The Transmission of Reliable and Unreliable Information*

Thomas Graeber Shakked Noy Christopher Roth

January 17, 2026

Abstract

Information often spreads and influences beliefs regardless of its reliability. We show that this occurs in part because indicators of reliability are disproportionately 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 more than twice as much as the claim itself. Reliable and unreliable information, once filtered through transmission, impact listener beliefs similarly, substantially reducing the efficiency of downstream decisions. 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 experiments, a large corpus of everyday conversations, and economic TV news demonstrate that reliability information is less likely to be cued during transmission and that attempts to retrieve it face greater interference in memory.

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

* An earlier version of this paper was titled "Lost in Transmission". We thank the editor 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 exemption 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. Listeners’ beliefs give us quantitative measures of the information content of original and transmitted messages, allowing us to estimate loss of different types of information using methods described below.

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.[1] 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.[2] 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 about three times 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 directly examine the transcripts of transmitted messages. Nearly all of the transmitted 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 $100 \times [(1.37 - 0.88)/1.37] \approx 34\%$ loss of sensitivity to variation in the level of the original prediction, induced by transmission.

By contrast, loss of reliability information is nearly three times as large. Among listeners who hear the original messages, switching from a weak-reliability message 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 91% 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, reliable messages are perceived as

[1] Robustness experiments show that our findings also hold when original messages includes numerical expressions for level and reliability.

[2] Transmission under these baseline incentives depends on which content transmitters believe is relevant for updates. We directly study these beliefs as well as several alternative incentive schemes.

less reliable, but unreliable messages are perceived as more reliable. Similarly, high forecasts are perceived as predicting a less high level, and low forecasts as predicting a less low level.

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 direction 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.

Economic implications. These transmission-induced distortions are economically consequential. In an additional experiment, we reproduce our main experimental design in a new setting: transmitters convey narratives about imminent earnings announcements for two real firms to listeners who make incentivized investment decisions relating to those firms. We replicate our main findings—here, transmission causes 15% loss of level information and 77% loss of reliability information—and further calculate that, under plausible assumptions, the loss of reliability information reduces expected investment surplus by 4.3%, while the loss of level information reduces it by 2.3%.

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.[3] 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 contribute to the spread of unreliable news and misinformation.

The transmission loss we document also has a distinct consequence in models where agents receive and pass on information. 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. In simulation exercises with parameters calibrated to match results from our experiments, we show that the transmission bias we document both compounds updating biases and shifts the distribution of efficiency losses onto sophisticates.

Robustness experiments. A potential concern is that our main finding might be partly driven by the fact that transmitters *have to* pass on each message. In practice, individuals might

[3]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.

simply choose not to transmit information they perceive as less credible. To test this, we replicate our investment experiment while allowing transmitters to forego recording a voice message. We find that transmitters choose to record a voice message 81% of the time, and that transmitters are not less likely to pass on low-reliability messages. The insensitivity plausibly reflects two factors. First, using open-text elicitations, we show that transmitters are not thinking about reliability when deciding whether to pass on a message (foreshadowing our later mechanism results). Second—both in our experiment and the real world—declining to pass on a piece of information may beneficially avoid the risk of accidentally portraying unreliable information as reliable (or vice versa), but it also means wholly dropping the information itself. This severely limits the efficacy of declining to pass on information as an antidote to the distortions we document.

While real-world communication is typically qualitative, many important settings involve transmission of quantitative information. In an additional experiment, we replicate our main results when the original forecasts include quantitative level and reliability statements (a percentage point estimate and percentage confidence level). This experiment also addresses concerns that our baseline results are driven by some extraneous difference between the way level and reliability information are communicated in our design, for example, that qualitative level manipulations feel sharper or more binary than qualitative reliability manipulations.

We also replicate the results from our baseline design using an alternative incentive scheme in which transmitters are incentivized to record messages that reproduce the full shift in belief distributions induced by the original messages (rather than simply the shift in mean beliefs as in our baseline incentive scheme), mitigating concerns that incentives based on mean beliefs unduly focus transmitters on level 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 examining participants' perceived benefits of communicating level versus reliability information and report two pieces of evidence. After transmitters in our main experiment record their messages, we ask a subset of respondents how important it is to pass on level and reliability information to maximize the likelihood of obtaining the incentive payment. Respondents, on average, deem them equally important, correctly anticipating that listeners' belief updates are highly sensitive to both our level and reliability manipulations. Second, we conduct an additional experiment that explicitly and equally incentivizes transmitters to pass on level and reliability information, effectively holding constant beliefs about the relative benefits of transmitting the two dimensions. Even under these more conservative incentives, we find pronounced differential information loss, at about 30% for level information and 70% for reliability.

Next, we ask whether the perceived cognitive costs of transmitting reliability information are higher. We conduct an additional 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. These beliefs do not change much *ex post*, after participants have experienced the task. This suggests that higher perceived difficulty of transmitting reliability information cannot account for differential information loss.

Finally, we extend our analysis of mechanisms beyond perceived benefits and costs to embrace the potential constraints memory introduces into the transmission process, outside of the transmitter's awareness. Leveraging a standard distinction in memory research (see, e.g., Kahana, 2012), we distinguish between *cued recall* of specific pieces of information from the original message once explicitly prompted for them, and *free recall* of information that occurs while transmitters record their message ("what comes to mind").

Starting with cued recall, we analyze memory loss among transmitters by eliciting their beliefs about the level and reliability of the predictions in the original recordings after they have recorded their messages. We find that transmitters' post-transmission beliefs are just as sensitive to variations in the original recordings as the beliefs of listeners directly hearing original recordings. This indicates minimal memory loss among transmitters in cued recall.

However, even though transmitters remember reliability information when prompted, reliability information may not come to mind *when completing their recordings*, which is a cognitively taxing task. 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 even easier to transmit. We conduct an additional experiment to directly test the hypothesis that reliability does not come to mind unless specifically cued. This experiment replicates our previous designs while ramping up the during-recording salience of level and reliability information. We show salient text on the recording screen reminding respondents to communicate both level and reliability. In this experiment, differential information loss is eliminated entirely. Our findings hence reveal that important information may fail to be transmitted even if it is explicitly known to be important and remembered when directly prompted.

Observational evidence. We conclude from our mechanism experiments that reliability information is lost in transmission largely because it fails to come to mind during the cognitively taxing process of transmission. The final section of the paper considers potential reasons for this and brings in observational evidence from everyday conversations and economic TV news.

Drawing on a workhorse framework of associative episodic memory (Bordalo et al., 2025b; Kahana, 2012), we distinguish between two reasons reliability could be less likely to come to mind during the transmission process. First, real-world requests for transmission may only rarely *cue* reliability information, meaning transmitters are not used to retrieving reliability. Second, reliability information may tend to be communicated in more generic ways (e.g., "I am unsure") than corresponding claims, making reliability statements more similar to each other. Attempts to retrieve a specific piece of reliability information may hence face more *interference*.

Evidence from the British National Corpus, a large collection of everyday conversations in English, supports the first possibility. Requests for information virtually never explicitly ask about the reliability of the requested information, and indirect cues of reliability (in the form of reliability markers or indicators) appear in only about a quarter of requests. This scarcity of reliability cues matters: in three different contexts, we show that reliability is much more likely to be transmitted when directly or indirectly cued. This is true for irrelevant cues randomly varied in our experiment; it is true on TV news, where expressions of uncertainty in economic segments are substantially more responsive to a benchmark measure of true economic uncertainty on days where the preceding day's coverage featured a higher volume of uncertainty language that could serve as an indirect cue; and it is true in the British National Corpus, where answers to questions containing indirect reliability cues are more likely to contain reliability indicators themselves.

The interference hypothesis also finds support in the British National Corpus: expressions of reliability information we extract using an LLM are about 40% more semantically similar to each other than expressions of level information are to each other. This is because reliability information is often communicated through generic language ("maybe," "I'm quite sure"), while level information typically comes in the form of rich, contextually specific arguments and examples.

Related literature. This paper is connected to work in various fields. Our focus on the transmission of qualitative stories about economic variables relates to a growing literature on the diffusion of qualitative information in the form of narratives (Hirshleifer, 2020; Shiller, 2017).[4] 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., 2024b; Kendall and Charles, 2025). We relate to work by Graeber et al. (2024a), who study how explanations shape the contagion of truths and falsehoods. Serra-Garcia (2025) studies how incentives to attract attention affect the transmission of scientific information. Thaler (2025) studies how strategic incentives shape the supply of false messages in a politicized context, and Thaler et al. (2025) study how individuals strategically use the imprecision of language. Our experiments identify which kinds of information are more likely to be successfully passed on from one person to another through spoken communication. 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.

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) and memory (Bordalo et al., 2023, 2021, 2025b). Our work relates in particular to a growing literature which emphasizes an important role of cue-based belief formation (Bordalo et al., 2025a; Conlon and Kwon, 2025; Gennaioli and Shleifer, 2010). Our paper differs from this literature in its focus on the cognitive foundations of the verbal transmission of information, and how they affect the *supply* of information. Our evidence from both controlled experiments and observational data suggests an important role for

[4]In his presidential address to the American Finance Association, Hirshleifer (2020) argues that "a key, underexploited building block of social economics and finance is social transmission bias: systematic directional shift in signals or ideas induced by social transactions."

selective memory in driving the differential loss of reliability information. The experimental and field evidence suggest that sparse reliability cues and higher interference may make information about reliability less likely to come to mind. Our findings are consistent with work in psychology demonstrating that, in fluent processing modes, reliability is rarely scrutinized, while epistemic vigilance is selectively triggered by cues such as inconsistency or explicit uncertainty (Kahneman, 2011; Sperber et al., 2010). They also connect to Stanovich (2018) and others' prominent framework of "miserly" processing, arguing that we naturally tend to substitute easy answers for hard ones as a way of conserving 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 when we do. Our findings suggest that, in the case of verbal information transmission, the bottleneck primarily lies in the former (detecting the need to override the default response) rather than the latter: 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; 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). Conlon et al. (2025) show in the context of a classic balls-and-urns belief updating problem that people are much less sensitive to quantitative information discovered by others, compared to equally-relevant information they discover themselves. We differ from this literature in our focus on (i) the transmission of qualitative information in the form of spoken narratives, (ii) the investigation of underlying cognitive mechanisms that shape transmission of different kinds of information, and (iii) studying how transmission shapes economic behaviors. A key conceptual distinction relative to prior work is that belief updating biases primarily have private consequences for the individuals subject to them, whereas biases in information transmission directly generate externalities by systematically shaping the beliefs and decisions of others.

Finally, information transmission has also been the subject of work outside of economics. For example, Carlson (2018, 2019) finds that political information is partially lost when people transmit it in writing. Similar "chain of transmission" paradigms have also been used to study how culture shapes the effects of transmission on content (e.g., Mesoudi and Whiten, 2008). In the cognitive sciences, interest in information transmission reaches back at least to Bartlett's seminal 1932 studies on *serial reproduction* of stories from memory (Bartlett, 1995). Work in these fields does not examine economic information or the differential transmission of information about level and reliability, nor does it study transmission in a real-world setting with field data.

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 transmitted versions before forming their beliefs.[5]

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.[6] 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.

[5]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`.

[6]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.

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 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.[7] 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.

[7]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.

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 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 E.

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 information about 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. First, we randomize whether the message argues for an increase or a decrease in the level of the variable (*level manipulation*). Second, we randomize whether the message is reliable or unreliable (*reliability manipulation*).

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).[8]

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.[9]

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

[8]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.

[9]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."

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.

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 our baseline results (Section 3), we find no evidence that transmission has any average effect on 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 object. 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 key outcome variables: state belief movements (the respondent's posterior about the economic variable minus their prior); 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 are not informed and simply capture any 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 the following question:

> 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.[10] 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.[11] Results based on incentivized versus unincentivized message beliefs are virtually identical, as shown in Appendix Figure A8.

---

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

[11]Responses are incentivized with the following formula for beliefs about the originator's prediction and reliability, respectively: Probability of winning $20 [in %] $= 100 - \alpha$(Response [in %] $-$ 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

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 A3 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 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, which we discuss while presenting them.

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. The figure illustrates that the different coders and the large language model come to similar conclusions.[12] See Appendix Table A2 for some examples of transmitted messages and their hand-codings.

were provided by us and there is no corresponding originator belief. The differing α's simply account for the differing units and standard deviations of level and reliability beliefs—level beliefs have a standard deviation of 8.8 and reliability beliefs have a standard deviation of 24.5.

[12]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.

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).

a) Percent of Scripts Containing Statements about Level and Reliability

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

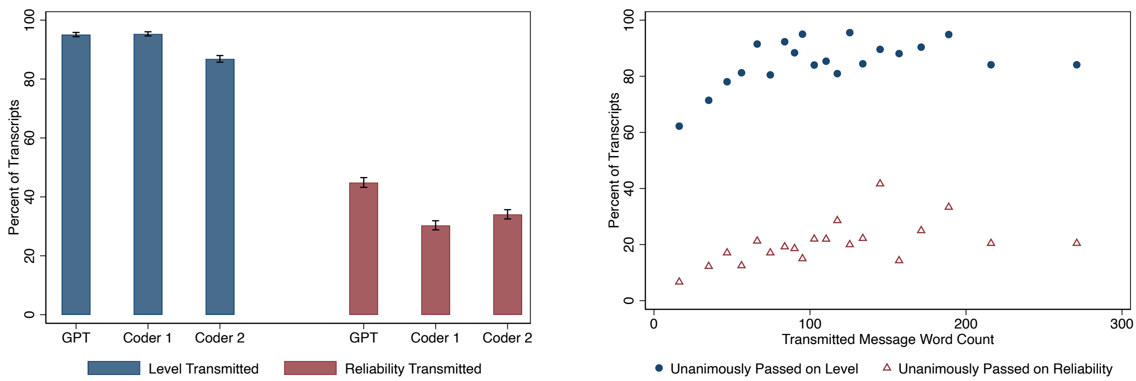


Figure 1: This figure presents data from our baseline experiment (Belief Movement Incentives). Panel (a) shows the percent of transcripts that are coded as conveying any information about the level and reliability of the original forecast, separately by two human coders and GPT-4, with standard error 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.

Transmitted recordings also include many disfluencies—hesitations, "um" statements, self-corrections, and so on—which could influence how listeners perceive the reliability of the original 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 about three times more likely to mention the level of the original prediction than to mention its reliability.*

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 information loss; or perhaps reliability is passed on in some 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 information.

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 $100 \times [(1.37 - 0.88)/1.37] \approx 34\%$ loss of sensitivity to level information. In other words, listeners who hear transmitted recordings are 34% 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 LevelBelief_i is the listener's belief about the level of the original prediction (z-scored at the topic by reliability manipulation type level); HighLevel_i is a dummy for the original forecast having a high level; and 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.[13] 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*

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

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.

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.*

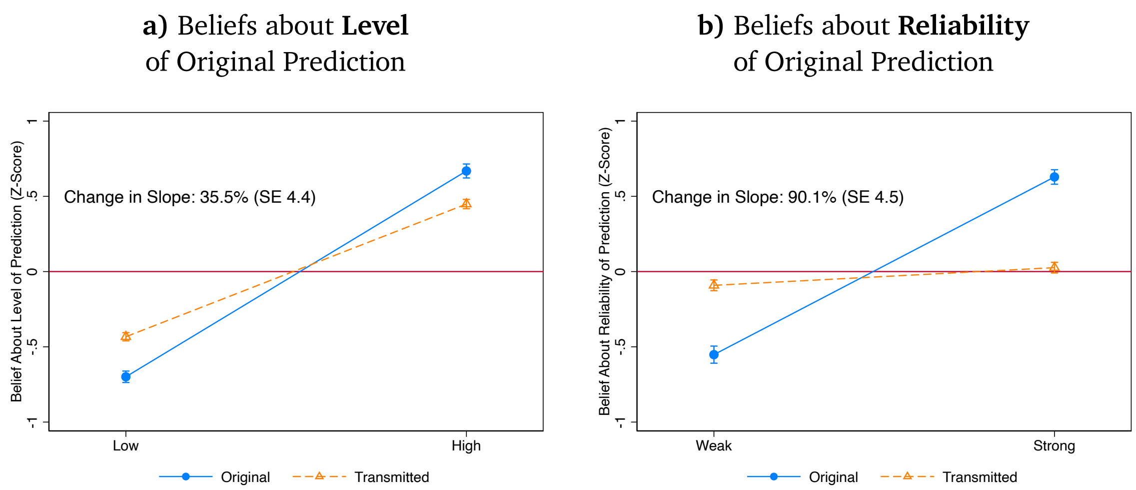


Figure 2: Slopes in the two panels differ at $p < 0.001, \chi^2 = 74.5$. This figure presents data from our baseline experiment (Belief Movement Incentives). It shows listeners' beliefs about the level and reliability of the prediction in the original message, separately by whether the original message 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 Tables A5 and A6 for regression tables. $N = 1,510$ listeners and 540 transmitters.

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

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 C 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 30-50% 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). We now examine these state belief updates.

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 in Panel (b) of Figure 3).

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.[14]

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.

[14]While reliability information loss is stronger than level information loss, this does not mean that the reliability effect will dominate; Panel (a) of Figure 3 shows that switches from high to low level matter about twice as much in the belief updating process as switches from weak to strong reliability.

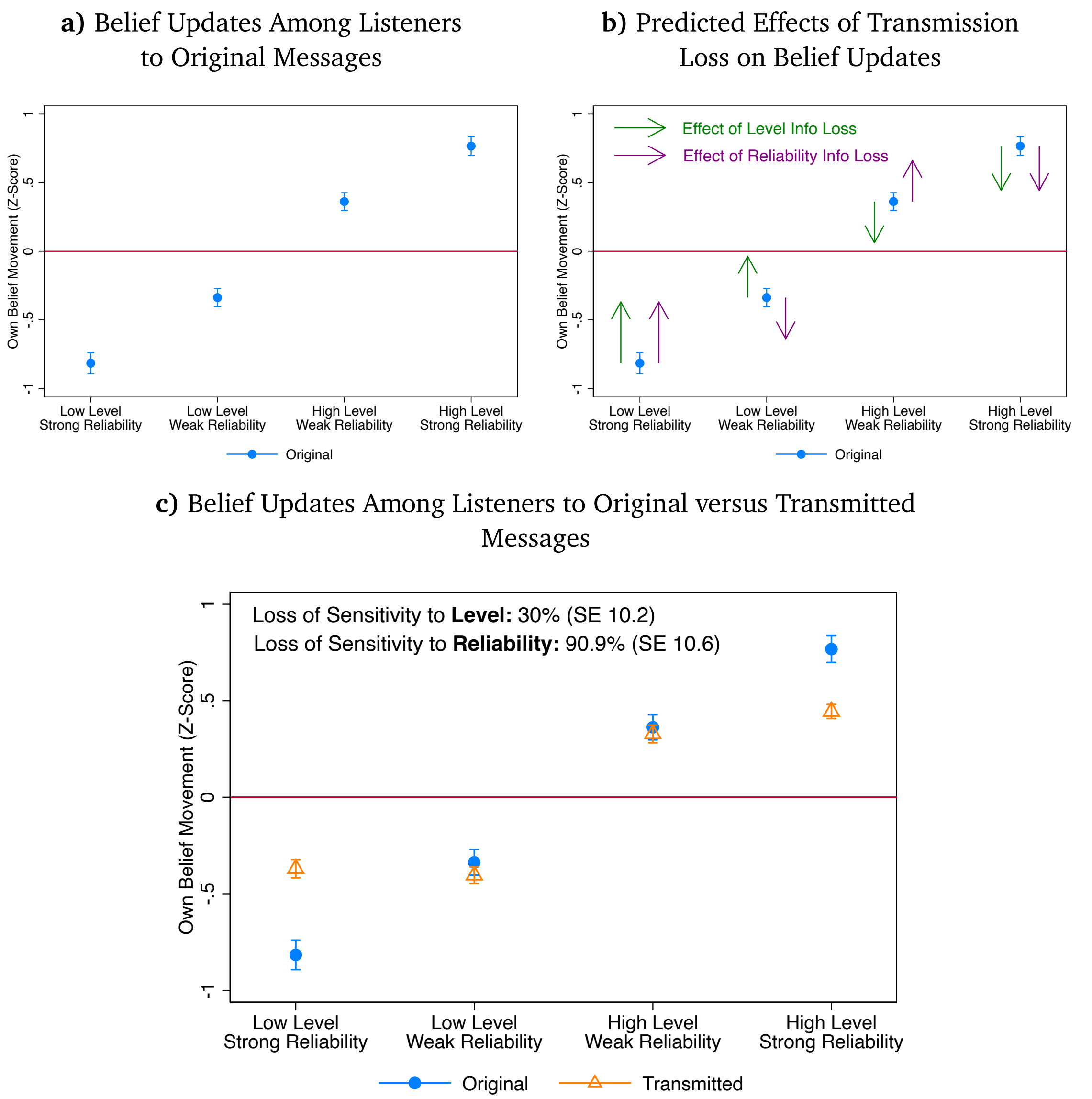


Figure 3: This figure shows average belief movements (posterior minus prior) about the economic variable from our baseline experiment (Belief Movement Incentives). Panel (a) shows average belief movements about the economic variable across the four different level-reliability conditions, only for listeners 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: $\text{Belief Update}_i = \alpha_0 + \alpha_1 \text{HighLevel}_i + \alpha_2 \text{StrongReliability}_i + \alpha_3 \text{Transmitted}_i + \alpha_4(\text{HighLevel}_i \times \text{Transmitted}_i) + \xi_i$. The loss of sensitivity to reliability information is calculated from a regression of the form $\text{Belief Update}_i \times (2 \times \text{HighLevel}_i - 1) = \gamma_0 + \gamma_1 HighLevel_i + \gamma_2 \text{StrongReliability}_i + \gamma_3 \text{Transmitted}_i + \gamma_4(\text{HighLevel}_i \times \text{Transmitted}_i) + \zeta_i$, 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 A4 gives regression versions of these results. 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.

Implications of transmission loss. Summing up, Figure 3 shows that transmission-induced information loss has two impacts on downstream belief updates. First, 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.[15] This means that transmission *reduces the average impact of new information on beliefs*, implying that if a population starts with polarized priors, new information will cause less belief convergence in the presence of verbal diffusion of the information. Second, 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 the same amount from weak- and strong-reliability messages. This means that transmission *increases the relative influence of weak-reliability messages* on overall belief updates: through transmission, information about the quality of messages gets garbled.

Result 3. *Verbal transmission weakens the average effect of new information on beliefs. It also increases the relative influence of weak-reliability information compared to strong-reliability information.*

3.4 Economic Consequences

3.4.1 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 application where both level and reliability beliefs translate directly into consequential choices. Moreover, financial investment decisions are a context where the transmission of qualitative narratives has been argued to be particularly consequential (Hirshleifer, 2020; Shiller, 2017), especially given the recent rise of social-media-driven stock-market and crypto investing.

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,

[15]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 ≈ 0.

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. Our results, summarized in Panels (a) and (b) of Figure 4, 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 greater noisiness of the investment outcome. Loss of certainty information is strong enough in this experiment that, unlike in our baseline results, we observe a transmission-induced *overreaction* in the low level × low certainty quadrant; investment is too skewed in the direction of the message because its low certainty level has been lost in transmission.

Interpretation of findings. Certainty beliefs are distorted less symmetrically in this design than our baseline, 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 later *high salience* mechanism experiment, 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 6 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.

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 48% and 35%—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.[16] 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 4 adjusts transmitted reliability beliefs upwards by 0.27 SDs to account for this; after this adjustment, reliability loss no longer appears asymmetric. Appendix Figure A13 shows that this adjustment makes a big difference in this experiment, but not in our baseline experiment.

Economic implications: back-of-the-envelope calculation. What are the economic implications of the loss of level and certainty information in this context? We conduct a simple 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).

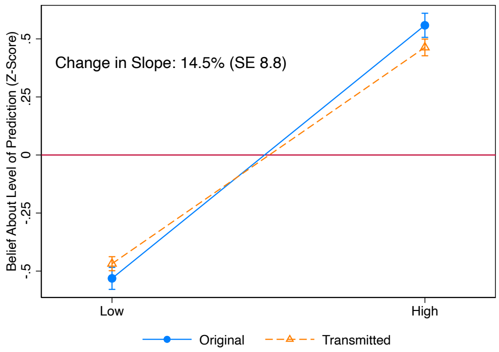
We use our estimated utility function to calculate the welfare distortions implied by the level and certainty information loss we document. Eliminating both distortions would improve investment surplus by 19.5% in dollar terms and 6.5% in utility terms. Eliminating certainty information loss would improve surplus by 7.6% in dollar terms and 4.3% in utility terms (since the importance of reliability information is increasing in risk aversion).[17]

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.

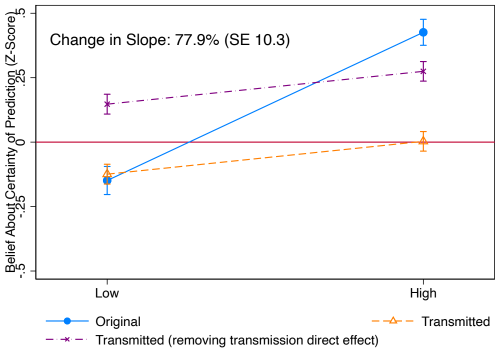
[16]When making these comparisons, we control for whether the original message was high or low reliability to account for the fact that the latter group can be selected on reliability.

[17]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.

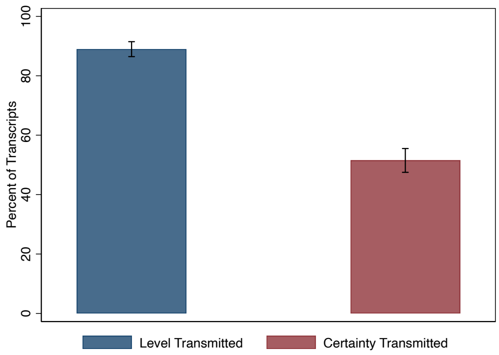
a) Beliefs About **Level** of Original Prediction



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)

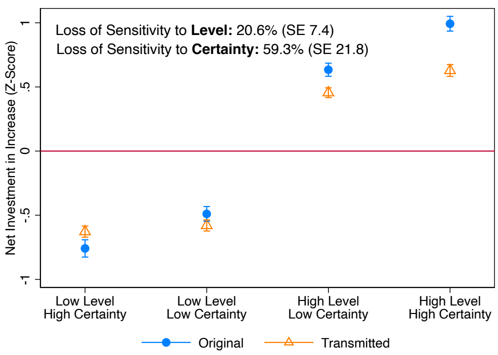


Figure 4: Slopes in Panels (a) and (b) differ at $p < 0.001$, $\chi^2 = 20.82$. This figure presents data from our version of the baseline experiment that uses earnings-announcement narratives and listener-investment-based incentives. Panels (a) and (b) replicate Figure 2, showing listeners' beliefs 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 and our two human coders 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.

3.4.2 Diffusion of Information and Externalities of Transmission Loss

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. 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—shifts the distribution of efficiency losses onto sophisticated agents who *would* react accurately to reliability information had it reached them.

Assumptions underlying the simulation. To demonstrate this point quantitatively, we perform a simple 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 τ_s is itself drawn from a gamma distribution with mean μ_τ and variance σ_τ^2. 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 3.5).

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 $\eta \sim N(0, \sigma_\eta^2)$. We assume that all recipients of a transmitted message account for the added transmission noise η. (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 η 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 σ_η^2 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 κ 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 5 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.

a) Belief Updates versus Signal Precision **b)** Welfare Losses

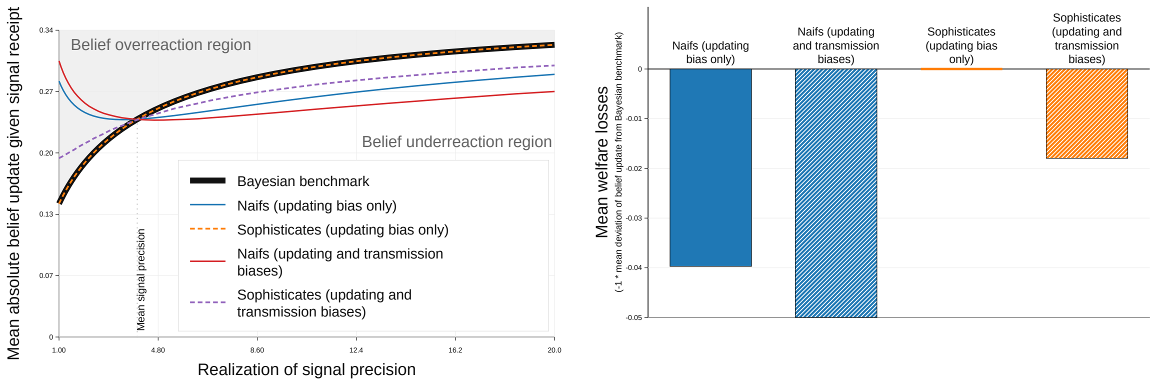


Figure 5: This figure presents results from our simulation exercise. Panel (a) plots the mean absolute magnitude of belief updates for a range of values of the signal precision, separately for naïfs and sophisticates and separately by whether we assume naïfs exclusively exhibit an updating bias or both 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.

3.5 Robustness

We now summarize the design and results of three preregistered experiments that probe the robustness of our main results to alternative design choices. Full descriptions of these experiments and associated results are available in Appendix B.

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 3.4 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 A17 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 3.4.

The insensitivity of transmitters' sharing decisions to the certainty level of the original message reflects two factors. First, foreshadowing our mechanism results, the certainty or reliability of the original information is not top-of-mind for transmitters deciding whether to record a message. At the end of the task, we ask transmitters an open-text question about which factors influenced their decision to record a message, and use GPT-5.2 to extract the most common reasons mentioned. Appendix Table A7 displays the results. By far the most common sentiment expressed by transmitters is a generic desire to be helpful to the listener, without any mention about the reliability or certainty of the original information; variants of this reason are given by 69% of transmitters who chose to pass on at least one message. Only 15% of these transmitters mention the high certainty or reliability of the original message as a motivation for passing on the information. Conversely, among transmitters choosing *not* to pass on at least one message, the dominant reason, appearing in 55% of responses, is a fear of failing to pass on the information adequately due to a personal lack of expertise.[18] Only 19% of these transmitters mention the low certainty or reliability of the original message as a reason not to pass it on.

This finding that the reliability of information does not come to mind during extensive-margin information-sharing decisions is consistent with evidence on the sharing of fake news on social media collected by Pennycook et al. (2021). The authors find that social media users are equally likely to share true and false headlines they are exposed to, even though they are able to distinguish true and false headlines when prompted; moreover, their sharing decisions become more correlated with accuracy when their attention is nudged towards a source's accuracy.

Second, both in our experimental context and 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

[18]For example: "I don't know anything about investing, so therefore, it would be in the best interest of the participant to make their own investment decision." Or "I did not feel confident in relaying information. There were a lot of words I did not understand and I felt I would screw up the message if I tried to record it. Because the content was unfamiliar, I was not certain I would remember much." Or: "To articulate all the details and reasonings behind the investment and show what it is all about I don't think I could have done that very well and may have confused the other participant."

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 exactly the same way at one point in the transcripts (through a single numerical percentage, e.g., an 8% 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 A18 show that differential loss *strengthens* under this design: level information loss halves, to 12.8%, while reliability loss is unchanged. A formal test of equality of the two information loss statistics rejects the null at $p < 0.001, \chi^2 = 21.7$.

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 A19 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 $p < 0.01, \chi^2 = 6.91$.

4 Mechanisms Underlying Differential Loss

What drives the differential loss of reliability and level information? In this section, we systematically 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. If differential loss does not result from a deliberate cost-benefit tradeoff, the reason may be one that the decision-maker herself considers suboptimal.[19] Specifically, (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.

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 several additional pieces of evidence from our baseline experiment. Participants in our main transmitter survey are randomized into seeing one of three sets of supplementary questions. First, we test for the role of biased beliefs about the relevance of reliability versus level information. In particular, participants may (mistakenly) believe that passing on reliability information would not affect listeners' belief updates and hence their probability of receiving the bonus payment. At the end of the transmitter experiment, we ask one-third of respondents how much passing on the reliability and level of the speaker's prediction increases the likelihood of receiving a bonus. We find that respondents believe that passing on reliability information is roughly equally likely to increase their chance of receiving a bonus as passing on level information: the average response is 71% for level and 68% for reliability. This is true even

[19]Here, we mean suboptimal not relative to a fully unconstrained, rational decision-maker. Rather, we use a subjective notion of optimality given the decision-maker's perception of her own constraints. The constraints that she is aware of enter her constrained optimization, reflected in her perceived benefits and costs. Additionally, however, there may be uninternalized constraints that she is not aware of, which affect behavior but are not accounted for in the decision-maker's subjective tradeoff, and hence suboptimal in that precise sense.

among respondents whom we classify as *not* passing on reliability information in their recordings (averages of 73% versus 66%).

Second, to test whether respondents are aware that they are omitting specific information, we ask another one-third of respondents explicit questions about whether they included level information and whether they included reliability information in their recordings. In line with our findings from the transcripts analysis, we find that 64% of respondents admit to not passing on reliability information, and 31% state they did not pass on level information.[20]

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,[21] strongly suggesting that respondents ignoring or misremembering incentives cannot explain the patterns in our data.

Taken together, these separate pieces of evidence from the baseline transmitter study show that people infer from the incentive scheme that reliability is as important to pass on as the level, that they do not forget the incentive scheme over the course of the experiment, and yet they admit to not passing on reliability in their actual recordings. This provides a first sign that the differential loss of reliability information is not due to explicit beliefs about lower benefits of transmitting reliability.

4.1.2 Additional Evidence: Incentives for Content Transmission

To more directly probe the importance of the perceived benefits of transmitting level versus reliability information, we conduct an additional experiment. 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 this supplementary experiment, transmitters are directly incentivized to pass on all of the original message's content, with 50% of respondents explicitly told to pass on level and reliability information. We still observe large differential information loss, albeit slightly smaller in magnitude than in our main results.

[20]The fact that 31% of respondents report not passing on level, despite our coders classifying almost everyone as passing on level, may suggest that these respondents do not correctly understand these concepts. However, first, our baseline incentives make no mention of level and reliability (instead, holistic transmission of relevant information is incentivized), so there is no need to understand (and no room to misunderstand) the level/reliability distinction. Second, when we restrict to people who said they passed on level in this question, we see the same differential information loss (27% for level and 84% for reliability).

[21]These questions are: (1) Which of the following is true? To maximize my earnings, ... (A) I should imitate the original recording, but in a different accent or voice. (B) I should describe the general topic of the original message without being specific about its contents. (C) I should pass on all information from the original message that I think will influence how people change their beliefs. And (2): Which of the following is true? I will be paid based on... (A) How many questions I can answer correctly about the original recording. (B) How close the average belief change induced by my recording is to the average belief change induced by the original recording. (C) I will be paid based on how similar other respondents say my recording is to the original recording.

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 one in ten transmitters will be selected for bonus eligibility and that, if selected, 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, unlike the baseline scheme, 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. In the other treatments, transmitters were not introduced to this distinction before producing their own recordings.

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 A14 shows results, pooling the implicit and implicit 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 34% vs. 91% 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 A15 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.

Note on incentive robustness. This constitutes the fourth incentive scheme under which we replicate our main finding of differential reliability information loss: our baseline experiment (mean belief movement incentives), our investment experiment (investment payoff incentives), our belief distribution robustness experiment (belief distribution movement incentives), and now both implicit and explicit content-transmission incentives.

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, suggests that transmitters are not even *trying* to transmit reliability, suggesting that anticipated costs are more likely to be relevant than experienced ones.

Design. As a direct test of the initially *anticipated* costs of transmitting level versus reliability information, 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. 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.

The setup of this experiment closely mirrors the *explicit incentives* treatment presented in Section 4.1, where transmitters were told that an external evaluator will compare the transcript of their message to the transcript of the original recording and separately rate how well the level and reliability of the original recording were communicated. Departing from that design, respondents here *choose* which of the evaluator's two responses will determine their bonus payment, and are told that they should focus purely on transmitting that dimension of the original message. Moreover, we elicit respondents' perceived difficulty of transmitting level versus reliability information, both before and after they actually create their recordings.

Logistics. We conducted this experiment with 97 respondents on Prolific in November 2023.

Results. Panel (a) of Appendix Figure A16 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 differential information loss. Virtually all respondents pass on the characteristic they chose.

Heterogeneity. There is no heterogeneity in perceived costs that could generate the pattern of differential information loss we observe. For example, suppose that the 60% of people choosing to transmit reliability are capable of transmitting both types of information in the main experiment, but the 40% 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 in the data: the groups choosing to transmit level versus reliability information give similar average difficulty ratings and have similar 15-point average difficulty gaps between the parameter they choose to transmit and the other parameter.

Summary. The evidence summarized above suggests that the differential loss of reliability information in our experiment does not reflect a belief that reliability information is less important to pass on or a belief that it is more costly to transmit. At the same time, this should not be taken to mean that people in all practical circumstances accurately assess the importance of passing on reliability information and find it easy to transmit. Rather, these exercises provide an existence proof that at least one other mechanism is important in driving differential loss of reliability information. Insofar as people in the real world underestimate the importance of reliability or find it harder to transmit, this means our main estimates provide a lower bound on the degree of differential information loss.

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 Memory Constraints and 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. In particular, one possibility is that reliability information simply *does not come to mind* in the cognitively challenging moment of transmission.

4.3.1 Conceptual Discussion: Cued Versus Free Recall

To structure our investigation, we follow the canonical distinction in memory research between *cued recall* and *free recall* situations (e.g., Kahana, 2012). In cued recall, people are given prompts related to the specific piece of information to be retrieved, and these prompts guide the retrieval process. In the free recall paradigm, researchers test whether and which information people recall in the absence of specific cues or prompts related to the target piece of information.

In our context, we apply these concepts to the recall of level and reliability information. On the one hand, transmitters may generally struggle to retrieve from memory the reliability information contained in the original messages, preventing them from passing it on to listeners. To test for this possibility, in a *cued recall* intervention, we ask transmitters about the level and reliability information in the original messages, after they have completed their tasks.

On the other hand, reliability information might be accessible from memory if actively sought out but may not come to mind automatically during transmission. While the transmission task prompts transmitters to recall the original messages, they are not explicitly prompted (on the transmission task page) to recall the level and reliability information contained in those messages. Consequently, the transmission process is best characterized as a free recall setup with respect to retrieving level and reliability information. To test for the role of constraints in free recall, we design an additional experiment that strongly increases the salience of reliability and level information *at the time of recording*, possibly increasing the ease with which reliability information comes to mind. In effect, this manipulation turns the free recall situation of the recording into a cued recall setting.

4.3.2 Memory Constraints in Cued Recall

We start by analyzing the beliefs of transmitters in the baseline experiment, measured after they complete their recordings.[22] Specifically, we present transmitters with the same set of three beliefs questions we pose to listeners, i.e., we elicit transmitters' state beliefs as well as their message beliefs (see Section 2).[23]

Appendix Figure A12 demonstrates that there is virtually no memory loss about the original message's reliability among transmitters: 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. If anything, there is more memory

[22]This should provide us with a lower bound for the role of memory constraints in explaining the transmission results as beliefs are elicited after and not during the recording.

[23]A random 50% of transmitters also give their priors about the two states before hearing the recordings, allowing us to calculate state belief updates.

loss for level information than reliability information.

These data also allow us to characterize differential loss *accounting for memory constraints*: we compare the sensitivity of listeners hearing transmitted recordings to the sensitivity of *transmitters* (instead of the sensitivity of *listeners hearing original messages*, as in our baseline analyses). We still find strong differential information loss, with reliability information loss of 87.2% and level information loss of 7.1%.

This evidence establishes that transmitters, when explicitly prompted, recall reliability information to the same degree as listeners. However, as pointed out above, the actual process of recording resembles a free recall situation rather than cued recall. Hence, this leaves open the possibility that reliability information simply does not come to transmitters' minds when recording their voice messages.

4.3.3 Memory Constraints in Free Recall

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 turning the free-recall recording task into a cued-recall situation. The design also allows us to assess whether emphasizing one dimension of information crowds out the transmission of the other.

Design. 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 reminder is presented in large, red font.

Logistics. This experiment was conducted on Prolific in November 2023 with 244 transmitters and 1,010 listeners.

Results. Figure 6 summarizes the results of the *high salience* experiment. Consistent with the hypothesis that reliability information is more likely to come to mind when it is explicitly cued during transmission, we observe a large reduction in reliability information loss. At the same time, this reduction is accompanied by a statistically significant increase in level information loss.

Panel (a) shows that transmitters in the *high salience* experiment 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%.

Analyzing message beliefs, Panels (b) and (c) of Figure 6 show that reliability information loss decreases to 39% (from 90% in our baseline experiment, $p < 0.001$), while level information loss increases from 36% to 53% ($p<0.05$), plausibly reflecting crowding-out of level information as transmitters talk more about reliability. This suggests 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.

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.

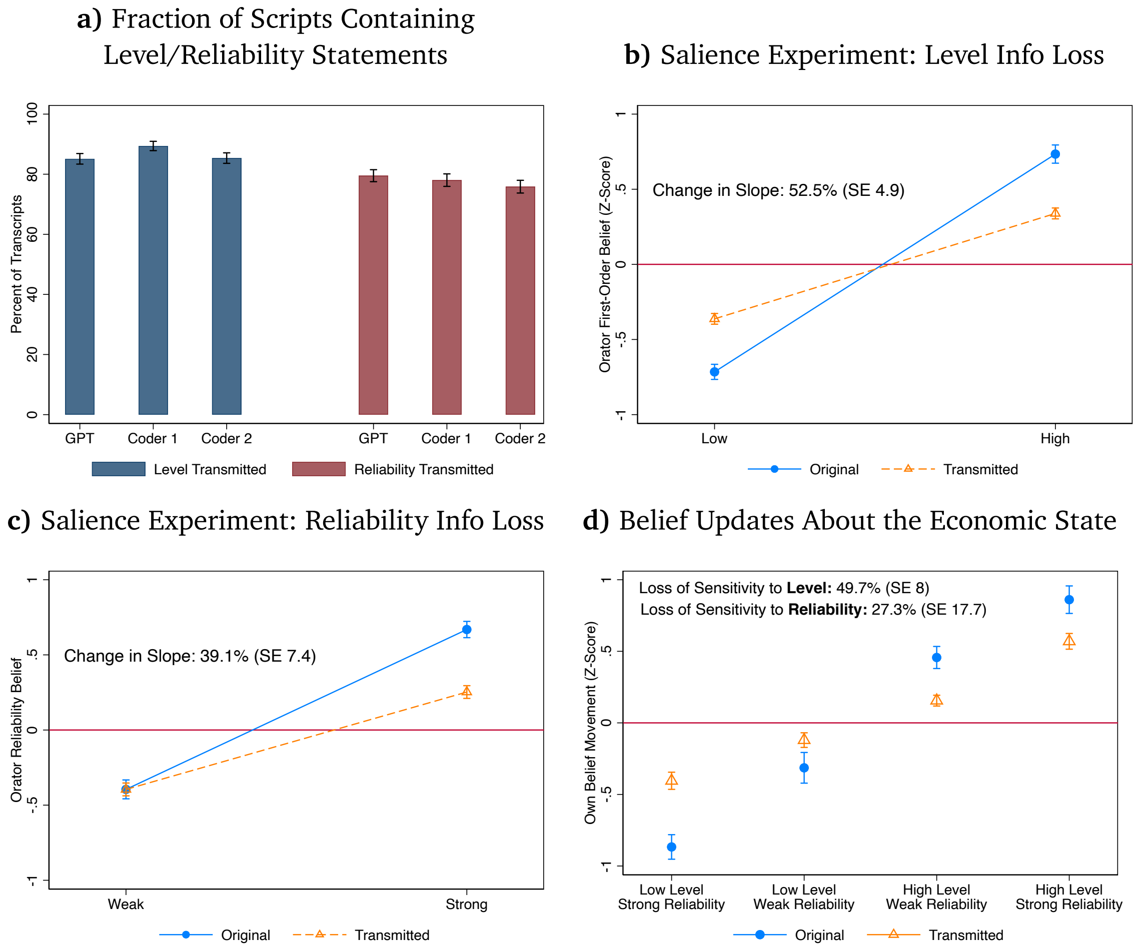


Figure 6: This figure presents data from the *high salience* experiment. Panels (a) and (b) replicate Figure 2, showing beliefs about the original message's level and reliability, separately by whether level is low/high, reliability is weak/strong, and the listener is hearing an original or transmitted message. Panel (c) replicates Panel (a) of Figure 1, showing which fraction of transmitted scripts contain statements about the level or reliability of the original prediction. 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, the *high salience* 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 6 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.

Result 5. *Reliability information is lost in transmission largely because it fails to come to mind during transmission. Reminding transmitters to consider reliability at the moment of transmission substantially reduces reliability information loss, but does so by taxing a common cognitive resource,* *leading to increased loss of level information.*

Relationship to the psychology of reliability neglect. As previously mentioned, our finding that people neglect to transmit reliability information complements previous evidence that people neglect reliability when updating their beliefs (Griffin and Tversky, 1992). In the classical distinction between representational and computational stages of decision-making (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 seems 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.

5 Why is Reliability Less Likely to Come to Mind?

Our mechanism findings raise the question of *why* reliability information is less likely to come to mind than level information, absent explicit reminders. In this section, we introduce two potential explanations and show suggestive field evidence supporting both of them.

5.1 Framework

We interpret the transmission process through the lens of associative memory models of retrieved context (e.g., Bordalo et al., 2025b; Kahana, 2012). When a transmitter encounters an original message, the message is stored in the transmitter's memory as a trace that includes both substantive content (e.g., facts, arguments, and examples contained in the message) and meta-features (the message's source and credentials). When the transmitter later encounters a cue that triggers transmission—the transmission task, in our experiments, or a question or conversation in real life—the cue activates stored traces with probabilities that are increasing in the similarity between the cue and the trace, and decreasing in the interference between the trace and other traces (Bordalo et al., 2024).

In this class of models of episodic memory, the probability of retrieval is affected both by characteristics of the cue (via similarity) and characteristics of the memory database (via interference). This, in turn, suggests two classes of reasons why reliability information might come to mind less than level information in practice.

Cues. 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 cuing either level or reliability) to naturally cue level information, but not reliability.

Interference. Predictions about the levels of economic variables are typically supported by, and communicated through, rich contextually-specific claims, arguments, and examples. For example, the level claims in our messages about housing were supported by arguments about residential construction permits, interest rates, and geographic movement of populations. This specificity and richness means that different level claims from different contexts are strongly distinct from each other, leaving little scope for interference between them in the retrieval process.

By contrast, reliability is much more likely to be communicated using semantically generic terms such as certainty or uncertainty prefixes, statements about confidence levels, or the name of a source. This creates much more potential for interference between different pieces of reliability information that could block retrieval. Indeed, a longstanding literature on "source monitoring errors" in memory finds that individuals have difficulty accurately attributing particular pieces of memory to a source (Hovland and Weiss, 1951; Johnson, 1997). Several pieces of evidence connect these source monitoring errors to interference between sources in the memory bank (Henkel and Franklin, 1998; Johnson et al., 1993; Lindsay and Johnson, 1989). Put more simply, richly detailed arguments concerning levels may come to mind more easily than short, generic reliability statements because of the greater distinctiveness of the former.

5.2 Evidence on Cues

Almost all questions, by virtue of asking about some specific piece of information, will cue level information. By contrast, questions could cue reliability information either directly (via explicit requests for certainty or uncertainty information) or indirectly (by containing their own certainty or uncertainty markers or other features that remind the person being questioned about reliability).

5.2.1 Direct Cues of Reliability

Our *high salience* experiment in Section 4.3 showed that explicit cues of reliability are effective at inducing transmission of reliability information. However, evidence from the field suggests that everyday questions almost never explicitly request reliability information.

Specifically, we examine 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 (OpenAI's GPT4o) to annotate each of the ∼800,000 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) or secondhand information that the speaker would have to source from somewhere else (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 a request for reliability information, such as the interlocutor's certainty level or the source of the information.

Of the 70,974 questions our LLM identifies in the Corpus (8.5% of the conversational lines), only 102 are classified as containing an explicit request for information about the certainty or reliability of the interlocutor's answer. When we restrict to the 19,863 questions that are classified as being about information that the questioner's interlocutor would only have secondhand access to—i.e., requests for the transmission of secondhand information—only two contain such requests. Explicit requests for reliability information are hence vanishingly rare in everyday spoken conversation.

5.2.2 Indirect Cues of Reliability

Requests for transmission may also include indirect cues of reliability information—for example, requests may themselves include reliability markers. Do such indirect cues matter in the sense of successfully raising the rate at which reliability information is transmitted?

Indirect cues matter: evidence from our experiment. Our baseline transmitter experiments seeded random variation in exposure to indirect 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. Since transmitters heard two recordings back-to-back before completing their transmission tasks, those who heard two with certainty- or uncertainty-denoting prefixes were exposed to strictly more indirect reliability cues than those who heard one recording with certainty- or uncertainty-denoting prefixes and one with no prefixes.

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 indirect and irrelevant cues in the *other* recording increases the probability of transmitting reliability information for *this* recording by 10 percentage points, or nearly 50 percent—a substantial effect of indirect reliability cues. Column (2) confirms, using our 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). Column (3) shows, by contrast, that transmitters do not mistakenly adopt reliability indicators inconsistent with the original message. We return to Columns (4) and (5) later in this section.

Table 1: Indirect Reliability Cues Increase Transmission of Reliability

	Experimental Results: Reliability Passed on (Handcoding)			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				0.109***	0.101***
				(0.004)	(0.007)
Constant	0.130***	0.274***	0.123***	0.274***	0.298***
	(0.028)	(0.037)	(0.027)	(0.002)	(0.004)
Observations	702	702	702	70974	19863

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 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 contained any reliability indicators.

Indirect cues matter: evidence from economic cable news. We now turn to a high-stakes field setting for an additional test of whether indirect cues matter. Specifically, we measure the transmission of economic uncertainty through cable TV news segments and quasi-experimentally test whether uncertainty cues on preceding days increase the fidelity of this transmission. For the sake of brevity, we here describe the main ideas behind our methodology and the main results, deferring a detailed exposition and robustness checks to Appendix D.

To measure the transmission of economic uncertainty, we combine a newspaper-based benchmark measure of economic uncertainty (Baker et al., 2016) 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.

We test whether higher quantities of uncertainty language on a channel's news segments *yesterday*—which we treat as indirect cues of reliability—are associated with a greater responsiveness of uncertainty language in economic segments to (our benchmark of) true economic uncertainty *today*. Since the outcome we are interested in is how the *slope* between TV and newspaper language varies over time, neither serial correlation in uncertainty language on cable news nor any mechanical relationship between newspaper and TV language threatens our empirical strategy. Our main specification includes channel fixed effects interacted with calendar month and economic coverage quantities.

Appendix Figure A23 shows that on days with above-median volumes of uncertainty language on yesterday's news, uncertainty language in today's economic segments is substantially more sensitive to our benchmark measure of economic uncertainty. Recent indirect cues of uncertainty hence seem to make news commentators more attentive to uncertainty, increasing the fidelity of their transmission of uncertainty. This analysis is not conclusive: one alternative interpretation is that the act of communicating uncertainty on preceding days shifts commentators' beliefs about the importance of transmitting uncertainty information, rather than acting as an unconscious cue.

Appendix Table A9 presents regression versions of these results and subjects them to a battery of robustness and placebo tests. These include specifications that add interactions between channel fixed effects and lagged measures of benchmark economic uncertainty and economic coverage; specifications that control for calendar *day* fixed effects, zooming in on variation across channels in channel-specific uncertainty language yesterday; specifications that exclude the Covid years; and placebo tests that use uncertainty language on *other* channels and uncertainty language in *non*-economic coverage. Results remain highly robust across all of these checks.

Indirect cues in everyday conversations. Finally, we return to the British National Corpus. While virtually none of the 70,974 questions identified by our LLM contain explicit requests for reliability information, 26% of them are classified by our LLM as containing indirect cues of reliability in the form of certainty, uncertainty, or reliability markers. Column (4) of Table 1 shows that a question containing such a marker is 10 percentage points more likely to be followed by an answer that contains a reliability indicator, compared to a question that contains no such indicators. This remains true in Column (5), where we restrict to questions requesting second-hand information.

This analysis is descriptive: it could also reflect reliability cues shifting interlocutors' beliefs about the importance of transmitting reliability information, consistent with our earlier observation that cost-benefit considerations may be important in the real world even if they do not appear relevant in our experiment. However, the results of this analysis are consistent with the evidence from our experiments and TV news: indirect cues appear to help trigger reliability transmission. Despite the apparent efficacy of these indirect cues, they remain relatively uncommon in the Corpus's conversations: while all of the questions we analyze (by definition) request some piece of level information, only a quarter contain any indirect reliability cues.

Cues: Conclusion. One hypothesis is that transmission contexts are naturally associated with level but not reliability information because level information is frequently cued in such contexts but reliability information is not. In this section, we have shown that in everyday conversational contexts where level information is requested, reliability cued are exceedingly rare, and is implicitly cued only about a quarter of the time. We have also shown that cues matter: both explicit and implicit cues of reliability induce greater transmission of reliability information in our experiment, in everyday conversations, and in TV news.

Cues: Caveats. Importantly, we do not claim that cost–benefit considerations are 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 and higher interference for reliability—are pervasive in naturally occurring communication.

5.3 Interference

Our second hypothesis concerns interference: if expressions of reliability tend to be more similar to each other than expressions of level, reliability information will be harder to retrieve because retrieval attempts will face more interference.

To test this hypothesis, we return to the British National Corpus.[24] We take the set of conversational lines containing reliability markers and compute the average cosine similarity of 100,000 randomly selected pairs. We then take the set of conversational lines containing statements about substantive claims (which we treat as level statements) and similarly compute the average cosine similarity of 100,000 randomly selected pairs.[25]

The leftmost columns of Figure 7 show that pairs of conversational lines containing reliability indicators are about 15% more semantically similar to each other than pairs of lines containing level statements.[26] This difference is attenuated by the fact that many of the conversational lines containing reliability indicators *also* contain level statements, and many of the lines containing level statements also contain extraneous words. To narrow down to our objects of interest, we

[24]Because the Stanford Cable TV News Analyzer, which we used for our TV analysis, does not allow us to access raw transcripts (just query them for specific words), we cannot perform this exercise in that dataset.

[25]We exclude trivial statements, such as "My name is John".

[26]Cosine similarity scores range between 0, representing complete orthogonality, and 1, representing identity; the baseline semantic similarity of both types of statements is between 0.15 and 0.20, indicating relatively low similarity, due to the diversity of conversation topics in our corpus.

use our LLM to extract the core snippets of lines containing expressions of reliability or level statements. When comparing the semantic similarity of randomly chosen pairs of these subsets, our result strengthens substantially: the rightmost columns of Figure 7 shows that expressions of reliability are about 40% more semantically similar to each other than statements about level.

This is not mechanical: in principle, expressions of reliability could be just as semantically diverse as expressions of level statements. This could be true if, for example, reliability is communicated via rich contextual information about a source ("the person who told me this was a bit drunk and seemed to have a poor memory"). However, in practice, the fact that reliability is often communicated through a stock set of modifiers ("maybe," "I'm not so sure") causes considerable semantic similarity across reliability expressions. This makes each reliability statement less distinctive and hence potentially harder to recall.

a) Reliability Claims are More Semantically Similar to Each Other than Level Claims (British National Corpus)

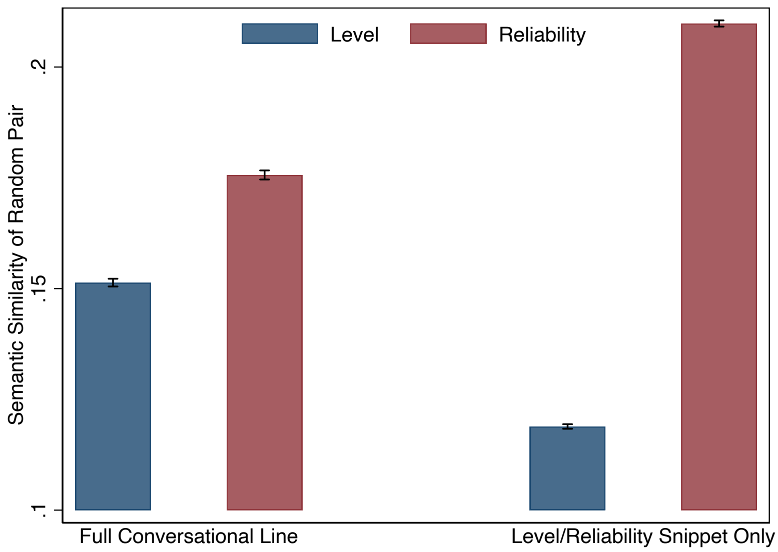


Figure 7: This figure presents results from the British National Corpus. The sample for the Level bars is the set of conversational lines our LLM classifies as being statements about a subject that could conceivably involve some uncertainty. The sample of the Reliability bars is the set of conversational lines our LLM classifies as containing reliability indicators. In the leftmost columns, we use the text of the full conversational line; in the rightmost column, we use only the snippet containing the level statement or reliability markers, as extracted by the LLM. Each bar shows the average cosine similarity of 100,000 randomly selected pairs from within the relevant set, calculated using a BERT-based sentence-embedding model.

Result 6. *A simple associative memory framework suggests that reliability information might be less likely to come to mind either because it is not naturally cued in transmission contexts or because attempts to retrieve reliability information face greater interference. Evidence from everyday conversations suggests that requests for transmission virtually never explicitly cue reliability information and only indirectly cue it about a quarter of the time. This matters: evidence from our experiments, everyday conversations, and economic TV news suggests that reliability is much more likely to be* *transmitted when it is directly or indirectly cued in the transmission request. Meanwhile, evidence from everyday conversations suggests that reliability statements are much less semantically distinctive from each other than level statements, supporting the idea that reliability retrieval may face greater interference.*

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. Participants in our experiments are tasked with listening to audio clips discussing economic variables, and conveying the information in the clips through voice messages. Other participants listen to either the original recorded voice messages or transmitted versions of those messages, and then state incentivized beliefs. 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. 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 may be less likely to come to mind either because it is less likely to be cued in everyday contexts or because attempts to retrieve reliability information face greater interference. An examination of everyday conversations and economic TV news yields support for both of these possibilities.

Our results have direct economic consequences. In an incentivized investment experiment, we show that transmission-induced losses of uncertainty information translate into systematically distorted portfolio choices. Investment decisions become markedly less sensitive to the certainty of the underlying information. This selective loss of uncertainty reduces expected investment surplus, even when transmitters face strong incentives to communicate payoff-relevant information faithfully.

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.

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.

__ , 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.

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.

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

Bordalo, Pedro, Giovanni Burro, Katie Coffman, Nicola Gennaioli, and Andrei Shleifer, "Imagining the Future: Memory, Simulation and Beliefs about Covid," *Review of Economic Studies*, 2024.

__ , __ , 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.

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

__ , Nicola Gennaioli, 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.

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.

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.

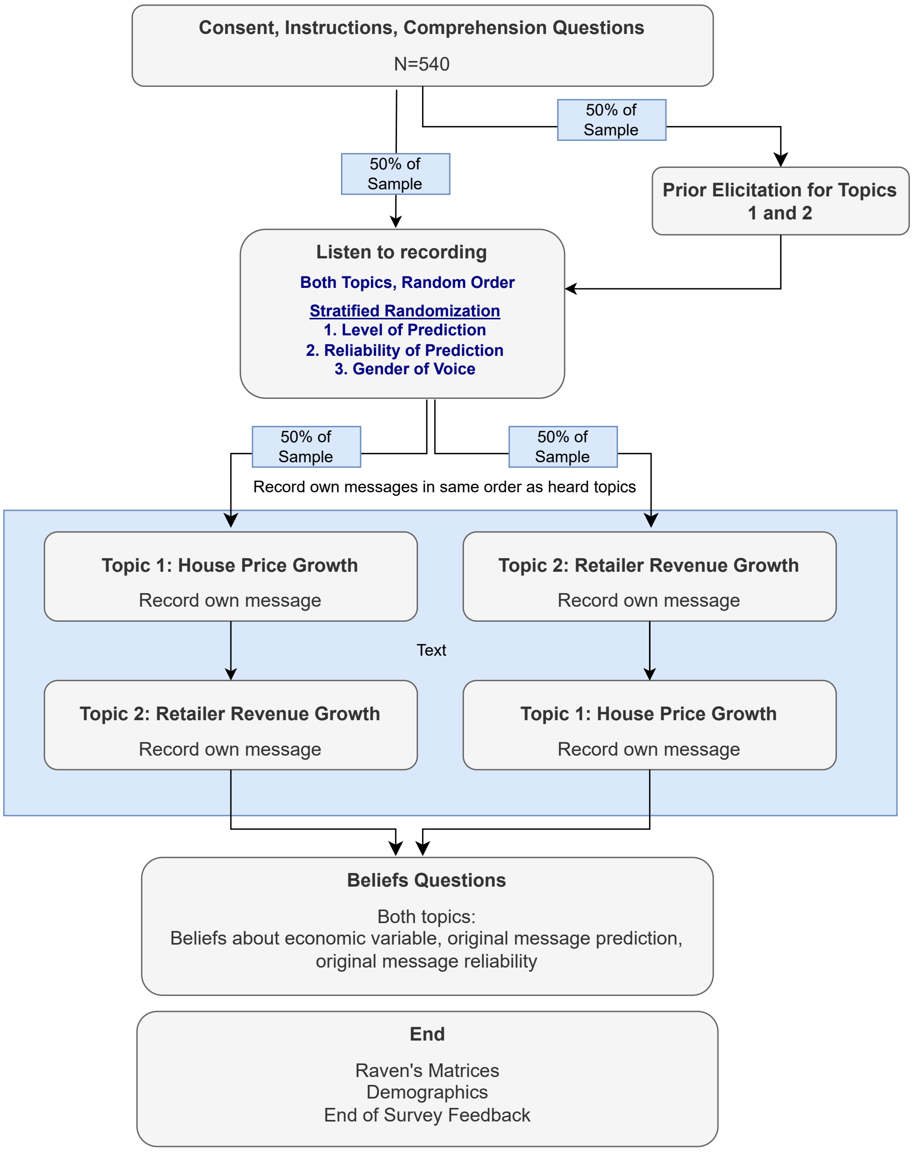
Tversky, Amos and Daniel Kahneman, "Judgment under Uncertainty: Heuristics and Biases: Biases in judgments reveal some heuristics of thinking under uncertainty.," *science*, 1974, *185* (4157), 1124–1131.

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.

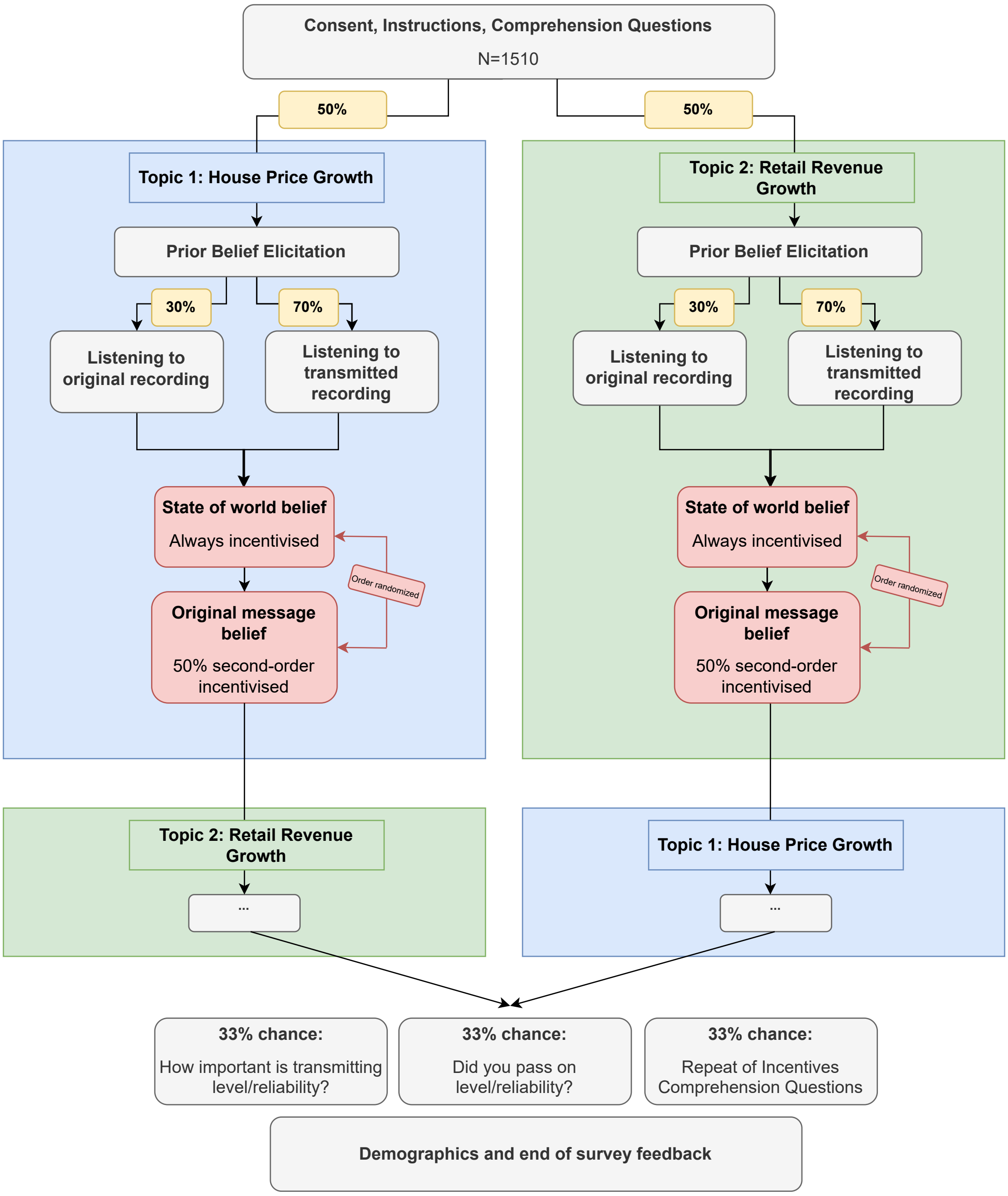
A Additional Exhibits

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
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 (287)	Incentives tied to listener investment payoff	Speech content; beliefs about originator level and reliability	17533
Listener	Investment incentives	Prolific (1299)	Original vs. transmitted recording	Investment choice; beliefs about state; beliefs about originator level and reliability	17533
Extensive margin experiment					
Transmitter	Investment incentives	Prolific (298)	Optional transmission decision	Transmission choice; speech content (if transmitted)	17540
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

Notes. All experiments implement a 2 × 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.

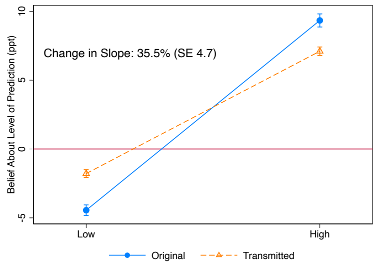
Appendix Table A2: Example Transmitted Messages

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.

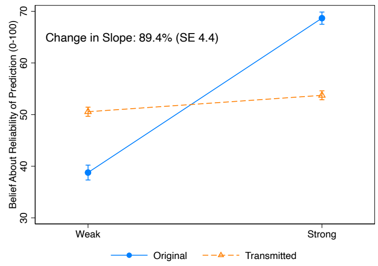
A.1 Additional Figures

A.1.1 Baseline Experiment: Belief Movement Incentives

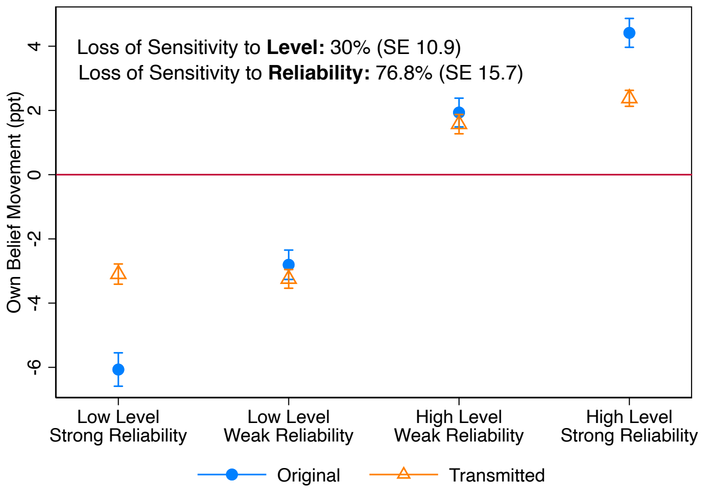
a) Level Message Beliefs: Not Z-Scored



b) Reliability Message Beliefs: Not Z-Scored

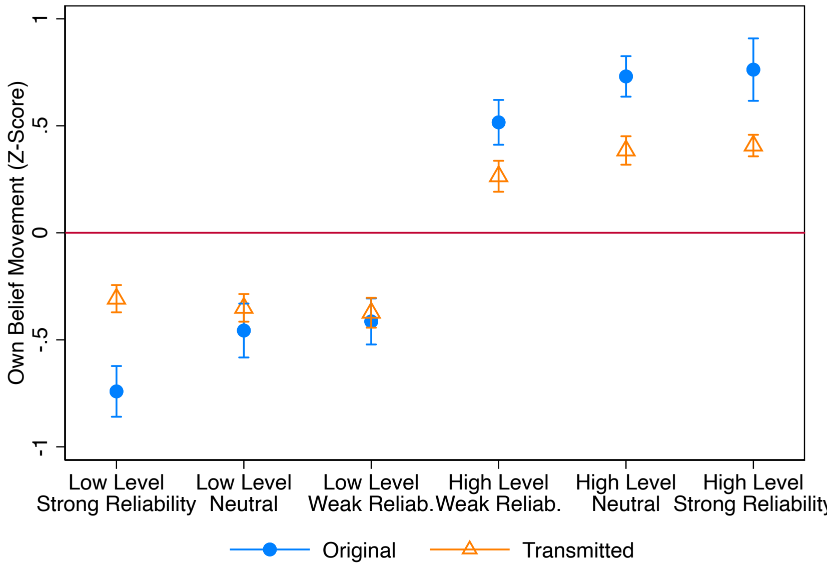


c) State Belief Updates: Not Z-Scored



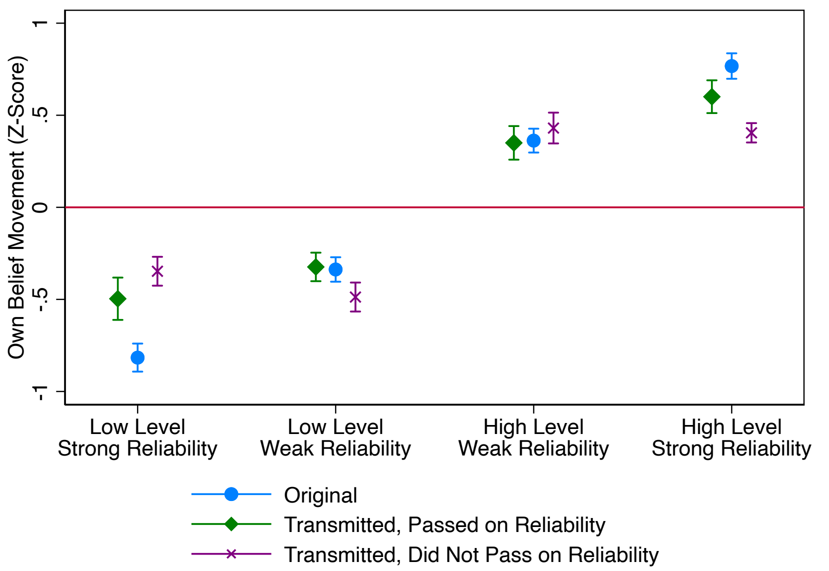
Appendix Figure A3: This figure presents data from our baseline experiment (Belief Movement Incentives). It replicates Figures 2 and 3 using raw (rather than z-scored) beliefs.

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



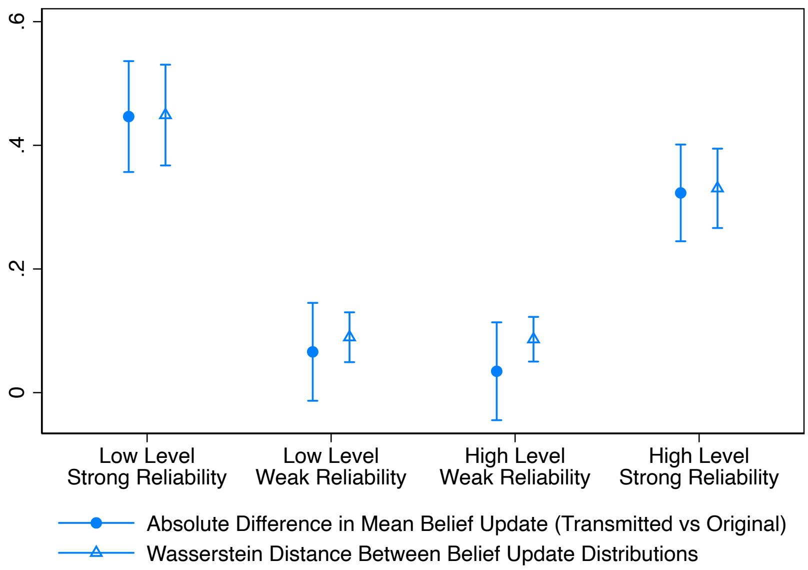
Appendix Figure A4: This figure presents data from our baseline experiment (Belief Movement Incentives). It is an alternative version of Panel (c) of Figure 3. It shows the average belief updates of listeners, restricting to the Modular reliability manipulation, which has a weak-reliability, strong-reliability, 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



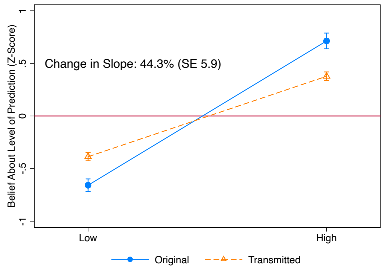
Appendix Figure A5: This figure presents data from our baseline experiment (Belief Movement Incentives). It is an alternative version of Panel (c) of Figure 3. It shows the average belief updates of listeners, splitting listeners who hear transmitted recordings by whether the transmitted recording is 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

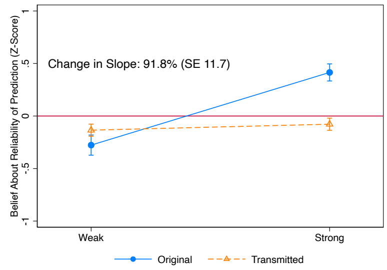


Appendix Figure A6: This figure presents data from our baseline experiment (Belief Movement Incentives). It is an alternative version of Panel (c) of Figure 3. It shows the absolute mean difference in belief updates between listeners who directly heard the original messages versus those who heard transmitted 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.

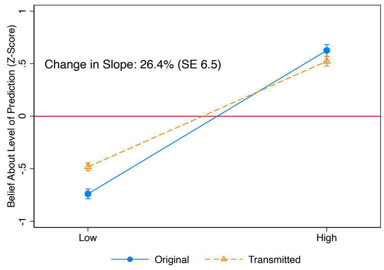
a) Level Info Loss: Modular Manipulation



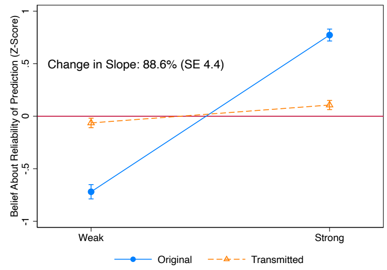
b) Reliability Info Loss: Modular Manipulation



c) Level Info Loss: Naturalistic Manipulation

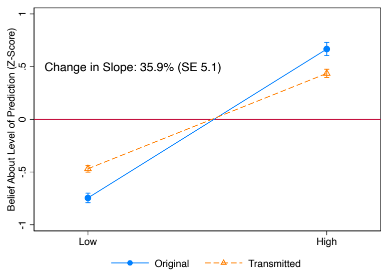


d) Reliability Info Loss: Naturalistic Manipulation

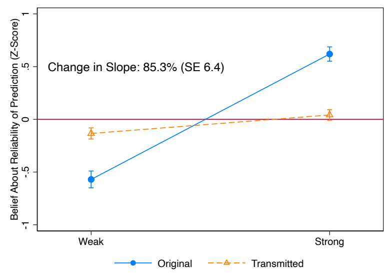


Appendix Figure A7: This figure presents data from our baseline experiment (Belief Movement Incentives). It replicates Figure 2, showing beliefs about the level and reliability of the original prediction, separately by respondents in our *modular* versus *naturalistic* reliability manipulations.

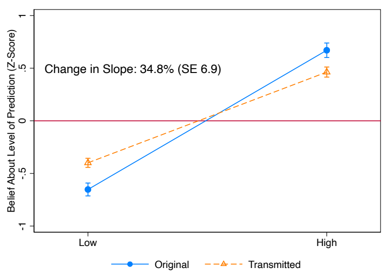
a) Level Info Loss: No Incentives



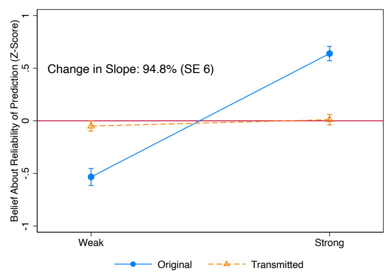
b) Reliability Info Loss: No Incentives



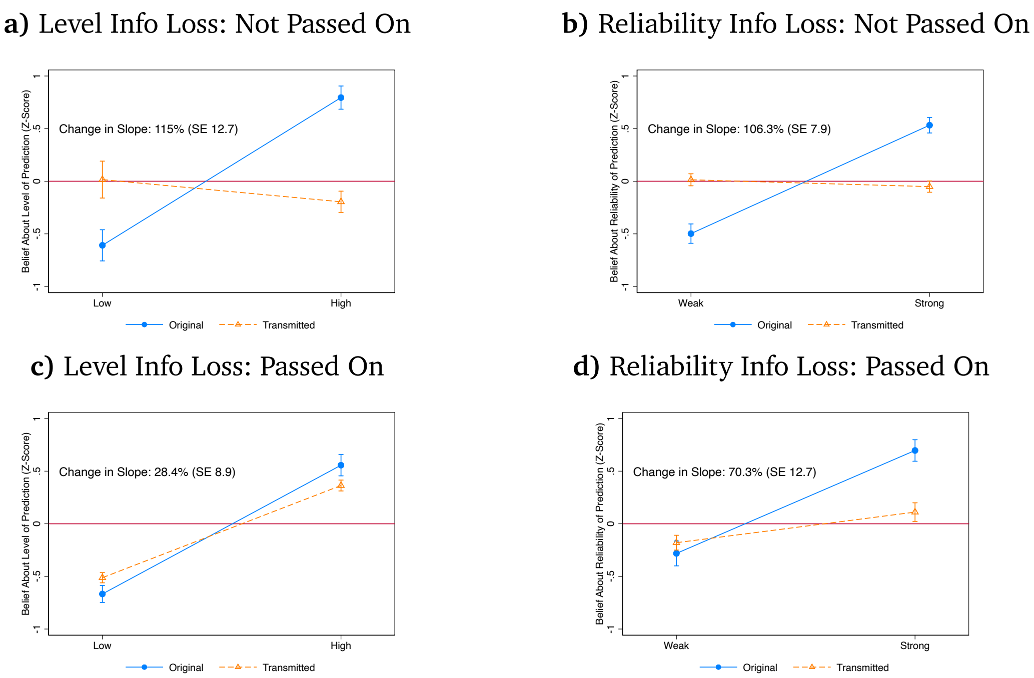
c) Level Info Loss: Second-Order Incentives



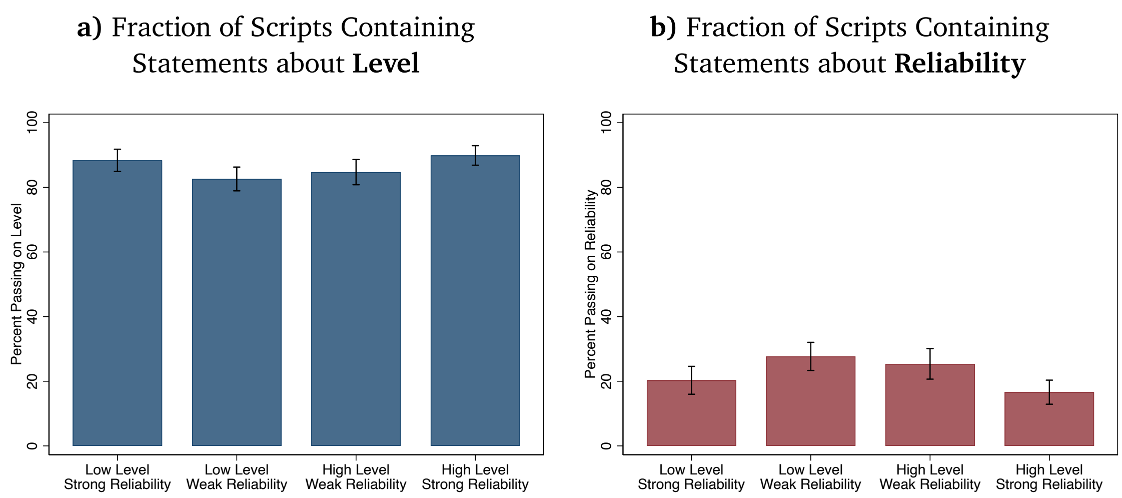
d) Reliability Info Loss: Second-Order Incentives



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.



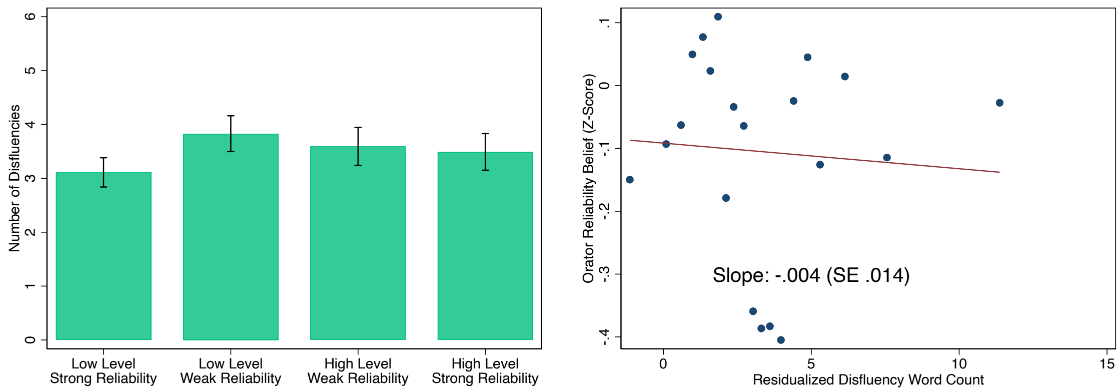
Appendix Figure A9: This figure presents data from our baseline experiment (Belief Movement Incentives). It replicates Figure 2, showing beliefs about the level and reliability of the original prediction. Panels (a) and (b) restrict to recordings that both human coders and GPT-4 unanimously agree do *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.



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

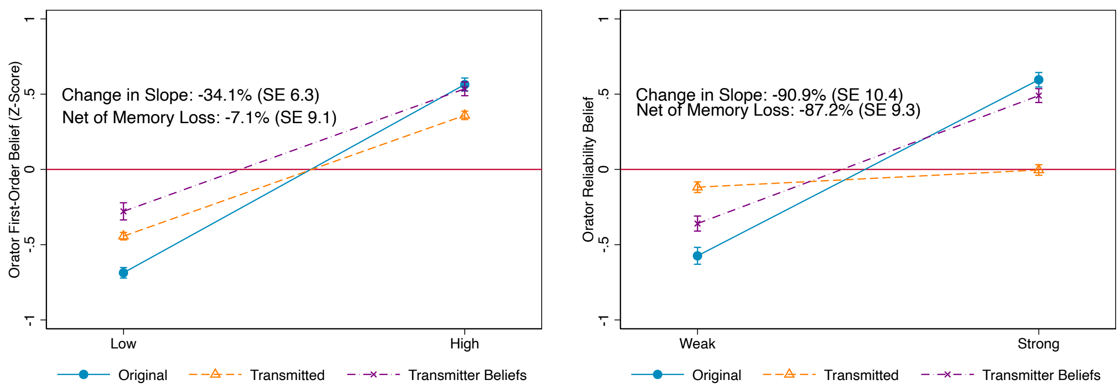
b) Disfluencies Do Not Predict Listener Beliefs



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 original transcript. 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, controlling for the transmitted message's overall word count and topic fixed effects.

a) Transmission of **Level**

b) Transmission of **Reliability**

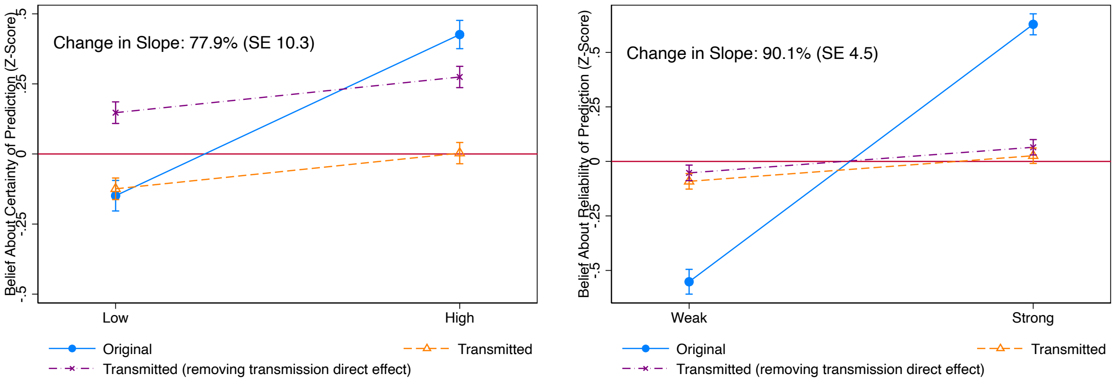


Appendix Figure A12: This figure presents data from our baseline experiment (Belief Movement Incentives). It replicates Figure 2 but adds a line representing the beliefs of the *transmitters* who create the recordings. The "net of memory loss" statistics compare the orange line to the purple line instead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

A.1.2 Supplementary Experiment: Investment Task

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

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



Appendix Figure A13: Panel (a) directly replicates Panel (b) of Figure 4. Panel (b) replicates this figure using data from our baseline experiment. In both figures, 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 both cases.

Back of the envelope welfare calculation. We proceed as follows. We assume listeners have homogeneous CRRA utility with coefficient γ 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*.

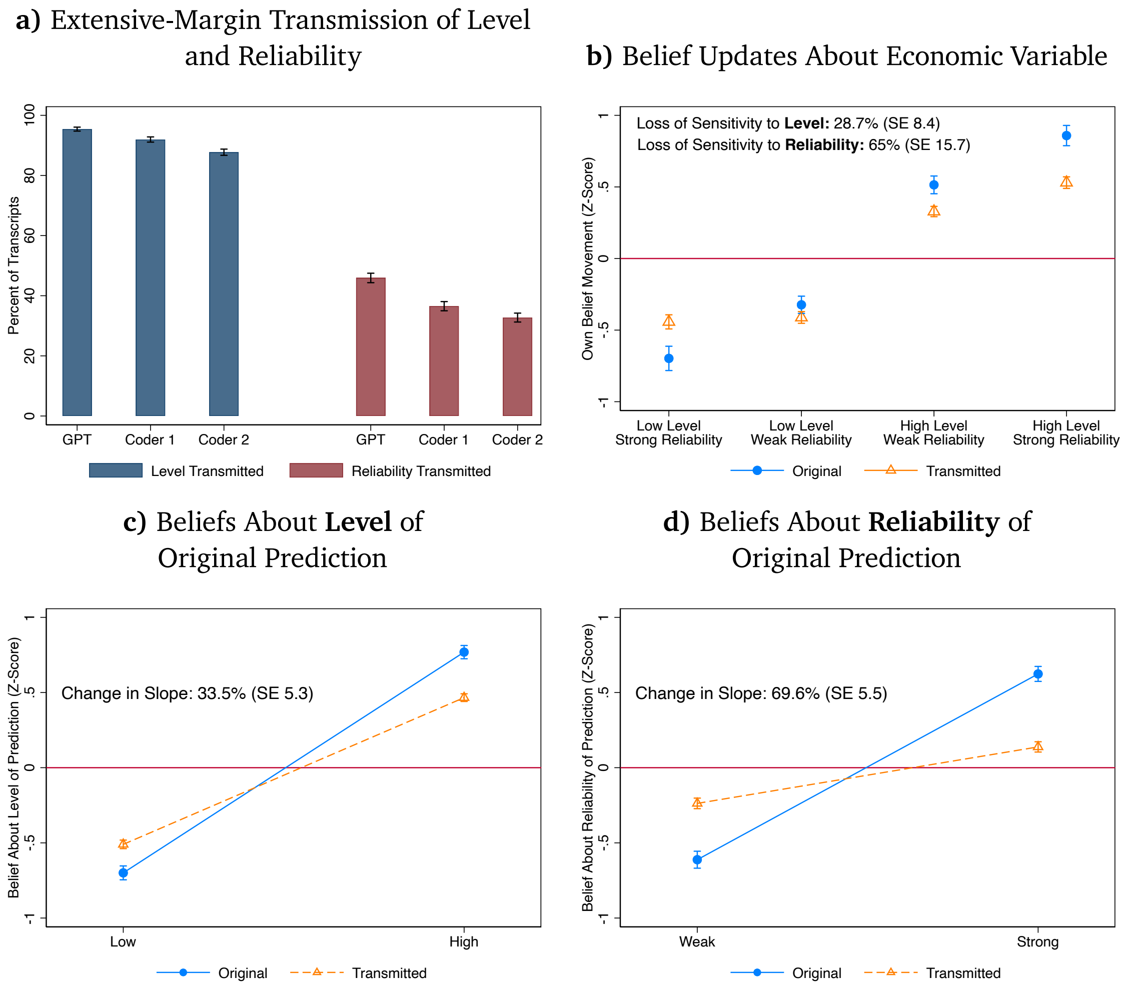
Because we want to compute counterfactuals in which we separately eliminate level and reliability information loss, we map listeners' message beliefs (rather than state beliefs) into investment choices via an implied posterior distribution over $\{-10, 0, 10\}$. We interpret the listener's level message 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 calculations, we assume in each level $\times$ 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.

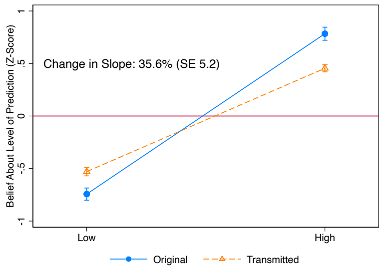
To compute efficiency losses under belief attenuation, we compare optimal portfolios under the ground-truth correct beliefs to the optimal portfolios chosen given beliefs attenuated by the amounts estimated in our experiment: either attenuating beliefs with respect to variation in level (by shrinking 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

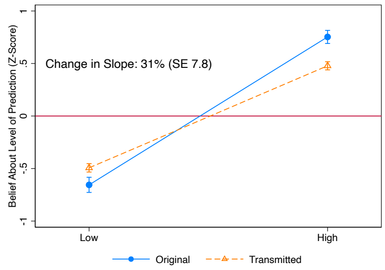


Appendix Figure A14: This figure presents data from our version of the baseline experiment that uses explicit incentives for transmission of level and reliability. Panel (a) replicates Figure 1. It shows the fraction of transmitted messages classified by GPT-4 and our two human coders as containing statements about the level or reliability of the original forecast. Panel (b) replicates 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 A15 splits these graphs by implicit versus explicit incentives.

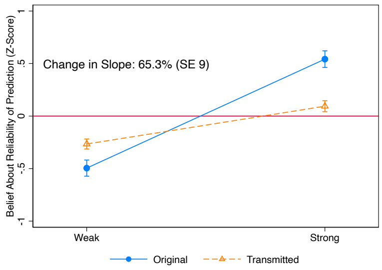
a) Level: Explicit Incentives



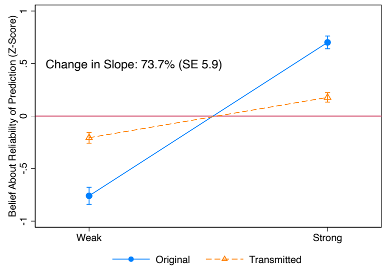
b) Level: Implicit Incentives



c) Reliability: Explicit Incentives



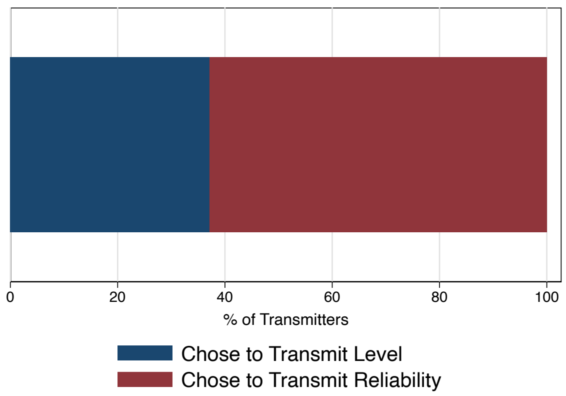
d) Reliability: Implicit Incentives



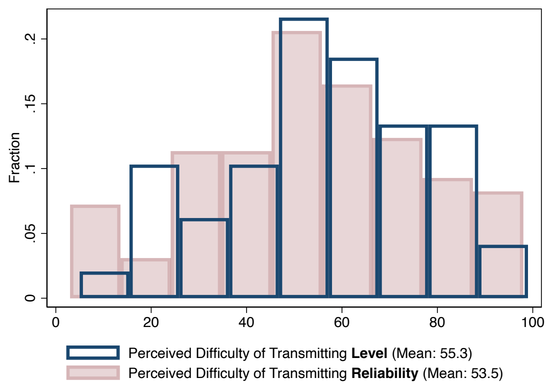
Appendix Figure A15: This figure replicates Figure 2 in the Content Transmission Incentives data, separately by respondents randomized into the explicit and implicit transmission incentives.

A.1.4 Supplementary Experiment: Choice of Incentives

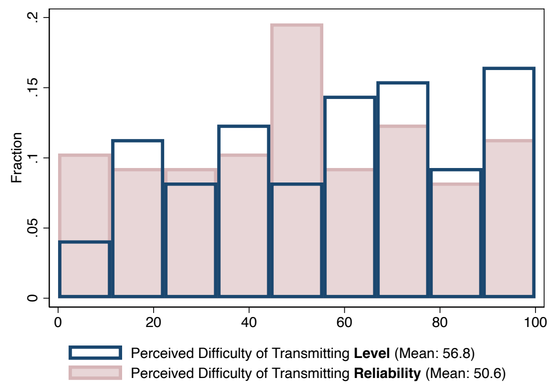
a) Transmitters' Choices of Incentives



b) Beliefs About Difficulty of Transmission (**Pre**-Transmitting)



c) Beliefs About Difficulty of Transmission (**Post**-Transmitting)



Appendix Figure A16: 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.2 Summary Statistics

Appendix Table A3: Summary Statistics: Listener and Transmitter Experiments

	Belief Movement Incentives		Content Transmission Incentives		Incentive Choice	Salience	
	Transmitters	Listeners	Transmitters	Listeners	Transmitters	Transmitters	Listeners
Age	43	40	37	38	43	37	38
Female	.52	.49	.52	.49	.52	.52	.52
Employed	.79	.78	.8	.75	.81	.8	.78
Education: BA+	.61	.6	.59	.56	.64	.66	.63
Race: White	.67	.66	.73	.72	.73	.57	.61
Race: Black	.21	.17	.12	.14	.19	.24	.21
Observations	540	1510	501	1509	97	244	1010

A.3 Regression Tables

A.3.1 Belief Movement Incentives

Appendix Table A4: 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 A5: 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 A6: 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 3.4, 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)."

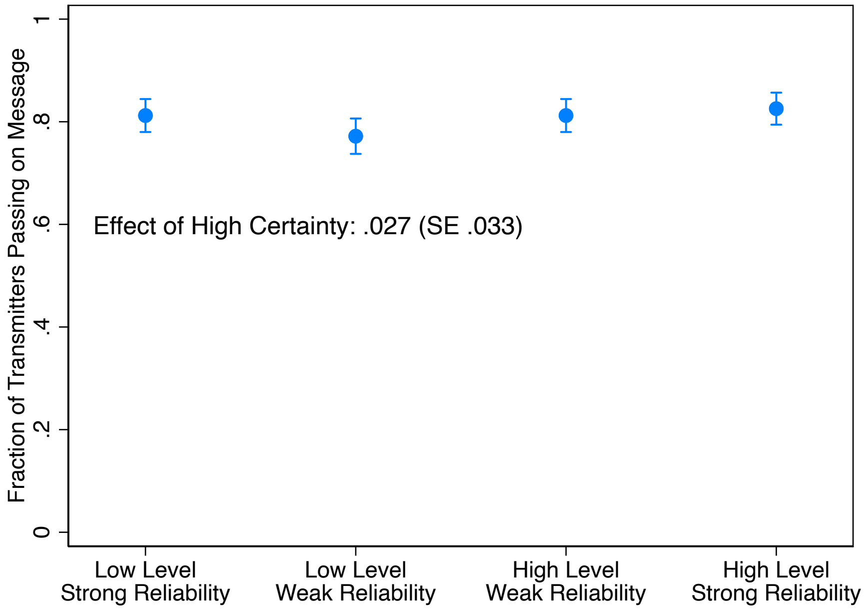
If transmitters opt out, the matched listener makes the investment decision without receiving any message.

Logistics. The additional transmission experiment was run with 298 U.S. respondents from Prolific in December 2024.

Results. We examine whether the decision to record a message depends on the level and reliability of the original message. Figure A17 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 A8 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 A7 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 A17: 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 A7: 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 third 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.

Appendix Table A8: 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 E 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 A18 Panel (a) 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.[27]

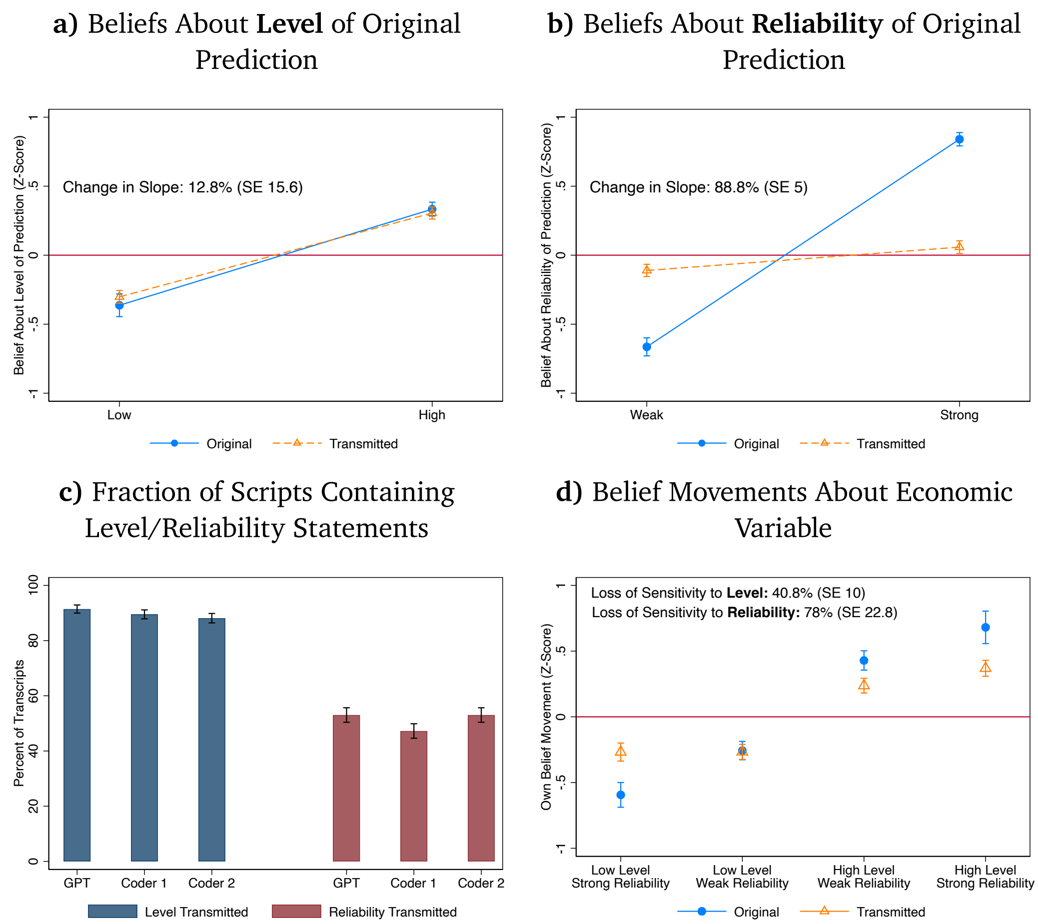
However, the increased fraction of transmitters mentioning reliability does not translate into a reduction in differential information loss according to our belief-based measures. Panels (c) and (d) show that differential loss *strengthens*: level information loss halves, to 12.8%, while reliability loss is unchanged. A formal test of equality of the two information loss statistics rejects the null at $p < 0.001$, $\chi^2 = 21.7$. Appendix Figure A18 Panel (b) 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,[28] 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.

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

[28] 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 A18: 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. $N = 181$ listeners 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 instructions (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).

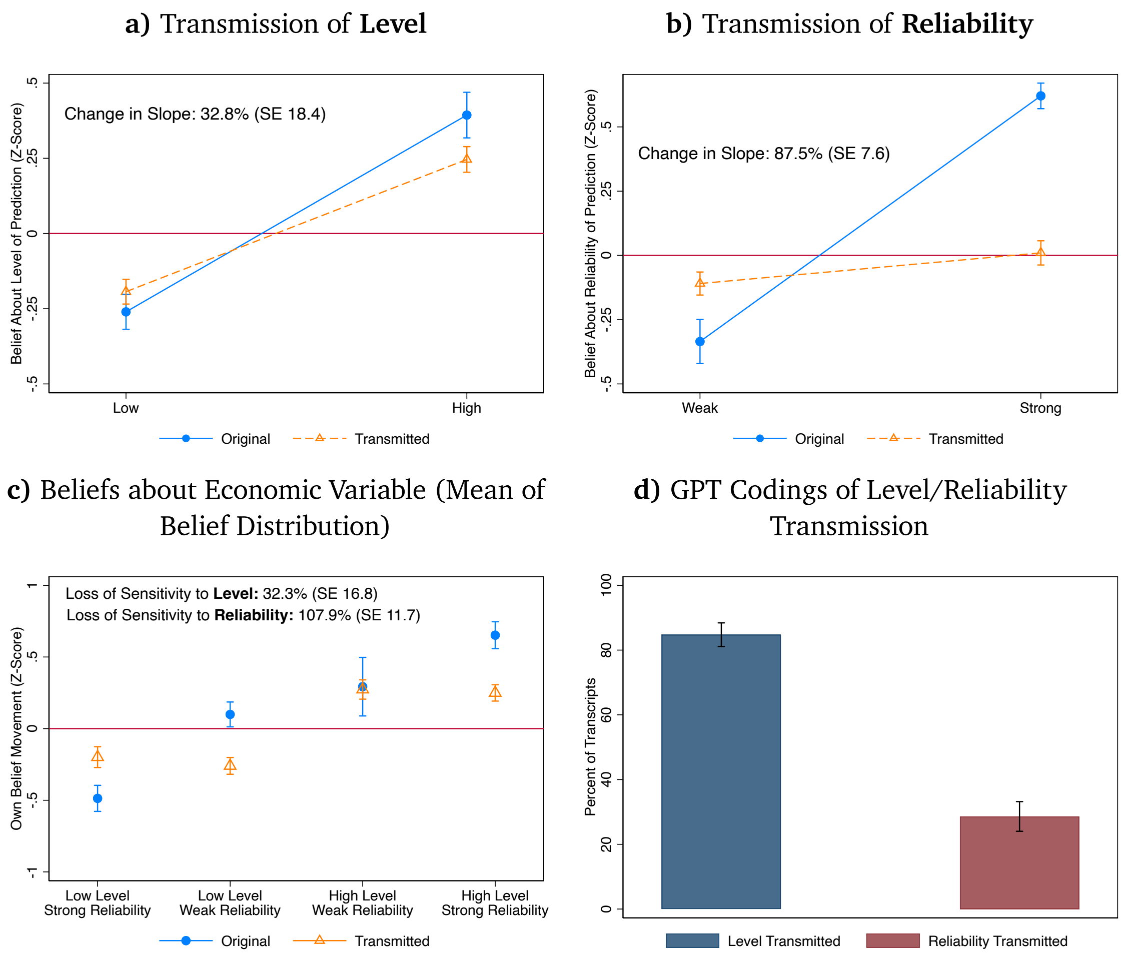
Transmitters' incentives are directly tied to how closely the beliefs induced by their message match those induced by the original recording. As stated in the instructions, *"your likelihood of receiving a bonus payment depends on how close the average probability distribution induced by your message 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 A19 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 32.8 percent relative to original messages (SE 18.4). By contrast, sensitivity to reliability collapses almost entirely, with an 87.5 percent reduction in the slope (SE 7.6), 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 32.3 percent (SE 16.8) and reduces sensitivity to reliability by 107.9 percent (SE 11.7), 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 transmitter recordings indicate that level information is transmitted in the vast majority of messages, whereas explicit references to reliability are rare. Taken together, Appendix Figure A19 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.



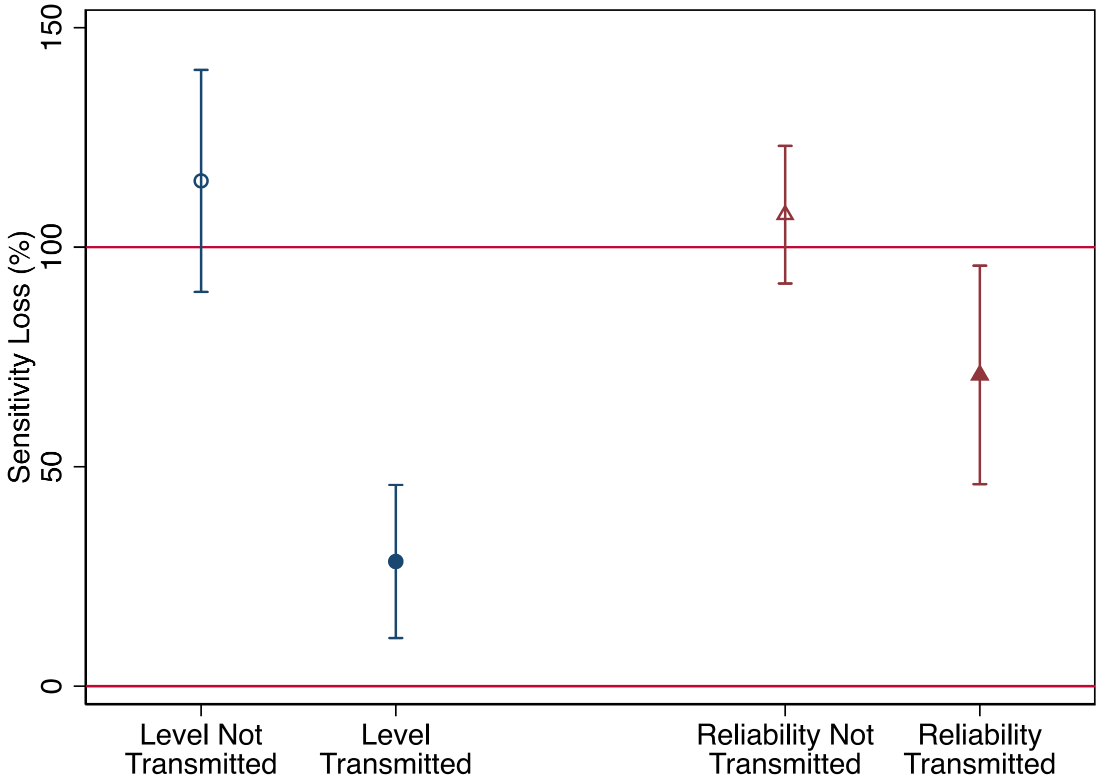
Appendix Figure A19: This figure presents results from our Belief Distribution Movement Incentives experiment, which replicates our baseline experiment with the following change: (i) we elicit beliefs about the state of the world through an interface that asks for a subjective probability distribution 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 changed 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 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. Panel (b) of Figure 1 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 70.9% (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 A22 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 transmitters' 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.

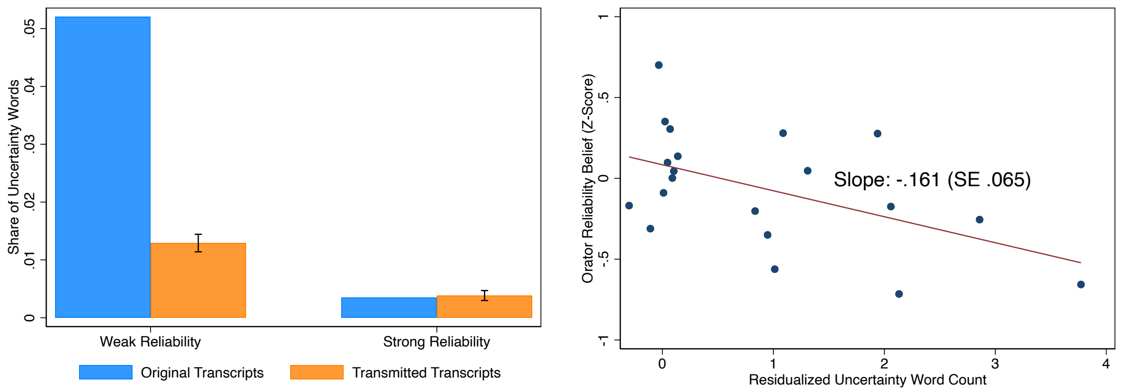
Appendix Figure A20: Information Loss Conditional on
Presence of Statements about Level/Reliability



Appendix Figure A21: 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 A22: 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.

D Details on TV Uncertainty Analysis

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.[29]

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:

$$\text{EconUncCable}_{t,c} = \alpha\text{EPU}_t + \beta\text{AllUncCable}_{t-1,c} + \gamma\text{EPU}_t \times \text{AllUncCable}_{t-1,c} + X_{t,c} + \varepsilon_{t,c} \quad (2)$$

where $\text{EconUncCable}_{t,c}$ is the frequency of appearances of uncertainty words in proximity to economic words on day t on channel c. EPU_t is Economic Policy uncertainty index at day t. $\text{AllUncCable}_{t-1,c}$ 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 γ, which captures the effect of recent uncertainty in cable news on the responsiveness of $\text{EconUncCable}_{t,c}$ 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 $\text{AllUncCable}_{t-1,c}$ and the responsiveness of $\text{EconUncCable}_{t,c}$ 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 γ 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 $\text{AllUncCable}_{t-1,c}$ and the responsiveness of $\text{EconUncCable}_{t,c}$ to EPU_t which

[29]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.

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.

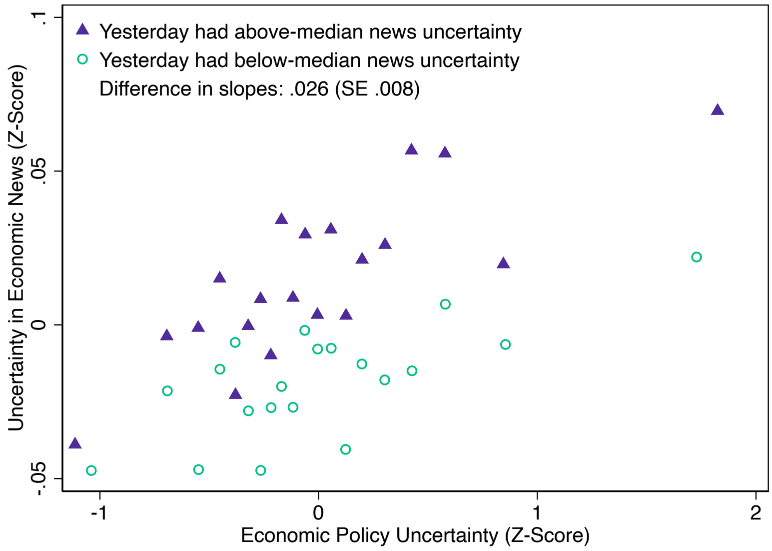
Figure A23 Panel (a) plots channel-specific economic uncertainty on cable TV news against the newspaper-based Economic Policy Uncertainty (EPU) index, separately for days following above- versus below-median uncertainty coverage on the same channel. It shows that expressions of economic uncertainty on 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 A9. 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 A9 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 A23: 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 D for details.

Appendix Table A9: 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

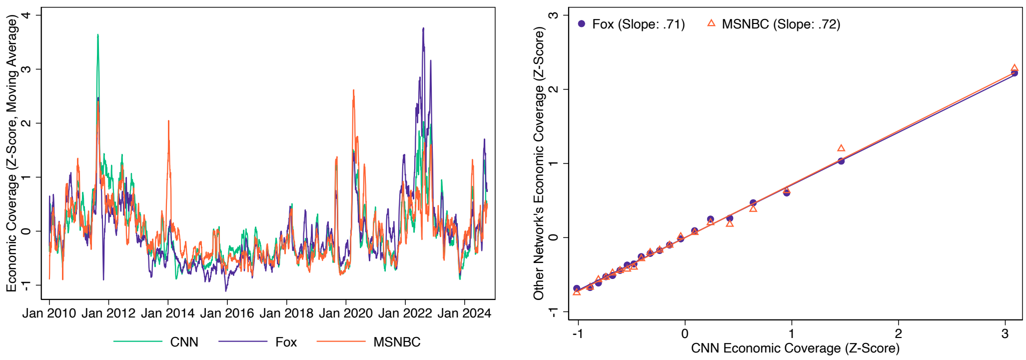
of the above words that occur *without* any of the economic words

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 A24 plots the time-series of our measure of the quantity of economic coverage. It picks out major events, including the September 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

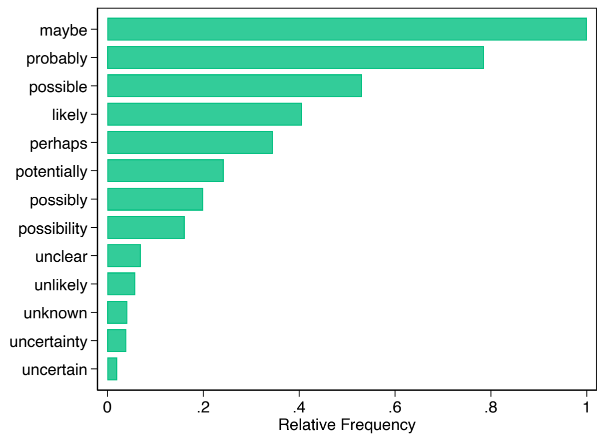
b) Correlation Between Channels' Economic Coverage



Appendix Figure A24: 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 A25 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

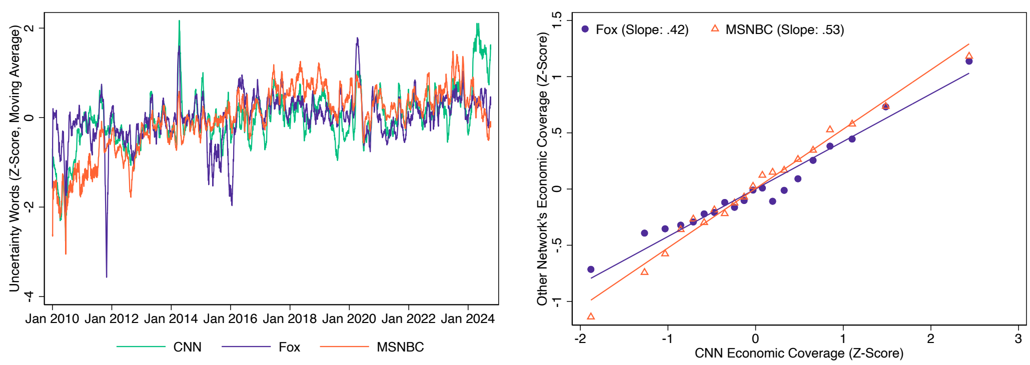


Appendix Figure A25: This figure plots a 30-day moving average of a z-score of our measure of the amount of economic coverage on cable news.

Figure A26 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 A26: 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.

E 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.

E.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 module 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] will strongly increase substantially 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] are 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 to never 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] increase 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 lead 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.

E.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