliability rather than neutral. The dependent variable is an indicator for whether the other recording heard by the same transmitter contained reliability prefixes. Standard errors are clustered at the transmitter level.

## A.2 Summary Statistics

Appendix Table A4: Summary Statistics: Listener and Transmitter Experiments

| | Baseline | | Content Trans. | | Incentive Choice | High Salience | | Quant. Scripts | | Belief Dist. | Investment |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Trans. | List. | Trans. | List. | Trans. | Trans. | List. | Trans. | List. | List. | List. |
| Age | 43 | 40 | 37 | 38 | 43 | 37 | 38 | 36 | 40 | 44 | 43 |
| Female | .52 | .49 | .52 | .49 | .52 | .52 | .52 | .5 | .5 | .5 | .5 |
| Employed | .79 | .78 | .8 | .75 | .81 | .8 | .78 | .73 | .81 | .8 | .75 |
| Education: BA+ | .61 | .6 | .59 | .56 | .64 | .66 | .63 | .61 | .69 | .62 | .57 |
| Race: White | .67 | .66 | .73 | .72 | .73 | .57 | .61 | .54 | .54 | .75 | .71 |
| Race: Black | .21 | .17 | .12 | .14 | .19 | .24 | .21 | .21 | .36 | .16 | .15 |
| Observations | 540 | 1510 | 501 | 1509 | 97 | 244 | 1010 | 181 | 807 | 892 | 1196 |

*Note*: The "Observations" row reports the number of respondents with nonmissing demographic variables used to compute the summary-statistics rows in this table.

## A.3 Regression Tables

### A.3.1 Belief Movement Incentives

Appendix Table A5: Belief Updates About State of the World

| | (1) Pooled | (2) Modular Only | (3) Naturalistic Only | (4) High Transmitter IQ | (5) Low Transmitter IQ |
|---|---|---|---|---|---|
| **Low Level × Strong Reliability** | -0.816*** | -0.741*** | -0.879*** | -0.845*** | -0.737*** |
| | (0.132) | (0.243) | (0.145) | (0.112) | (0.195) |
| **Trans. × Low Level × Strong Reliability** | 0.446*** | 0.433* | 0.454*** | 0.494*** | 0.323 |
| | (0.141) | (0.252) | (0.162) | (0.127) | (0.213) |
| **Low Level × Weak Reliability** | -0.338*** | -0.414*** | -0.293*** | -0.381*** | -0.254** |
| | (0.043) | (0.058) | (0.040) | (0.043) | (0.107) |
| **Trans. × Low Level × Weak Reliability** | -0.066 | 0.041 | -0.129* | -0.038 | -0.122 |
| | (0.062) | (0.095) | (0.070) | (0.070) | (0.132) |
| **High Level × Weak Reliability** | 0.362*** | 0.516*** | 0.263*** | 0.380*** | 0.324*** |
| | (0.070) | (0.078) | (0.008) | (0.081) | (0.072) |
| **Trans. × High Level × Weak Reliability** | -0.035 | -0.252** | 0.105 | 0.015 | -0.165 |
| | (0.086) | (0.107) | (0.067) | (0.100) | (0.116) |
| **High Level × Strong Reliability** | 0.767*** | 0.762*** | 0.770*** | 0.709*** | 0.880*** |
| | (0.089) | (0.122) | (0.120) | (0.083) | (0.146) |
| **Trans × High Level × Strong Reliability** | -0.323*** | -0.355*** | -0.301** | -0.165* | -0.633*** |
| | (0.097) | (0.131) | (0.133) | (0.097) | (0.156) |
| **Nb. obs** | 2,509 | 1,272 | 1,237 | 1,690 | 819 |

Standard errors in parentheses
* p<0.1, ** p<0.05, *** p<0.01

*Note*: This table presents data from the baseline experiment with belief movement incentives. It shows regressions of respondents' belief updates (posterior minus prior, z-scored at the topic × reliability randomization type level) on dummy variables representing the four quadrants of our 2 × 2 level-reliability randomization, with no constant. Standard errors are two-way clustered at the voice recording by listener level. Column (1) does this for our full pooled sample, Column (2) for our subsample hearing the modular reliability manipulation, and Column (3) for the naturalistic reliability manipulation. Columns (4) and (5) split transmitters by above/below median performance on the Raven's Matrix questions they answer at the end of the survey, which we use as a measure of IQ.

Appendix Table A6: Beliefs About Level of Original Message's Prediction

| | (1) Pooled | (2) Modular Only | (3) Naturalistic Only | (4) High Transmitter IQ | (5) Low Transmitter IQ |
|---|---|---|---|---|---|
| **High Level** | 1.368*** | 1.371*** | 1.366*** | 1.299*** | 1.507*** |
| | (0.064) | (0.091) | (0.084) | (0.075) | (0.102) |
| **High Level × Transmitted** | -0.486*** | -0.607*** | -0.360*** | -0.324*** | -0.816*** |
| | (0.078) | (0.111) | (0.105) | (0.092) | (0.128) |
| **Transmitted** | 0.266*** | 0.271*** | 0.257*** | 0.211*** | 0.387*** |
| | (0.052) | (0.075) | (0.072) | (0.064) | (0.087) |
| **Constant** | -0.699*** | -0.658*** | -0.739*** | -0.673*** | -0.758*** |
| | (0.044) | (0.063) | (0.060) | (0.053) | (0.073) |
| **Nb. obs** | 2,509 | 1,272 | 1,237 | 1,690 | 819 |

Standard errors in parentheses

* p<0.1, ** p<0.05, *** p<0.01

*Note*: This table presents data from the baseline experiment with belief movement incentives . It shows regressions of respondents' beliefs about the level of the prediction in the original message on a dummy for the original message being in the high-level condition, a dummy for the respondent hearing a transmitted version of the message, and the interaction of those dummies. Standard errors are clustered at the listener by voice recording level. Column (1) does this for our full pooled sample, Column (2) for our subsample hearing the modular reliability manipulation, and Column (3) for the naturalistic reliability manipulation. Columns (4) and (5) split transmitters by above/below median performance on the Raven's Matrix questions they answer at the end of the survey, which we use as a measure of IQ.

Appendix Table A7: Beliefs About Reliability of Original Message's Prediction

| | (1) Pooled | (2) Modular Only | (3) Naturalistic Only | (4) High Transmitter IQ | (5) Low Transmitter IQ |
|---|---|---|---|---|---|
| **Strong Reliability** | 1.181*** | 0.692*** | 1.492*** | 1.143*** | 1.261*** |
| | (0.112) | (0.132) | (0.046) | (0.112) | (0.174) |
| **Strong Reliability × Transmitted** | -1.063*** | -0.635*** | -1.322*** | -1.020*** | -1.156*** |
| | (0.123) | (0.154) | (0.080) | (0.126) | (0.197) |
| **Transmitted** | 0.461*** | 0.142 | 0.655*** | 0.452*** | 0.478*** |
| | (0.095) | (0.129) | (0.059) | (0.099) | (0.147) |
| **Constant** | -0.552*** | -0.277** | -0.719*** | -0.528*** | -0.600*** |
| | (0.088) | (0.115) | (0.036) | (0.090) | (0.131) |
| **Nb. obs** | 2,079 | 842 | 1,237 | 1,411 | 668 |

Standard errors in parentheses

* p<0.1, ** p<0.05, *** p<0.01

*Note*: This table presents data from the baseline experiment with belief movement incentives. It shows regressions of respondents' beliefs about the reliability of the prediction in the original message on a dummy for the original message being in the strong-reliability condition, a dummy for the respondent hearing a transmitted version of the message, and the interaction of those dummies. Standard errors are clustered at the listener by voice recording level. Column (1) does this for our full pooled sample, Column (2) for our subsample hearing the modular reliability manipulation, and Column (3) for the naturalistic reliability manipulation. Columns (4) and (5) split transmitters by above/below median performance on the Raven's Matrix questions they answer at the end of the survey, which we use as a measure of IQ.

# B Robustness Experiments

## B.1 Extensive Margin of Transmission

**Design.** This experiment replicates the transmitter experiment from the investment design in Section 5.1, with one key modification: after listening to each original message, transmitters can choose whether or not to record a voice message to pass on. To make the extensive-margin choice salient, transmitters are explicitly instructed that their recording—if made—will be the sole information available to a matched listener making an incentivized investment decision, and that they should only record a message if they believe it would be useful.

Specifically, transmitters are told:

> "You should aim to create voice messages that convey any information that would be relevant to someone deciding whether and how to invest in this company. If you feel there is no useful information to pass on, you can choose not to record a voice message. Note that the other participant will not have any information about the company other than your voice recording (if you choose to record one)."

If transmitters opt out, the matched listener makes the investment decision without receiving any message.

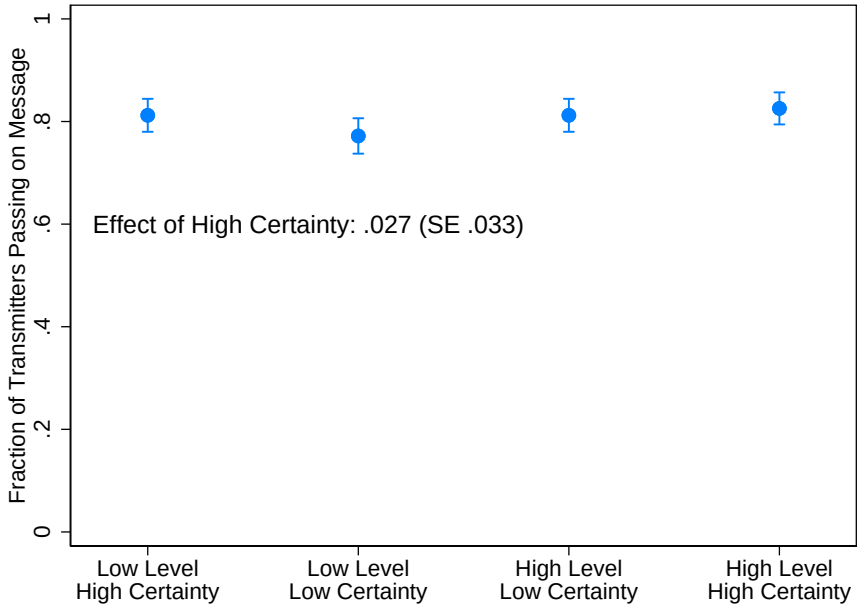
**Logistics.** The additional transmission experiment was run with 296 U.S. respondents from Prolific in December 2025.

**Results.** We examine whether the decision to record a message depends on the level and reliability of the original message. Figure A19 displays the fraction of transmitters who chose to pass on information across the four experimental quadrants. Transmission rates are high and stable, with approximately 80% of participants opting to record a message in every condition. To test for differences, we regress the transmission dummy on an indicator for high-certainty signals. The resulting coefficient of .027 (SE .033) is both small in magnitude and statistically insignificant, suggesting that information certainty does not meaningfully drive the decision to transmit information. Transmitters appear to view even noisy or weakly reliable signals as sufficiently useful to pass on to the next agent. Appendix Table A9 also shows that transmitters are not more likely to pass on surprising messages, where we interpret "surprising" as meaning a message that predicts strictly the opposite sign as the transmitter's prior.

Appendix Table A8 reports the distribution of reasons transmitters give for choosing whether or not to pass on a message. By far the most common sentiment expressed by

transmitters who pass on at least one message is a generic desire to be helpful to the listener, without any reference to the reliability or certainty of the original information; variants of this reason account for the majority of responses and are strongly positively associated with transmission. Mentions of high certainty or reliability of the original message are comparatively rare. Conversely, among transmitters who choose not to pass on at least one message, the dominant reason is pessimism about their own ability to transmit the information adequately, which appears in more than half of responses and is strongly negatively associated with transmission. References to low message certainty or a desire to avoid influencing the listener are substantially less frequent, indicating that non-transmission primarily reflects self-assessed communicative limitations rather than explicit concerns about message reliability.

**a)** Fraction of Transmitters Choosing to Record a Message




Appendix Figure A19: This figure presents data from our extensive-margin transmission experiment. It shows the fraction of transmitters choosing to record a message to pass on to their matched listener, separately by the level × certainty quadrants randomized in the original message. It also displays the coefficient from a regression of a dummy variable for the transmitter choosing to pass on the message on an indicator for the original message being high-certainty. Error bars represent 1 SE in either direction.

Appendix Table A8: Reasons for passing on/not passing on a message

| Description | Mean (All) | Mean (Didn't Transmit At Least One) | Coef. [SE] |
|---|---|---|---|
| Original message highly certain | 0.14 | 0.12 | 0.077 [ 0.049] |
| Original message uncertain | 0.09 | 0.19 | -0.213** [ 0.086] |
| Desire to help listener | 0.62 | 0.22 | 0.338*** [ 0.045] |
| Desire to avoid influencing listener | 0.09 | 0.19 | -0.317*** [ 0.098] |
| Pessimism about personal ability to transmit info | 0.25 | 0.55 | -0.375*** [ 0.058] |

*Note*: This table presents results from our extensive-margin transmission experiment. After choosing whether or not to record both voice messages, transmitters are asked an open-ended question about what factors influenced their decisions. This table presents 5 categories of reasons coded from these open-text responses using GPT-5.2. The second column displays the share of responses coded as containing this reason among all transmitters; the third contains the share of responses coded as containing this reason among transmitters who chose not to transmit at least one of the two messages; and the fourth column displays the coefficient from a regression of the transmitter's probability of sending a message (the number of messages sent divided by 2) on an indicator for the transmitter's message containing this reason. The sample consists of the 296 message-level observations for which the prior measure needed to define this sign comparison is nonmissing.

Appendix Table A9: Are Transmitters More Likely to Transmit Surprising Information?

| | Chose to Record Message (1) |
|---|---|
| Prior is Strictly Opposite-Signed to Original Message | -0.055 |
| | (0.055) |
| Observations | 296 |

Standard errors in parentheses
* p<0.1, ** p<0.05, *** p<0.01

*Note*: This table shows results from our Endogenous Transmission experiment. It shows a regression of an indicator for the transmitter choosing to record a voice message on an indicator for their prior about the economic state having strictly the opposite sign as the original message they heard (e.g., the transmitter's prior is that the company's earnings will be strictly lower than expected, and the original message says they will be strictly higher).

## B.2 Quantitative Communication

**Design.** The experimental design is virtually identical to our baseline but adds quantitative information about both level and reliability to the original scripts. Quantitative information about the level is conveyed by providing a point estimate of the change in revenue growth. Quantitative reliability information is communicated via a probabilistic confidence statement. The quantitative statements are added to the final part of the script, where the speaker sums up their forecast and confidence level. In the context of a high reliability revenue growth forecast, quantitative information is conveyed as follows:

> Overall, I am confident this means that the revenue growth of this company will definitely fall strongly over the forthcoming year, by about 8%. I am more than 90% confident in this forecast.

In the low reliability revenue growth forecast quantitative information is presented as follows:

> Overall, I think it is conceivable that this means that the revenue growth of this company will imaginably fall strongly over the forthcoming year, by about 8%. That said, I am only 10% confident about this forecast.

The quantitative forecast was an 8% increase or decrease in the case of revenue growth and a 10% increase or decrease in the case of home price growth; confidence levels were either 10% or 90%. See Appendix Section F for the full set of quantitative scripts.

**Logistics.** The additional transmission and listener experiments were run with 181 and 834 U.S. respondents from Prolific, respectively, in June 2024.

**Results.** Figure A20 Panel (c) shows that the differential loss of reliability indicators in transcripts remains highly statistically significant ($p < 0.01$) and economically sizable. The level-reliability gap is somewhat smaller than in our baseline experiment: about 50% of transmitters now mention reliability in their messages, compared to 30% in our baseline experiment, suggesting that numerical confidence statements increase the salience of reliability information or make it easier to transmit. As before, over 90% of respondents mention the level.[2]

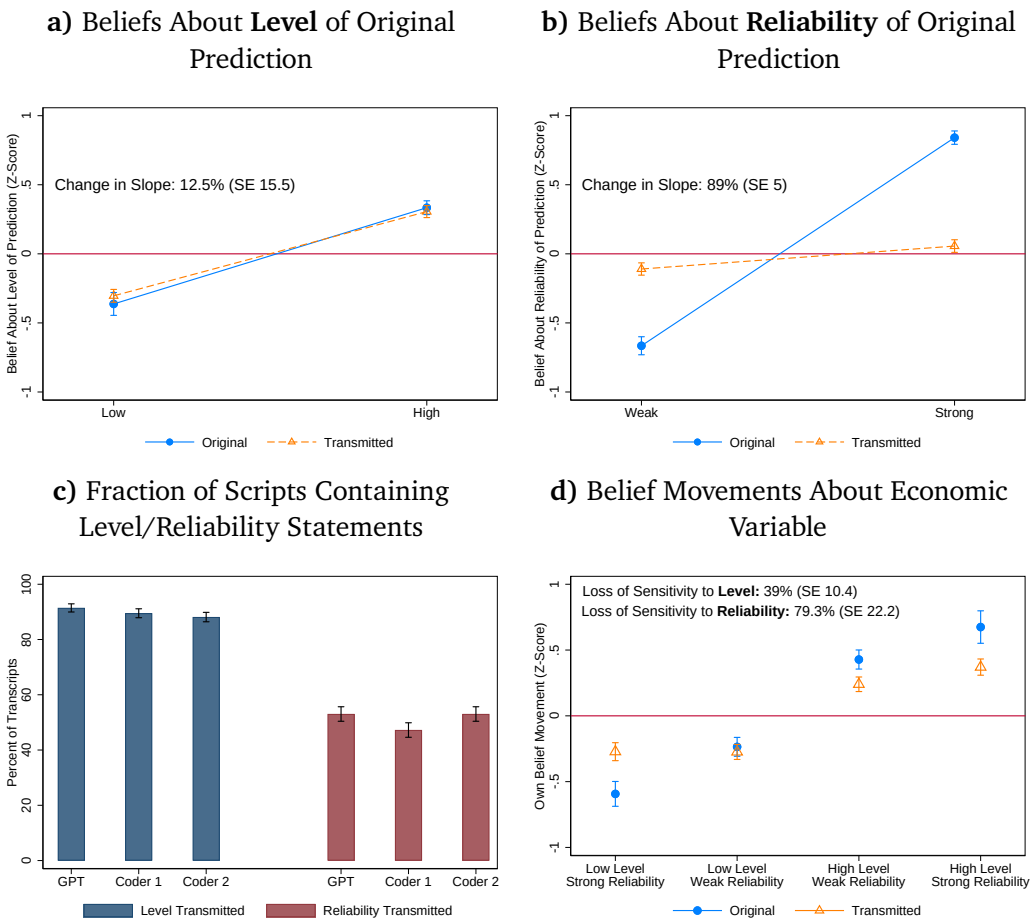
However, the increased fraction of transmitters mentioning reliability does not translate into a reduction in differential information loss according to our belief-based measures. Panels (a) and (b) show that differential loss *strengthens*: level information loss falls to 12.5%, while reliability loss remains 89.0%. A formal test of equality of the two information loss statistics rejects the null at $p < 0.001, \chi^2 = 22.29$. Appendix Figure A20 Panel (d) shows that this is also true when analyzing listeners' state belief updates instead of message beliefs.

Differential loss strengthens despite the increase in the fraction of scripts mentioning reliability because the quantitative scripts increase the impact of our original recordings on the reliability beliefs of listeners directly hearing them,[3] so that the omission of reliability information has a greater impact on beliefs than in our baseline experiment. (In other words, a given fraction of scripts omitting reliability information has a larger impact on information loss as measured by message or state beliefs, because of the stronger first-stage effect of the original recordings on beliefs.)

Taken together, these findings demonstrate that the differential information loss persists when both level and reliability information are also conveyed quantitatively.

[2]About 45% of transmitters pass on the level number and 25% pass on the reliability number.

[3]Among listeners hearing original recordings, the reliability manipulation in this experiment generates a 44-point gap in reliability beliefs on a scale of 0-100, compared to 30 points in the baseline.




Appendix Figure A20: This figure presents data from our version of the baseline experiment that uses quantitative scripts. Panels (a) and (b) replicate Figure 2, showing listeners' beliefs about the level and reliability of the prediction in the original message, separately by whether the original message is low- vs high-level or weak- vs strong-reliability, and separately by whether the listener hears the original message or a transmitted version of it. Dots are mean beliefs and bars are standard error bars (1 SE each direction). Panel (c) replicates Figure 1. It shows the fraction of transmitted messages classified by GPT-4 and our two human coders as containing statements about the level or reliability of the original forecast. Panel (d) replicates Figure 3. $N = 181$ transmitters and 834 listeners.

## B.3 Belief Distribution Movement Incentives

**Design.** To address the concern that eliciting only mean beliefs may under-incentivize the faithful transmission of uncertainty, we replicate our main design using an alternative incentive scheme that elicits listeners' *full subjective belief distributions*. Instead of reporting a single point estimate, listeners allocate probability mass across discrete outcome bins using a graphical interface following Crosetto and De Haan (2023). The interface enforces that probabilities sum to 100 percent and allows participants to express dispersion, skewness, and multi-modality.

To increase the salience of the distributional incentives for transmitters, the instruc-

tions (i) explicitly explain that listeners will report beliefs using this histogram interface and (ii) introduce participants to the interface (including an example distribution), illustrating how probability mass can be allocated across outcome bins.

In addition, listeners are walked through the interface and complete practice items using it before providing incentivized beliefs (see the interface-introduction screens).

Transmitters' incentives are directly tied to how closely the beliefs induced by their message match those induced by the original recording. As stated in the instructions, *"your likelihood of receiving a bonus payment depends on how close the average probability distribution induced by your message is to the average probability distribution induced by the original recording."* The distance between the two distributions is defined explicitly: *"the distance between distributions is measured based on the sum of squared deviations in the probability mass assigned to different outcome bins."* A smaller distance implies a higher probability of receiving a bonus payment, rewarding transmitters for preserving both the central tendency and the dispersion of beliefs conveyed by the original message. Apart from the belief elicitation method and incentive calculation, all other aspects of the transmitter task are identical to the baseline design.

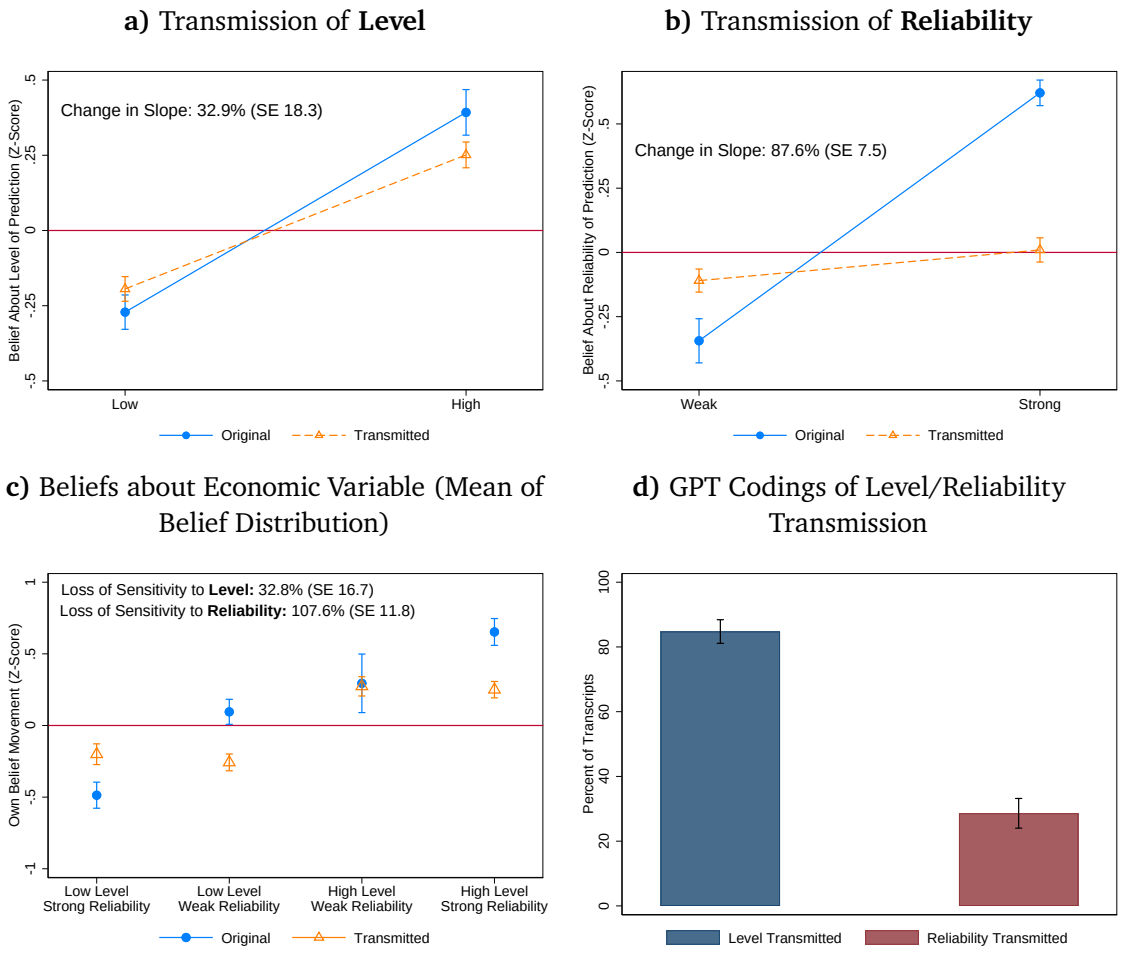
**Logistics.** The additional transmission and listener experiments were run with 225 and 892 U.S. respondents from Prolific, respectively, in December 2025.

**Results.** Appendix Figure A21 reports results from the Belief Distribution Movement Incentives experiment, which replicates our baseline design while incentivizing transmitters to match the full shift in listeners' belief distributions rather than only the mean. Panels (a) and (b) replicate the extensive-margin transmission results for belief level and reliability. Transmission leads to a modest attenuation in sensitivity to level: the slope with respect to level is reduced by 32.9 percent relative to original messages (SE 18.3). By contrast, sensitivity to reliability collapses almost entirely, with an 87.6 percent reduction in the slope (SE 7.5), indicating that reliability information is largely lost in transmission even under stronger incentives.

Panel (c) shows that these distortions carry through to economically relevant belief updates. Mean belief movements respond strongly to both level and reliability when listeners hear original messages, but transmission sharply reduces responsiveness to reliability while leaving responsiveness to level comparatively intact. Quantitatively, transmission reduces sensitivity to level by 32.8 percent (SE 16.7) and reduces sensitivity to reliability by 107.6 percent (SE 11.8), implying that transmitted messages induce nearly identical belief movements regardless of whether the underlying information is weak or strong.

Panel (d) provides direct evidence on the mechanism. GPT-based codings of trans-

mitter recordings indicate that level information is transmitted in the vast majority of messages, whereas explicit references to reliability are rare. Taken together, Appendix Figure A21 shows that the selective loss of reliability information is not an artifact of mean-based incentives: even when transmitters are rewarded for matching the entire belief distribution, verbal diffusion preserves level cues but systematically strips away information about uncertainty.




Appendix Figure A21: This figure presents results from our Belief Distribution Movement Incentives experiment, which replicates our baseline experiment with the following change: (i) we elicit beliefs about the state of the world through an interface that asks for a subjective probability distribution over different states of the world; (ii) instead of being incentivized based on how close the mean belief change induced by their message is to the mean belief change induced by the original message, transmitters are incentivized based on how close the shift in *distribution of beliefs* induced by their message is to the shift induced by the original message. Panels (a) and (b) replicate Figures 2, Panel (c) replicates Figure 3, and Panel (d) replicates the GPT-coded bars in Panel (a) of Figure 1. $N = 225$ transmitters and 892 listeners.

# C Diffusion of Information and Externalities of Transmission Loss

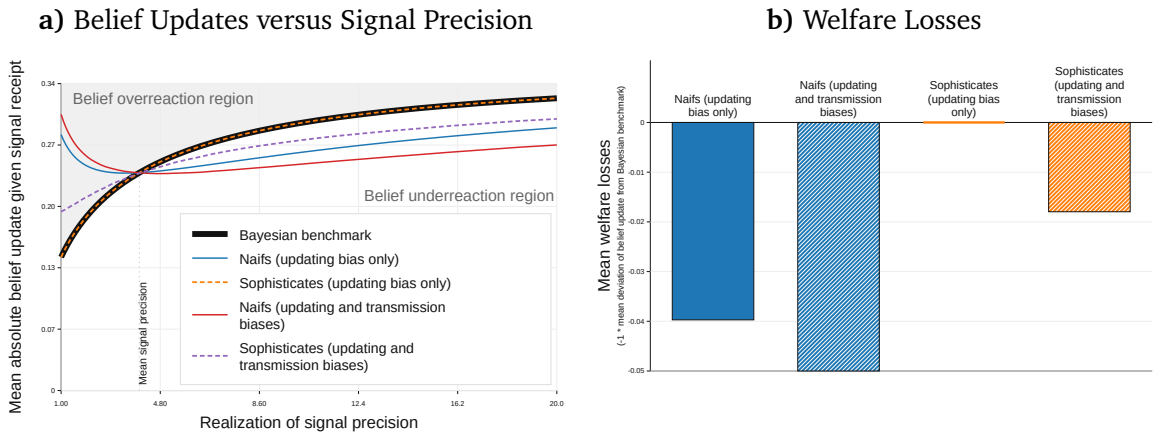
This section describes the results of simple simulation exercises showing how transmission-induced loss of reliability information both *compounds* insensitivity to reliability information in belief updating (Augenblick et al., 2025; Griffin and Tversky, 1992) and generates distinct consequences.

**Assumptions underlying the simulation.** We perform a simulation exercise using parameters calibrated to match results from our experiments. We consider a population of agents with a common prior over a state $\theta \sim N(\mu_0, \frac{1}{\tau_0})$. Half of the agents are naïve and half are sophisticated. Half of the population are randomly exposed to the realization of an informative signal about the state, $s = \theta + \varepsilon$, where $\varepsilon \sim N(0, \frac{1}{\tau_s})$, and $\tau_s$ is itself drawn from a gamma distribution with mean $\mu_\tau$ and variance $\sigma^2_\tau$. Agents exposed to the signal then pass it on to a random peer who was not initially exposed, with an 80% probability (matching the results of our extensive-margin transmission experiment described in Section 5.2).

We consider two forms of naïvete. Naïfs have an *updating bias* if they perform Bayesian updating on the signal realization using an attenuated perception of the signal's precision shrunk towards its mean precision, $\hat{\tau}_s = \mu_\tau + \kappa(\tau_s - \mu_\tau)$ with $\kappa \in [0, 1]$, in the spirit of Augenblick et al. (2025). Naïfs have a *transmission bias* if, when passing on the signal, they faithfully transmit the signal realization but pass on the signal precision with added noise, $\tilde{\tau}_s = \tau_s + \eta$, where $\eta \sim N(0, \sigma^2_\eta)$. We assume that all recipients of a transmitted message account for the added transmission noise $\eta$. (Assuming that naïfs ignore this noise would yield very similar results.) Sophisticated recipients then make a fully Bayesian update about the state using their posterior about the precision; naïfs add the updating bias attenuation and then perform a Bayesian update. The transmission noise $\eta$ is a reduced-form way of representing the fact that once filtered through transmission, the reliability of the original message becomes difficult to discern, for example because most of the original reliability indicators get dropped. Consequently, the reliability perceptions of recipients of transmitted messages are compressed towards the default reliability belief, exactly as observed in Figure 2 Panel (b).

We calibrate the transmission noise $\sigma^2_\eta$ so that naïve transmission attenuates the sensitivity of recipients' belief updates to the signal precision by 90%, matching the results of Figure 3; to level the playing field, we calibrate $\kappa$ so that the updating bias is equally consequential (accounting for the fact that it affects twice as many naïfs as the transmission bias).

**Simulation results.** Panel (a) of Figure A22 shows mean absolute belief updates as a function of different (realizations of) signal precisions. The black line shows the Bayesian benchmark: belief updates are increasing in the signal precision. When naïfs are only subject to an updating bias, sophisticates' belief updates remain perfectly Bayesian (orange dashed line); naïfs, on the other hand, over-react to low-precision signals and under-react to high-precision signals (blue line). Adding a transmission bias to naïfs has two effects: it increases the severity of naïfs' under- and over-reactions (red line), while also creating a pattern of under- and over-reaction among sophisticates, who now have attenuated perceptions of the signal precision (purple dashed line). Commensurately, Panel (b)—which shows the mean deviation of belief updates from the Bayesian benchmark, that we use as a proxy for welfare—shows that the addition of transmission biases both compounds welfare losses among naïfs and introduces welfare losses among sophisticates, shifting the distribution of lost welfare.




Appendix Figure A22: This figure presents results from our simulation exercise. Panel (a) plots the mean absolute magnitude of belief updates for a range of values of the signal precision, separately for naïfs and sophisticates and separately by whether we assume naïfs exclusively exhibit an updating bias or both an updating and transmission bias. Naïfs' belief updates are nonmonotonic in the signal precision because lower signal precisions imply more dispersed signal realizations. Panel (b) plots mean "welfare losses" for each group in each case, where the welfare loss is just $-1\times$ the mean absolute deviation between the belief update and the correct belief update, averaging across the range of signal precisions we consider.

**Robustness to alternative parameter values.** Table A10 tests the robustness of the simulation results to variations in several of the key parameters: the probability that a message is transmitted at all; the share of the population that is naïve; the sensitivity of naïfs (relative to sophisticates) to reliability in belief updating; and the amount of transmission noise added to reliability information by naïfs. Results show that the key qualitative patterns documented in our simulation hold across parameter values:

Appendix Table A10: Robustness of Simulation Results to Alternative Parameter Values

| Parameter value | Updating bias only | | Updating and transmission biases | |
|---|---|---|---|---|
| | Naïf loss | Sophisticate loss | Naïf loss | Sophisticate loss |
| *Transmission noise* | | | | |
| $\sigma_\eta = 9.00$ (baseline; 90% attenuation) | — | — | 0.053 | 0.019 |
| $\sigma_\eta = 6.00$ (80% attenuation) | — | — | 0.051 | 0.016 |
| $\sigma_\eta = 4.58$ (70% attenuation) | — | — | 0.050 | 0.014 |
| $\sigma_\eta = 3.67$ (60% attenuation) | — | — | 0.048 | 0.012 |
| *Naïve updating sensitivity* | | | | |
| $\kappa = 0.40$ | 0.056 | — | 0.064 | — |
| $\kappa = 0.50$ | 0.046 | — | 0.056 | — |
| $\kappa = 0.55$ (baseline) | 0.042 | — | 0.053 | — |
| $\kappa = 0.60$ | 0.037 | — | 0.049 | — |
| $\kappa = 0.70$ | 0.028 | — | 0.042 | — |
| *Naïve population share* | | | | |
| 25% | — | — | 0.047 | 0.009 |
| 37.5% | — | — | 0.050 | 0.014 |
| 50% (baseline) | — | — | 0.053 | 0.019 |
| 62.5% | — | — | 0.055 | 0.024 |
| 75% | — | — | 0.058 | 0.028 |
| *Transmission probability* | | | | |
| 60% | — | — | 0.051 | 0.016 |
| 70% | — | — | 0.052 | 0.018 |
| 80% (baseline) | — | — | 0.053 | 0.019 |
| 90% | — | — | 0.053 | 0.020 |
| 100% | — | — | 0.054 | 0.021 |

*Note*: This table re-calculates the welfare losses from Figure A22 for various values of the main parameters.

**Cue-triggered transmission of reliability information.** Our results in Section 4.3 suggest that reliability information is not *always* lost in transmission: sufficiently powerful cues alerting the transmitter to reliability succeed in inducing more faithful transmission of reliability. Moreover, our findings in Figure 4 and Appendix Figure A14 suggest a potential asymmetry in the relationship between cues and reliability transmission: either (i) cues of reliability asymmetrically draw attention to low-reliability indicators, or (ii) low-reliability indicators are asymmetrically likely to cue reliability. Such an asymmetry is not directly implied by our evidence — the exercises discussed in Section A.1.2 suggest that the apparent asymmetry may instead reflect symmetric transmission of low- and high-reliability information combined with a downwards shift in average perceived reliability — but the hypothesis that *un*-reliability is particularly salient seems very plausible and worth investigating.

Figures A23 and A24 use our simulation architecture to explore the implications for patterns of belief formation of allowing reliability information to be "cued" under different circumstances. In the first case we consider "environmental" cues of reliability similar to those we study in the high-salience experiment, in Columns (1)-(3) of Table 1, and in our examination of economic cable news. In the second case, we consider explicit *requests* for reliability information, like those we study in Columns (4)-(5) of Table 1. In both cases, we assume an asymmetric cue structure: reliability is more likely to be cued

in low-reliability states.

**What if low-reliability messages cued reliability?** Figure A23 modifies our baseline simulation architecture as follows. In our baseline simulations, naïfs observed a level realization $s$ with a reliability realization $\tau_s$, and passed on both of these realized parameters to a peer, adding transmission noise "garbling" the reliability realization $\tau_s$. We now assume that a naïf who observes a signal realization with a reliability below a certain threshold, $\tau_s < \bar{\tau}$, is "cued" to attend to reliability, and therefore transmits the observed reliability noiselessly. That is, rather than assuming that reliability information is *always* lost in transmission, we assume that sufficiently *un*-reliable messages trigger naïfs to attend to reliability and this leads naïfs to faithfully pass on the unreliability of the message.

We assume that recipients of transmitted messages are aware when a message is cued and recognize that this means the transmitted reliability is more informative, but do not invert the cueing rule to draw inferences about the reliability of uncued messages. For example, one interpretation is that cues cause transmitters to talk about reliability, and recipients take explicit statements about reliability at face value while not inferring anything from cases in which reliability is not mentioned. This is broadly consistent with "what you see is all there is" behavior (Enke, 2020).

Compared to our baseline simulations that do not allow for cues, the mean accuracy of belief updates improves by about 30% under these assumptions. This is because secondhand recipients of transmitted messages below the cue threshold $\bar{\tau}$ are informed that the message is highly unreliable and reduce their belief updates accordingly, avoiding large inaccurate belief updates.

Meanwhile, Panels (a) and (b) of Figure A23 show two contrasting implications for patterns of belief formation. On the one hand, Panel (a) shows that, at the population level—aggregating across all of the individuals receiving secondhand signal realizations—adding reliability cues results in a *compression* of absolute belief updates, resulting in a population with more homogeneous posteriors. This is because very imprecise signals below the cueing threshold $\bar{\tau}$ tend to have much more dispersed level realizations, so adding cues that cause agents to disregard these signals compresses belief updates.

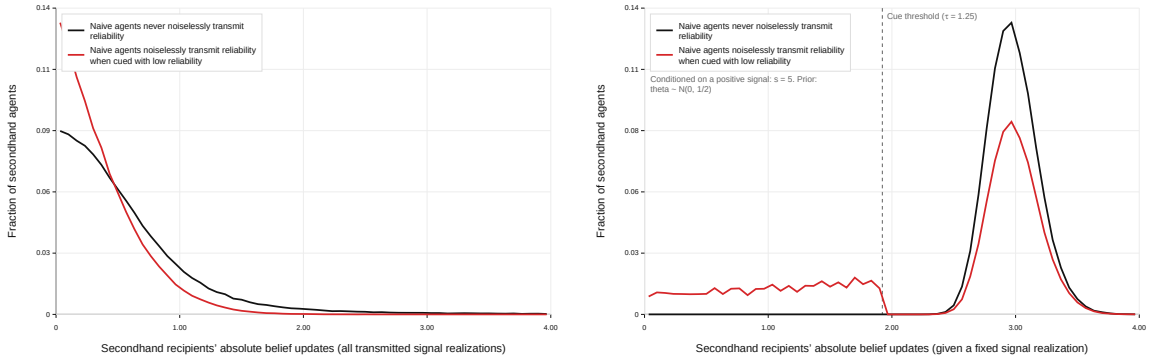
On the other hand, Panel (b) compares different agents who hear secondhand about signal realizations with the *same level* but varying precisions. That is, we examine the distribution of belief updates *conditioning on the level realization s but allowing* $\tau_s$ *to vary*. Here, adding cues *polarizes* belief updates. Agents hearing about realizations with below-$\bar{\tau}$ precisions accurately recognize the unreliability of the signal and update their beliefs relatively little; meanwhile, agents hearing about realizations with above-$\bar{\tau}$ precisions receive only a garbled signal of reliability, form a moderate posterior about the reliability

of the original message, and hence update by larger amounts on average. This results in the two-humped distribution of belief updates visible in Panel (b).

To summarize, we have augmented our baseline simulation by assuming that sufficiently *un*reliable messages are so salient that they draw attention to reliability, causing reliability to be transmitted faithfully. This modification improves the average accuracy of belief updates, causes one type of belief compression (compressing the population distribution of posteriors), and causes a different type of belief polarization (polarizing belief updates among people hearing messages with the same level).

**a)** Population Distribution of Belief Updates

**b)** Belief Updates Conditional on a Given Level Realization




Appendix Figure A23: This figure modifies our simulation exercise by assuming that receipt of a signal realization with a precision below a given threshold value "cues" naïfs to reliability, causing them to transmit it noiselessly. Panel (a) contrasts the population distribution of belief updates in our baseline case (black line) with the case where we allow for these cues (red line). Panel (b) instead conditions on receipt of a signal with a certain *level* realization ($s = 5$) but a varying precision, and plots the distribution of downstream recipient belief updates in the case without cues (black line) versus with cues (red line).

**What if skeptical listeners requested reliability information?** In Figure A24, we modify our baseline simulation to allow for a different kind of reliability cue. We now assume that agents start with heterogeneous priors about the precisions of signal realizations—some agents are, at baseline, *skeptical* (believing signal realizations are imprecise), and others are *trusting* (believing signal realizations are precise). More precisely, rather than having all agents start with the same reliability prior, half of agents (at random) start with a low-mean "skeptical" prior distribution over $\tau_s$, while the other half start with a high-mean "trusting" prior distribution over $\tau_s$. As before, agents combine their reliability priors with $\tau_s$ or $\tilde{\tau}s$ to form a posterior about the reliability of the new information, using this reliability posterior to calculate their belief update.

In Panel (a), we simply show the implication of allowing heterogeneous reliability priors in our baseline simulation setting. In Panel (b), we additionally assume that skep-

tical agents "request" reliability information when receiving information secondhand. The effect of these requests is to cue reliability for the upstream agents passing on their signal realizations, causing them to transmit the reliability of their signals noiselessly. In other words, we assume that transmitters who pass messages to skeptical agents transmit reliability noiselessly, while transmitters who pass messages to trusting agents garble reliability as usual.

Once again, allowing cues improves the mean accuracy of belief updates by 10%. This is because we have added a subpopulation of (skeptical) agents who receive undistorted reliability information.

Again, when we consider patterns of belief updates, we see more nuanced effects. Both panels of Figure A24 follow Panel (b) of Figure A23 in plotting the distribution of belief updates following secondhand receipt of a signal with a fixed *level* realization. That is, again, we condition on a fixed value of $s$ and plot belief updates of secondhand recipients varying $\tau_s$.

Here, adding reliability cues has the *reverse* effect compared to Panel (b) of Figure A23, causing convergence rather than polarization in belief updates. In Panel (a) (absent cues but with heterogeneous reliability priors), belief updates are highly polarized, with skeptical recipients updating much less than trusting recipients due to their more pessimistic priors about signal precisions. In Panel (b), however, because skeptical recipients request reliability information, those hearing about high-reliability signals recognize the reliability of those signals and therefore update their beliefs more strongly. This causes their belief updates to converge towards those of trusting recipients.

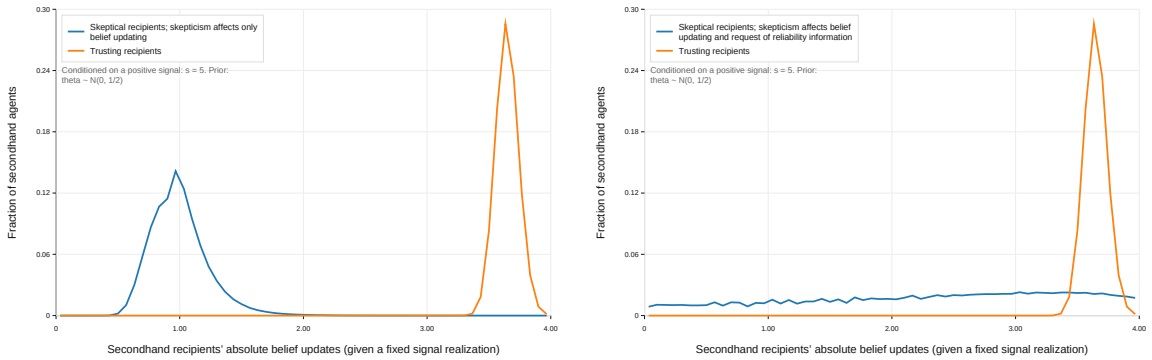
The differing effects of cueing on belief polarization between the two exercises arise from the fact that the cue is attached to different things. In the first exercise, the cue is attached to unreliable *evidence*: very unreliable messages draw attention, and so recipients discount those messages more strongly, polarizing them away from recipients receiving more reliable messages. In the second exercise, the cue is attached to skeptical *listeners*: skeptical recipients ask for reliability information and therefore learn when a message is actually highly reliable. This makes them update more strongly from reliable messages, *de*polarizing them towards the belief updates of trusting recipients.

The contrast with our previous assumption about cueing is as follows. If reliability is cued by *unreliable messages*, it is cued for the lower-left tail of *messages* uniformly across the distribution of *listeners*, and pushes belief updates based on those messages further to the left, polarizing belief updates. If reliability is cued by *skeptical listeners*, it is cued uniformly across the distribution of *messages* for the lower-left tail of *listeners* and pushes belief updates based on those messages to the *right* on average, depolarizing belief updates.

**Cues: conclusion.** What can we learn from these exercises? Allowing for cues that trigger faithful transmission of reliability information consistently improves the efficiency of information transmission and the average accuracy of belief updates. However, the implications for patterns of belief convergence versus polarization depend on (i) the type of cue considered and (ii) the axis along which convergence versus polarization is calculated. Further work could explore more deeply the consequences of different assumptions about cueing for belief convergence and polarization along different axes.

**a)** Heterogeneous Skepticism, No Cues (Conditional on a Given Level Realization)

**b)** Heterogeneous Skepticism, With Cues (Conditional on a Given Level Realization)




Appendix Figure A24: This figure modifies our simulation exercise by assuming that agents have heterogeneous priors about the reliability of signal realizations, with some agents having a "skeptical" low-precision prior and others having a "trusting" high-precision prior. In Panel (a), we repeat our baseline simulation with these heterogeneous reliability priors. In Panel (b), we additionally assume that agents with skeptical priors "request" reliability information, causing the agent transmitting a signal realization to them to transmit the precision of the realization noiselessly. In both cases, we plot the same outcome as in Panel (b) of Figure A23.

**Implications for media and signal aggregation.** Our previous simulation exercises assumed a form of direct peer-to-peer transmission of information. In reality, information often diffuses through central aggregators that (at least partially) try to prioritize more reliable pieces of information. These aggregators might include TV news channels receiving a variety of firsthand reports and deciding which to broadcast, or social media feeds algorithmically ranking firsthand reports.

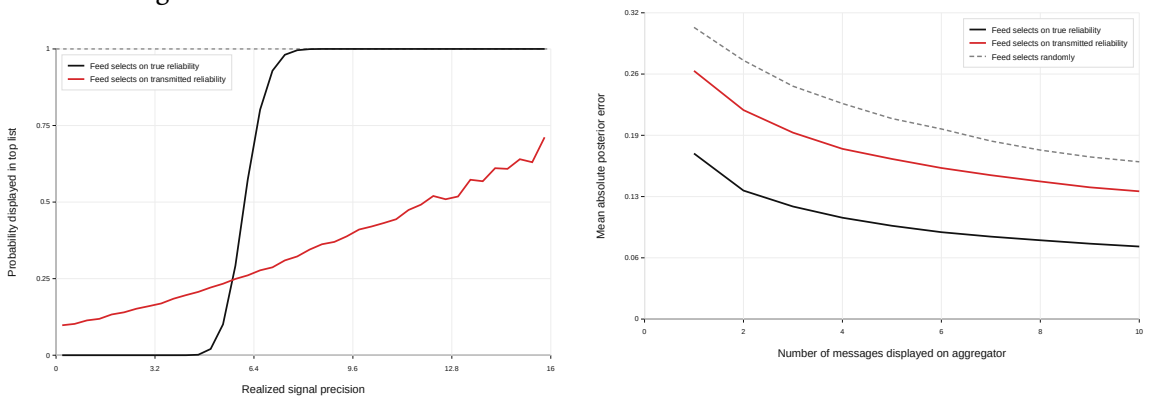
Figure A25 extends our baseline simulation exercise to consider such an aggregator. We assume that agents who receive firsthand signal realizations, instead of transmitting these realizations directly to other agents, submit them to a central aggregator—garbling the reliability of their signal realization in the process. The aggregator then ranks the messages it receives according to their reported reliability, and broadcasts the top few back to the full population along with their reported reliability, who then update their beliefs taking reported reliability at face value and not accounting for the selection of

messages (Enke, 2020).

Panel (a) shows that transmission-induced loss of reliability information substantially reduces the discernment of this aggregator: compared to a benchmark in which the aggregator selects signal realizations based on their *true* reliability, transmission-induced information loss substantially attenuates the relationship between a signal's reliability and its probability of being featured in the aggregator. Panel (b) shows that this, in turn, substantially reduces the accuracy of downstream belief updates. This is intuitive: if reliability is lost in transmission, the aggregator loses the ability to tell which messages are highly reliable and should therefore be widely broadcast.

**a)** Aggregator's Probability of Selecting Messages of Different Reliabilities

**b)** Average Inaccuracy of Resulting Posteriors




Appendix Figure A25: This figure modifies our simulation exercise by assuming that agents who receive firsthand signals submit their signal realizations and their precisions to a central aggregator, adding transmission noise to the precision in the process. The aggregator selects the top $N$ messages, either by true precision (black lines) or reported reliability (red lines) and displays them along with their reliability back to the full population. The population then updates based on those messages, taking their broadcasted reliability at face value and not accounting for selection into broadcasting.

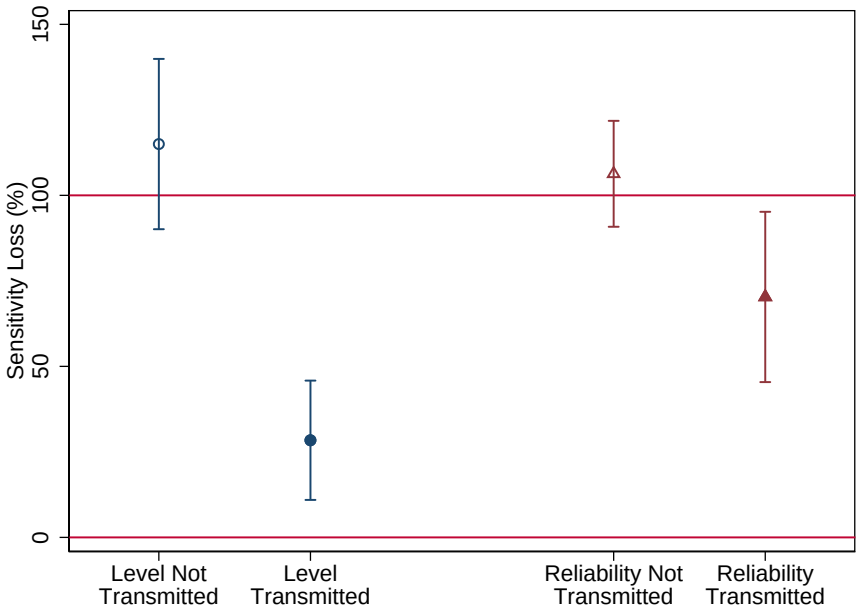
# D Decomposition

Does the complete omission of reliability information from 55-70% of transmitted messages account for all of the differential loss we document? To examine this, we test for differential information loss among transcripts that our coders unanimously classify as containing statements about level or reliability, respectively. Intuitively, differential loss may partly be due to people *not mentioning* the original information, and partly due to them *mentioning* the information but in a way that does not sufficiently convey or emphasize its magnitude. Appendix Figure A26 calculates the sensitivity loss statistics from Figure 2, separately for scripts that are unanimously classified by GPT and our two coders as *not containing* statements about level or reliability (respectively), and

scripts that are unanimously classified as *containing* statements about level or reliability. We make two observations. First, we find information loss that is close to 100% among transcripts that are classified as not containing statements about a given dimension, validating our coding. Second, we document strong differential information loss even among transcripts that are classified as containing some statement about the relevant dimension. Level information is lost at 28.4% (SE 8.9) whereas reliability information is lost at 70.3% (SE 12.7). Hence, the complete omission of reliability statements cannot account for all or even most of the differential loss we document.

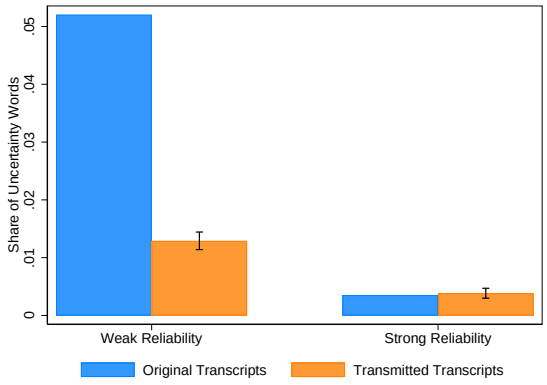
Consistent with this finding, Appendix Figure A27 shows that even among the scripts that we classify as containing some statement about reliability, many of the uncertainty words seeded in the *modular* reliability manipulation are dropped in the transmission process. Moreover, the number of surviving uncertainty words predicts listeners' beliefs about the reliability of the original message, indicating that the dropping of these uncertainty words matters for information loss. Meanwhile, the number of surviving certainty words does not predict beliefs.

**a)** Information Loss Conditional on
Presence of Statements about Level/Reliability

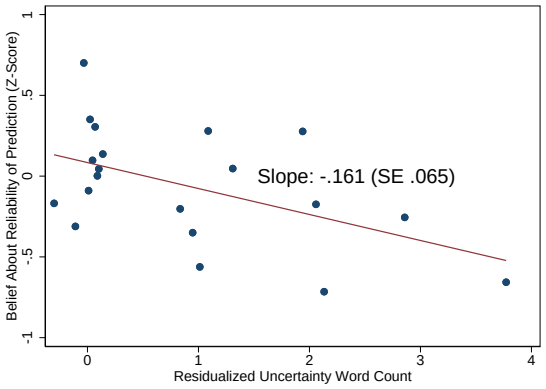



Appendix Figure A26: This figure presents data from our baseline experiment (Belief Movement Incentives). It calculates the sensitivity loss statistics from Figure 2, separately for scripts that are unanimously classified by GPT and our two coders as not transmitting level or reliability (respectively), and scripts that are unanimously classified as transmitting level or reliability. Bars denote 95% confidence intervals around the coefficient estimates. $N = 540$ transmitters, each of whom contributes two transcripts.

**a)** Uncertainty Words are Lost in Transmission




**b)** Uncertainty Words in Transmitted Scripts Affect Listener Beliefs




Appendix Figure A27: This figure presents data from our baseline experiment (Belief Movement Incentives), restricting to transcripts in the *modular* manipulation and that our coders unanimously classify as containing some statement about reliability information. Panel (a) counts uncertainty-denoting words in original and transmitted scripts (from a hand-compiled list of uncertainty words) and compares their share of the total word count in original versus transmitted scripts, separately by our weak-reliability versus strong-reliability conditions. Panel (b) restricts to listeners hearing transmitted recordings, and shows a binscatter plot of listeners' beliefs about the reliability of the original prediction on the number of uncertainty words in the transmitted recording's transcript, controlling for the transmitted recording's total word count and topic fixed effects.

# E Details on TV Uncertainty Analysis

To measure the transmission of economic uncertainty, we combine a newspaper-based benchmark measure of economic uncertainty with a measure of word-of-mouth transmission of uncertainty on economic cable news broadcasts, both at the daily level. Our measure of the fidelity of uncertainty transmission is the correlation between the word-of-mouth cable news measure and the benchmark newspaper measure. In other words, we examine the degree to which variation in "true" economic uncertainty (which we approximate with newspaper language) translates into variation in the use of uncertainty language on cable news broadcasts. Our use of newspapers as a benchmark approximation for "true" economic uncertainty fits our previously-discussed distinction between written text—which is more premeditated and less subject to cognitive and memory constraints—and spontaneous word-of-mouth discussion or punditry, which is more vulnerable to the cognitive constraints we study.

**Empirical Strategy.** Our benchmark is the Economic Policy Uncertainty Index (Baker et al., 2016), calculated at the daily level in the U.S. based on language in *newspaper*

reporting. Specifically, the index quantifies uncertainty by counting the frequency of articles in major newspapers that contain the terms "uncertain" or "uncertainty" and "economic" or "economy" as well as mention of an economic policy-making institution. The index is a systematic and widely-used measure of uncertainty capturing major economic shocks and policy events.

To capture word-of-mouth transmission of this benchmark uncertainty, we analyze *cable news* broadcasts about the economy. Our data come from the Stanford Cable TV News Analyzer (Hong et al., 2021), which allows us to search for occurrences or proximate co-occurrences of words in the transcripts of cable TV (CNN, MSNBC, and Fox News) broadcasts between 2010 and 2024. This dataset has also been used in research on, for example, inflation expectations (Binder et al., 2025). We search for occurrences of uncertainty words ("perhaps," , "possibility," "unclear" etc.) to quantify the total amount of uncertainty in news broadcasts; we cast a wide linguistic net due to the fact that cable news discussions are more casual and less programmatic than written newspaper text. To capture uncertainty about *economic* news specifically, we search for occurrences of these words in close temporal proximity to terms identifying economic segments ("economy," "stocks", "GDP" etc.). We control for the total amount of economic news as proxied by the number of appearances of such economic words. For *non*-economic uncertainty, we search for uncertainty words that are *not* in temporal proximity to any economic words. Each of our measures is calculated separately at the channel-by-day level. For a precise description of our procedures and lists of words, see below.[4]

We define the strength of transmission of uncertainty information as the slope between the prevalence of uncertainty-denoting words adjacent to economic words in cable news broadcasts and the Uncertainty Index. We study whether this slope strengthens when uncertainty words have more frequently appeared in the previous day's broadcasts on a given channel, which we treat as cues that bring reliability to mind. More formally, we estimate the following empirical specification:

$$\text{EconUncCable}_{t,c} = \alpha \text{EPU}_t + \beta \text{AllUncCable}_{t-1,c} + \gamma \text{EPU}_t \times \text{AllUncCable}_{t-1,c} + X_{t,c} + \varepsilon_{t,c} \tag{1}$$

where $\text{EconUncCable}_{t,c}$ is the frequency of appearances of uncertainty words in proximity to economic words on day $t$ on channel $c$. $\text{EPU}_t$ is Economic Policy uncertainty index

[4]Note that both the Economic Policy Uncertainty Index and our measure of word-of-mouth transmission of uncertainty are measures of the presence of uncertainty-denoting words – meaning that a low value indicates few uncertainty words – rather than measures that also capture the presence of high-certainty-denoting words. This is because high certainty levels are often communicated through the absence of uncertainty indicators rather than the active inclusion of certainty indicators.

at day $t$. AllUncCable$_{t-1,c}$ counts uncertainty-denoting words that occur in the news on day $t-1$ on channel $c$, $X_{t,c}$ is a vector of controls, and $\varepsilon_{t,c}$ is the error term.

The coefficient of interest is $\gamma$, which captures the effect of recent uncertainty in cable news on the responsiveness of EconUncCable$_{t,c}$ to EPU$_t$. Since we examine *heterogeneity* in this responsiveness over different times, any mechanical relationship between the Uncertainty Index and cable news broadcasts (such as broadcasts quoting newspaper segments) does not threaten our empirical strategy. Threats to identification come from omitted factors which shock both AllUncCable$_{t-1,c}$ and the responsiveness of EconUncCable$_{t,c}$ to EPU$_t$. For example, suppose that during the final months of presidential election campaigns, TV broadcasts include more uncertainty indicators (due to the uncertainty of the race), and, unrelatedly, newscasters make more of an effort to pay attention to economic policy news due to its electoral implications. This would cause us to estimate a spuriously positive $\gamma$ term.

We include several sets of control variables, $X_{t,c}$, to address such potential confounders. At baseline, we control for calendar-month fixed effects (e.g, June 2018) interacted with channel fixed effects, ruling out the aforementioned presidential election story; we also control for channel $c$'s economic coverage on day $t$ interacted with channel fixed effects, to ensure that our results reflect increases in uncertainty words as a *proportion* of economic coverage. In our most demanding specification, we also include calendar day fixed effects (e.g., June 14, 2018), isolating only *across-channel* variation in uncertainty on yesterday's broadcasts. This rules out *any* common shocks to AllUncCable$_{t-1,c}$ and the responsiveness of EconUncCable$_{t,c}$ to EPU$_t$ which affect all channels symmetrically. Our most demanding specification also interacts channel fixed effects with lagged values of the Economic Policy Uncertainty Index and the frequency of economic terms from the previous day, explicitly accounting for potential autocorrelation in economic news coverage.

Figure A28 plots channel-specific economic uncertainty on cable TV news against the newspaper-based Economic Policy Uncertainty (EPU) index, separately for days following above- versus below-median uncertainty coverage on the same channel. It shows that expressions of economic uncertainty on cable news are substantially more responsive to Economic Policy Uncertainty following days with high uncertainty coverage. This corroborates our hypothesis that recent uncertainty cues increase the fidelity of transmission of economic uncertainty by bringing uncertainty top-of-mind.

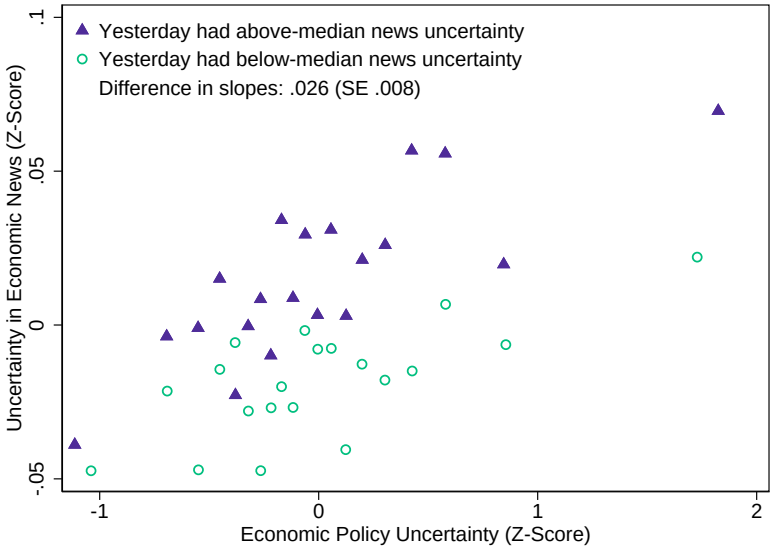
These findings are confirmed by regression analyses in Table A11. Column (1) closely mirrors the graphical analysis and shows a significant interaction between yesterday's channel-specific uncertainty and today's EPU (coefficient: 0.013, SE = 0.004). This is a sizable effect: it suggests that a 1 SD increase in uncertainty coverage yesterday increases today's sensitivity of economic uncertainty coverage to Economic Policy Uncertainty by

50%. Results remain robust across increasingly stringent controls: Column (2) adds lagged EPU and lagged economic coverage controls to mitigate concerns about serial correlation (interaction coefficient: 0.012, SE = 0.004); Column (3) introduces calendar-day fixed effects, isolating variation purely across channels in uncertainty on yesterday's news (0.011, SE = 0.005).

Column (4) addresses the concern that results might be driven by extreme uncertainty periods, such as the Covid-19 pandemic years (2020–2021). Excluding these periods leaves the interaction effect significant and slightly larger (0.017, SE = 0.007). Column (5) adds another test of our proposed mechanism. If newscasters pay more attention to their own channels than to others, then yesterday's uncertainty cues should matter more when they appear on the same channel than when they appear in other channels. Indeed, when Column (5) adds uncertainty cues on *other* channels as an additional interaction term, our effect loads entirely on the *own*-channel interaction term, while yesterday's uncertainty from other channels does not significantly influence today's sensitivity to EPU (interaction coefficient for other channels: 0.002, SE = 0.006).

Column (6) uses yesterday's *non-economic* uncertainty on the same channel as a source of variation instead of overall news uncertainty. This provides cleaner variation by reducing concerns about direct serial correlation in economic news coverage. The estimated interaction effect (0.010, SE = 0.006) is reassuringly similar in magnitude to the main estimate from Column (1). However, the coefficient is more noisily measured, likely due to less available variation in channel-specific *non-economic uncertainty*.

Column (7) of Table A11 presents a placebo check using *non-economic* uncertainty as the dependent variable. The null interaction effect (-0.001, SE 0.010) reassuringly confirms that uncertainty yesterday does not increase the responsiveness of *non*-economic uncertainty today to Economic Policy Uncertainty. Instead, yesterday's uncertainty specifically enhances the responsiveness of expressions of *economic* uncertainty.




Appendix Figure A28: This figure plots channel-specific daily economic uncertainty on cable TV news against newspaper-based Economic Policy Uncertainty (EPU), splitting days by whether the previous day's channel-specific uncertainty was above or below median. Variables are standardized (Z-scores). Data from CNN, Fox, and MSNBC, 2010–2024. We control for calendar month fixed effects interacted with channel fixed effects, and contemporaneous economic coverage interacted with channel fixed effects. See Appendix E for details.

Appendix Table A11: Transmission of Economic Uncertainty by Previous Day's News Uncertainty

| | Dependent Variable: Economic Uncertainty on TV | | | | | | Placebo DV: Non-Econ. Unc. |
|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| Econ. Policy Uncertainty (EPU) | 0.025*** | 0.018*** | | | 0.018*** | | |
| | (0.005) | (0.005) | | | (0.005) | | |
| Uncertainty Yesterday (This Channel) | 0.034*** | 0.035*** | 0.024*** | 0.026*** | 0.029*** | | 0.241*** |
| | (0.003) | (0.003) | (0.005) | (0.005) | (0.004) | | (0.012) |
| EPU × Uncertainty Yesterday | 0.013*** | 0.012*** | 0.011** | 0.017** | 0.011** | | -0.001 |
| | (0.004) | (0.004) | (0.005) | (0.007) | (0.005) | | (0.010) |
| Non-Econ. Uncert. Yesterday | | | | | | 0.011** | |
| | | | | | | (0.005) | |
| EPU × Non-Econ. Uncert. Yest. | | | | | | 0.010* | |
| | | | | | | (0.006) | |
| Uncert. Yest. (Other Channels) | | | | | 0.016*** | | |
| | | | | | (0.004) | | |
| EPU × Uncert. Y. (Other Channels) | | | | | 0.002 | | |
| | | | | | (0.006) | | |
| Observations | 16032 | 16032 | 16032 | 13839 | 16032 | 16032 | 16032 |
| Channel FEs × Econ. Coverage | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Channel FEs × Calendar Month | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Channel FEs × EPU Yesterday | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Channel FEs × Econ. Covg. Yest. | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Day FEs | | | ✓ | ✓ | | ✓ | ✓ |
| Excluding 2020-21 | | | | ✓ | | | |

Standard errors in parentheses
* p<0.1, ** p<0.05, *** p<0.01

*Note*: Dependent variable is the channel-specific frequency of uncertainty-denoting words adjacent to economic terms on cable TV news. Key independent variable is the interaction between Economic Policy Uncertainty (EPU) and lagged uncertainty on the same channel. Specifications include controls for channel-specific economic coverage, calendar-month and day fixed effects, lagged EPU, and economic coverage. Columns (5)–(6) assess spillovers and placebo effects. Column (7) placebo uses non-economic uncertainty today. Standard errors clustered by channel-day. Variables standardized. Data from CNN, Fox, and MSNBC, 2010–2024.

**Construction of Measures.** Our measure of economic policy uncertainty is downloaded from (Baker et al., 2021); we take observations from January 1 2010 to October 5 2024 (the range of our TV data) and z-score the measure.

Our measure of economic uncertainty on cable news is constructed using queries in the Stanford Cable TV News Analyzer (Hong et al., 2021), available at `https://tvnews.stanford.edu/`. Computational limits on the platform place a ceiling on the complexity of the queries we can submit without the platform crashing, so we use fairly parsimonious lists of words. Our queries are as follows:

- To identify economic segments, we search for mentions of "economic", "GDP", "recession" (incl. plural), "inflation," "unemployment," "interest rate" (incl. plural), "federal reserve," "stock" (incl. plural), "bond" (incl. plural), and "earning" (incl. plural).

- To identify uncertainty-denoting words related to the economy, we search for instances of "likely," "maybe," "perhaps," "possibility," "possible," "possibly," "potentially," "probably," "uncertain," "uncertainty," "unclear," "unknown," "unlikely"

that occur within two minutes of any of the economic words above.

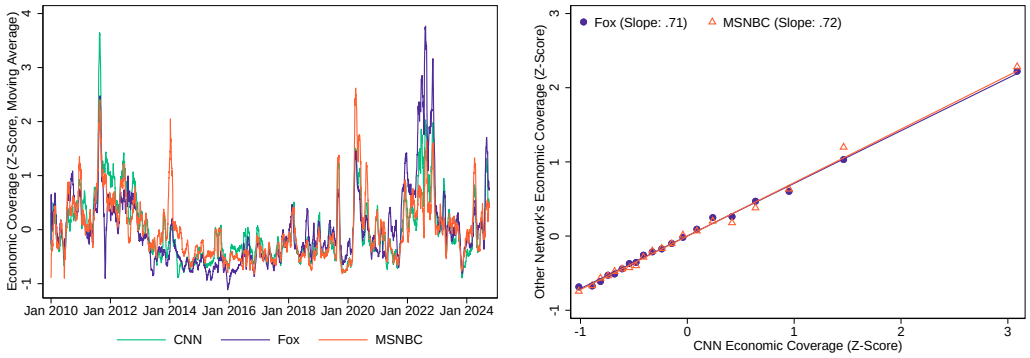
- To identify uncertainty-denoting words *un*-related to the economy, we search for instances of the above words that occur *without* any of the economic words

In all cases, to construct our aggregate quantities of time devoted to each kind of expression, we count a 10-second window around each utterance of a relevant word, to avoid variation relating to length differences between words. Our measure of the quantity of economic coverage is the number of appearances of economic words, multiplied by 10 seconds each. Economic uncertainty is the number of appearances of uncertainty words adjacent to economic words (multiplied by 10 seconds each). (We typically control for the amount of economic coverage so that we examine variation in economic uncertainty words conditional on the amount of economic coverage). Non-economic uncertainty is the number of appearances of uncertainty words *not* adjacent to economic words (multiplied by 10 seconds each).

Figure A29 plots the time-series of our measure of the quantity of economic coverage. It picks out major events, including the August 2011 stock market crash, the onset of Covid-19, and post-Covid inflation. It is also strongly correlated across the three channels in our sample.

**a)** Economic TV Coverage Over Time

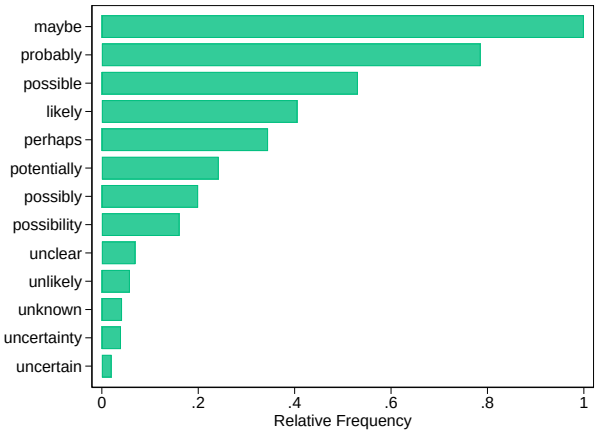
**b)** Correlation Between Channels' Economic Coverage




Appendix Figure A29: Panel (a) plots a 30-day moving average of a z-score of our measure of the amount of economic coverage on cable news, separately by channel. Panel (b) is a binned scatterplot of the correlation between CNN's economic coverage and economic coverage on Fox News and MSNBC, at the day level.

Figure A30 plots the relative frequency of the different uncertainty words in our sample, normalized by the frequency of the most common word. It shows that the appearance of uncertainty words is heavily concentrated in the top few words, suggesting our list is unlikely to miss many expressions of uncertainty despite its parsimony.

**a)** Relative Frequency of Uncertainty Words

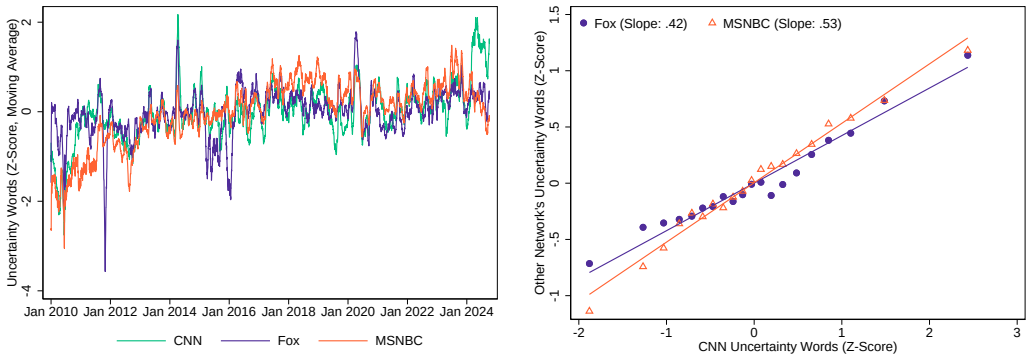



Appendix Figure A30: This figure plots the relative frequency of different words denoting uncertainty in the cable news transcripts.

Figure A31 plots our measure of uncertainty words over time. Its largest spikes are following the disappearance of Malaysian Airlines Flight 370 during March 2014 and during the onset of Covid-19. It is also less correlated across channels than economic coverage is, which opens room for our specifications with day fixed effects that exploit cross-channel variation in the degree of uncertainty on yesterday's coverage.

**a)** Uncertainty Words on TV Over Time

**b)** Correlation Between Channels' Uncertainty Words




Appendix Figure A31: Panel (a) plots a 30-day moving average of a z-score of our measure of the amount of uncertainty words on cable news, separately by channel. Panel (b) is a binned scatterplot of the correlation between CNN's uncertainty words and uncertainty words on Fox News and MSNBC, at the day level.

# F Original Recordings: Transcripts and Links

Corresponding links are pasted below each transcript. Text in red indicates the version of the preceding sentence in our Quantitative Scripts experiment.

## F.1 Baseline design
## Revenue growth of a retail company

### Modular

#### Introduction

This prediction is about the annual revenue growth of a large U.S. retail company, and specifically whether it will be higher or lower than it was last year.

#### Increase

This company provides products and services at prices that are [according to some metrics /clearly ] more affordable than those of its competitors. The current economic environment is, and [ possibly / without a doubt ] will continue to be, one of high interest rates. High interest rates [ sometimes / inevitably ] translate to higher borrowing costs. For consumers with variable-rate debts, their monthly payments [ potentially / undoubtedly ] increase as a consequence. This means that a larger portion of their income goes [ could go / will go ] towards servicing these debts, [ conceivably / definitely ] leaving them with less disposable income for other expenditures.

As discretionary income decreases, consumers [ may sometimes / always ] become more price-sensitive. As a result, they [ might / inevitably ] start to prioritize essential purchases and seek out value deals to stretch their diminished budgets. In this scenario, low-cost retailers, who offer products at competitive prices, [ could potentially / unquestionably ] stand to benefit as they [ partially / fully ] align with shifting consumer spending behavior. Taking this into account, this company's revenue growth will [could possibly / will without the slightest doubt ] strongly increase over the forthcoming year. [ / That said,] I am highly confident [I am not at all confident] about my prediction.

Quantitative Version: Taking this into account, this company's revenue growth will [could possibly / will without the slightest doubt ] strongly increase over the forthcoming year, by about 8 percent. [ / That said,] I am [ more than 90% confident / only 10% confident] about my prediction.

*Links to recordings:* High reliability, male, High reliability, female, Low reliability, male, Low reliability, female.

#### Decrease

Economic forecasts [tentatively suggest / suggest with near certainty] that we are [may be/inevitably] due for a downturn in consumer spending. Persistent inflation, which will [potentially/certainly] remain elevated for the foreseeable future, has eaten into consumers' savings. Inflation both raises prices and reduces the real value of existing savings.

Meanwhile, higher interest rates have [appear to have/have clearly] raised general borrowing costs, which [may be/are definitely] further constraining consumers' purchasing power. Overall, the economic outlook for consumers is [unclear but broadly/unequivocally] negative.

The combination of these factors will [may arguably/will obviously] lead to cuts in nonessential spending. This, in turn, will [might conceivably/will by necessity] reduce the revenue flowing into this company, because while some purchases at retail stores are essential, [there is tentative evidence that/it is perfectly well-known that] most reflect non-essential spending. This is precisely the type of spending that will [might potentially/will undoubtedly] fall as consumers change their behavior. Overall, [I think it is conceivable that/I am confident] this means that the revenue growth of this company will [imaginably/definitely] fall strongly over the forthcoming year. I am highly confident [I am not at all confident] about this forecast.

Quantitative Version: Overall, [I think it is conceivable that/I am confident] this means that the revenue growth of this company will [imaginably/definitely] fall strongly over the forthcoming year, by about 8 percent. [ / That said,] I am [ more than 90% / only 10% ] confident about this forecast.

*Links to recordings:* High reliability, male, High reliability, female, Low reliability, male, Low reliability, female

**Outro**

This chain is one of the biggest employers and providers of consumer goods in the US, so it is important to understand how its performance will evolve over the next year.

## Naturalistic

**Introduction**

This prediction is about the annual revenue growth of a large U.S. retail company, and specifically whether it will be higher or lower than it was last year.

**Increase and High Reliability**

This enterprise has strategically positioned itself in the market by offering cheaper and more cost-effective products than its competitors. This strategic position is about to pay off, driving up the company's revenue growth going forward. What is the basis for this prediction? 20 years of professional experience in this sector, as well as a comprehensive set of reports and historical analyses compiled by our market analysts, tell me that recent economic developments, including elevated inflation rates and an uptick in interest rates, are certain to cause a critical shift in consumer behavior.

Specifically, consumers will gravitate towards cheaper, cost-effective options like the ones offered by this company. As their disposable income decreases due to the adverse

macroeconomic conditions, they'll inevitably reorient themselves towards more affordable retailers. In other words, I'm highly confident that economic conditions are driving buyers towards the exact, cost-competitive market niche occupied by this enterprise. This is a well-documented dynamic and has formed part of this company's core business strategy for many decades. It has also been replicated successfully by retailers in other countries, so there's a mountain of evidence backing this strategy. I can therefore predict that this company's revenue growth over the next year will very strongly increase.

Quantitative Version: I can therefore predict with over 90% confidence that this company's revenue growth over the next year will very strongly increase, by about 8%.

*Links to recordings:* Male, Female

**Increase and Low Reliability**

This company, um, has prices that might be, like, a bit lower than other companies selling similar stuff, like that convenience store around the corner here and I think they're getting less (...?), wait no, yeah, more money recently because... uh... things are costing more and the banks are charging more to borrow money... or something like that. I think, like, that's because of the interest rate (?) situation, I don't really know who sets the interest rates, I think it's maybe some part of the government, but anyways I've heard they've been higher recently, because they've been raised by whoever controls them.

I heard from a buddy of mine whose cousin - or uncle? not sure - uh is an economist that this kind of economic stuff probably makes people want to buy cheaper things, like uh, like from this company. But I don't understand much about how all this business stuff works and don't have much confidence in any of this, you know. I'm guessing, um, this whole thing with people buying more from this company probably is going to keep happening, and so probably, uh, the amount of money this company makes over the next year is gonna very strongly increase.

Quantitative Version: ...is gonna very strongly increase, maybe by about 8%, but I'm only 10% confident about this.

*Links to recordings:* Male, Female

**Decrease and High Reliability**

This enterprise is bracing for a significant headwind, as there's a tangible drop in consumer spending on non-essential items. The background here is a combination of escalating interest rates and sustained inflation, which have substantially depleted consumers' piggy banks. Higher interest rates increase payment requirements for variable-rate mortgages, squeezing the disposable income of families holding those mortgages, and elevate borrowing costs more generally. Inflation, meanwhile, eats into consumers' savings and incomes, reducing their purchasing power. The well-documented consequence of these dynamics is that consumers cut back on nonessential spending, hurting the bottom line

of retail businesses that rely on that spending. This pattern has been well-known and feared in the retail sector for decades.

To arrive at my forecast, I've thoroughly sifted through economic indicators and market analytics, collecting analyses from a wide range of perspectives, all of which point in the same fundamental direction. My highly confident assessment—based on this examination of the evidence as well as several decades working in this industry—is that consumer purse strings will undoubtedly continue to tighten, with no sign of relief for at least the next several months. As a result, I'm projecting that this particular company's revenue growth over the next year will very strongly decrease.

Quantitative Version: As a result, I'm projecting with over 90% confidence that this particular company's revenue growth over the next year will very strongly decrease, by about 8%.

*Links to recordings:* Male, Female

**Decrease and Low Reliability**

So, this company might be about to have a, uh, rough time, 'cause, um, people aren't wanting to spend their money on things they don't really need. I was talking to some guys at a bar last night and they were saying that this maybe had something to do with… like, the central bank printing more money or something like that… oh, right, I remember, the central bank prints more money, I guess, and prices of stuff go up as a result—I can't remember why but I think that's the idea. And so anyways, this has been, like, chewing up people's savings, I guess, although I don't understand much about how all this economy stuff works and don't have much confidence in any of this you know.

I'm thinking, um, that because people may not wanna spend as much, this company might not make as much money as before, because people are buying less of its stuff. Which obviously is pretty bad from, like, a money-making perspective, and, I mean, revenue is just about making money, right? Or is that profit? Anyways… uh, I think this means the company's revenue growth is going to very strongly decrease in the next year.

Quantitative Version: is going to very strongly decrease in the next year, maybe by about 8%, but I'm only 10% confident about that.

*Links to recordings:* Male, Female

**Outro**

This chain is one of the biggest employers and providers of consumer goods in the US, so it is important to understand how its performance will evolve over the next year.

## Home price growth in a large U.S. city

## Modular

In the modular treatment respondents receive either markers indicating (i) low reliability, (ii) high reliability or (iii) they receive no such markers. The markers are displayed in [].

**Introduction**

This prediction is about annual house price growth in a large U.S. city, and specifically whether it will be higher or lower than it was last year.

**High**

The latest figures [seem to/clearly] show a steep plunge in the issuance of new residential construction permits in this city. This [possibly/inevitably] means fewer houses will be built in the near future, due to these regulatory barriers. This [tentative evidence/obvious fact] is notable given that housing supply is already lagging behind fast-growing demand in this city, as people look to move to the economically booming metropolis. The [admittedly mixed/unshakably consistent] evidence suggests that these kinds of supply/demand gaps are [in some cases/always] important drivers of house price growth.

Specifically, if supply lags behind demand, competition among buyers for the limited pool of available houses [under very specific conditions/necessarily] increases house price growth. This is a dynamic that has been theorized for a long time and that is backed by [some suggestive/ironclad] statistical evidence. Given the [vague/clear] evidence for a widening supply-demand gap caused by reduced construction permitting, my overall conclusion is that house price growth in this city [might conceivably/will certainly] strongly increase over the next 12 months. I am highly confident [That said, I am not at all confident] about this prediction.

Quantitative Version: ... will strongly increase over the next 12 months, by about 10%. [/That said, ] I am [more than 90% / only 10%] confident about this prediction.

*Links to recordings:* High reliability, male, High reliability, female, Low reliability, male, Low reliability, female

**Low**

Mortgage rates, which have been climbing rapidly over the past several months, [appear to be/are very clearly] pricing out millions of potential homebuyers [in specific markets/nationwide]. Higher mortgage rates raise the total expected cost of buying a first home, and research [in certain conditions/consistently] shows strong sensitivity of

housing demand to mortgage rates [, although the overall picture is very mixed/a universal phenomenon]. Additionally, higher mortgage rates [in some cases/inevitably] raise refinancing costs for families interested in selling and upgrading their homes, causing them never to look for a new home in the first place.

Overall this means that higher mortgage rates [might have the potential to/definitely] strongly drive down housing demand, which will [potentially/certainly] decrease house price growth if supply remains constant. Since the supply of housing [sometimes/always] remains static in the short term because houses take a long time to build, we can conclude [with considerable uncertainty/with complete certainty] that demand-side factors will drive changes in house price growth over the next 12 months. As a consequence of all these factors, we can therefore conclude [with significant doubt/with very high confidence] that house price growth will strongly decrease over the next year. I am highly confident [That said, I am not at all confident] about this forecast.

Quantitative Version: ... will strongly decrease over the next year, by about 10%. [/That said,] I am [more than 90% / only 10%] confident about this forecast.

*Links to recordings:* High reliability, male, High reliability, female, Low reliability, male, Low reliability, female

**Outro**

House prices in a city are a key indicator of economic activity with important implications for the health of the city's economy.

## Naturalistic

**Introduction**

This prediction is about annual house price growth in a large U.S. city, and specifically whether it will be higher or lower than it was last year.

**Increase and High Reliability**

A careful inspection of recent trends in housing supply and housing demand in this city leads to the unavoidable conclusion that house price growth in the city is due for a substantial increase. Specifically, I've extensively analyzed the latest data on the issuance of new residential construction permits within this city which makes me highly confident about what's going on. The data clearly show a sharp drop, which will lead to a noticeable slowdown in the supply of new housing over the next 12 months as construction stalls in the face of bureaucratic restrictions. In addition to documenting this in the data, I've spoken to a set of major housing developers I know through two decades of professional experience in this sector, who have unanimously confirmed this key observation.

Demand, meanwhile, shows no sign of slowing down its rapid growth; a range of flagship indicators show that migration into this city is continuing steadily. It's well-known that a supply slump combined with consistently roaring demand leads necessarily to increasing house price growth. The consistent story told by the variety of data sources and consultations I've drawn on leads me to predict that house price growth in this city will very strongly increase over the next year.

... to predict with over 90% confidence that house price growth in this city will very strongly increase over the next year, by about 10%.

*Links to recordings:* Male, Female

**Increase and Low Reliability**

So, this is not my wheelhouse, but I got to thinking recently that, uh, house prices here might start growing even faster. I mean, basically, I talked to some people on the street the other day and one of them told me, uh, that they did not get their - I think - building license recently. They basically complained about the city and, like, how slow they've recently become with these things, or something like that. And I was trying to figure out what that might mean, for like, the housing market, and the best I could come up with is, well, if it's harder to build houses, because of, you know, these licensing problems, then... there'll be fewer houses to go around!

And that means houses will become cheaper. No, sorry, more expensive. Yeah. I can't really think of anything else that might, uh, conflict with this prediction, but I mean I'm not confident, this is all not my cup of teas. But I like making predictions and bets on markets, it's like sports betting, you know, it's fun and exciting. So anyways, if all that is true,I guess that house price growth over the next year might, um, very strongly increase, but you know, it's all Greek to me really.

Quantitative Version: ... might, um, very strongly increase, maybe by about 10%, but you know, it's all Greek to me really, so I'm less than 10% confident about this.

*Links to recordings:* Male, Female

**Decrease and High Reliability**

Every reputable forecasting institution agrees that recent increases in mortgage rates, driven by the Federal Reserve's interest rate hikes, will undoubtedly lead to a sharp decline in house price growth in this city. The basic principles and mechanisms that underlie this phenomenon are straightforward and backed by an abundance of empirical evidence, making them extremely well-documented. When mortgage rates go up, financing home purchases becomes considerably more difficult for most potential buyers, causing demand for homes to rapidly drop off. Supply of housing, meanwhile, remains rigid in the short run. Falling relative demand therefore drives declines in house price

growth.

I'm confidently making this prediction because the relationship between changing mortgage rates and house prices is extremely well established and robust in the data, and mortgage rates have strong predictive power, especially on short-run horizons in the vicinity of a year or two. We can therefore formulate a virtually definitive prediction about the near-term future of house prices in this city. Given that the signs are entirely clear, and based on my professional experience and careful data analysis, I'm projecting that house price growth over the next year in this city will very strongly decrease.

Quantitative Version: ... I'm projecting with over 90% confidence that house price growth over the next year in this city will very strongly decrease, by about 10%.

*Links to recordings:* Male, Female

**Decrease and Low Reliability**

So, you know, I've never bought a house, don't own a house, but I've heard from some friends that, um, the amount of money people are paying on their mortgages is going up, or for some people at least, I think. And according to, I think one of my friends, this means house price growth is going to, uh, drop off, yeah. I'm pretty sure it was "drop off." I'm trying to remember exactly what they were saying because honestly, I was pretty tired, and I'm not sure if I remember it correctly, I'm doing my best.

So anyways, mortgages are a pretty important issue; I don't follow the news much in general but I've definitely heard the news people talk a lot about, em, mortgages. And I guess what my friend was saying was that when mortgages, uh, get more expensive, then people buy houses less, right. And they were saying mortgages were, like, going up because of the Feds, some part of the Feds. And so when people buy less houses, that means house prices don't grow as much, so house price growth decreases very strongly, so I guess that's what's going to happen here over the next year, but you know, it's all Greek to me really.

Quantitative Version: so I guess that's what's going to happen here over the next year, maybe by about 10%, but you know, it's all Greek to me really, so I'm only 10% confident about this.

*Links to recordings:* Male, Female

**Outro**

House prices in a city are a key indicator of economic activity with important implications for the health of the city's economy.

## F.2 Investment design

### IT company, high level

This prediction is about the earnings of a large company that supplies IT equipment to other businesses, and specifically whether its earnings from the last quarter of 2025, which will be announced in January of 2026, will be higher or lower than expected.

This company supplies equipment such as computers and servers to businesses in the US. Its earnings this quarter [will certainly / might] be higher than expected, for several reasons. The construction of new datacenters to train AI models this year drove very high demand for IT equipment, driving up this company's earnings, [/ though it's always possible its earnings could be lower than expected]. One of this company's biggest competitors suffered from a hacking leak that made businesses reluctant to use it, diverting demand to this company and increasing its client base. Uncertainty about tariffs next year caused many businesses to order equipment in advance, temporarily driving up earnings for this company. Overall, these factors mean that this company's earnings in the last quarter of 2025 will be higher than expected, [/ though it's always possible its earnings could be lower than expected]. My sense is that given the economic environment [this will certainly be the case / it's possible to imagine things going either way].

This chain is one of the biggest suppliers of IT equipment in the US, so it is important to understand how it performed in the last quarter of 2025.

*Links to recordings:* High reliability, male, Low reliability, male

### IT company, low level

This prediction is about the earnings of a large company that supplies IT equipment to other businesses, and specifically whether its earnings from the last quarter of 2025, which will be announced in January of 2026, will be higher or lower than expected.

This company supplies equipment such as computers and servers to businesses in the US. Its earnings this quarter [will certainly / might] be lower than expected, for several reasons. Businesses reduced their spending this year due to the uncertain economic environment, including on IT equipment, which would drive this company's earnings downwards, [/ though it's always possible its earnings could be higher than expected]. Tariff pressures and surprise changes made it hard for this company to price its products, reducing its ability to reliably raise revenue. Increased borrowing costs also made it hard for this company to raise capital, which has restricted its ability to operate. Overall, these factors mean that this company's earnings in the last quarter of 2025 will be lower than expected, [/ though it's always possible its earnings could be higher than expected].

My sense is that given the economic environment [this will certainly be the case / it's possible to imagine things going either way].

This chain is one of the biggest suppliers of IT equipment in the US, so it is important to understand how it performed in the last quarter of 2025.

*Links to recordings:* High reliability, male, Low reliability, male

**Buildings supply company, high level**

This prediction is about the earnings of a large American company that sells building materials, and specifically whether its earnings from the last quarter of 2025, which will be announced in January of 2026, will be higher or lower than expected.

This company sells building materials. Its earnings this quarter [will certainly / might] be higher than expected, for several reasons. Forecasts of home repair and remodelling demand suggest that such demand will be higher than expected in late 2025, providing a higher than expected revenue stream to this company, [/ though it's always possible its earnings could be lower than expected]. Mortgage rates decreased, driving additional building projects which will create demand for this company's products. Shipping costs for this company's products also fell unexpectedly in late 2025, bolstering profits. Overall, these factors mean that this company's earnings in the last quarter of 2025 will be higher than expected, [/ though it's always possible its earnings could be lower than expected]. My impression is that given the circumstances [this will certainly be the case / it's possible to imagine things going either way].

This chain is one of the biggest suppliers of building equipment in the US, so it is important to understand how it performed in the last quarter of 2025.

*Links to recordings:* High reliability, male, Low reliability, male

**Buildings supply company, low level**

This prediction is about the earnings of a large American company that sells building materials, and specifically whether its earnings from the last quarter of 2025, which will be announced in January of 2026, will be higher or lower than expected.

This company sells building materials. Its earnings this quarter [will certainly / might] be lower than expected, for several reasons. Other companies in the same industry have been warning that demand in late 2025 is softer than they had expected, suggesting a decrease in revenue streams, [/ though it's always possible its earnings could be higher than expected]. Construction activity in the US was down year-over-year, partly reflecting higher mortgage costs and disruptions to the construction workforce. Tariffs caused the prices of the raw materials that are used to create this company's building materials to rise. Overall, these factors mean that this company's earnings in

the last quarter of 2025 will be lower than expected, [/ though it's always possible its earnings could be higher than expected]. My impression is that given the circumstances [this will certainly be the case / it's possible to imagine things going either way].

This chain is one of the biggest suppliers of building equipment in the US, so it is important to understand how it performed in the last quarter of 2025.

*Links to recordings:* High reliability, male, Low reliability, male

# References

**Augenblick, Ned, Eben Lazarus, and Michael Thaler**, "Overinference from weak signals and underinference from strong signals," *Quarterly Journal of Economics*, 2025.

**Baker, Scott R, Nicholas Bloom, and Steven J Davis**, "Measuring economic policy uncertainty," *The quarterly journal of economics*, 2016, *131* (4), 1593–1636.

**__ , Nick Bloom, and Stephen J Davis**, "Economic policy uncertainty index for United States," Technical Report, Technical report, FRED: Federal Reserve Bank of St. Louis 2021.

**Binder, Carola, Pascal Frank, and Jane M Ryngaert**, "The Causal Effect of News on Inflation Expectations," Technical Report, National Bureau of Economic Research 2025.

**Crosetto, Paolo and Thomas De Haan**, "Comparing input interfaces to elicit belief distributions," *Judgment and Decision making*, 2023, *18*, e27.

**Enke, Benjamin**, "What you see is all there is," *The Quarterly Journal of Economics*, 2020, *135* (3), 1363–1398.

**Griffin, Dale and Amos Tversky**, "The Weighing of Evidence and the Determinants of Confidence," *Cognitive psychology*, 1992, *24* (3), 411–435.

**Hong, James, Will Crichton, Haotian Zhang, Daniel Y Fu, Jacob Ritchie, Jeremy Barenholtz, Ben Hannel, Xinwei Yao, Michaela Murray, Geraldine Moriba et al.**, "Analysis of faces in a decade of us cable tv news," in "KDD'21: Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining" 2021.