Lost in Transmission*

Thomas Graeber Shakked Noy Christopher Roth

July 23, 2024

Abstract

How does word-of-mouth transmission distort economic information? We pay participants to listen to audio recordings containing economic forecasts and accurately transmit the information through voice messages. Other participants listen to an original or a transmitted recording before stating incentivized beliefs. Across various transmitter incentive schemes, a forecast's reliability is lost in transmission at a far higher rate than the forecast's level. Reliable and unreliable information, once filtered through transmission, impact listener beliefs similarly. Mechanism experiments show that information about reliability is not perceived as less relevant or harder to transmit, but is less likely to come to mind during transmission.

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

* We thank Kai Barron, Doug Bernheim, John Conlon, Benjamin Enke, David Hirshleifer, Alex Imas, Frank Schilbach, Andrew Schotter, Michael Thaler and Florian Zimmermann for helpful comments and suggestions. We thank seminar audiences at the briq belief workshop, the Beliefs, Memory, Narratives Workshop, Brown, the Early Career Behavioral Economics Conference, the ECONtribute retreat, ETH Zurich, the Federal Trade Commission, Harvard Business School, MIT, Princeton, UC Santa Barbara and Stanford. We thank Simon Cordes, Pietro Ducco, Maximilian Fell, Frida Kueper, Malte Kornemann, Nicolas Röver, Constantin Schesch and Georg Schneider for outstanding research assistance. All of the data collections were pre-registered on the AEA RCT registry: `https://www.socialscienceregistry.org/trials/12119`. Roth acknowledges funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2126/1-390838866 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: Harvard Business School, `tgraeber@hbs.edu`. Noy: Massachusetts Institute of Technology, `snoy@mit.edu`. Roth: University of Cologne, ECONtribute, CEPR, NHH, and MPI for Research on Collective Goods Bonn, roth@wiso.uni-koeln.de.

1 Introduction

For many economic decisions, people rely on secondhand information obtained through conversations with family, friends, colleagues, acquaintances, and strangers (e.g., Lazarsfeld et al., 1968; Granovetter, 1973; Hirshleifer, 2020). Verbal transmission of information is imperfect: pieces can get lost, distorted, or added. As the game of *Telephone* illustrates, the meaning of even the simplest sentences is often garbled beyond recognition in the process of word-of-mouth transmission. The quality of economic decisions hence hinges not only on the quality of original information, but also on whether and how that information gets distorted as it passes through the chain of transmission.[1] In particular, if certain types of information are consistently more likely to be lost in transmission than others, systematic biases may emerge in the supply of information and predictably distort downstream beliefs and actions.

How does verbal transmission distort economic information? We conduct pre-registered, incentivized experiments with more than 5,000 participants that allow us to study the nature and consequences of verbal information transmission. In a *transmitter* experiment, participants listen to a one-minute message giving a qualitative forecast about an economic variable and are incentivized to then record themselves passing on the information they heard. Participants separately transmit forecasts about two variables: home price growth in a U.S. city and revenue growth of a U.S. retailer. In a subsequent *listener* experiment, a different set of participants listen to either the original message or a transmitted version of that message before stating incentivized beliefs about the relevant variable and about the characteristics of the original prediction. Comparing the beliefs of listeners hearing original messages to the beliefs of listeners hearing transmitted versions of those messages lets us characterize distortions introduced in the transmission process using simple quantitative measures. This belief-based notion of information distortions directly motivates the design of the incentives in the baseline transmitter experiment: transmitters are paid 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, friends share financial advice or economic expectations sourced from media consumption, doctors convey information about patients in shift-to-shift handoffs, and journalists convey news to the public.

To study whether some types of information are distorted by transmission more strongly than others, we focus on two key attributes of messages: the *level* of the prediction contained in a message—i.e., the realization of a signal—and the *reliability* of that prediction—i.e., the precision of that signal. This parsimonious but general taxonomy follows a basic characterization of messages in information economics. It is theoretically appealing because a Bayesian agent

[1] 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."

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

requires both a signal value and a signal precision to form an update.

Our experiments separately manipulate the level and reliability of the original forecasts seeded in the transmitter stage, allowing us to compare transmission-induced distortions of level information to distortions of reliability information. 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 two kinds of manipulations. The first changes the forecast accuracy expressed by the speaker by weaving certainty- or uncertainty-denoting prefixes (e.g., "definitely", "possibly") and explicit confidence statements (e.g., "I am highly confident", "I am not confident") into an otherwise-identical text. The second manipulation simultaneously changes 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, i.e. not using numbers, thereby mimicking how people naturally communicate in many real-world settings.[3] In the context of messages in natural language, the signal value (level) thus comprises substantive information about the predicted state of the world, whereas signal precision (reliability) captures information about the characteristics of the message itself, such as the quality of the underlying evidence or the credentials or confidence of the speaker. Our manipulations produce qualitatively Bayesian belief updates: when listening to the original messages, listeners update in the direction of the forecast's implied level, moderated by the forecast's reliability.

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

In our first set of analyses, we examine listeners' beliefs about the level and the reliability of the predictions in the original messages. We estimate the sensitivity of level beliefs to the experimental manipulation of level, as well as the sensitivity of reliability beliefs to the manipulation of reliability. We then compare the sensitivities of listeners who directly hear the original messages to the sensitivities of listeners who hear transmitted versions of those messages. The difference between the sensitivities of the two groups provides our main measure of transmission-induced *information loss* for level and reliability.

Consider the loss of level information. Among listeners who directly hear the original messages, switching from a low-level message to a high-level message shifts 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 $100 \times [(1.37 - 0.88)/1.37] \approx 34\%$ loss of sensitivity to variation in the level of the original prediction.

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 listen-

[3]We show that our findings also hold when communication includes numerical expressions for level and reliability.

ers 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 less reliable but unreliable messages are perceived as more reliable. Both effects have similar magnitudes, meaning that transmission does not change the average perceived reliability of messages. Similarly, forecasts in the high condition are perceived as predicting a less high level, and low forecasts as predicting a less low level. We show that such symmetric compression of both level and reliability towards intermediate (or prior) values can be microfounded in a simple model of noisy transmission, leveraging ideas from the recent literature on belief formation (e.g., Enke and Graeber, 2023; Augenblick et al., 2024; Ba et al., 2022).

In our second set of analyses, we examine listeners' belief updates about the economic variables discussed in the recordings. 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. We calculate that transmission causes the sensitivity of listeners' belief updates to our level manipulations to decrease by 30%, and their sensitivity to our reliability manipulations to decrease by 90%. These numbers are very similar to the information-loss statistics we calculated using listeners' beliefs about the level and reliability of the original predictions. In addition, the distortions of level and reliability information jointly result in a pronounced *shrinkage* of average belief updates towards zero.

In our final set of analyses, we abstract away from belief-based measures of information loss and directly examine the transcripts of transmitted messages. Human and machine coding of the content consistently reveals that while nearly all of the transmitted messages include some statement regarding the level of the original prediction, only about a third mention the original prediction's reliability or include other indicators of reliability, such as uncertainty prefixes. With an average of 114 words (8-10 sentences), transmitted messages tend to be only about half as long as the original messages. Yet even the longest 10 percent 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.

Notably, when we restrict to the one-third of transcripts that contain some statement about reliability, we still observe 70% loss of reliability information using our belief-based measures, indicating that even when reliability is mentioned, it is not fully communicated or emphasized.

While real-world communication is typically qualitative, many important settings involve transmission of quantitative information. In a robustness 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 pre-registered experiment additionally addresses concerns that our baseline results are driven by some extraneous difference between the way level and reliability are communicated in our baseline design, for example that qualitative level manipulations feel sharper or more binary than qualitative reliability manipulations.

There are two key implications of differential information loss, as illustrated by our model of a listener who decodes noisily transmitted messages. First, the loss of reliability indicators causes low quality information to excessively shape beliefs, while high quality signals are given too little weight. This effect of transmission may operate alongside and compound a distinct *updating* bias demonstrated in the context of quantitative signals: Griffin and Tversky (1992) and Augenblick et al. (2024) argue that, even conditional on knowing the precision or diagnosticity of a signal, people overinfer from weak signals and underinfer from strong ones. 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 the reliability indicators that do reach them may contribute to the spread of unreliable news and misinformation. Second, if a group of people with heterogeneous priors encounter a new piece of information filtered through verbal transmission, the average shrinkage of belief updates induced by transmission slows belief convergence and can sustain belief polarization.

We next ask what drives the differential information loss we document. On the one hand, reliability information could be disproportionately lost as the result of a deliberate tradeoff, either because the perceived benefits of transmitting reliability information are lower than for level information, or the perceived cognitive costs of transmitting reliability information are higher. On the other hand, differential loss could result not from a deliberate constrained optimization process, but 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. First, recall that transmitters in our main experiment are incentivized to record messages that induce downstream belief updates as close as possible to those induced by the original messages. After they 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. Second, we conduct an additional experiment that explicitly and equally incentivizes transmitters to pass on level and reliability information, effectively fixing beliefs about the relative benefits of transmitting the two dimensions. Even under this more conservative set of transmitter incentives, we find pronounced differential information loss, at about 30% for level information and 70% for reliability. These findings show that differences in beliefs about the benefits of transmitting level versus reliability information cannot account for much of the differential information loss we document.

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 of 63% choose to be incentivized based on their transmission of *reliability* information and 51% expect it to be easier to communicate. These beliefs do not change much *ex post*, after participants have experienced the task (52% expect reliability to be easier to communicate). 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 (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"). Transmitters may be unable to remember the reliability of the original message, even when explicitly asked about it (a failure of cued recall), or it may simply fail to come to mind during the transmission process (a failure of free recall).

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 about level and reliability 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, i.e., once they are specifically prompted about level and reliability information.

However, even though transmitters remember reliability information when prompted, reliability information may not come to mind *when completing their recordings*, i.e., in a free recall setting and facing significant cognitive constraints. Our combined 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, agreeing 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 combined findings on the memory channel reveal important differences between cued recall and less structured, free recall for transmission: plenty of information may fail to be transmitted even if it is explicitly known to be important and remembered when directly prompted.

We conclude from our series of mechanism experiments that reliability information is lost in transmission largely because it fails to come to mind during the cognitively taxing process of verbal transmission, and discuss potential reasons for this phenomenon. Moreover, we show that the quality of verbal transmission can be strongly improved through an intervention that reminds people *at the time of transmission* to also consider the reliability of information.

We close with a discussion of the implications and external validity of our findings, pointing to field evidence and real-world cases consistent with our main finding of differential loss of reliability information. The loss of reliability we document 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.

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 narratives (Shiller, 2017, 2020; Hirshleifer, 2020). Recent contributions in this literature have focused on the role of narratives for belief formation (Andre et al., 2022; Kendall and Charles, 2022; Bursztyn et al., 2023; Graeber et al., 2024b; Barron and Fries, 2023). Our experiments identify which kinds of information are more likely to be successfully passed on from one person to another through spoken communication. We relate to work by Graeber et al. (2024a), who study how explanations shape the contagion of truths and falsehoods.

We also relate to a literature on how belief formation is shaped by selective attention (Graeber, 2023; Ba et al., 2022; Hartzmark et al., 2021; Enke, 2020), complexity (Oprea, 2020; Enke and Graeber, 2023; Enke et al., 2023; Enke and Shubatt, 2023) and memory (Bordalo et al., 2021; Gennaioli and Shleifer, 2010; Bordalo et al., 2023, 2024). Previous research suggests that people pay insufficient attention to the "weight" (or precision, predictive validity) of evidence (relative to its "strength", or magnitude) when forming their beliefs in abstract and quantitative updating tasks (Griffin and Tversky, 1992; Massey and Wu, 2005). Our paper differs from this literature in its focus on how cognitive constraints shape the verbal transmission of information, and hence how they affect the *supply* of information.[4] Our results highlight an important role for selective memory in driving the differential loss of reliability information.

Our paper also contributes to a large literature on social learning (Weizsäcker, 2010; Mobius and Rosenblat, 2014; Banerjee, 1992; Bikhchandani et al., 1992; Bursztyn et al., 2014; Jackson and Yariv, 2007; Galeotti et al., 2010; Golub and Jackson, 2010; Golub and Sadler, 2016), information diffusion (Fehr et al., 2022; Banerjee et al., 2013, 2019; Chandrasekhar et al., 2022; Vespa and Weizsäcker, 2023; Han et al., forthcoming; Akçay and Hirshleifer, forthcoming), face-to-face interactions (Atkin et al., 2022; Battiston et al., 2021), and verbalization (Batista et al., 2024). Conlon et al. (2022) show in the context of a classic balls-and-urns belief updating paradigm 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, and (ii) the investigation of underlying cognitive mechanisms that shape transmission of different kinds of information.[5]

Finally, information transmission has also been the subject of work outside of economics. For example, Carlson (2019, 2018) finds that political information is partially lost when people trans-

[4]Thaler (2021) studies how strategic incentives shape the supply of false messages in a politicized context.

[5]Braghieri (2023) provides a theoretical framework to study the process of decoding for an agent who might have inaccurate beliefs about the information environment.

mit 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.

Our paper proceeds as follows: Section 2 describes the design of our baseline transmitter and listener experiments. Section 3 provides results on differential information loss. Section 4 provides evidence on the role of (i) expected benefits of transmission, (ii) anticipated costs of transmission and (iii) the role of memory constraints. Section 5 discusses interpretation and external validity. Section 6 concludes with a summary.

2 Baseline 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, a different set of respondents listen to either the original recordings or transmitted versions before forming their beliefs.[6]

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 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

[6]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`.

minutes.[7] 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. Chafe and Tannen, 1987; Akinnaso, 1982; Berger and Iyengar, 2013). Written text tends to be more formal, structured, premeditated, and requires 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, including tone, emphasis, and disfluencies such as pauses, repetitions, revisions, hesitations, or filler words.

Transmitter incentives. The design of our baseline transmitter incentives directly follows our conceptualization of a message's information content as *the average belief movement induced 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 10 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. Four remarks are in order. First, transmission under this scheme is guided by which elements of a message transmitters *believe* are most relevant for listeners' belief changes. Those beliefs may be biased, which would be a source of transmission distortions that we would want to capture. We examine these beliefs directly in Section 4.

[7]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 word: people are, over time, exposed to multiple pieces of information on various topics, before eventually relaying some information to others. In Section 4.3 we report evidence that this feature of the experiment is inconsequential: transmitter's beliefs about an original forecast at the very end of the experiment are similar to those of a listener immediately after hearing just that one forecast.

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.[8] 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.

Finally, while our experiment *requires* transmitters to pass on the information they hear, transmitters in real-world contexts can often decline to do so. For example, a transmitter may choose not to transmit information they are uncomfortable with, do not agree with, or think is not worth sharing. This extensive-margin decision of whether or not to pass on a message will also shape the supply of information. For simplicity, our design focuses on the intensive margin of transmission and examines loss of information conditional on an attempt to transmit the message. In Section 5, we discuss existing evidence on extensive-margin sharing decisions, which we argue suggests that our main implications would survive the addition of the extensive-margin decision.

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 as well as links to the audio recordings of the messages are available in Appendix C.

[8]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. However, in practice, we consider this to be extremely unlikely. Our data confirm this: we obtained no transmitter recordings indicating an attempt to communicate a predicted belief movement.

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 (Massey and Wu, 2005; Griffin and Tversky, 1992; Augenblick et al., 2024). 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 this 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).[9]

[9]As pre-specified, our main analysis focuses on comparisons between weak and strong reliability for simplicity. Appendix Figure A3 shows belief updates including the neutral-reliability condition.

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 exactly 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. Because both manipulations end up producing very similar results, we report all of our main results pooling both conditions, and show disaggregated results in Appendix Figure A5.

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 has a male or female voice. 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. We create the recordings using two human actors.

The different margins of randomization in the transmitter experiment 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.[10]

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

2.2 Listener Experiment

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

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

[10]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."

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 transmitted 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 attempt to pass on the original message. 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 of the prediction in the original message and the reliability of that prediction.

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.

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.[11] 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.[12] Results based on incentivized versus unincentivized message beliefs are virtually identical, as shown in Appendix Figure A6.

2.3 Sample and Procedures

We conducted our transmitter and listener 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 US respondents, respectively, in November 2023. Table A3 records summary statistics for all our experimental samples. All of the data collections were pre-registered on the AEA RCT registry: `https://www.socialscienceregistry.org/trials/12119`. As pre-registered, we drop recordings below the 5th percentile of recording

[11] 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.

[12] 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 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.

length or transcript word length (as a proxy for empty or content-less recordings). Following this restriction, our baseline transmitter experiment yields a total of 1,010 valid speech recordings. These were obtained by collecting speech recordings using the service Phonic, which we embed into Qualtrics.[13]

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, we analyze listeners' *message beliefs*, i.e., their beliefs about the characteristics of the original forecast. We show that when moving from original to transmitted messages, the loss of sensitivity of listeners' *reliability message beliefs* to our *reliability manipulations* is about three times as large as the loss of sensitivity of listeners' *level message beliefs* to our *level manipulations*. Second, we study listeners' *state belief updates*, i.e., their belief updates about the economic variables discussed in the recordings. We show that when moving from original to transmitted messages, the loss of sensitivity of listeners' *state belief updates* to our *reliability manipulations* is about three times as large as the loss of sensitivity of listeners' *state belief updates* to our *level manipulations*. Finally, we analyze the transcripts of transmitted messages and handcode them according to whether they contain statements about the level or reliability of the original forecast. We show that about three times as many transmitted messages contain statements about the level of the original forecast as contain statements about the forecast's reliability. Notably, however, our first two findings remain even when restricting to transmitted messages that do contain some statement about the reliability, indicating that even when reliability is mentioned, it is still not fully communicated.

Each of these methods of demonstrating differential information loss has its own advantages and drawbacks. Our analyses using message beliefs allow us to cleanly distinguish the level and reliability information contained in the original messages and separately track each type of information through transmission. However, message beliefs are not a very economically relevant object and may be difficult to interpret as outcomes. Our analyses using overall belief updates show differential information loss using a more economically relevant and interpretable outcome. However, the fact that overall belief updates are simultaneously determined by the level and reliability information that reach the listener mean that our analyses here could be contaminated by spillovers from one kind of information loss to the other. Finally, our analyses tracking mentions of level and reliability in transcripts are transparent and sidestep any potential issues with belief-based measures of information loss. However, these transcript analyses capture only part of the picture: our binary measure of "failure to mention reliability information" misses the differential information loss that occurs even when we condition on transcripts that do contain some mention of reliability. Overall, however, the fact that our finding of differential information

[13]We rely on an Amazon Web Services backend to feed the recordings into the Listener experiment.

loss is robust across these three separate measurement techniques bolsters our confidence in the reality of the finding.

3.1 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 1 presents results on message beliefs.

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

$$\begin{aligned} LevelBelief_i = \beta_0 + \beta_1 HighLevel_i + \beta_2 Transmitted_i \\ + \beta_3(Highlevel_i \times Transmitted_i) + \varepsilon_i, \end{aligned} \tag{1}$$

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.[14] The change in slope statistic is simply $-100 \times (\beta_3/\beta_1)$.

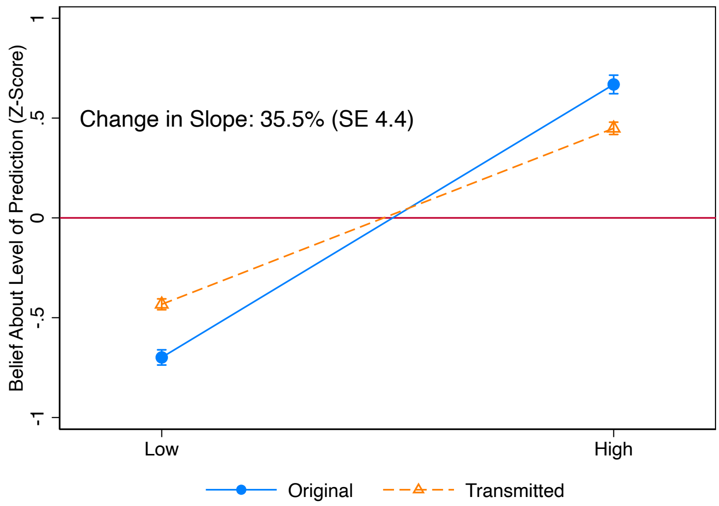
Panel (b) examines listeners' message beliefs about the reliability of the original predictions. Here, the sensitivity loss is nearly three times as strong. Listeners hearing the original messages believe the strong-reliability messages are 1.18 SDs more reliable than the weak-reliability messages on 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 1 further illustrates that, in both cases above, transmission has weakened the distinction between high- and low-level messages (or weak- and strong-reliability messages) by *symmetrically compressing listeners' beliefs towards an intermediate value*. This is compatible with the following dynamic: listeners hold an average prior that is located halfway between our two

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

manipulations; they update away from this prior when hearing a message; and the strength of this update is weakened by transmission. This weakening of belief updates would result if, for example, transmission introduced noise that obscured the original message's information content, letting us describe our result as *differential information loss*.

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



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

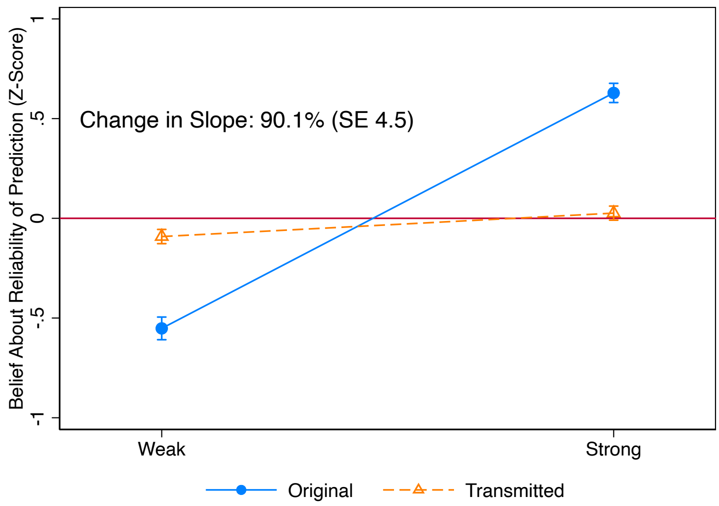


Figure 1: This figure presents data from our baseline experiment (Belief Movement Incentives). It shows listeners' beliefs about the level and reliability of the prediction in the original message, separately by whether the original message is low- vs high-level or weak- vs strong-reliability, and separately by whether the listener hears the original message or a transmitted version of it. Dots are mean beliefs and bars are standard error bars (1 SE each direction). $N = 1,510$ listeners and 540 listeners.

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 1. *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.*

We demonstrate differential information loss in two additional ways, using state belief updates and transcript analysis, in Sections 3.2 and 3.3. First, however, we formalize the interpretation of the message belief patterns displayed in Figure 1.

Formal interpretation. Figure 1 shows that transmission causes a symmetric attenuation of message beliefs to intermediate values. We can represent this symmetric attenuation by writing the level beliefs of listeners to transmitted messages, ℓ^t, as a weighted average of (i) the level beliefs of listeners to the corresponding original message, ℓ^o, and (ii) a default belief, ℓ^d, with ℓ^d falling between the ℓ^o induced by a high-level message and the ℓ^o induced by a low-level message.

$$\ell^t = \lambda_\ell \ell^o + (1 - \lambda_\ell)\ell^d \quad \text{with} \quad \lambda_\ell \in [0,1] \tag{2}$$

Similarly, we can write listeners' message beliefs about reliability, r^t, as

$$r^t = \lambda_r r^o + (1 - \lambda_r) r^d \quad \text{with} \quad \lambda_r \in [0,1]. \tag{3}$$

With a default belief of 0, our results from Figure 1 suggests a value of $\lambda_\ell \approx 1 - 0.34 = 0.66$ for level beliefs and $\lambda_r \approx 1 - 0.91 = 0.09$ for reliability beliefs.

This reduced-form characterization of the effects of transmission on message beliefs can be microfounded with various models from the existing literature. First, it can be captured by a model of noisy transmission that relies on the intuitions of noisy processing models popular in the recent literature (e.g., Enke and Graeber, 2023; Gabaix, 2019; Ba et al., 2022; Augenblick et al., 2024). This modeling approach provides a disciplined and tractable way to characterize common distortions in belief formation and other decision domains.

Under the noisy transmission interpretation, the defaults correspond to prior beliefs about the level and reliability of the original prediction, respectively. The transmission process adds zero-mean noise to the level and reliability expressed in the original message. The listener to a transmitted message combines their prior with the transmitted signal realization to form a Bayesian posterior about the level and reliability of the original prediction; the nonzero weight placed on priors due to the introduction of transmission noise results in the attenuation observed in Figure 1. The degrees of compression for level and reliability are captured by the *shrinkage factors* λ_ℓ and λ_r. These weights are determined by the variances of the transmission noise terms: more transmission noise means a lower λ and thus greater attenuation. *Differential* information loss results because transmission adds higher-variance noise to the reliability than the level. Under this model of noisy transmission, message beliefs emerge from (quasi-)Bayesian inference in a standard signal extraction problem. We relegate the derivation of the reduced-form equations (2) and (3) using a canonical noisy inference approach to Appendix A.[15]

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 weights λ reflect the difficulty of decoding each type of information.

While we consider the noisy transmission account to be compelling in our setting—transmission garbles messages in ways that add noise to the level and reliability communicated, inducing the listener to shrink to a prior level—we remain agnostic about which exact interpretation is the most accurate.

3.2 State Belief Updates

We now examine listeners' overall belief updates about the economic variables discussed in the original recordings, which we call *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 2. 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 2 shows that listeners to the original messages update twice as much on average 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.

However, because state belief updates are *simultaneously determined* by the perceived level and reliability of a piece of information, level and reliability information loss are actually not separately identified using state belief updates alone. To help interpret the effects of transmission on state belief updates, we therefore develop a simple illustrative model where a listener's state belief update is determined by the level and reliability information that reaches that listener, and show that the model's predictions about the consequences of differential information loss match our actual findings. The model also lets us think through the downstream consequences of differential information loss for the population's distribution of beliefs. Note that our findings do not depend in any way on the accuracy of this model; the model purely serves as a framework for interpreting the results.

[15]For illustrative purposes we derive our predictions in a model variant with a simplifying assumption that makes it less than fully Bayesian, as we explain in the Appendix. A fully Bayesian transmission model would yield the same predictions, at the cost of lower mathematical tractability.

Going from message beliefs to state beliefs. We model listeners' state belief updates as a function of their beliefs about the level and reliability of the original forecast (i.e., their message beliefs). We assume that listeners enter the experiment with a normal prior about the state,

$$\ell \sim \mathcal{N}(\ell^d, v). \tag{4}$$

Here, the mean of respondents' prior about the state, ℓ^d, coincides with their default message belief about the level, from Section 3.1.

We further assume that listeners view the level of the prediction in the original message, ℓ^o, as consisting of a noisy signal about the true state, with the precision of the noise term being equal to the reliability of the original prediction, r^o. Formally,

$$\ell^o = \ell + \varepsilon \qquad \text{with} \qquad \varepsilon \sim \mathcal{N}(0, 1/r^o). \tag{5}$$

Listeners to the original message observe ℓ^o and r^o, and form a Bayesian posterior estimate, $\hat{\ell}_1$, for the state:

$$\hat{\ell}_1 = \ell^d + \frac{v}{v + 1/r^o} \cdot (\ell^o - \ell^d) \tag{6}$$

Listeners to transmitted messages, by contrast, do not directly observe the level and reliability of the original message, ℓ^o and r^o. Instead, we here assume that they rely on their posteriors about those quantities, ℓ^t and r^t, given in equations (2) and (3). For listeners to transmitted messages, this yields the following posterior belief about the state, $\hat{\ell}_2$:

$$\hat{\ell}_2 = \ell^d + \frac{v}{v + 1/r^t} \cdot (\ell^t - \ell^d) = \ell^d + \frac{v}{v + 1/(r^d + \lambda_r(r^o - r^d))} \cdot \lambda_\ell \cdot (\ell^o - \ell^d) \tag{7}$$

The complete process of forming message and then state beliefs can be cast in terms of a unified two-stage noisy transmission model. In such a model, the original message is a noisy signal of the true state, the interpretation of the original message creates signals about the level and reliability, and the message transmission process adds additional zero-mean noise to those signals. We present a (slightly simplified) application of such a model to our transmission setup in Appendix A.

Comparative statics. Simple comparative statics about the effects of transmission on state belief updates can be derived from equations (6) and (7), by comparing the belief updates of listeners to original versus transmitted messages. Formally:

> *Prediction.* Suppose that listener belief updates follow Equation 7, that $\ell^o_L < \ell^d < \ell^o_H$ for low- and high-level messages L and H, and $r^o_W < r^d < r^o_S$ for weak- and strong-reliability messages W and S. Then absolute listener belief updates $|\hat{\ell}_2 - \ell^d|$ are increasing in λ_ℓ, holding λ_r fixed. Absolute listener belief updates are increasing in λ_r for strong-reliability messages and decreasing in λ_r for weak-reliability messages, holding λ_ℓ fixed.

The prediction follows immediately from Equation 7. For a more complete discussion of the assumptions underlying Equation 7 and hence this prediction, see Appendix A. Here, we discuss the prediction informally.

First, the prediction says that a stronger loss of level information (lower λ_ℓ) will uniformly shrink the absolute belief updates of listeners, by causing those listeners to place more weight on the default belief ℓ^d. This means that transmission causes listeners receiving *positive* level signals to update less positively, and causes listeners receiving *negative* level signals to update less negatively.

Second, loss of reliability information can either shrink or amplify absolute belief updates, depending on the reliability of the original message. In general, absolute belief updates should be larger the greater the perceived reliability of the original prediction. Transmission noise shrinks reliability perceptions towards the default reliability belief r^d. Hence for strong-reliability messages, where $r^o > r^d$, transmission-induced attenuation *reduces* the perceived reliability of the message. Transmission should hence shrink absolute belief updates from strong-reliability messages. By contrast, for weak-reliability messages, $r^o < r^d$, so transmission attenuation *increases* perceived reliability and amplifies absolute belief updates.

Experimental results. Panel (a) of Figure 2 displays the average state belief updates of listeners who directly hear original recordings, across the four categories of our level/reliability cross-randomization. In this figure we pool data from both topics, revenue and home price 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 2 shows the predicted effects of transmission on state belief updates given the comparative statics outlined above, and Panel (c) shows the actual effects of transmission, which match the predicted effects.

To understand the predictions and results, recall that the loss of *level* information should uniformly shrink listeners' belief updates towards zero (the mean belief update). This is because, as Panel (a) of Figure 1 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 2).

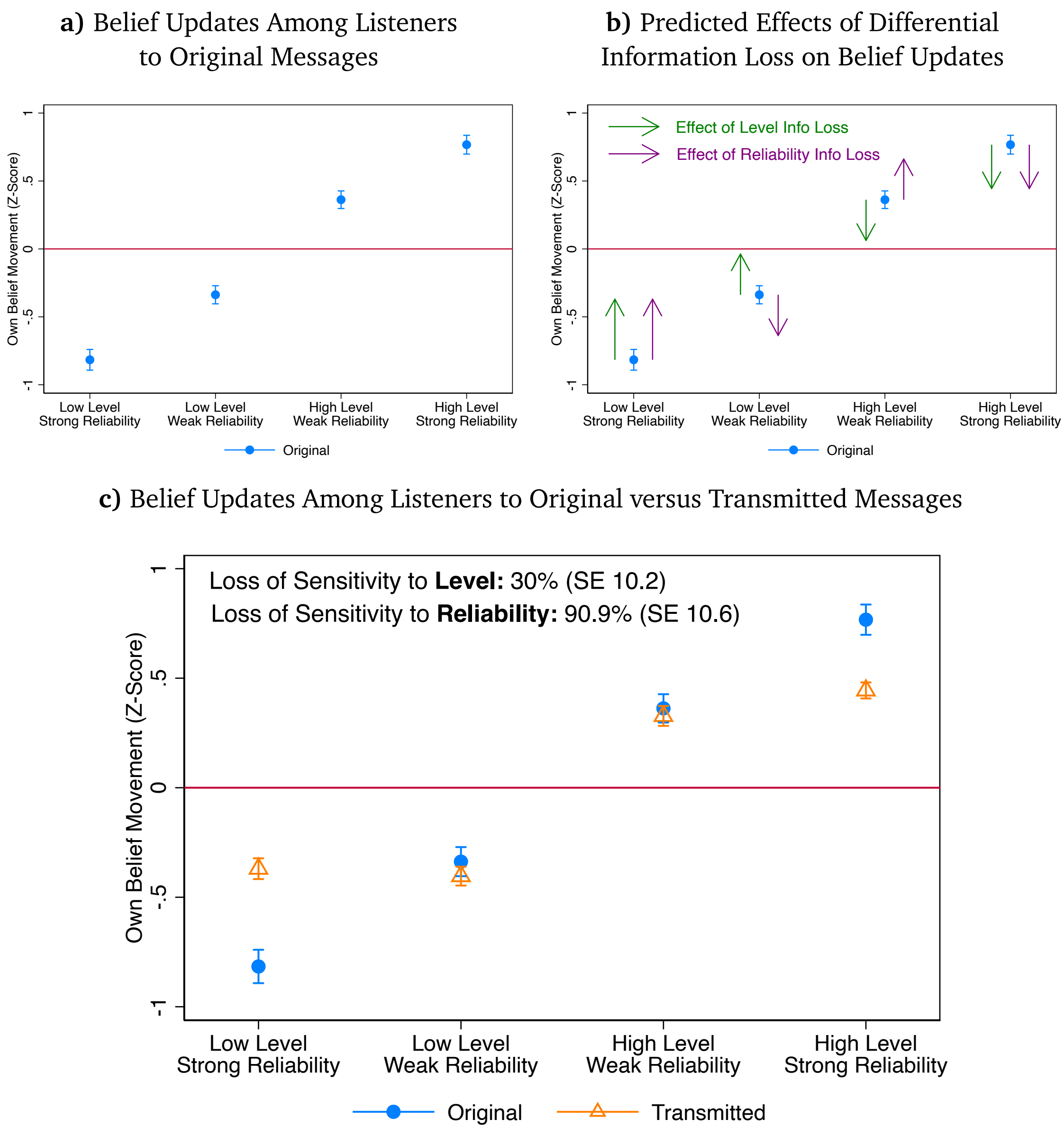


Figure 2: This figure shows average belief movements (posterior minus prior) about the economic variable from our baseline experiment (Belief Movement Incentives). Panel (a) shows average belief movements about the economic variable across the four different level-reliability conditions, only for listeners who directly hear the original messages. Dots are mean beliefs and bars are standard error bars (1 SE each direction). Panel (b) adds illustrative arrows. Panel (c) adds the corresponding beliefs of listeners hearing transmitted versions of the messages. $N = 1,510$ listeners and 540 transmitters. The loss of sensitivity to level information is calculated from a regression of the form: $\text{BeliefUpdate}_i = \alpha_0 + \alpha_1 \text{HighLevel}_i + \alpha_2 \text{StrongReliability}_i + \alpha_3 \text{Transmitted}_i + \alpha_4(\text{Highlevel}_i \times \text{Transmitted}_i) + \xi_i$. The loss of sensitivity to reliability information is calculated from a regression of the form $\text{BeliefUpdate}_i \times (2 \times \text{HighLevel}_i - 1) = \gamma_0 + \gamma_1 HighLevel_i + \gamma_2 \text{StrongReliability}_i + \gamma_3 \text{Transmitted}_i + \gamma_4(\text{Highlevel}_i \times \text{Transmitted}_i) + \zeta_i$, 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 A7 gives regression versions of these results. Figure A3 shows these results restricting to the Modular manipulation and including the neutral-reliability condition. Appendix Figure A4 shows empirical results validating the arrows in Panel (b).

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 1). 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 2). 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 2).

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.[16]

Panel (c) of Figure 2 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 A4 validates our comparative-static explanation of the empirical results by splitting transmitters according to whether they are coded as passing on reliability (see Section 3.3). 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.[17]

Summing up, what are the implications of transmission-induced information loss for the pattern of state belief updates? Two facts are evident from Panel (c) of Figure 2. First, averaging across all four conditions, listeners' absolute belief updates are 30% smaller when listening to transmitted messages, an effect that is entirely driven by the strong-reliability conditions.[18] 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 con-

[16]While reliability information loss is stronger than level information loss, this does not mean that the reliability effect will dominate; Panel (a) of Figure 2 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.

[17]These differences are small in magnitude, because (as we explain in Section 3.3) our binary codings do not capture the extent to which reliability information is passed on, which matters for belief updates.

[18]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.

trast, 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 2. *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.*

Two caveats apply to this result. First, the extent to which transmission weakens the average effect of new information on beliefs will depend on the proportion of real-world messages that fall into each of our four experimental buckets. Our experiment fixes the percentage of messages in each bucket at 25%; varying the percentages will vary the extent of this effect. This effect could even flip if real-world variations in reliability are such that the level and reliability effects do not cancel out in the middle two buckets, and the reliability effect dominates.

Second, one might object that when people encounter unreliable messages in the wild, they simply *do not pass them on*, rather than passing them on (as our experiment forced them to) and omitting information about their reliability. In this case, reliable and unreliable messages would not converge in influence, since the latter kind would not be passed on. We discuss this objection in Section 5 and point to field evidence that people that people *do* tend to pass on information they know to be unreliable while omitting mention of its unreliability.

3.3 Transcript Analysis

As a final way to demonstrate differential information loss, we abstract away from belief-based measures of information content and simply examine the transcripts of transmitted messages to see whether they contain statements about the level or reliability of the original forecasts.

Panel (a) of Figure 3 displays the share of transmitter transcripts classified as containing statements about the level of the original prediction or about the reliability of the original prediction. 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. 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.[19] See Appendix Table A2 for some examples of transmitted messages and their handcodings.

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). Panels

[19]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.

(c) and (d) show that this is true independent of the length of the transmitted message: even among transmitted messages that are 200-300 words long (longer even than the original messages), only 20% are unanimously agreed 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. This extra detail does not seem to matter for beliefs: conditional on whether level and reliability are passed on, longer messages are not associated with stronger transmitter belief updates or greater sensitivity to the original manipulations.

In Appendix Figure A9, we show that the fraction of scripts containing statements about level or reliability is also fairly stable across our four level × reliability conditions. In the weak-reliability conditions, transmitters are marginally more likely to say something about reliability and slightly less likely to say something about the level.

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 3 calculates the sensitivity loss statistics from Figure 1, 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 A7 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.

In addition to potential alterations in the substantive content, transmitted recordings may exhibit differences in non-content features. One notable difference is the increased presence of disfluencies in these recordings—hesitations, "um" statements, self-corrections, and so on. Theoretically, these disfluencies might influence how listeners perceive the reliability of the original forecast.[20] However, as demonstrated in Appendix Figure A7 Panel (c), there is no significant

[20]Relatedly, recent work by Gorodnichenko et al. (2023) shows that non-content features in the form

correlation between the presence of disfluencies and the original forecast's perceived reliability.

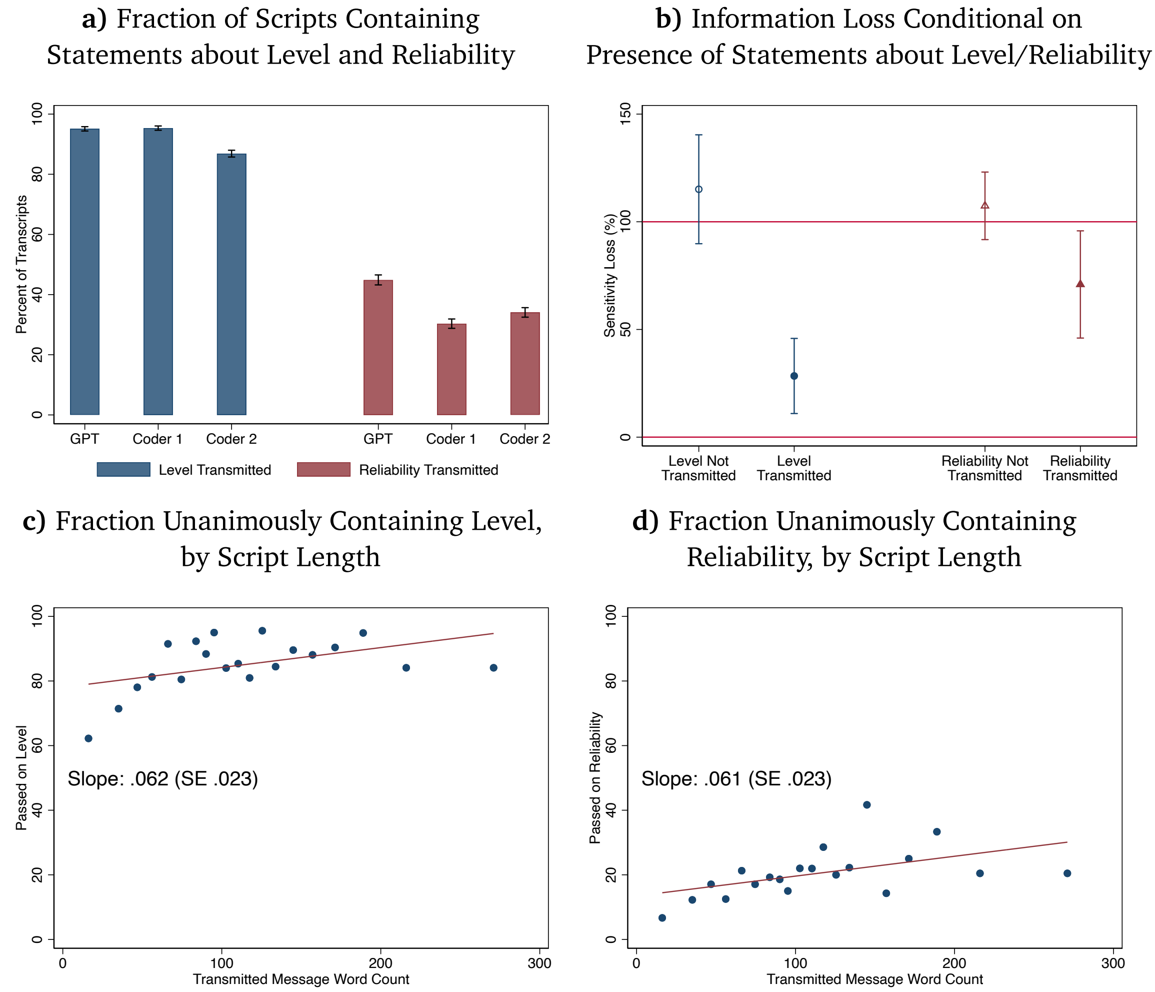


Figure 3: This figure presents data from our baseline experiment (Belief Movement Incentives). Panel (a) shows the fraction of transcripts that are coded as conveying any information about the level and reliability of the original forecast, separately by two human coders and GPT-4. Panel (b) calculates the sensitivity loss statistics from Figure 1, 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 are standard error bars in Panel (a); in Panel (b) they denote 95% confidence intervals around the coefficient estimates. Panels (c) and (d) show binscatter plots of indicators for being unanimously classified as containing statements about level versus reliability, regressed on the word count of the script. $N = 540$ transmitters, each of whom contributes two transcripts.

3.4 Robustness: 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

of emotions conveyed through voice significantly affect the interpretation of monetary policy announcements.

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, in an additional preregistered experiment.

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.

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 C for the full set of quantitative scripts.

Logistics. The additional transmission and listener experiments were run with 181 and 834 US respondents from Prolific, respectively, in June 2024. This collection was also pre-registered at `https://www.socialscienceregistry.org/trials/12119`.

Results Figure 4 shows that, if anything, this change to the scripts produces even stronger differential information loss: it halves level information loss, to 12.8%, while leaving reliability information loss 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 A11 shows that this is also true when analyzing listeners' state belief updates instead of message beliefs.

Interestingly, Appendix Figure A12 shows that a higher share of transmitters now mention reliability in their transmission: about 50% according to our human coders, compared to 30% in our baseline experiment, suggesting that numerical confidence statements increase the salience or ease of transmitting of reliability information. As before, over 90% of respondents mention the level. About 45% of transmitters pass on the level number and 25% pass on the reliability number. The increased fraction of transmitters mentioning reliability does not translate into a reduction in differential information loss according to our belief-based measures, in part because the quantitative scripts increase the impact of our original recordings on the reliability beliefs of listeners directly hearing them,[21] so that the omission of reliability information has a greater impact on beliefs than in our baseline experiment. (Hence a given fraction of scripts omitting reliability information has a larger impact on information loss as measured by message or state beliefs.)

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

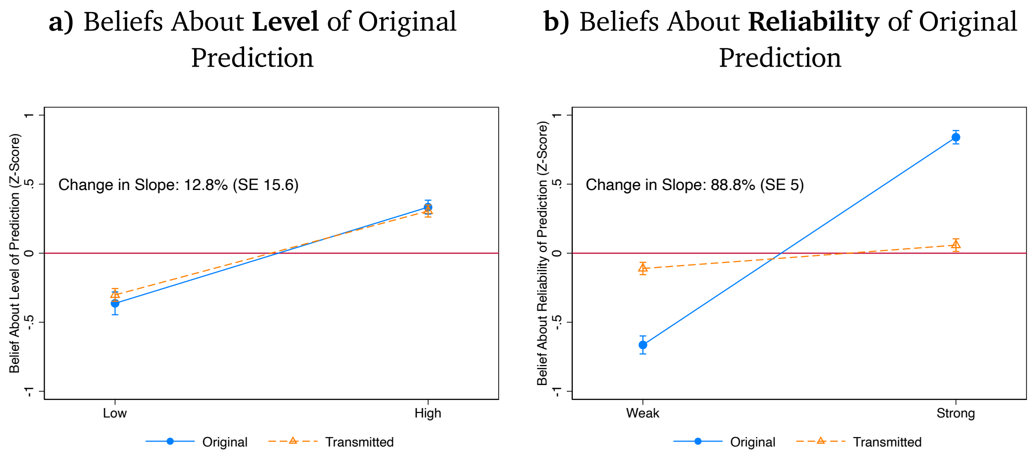


Figure 4: This figure presents data from our version of the baseline experiment that uses quantitative scripts. It shows listeners' beliefs about the level and reliability of the prediction in the original message, separately by whether the original message is low- vs high-level or weak- vs strong-reliability, and separately by whether the listener hears the original message or a transmitted version of it. Dots are mean beliefs and bars are standard error bars (1 SE each direction). $N = 181$ listeners and 834 listeners.

3.5 External Validity

We further discuss the external validity of our baseline findings in Section 5. Before that, we examine underlying mechanisms, which includes tests of alternative transmitter incentive schemes. These variations are informative about the generality of our results.

[21]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.

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 subjectively suboptimal.[22] 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 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

[22]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.

level information: the average response is 71% for level and 68% for reliability. Strikingly, this is true even 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.[23]

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,[24] 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.

[23]The fact that 31% of respondents report not passing on level, despite our handcoders 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).

[24]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 10 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 feelings 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, one about the level of the prediction in the message 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 US respondents from Prolific, respectively, in September 2023. This collection was also pre-registered at `https://www.socialscienceregistry.org/trials/12119`.

Main results. Figure 5 illustrates the sensitivity of listeners' message beliefs to the experimental manipulations of level and reliability, separately for original and transmitted recordings. In this figure, we pool data from the explicit and implicit incentive schemes. Panel (a) shows that listeners who hear a low-level original recording believe that the prediction is 1.47 SDs lower on average than listeners who hear a high-level original recording. Meanwhile, 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 0.98 SDs, indicating a 33.5% (SE 5.3) loss of sensitivity to level information. Panel (b) highlights a substantially more pronounced loss of sensitivity to reliability information: listeners who hear an original weak-reliability recording believe its reliability is 1.24 SDs lower on average than listeners who hear an original strong-reliability recording. The corresponding difference for beliefs about transmitted versions of these recordings is only 0.38 SDs, indicating a 69.6% (SE 5.5) loss of sensitivity to reliability information. 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 $p < 0.001$, $\chi^2 = 27.1$.

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

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

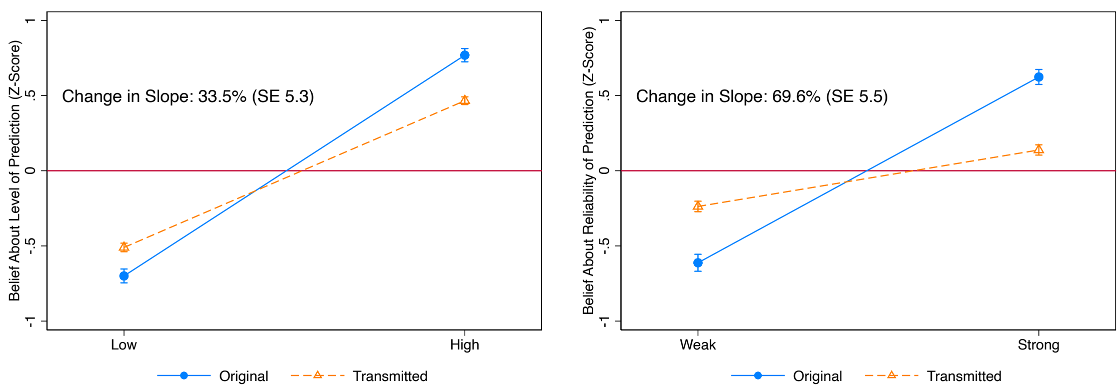


Figure 5: This figure replicates Figure 1 using data from the Content Transmission Incentive Experiments. It shows listeners' beliefs about the level and reliability of the prediction in the original message, separately by whether the original message is low- vs high-level or weak- vs strong-reliability, and separately by whether the listener hears the original message or a transmitted version of it. Dots are mean beliefs and bars are standard error bars (1 SE each direction). $N = 1,509$ listeners and 501 transmitters. Appendix Figure A13 shows belief updates about the economic variable from this experiment and Appendix Figure A14 splits these graphs by implicit versus explicit incentives.

Differences by transmitter incentives. Appendix Figure A14 shows that results are fairly similar across the implicit and explicit incentive schemes. The loss of sensitivity to level is 39.8% for explicit incentives and 32.1% for implicit incentives. The loss of sensitivity to reliability is 65.1% for explicit incentives and 73% for implicit incentives.

Script analysis. Appendix Figure A15 corroborates the belief patterns with an analysis of information conveyed in the transmitted voice messages. A script analysis highlights that even when respondents are explicitly incentivized to pass on all of the information, more than 50% fail to convey any reliability information.

Summary. Taken together, our two additional incentive manipulations paint a clear picture. Making the incentive to transmit reliability information successively more payoff-relevant and salient by moving from the baseline to the implicit and then the explicit scheme does decrease reliability information loss somewhat, but these (heavy-handed) manipulations have quantitatively moderate effects and substantial differential loss persists. Therefore, perceptions about the relative importance of transmitting level and reliability information are unlikely to explain our main finding of differential information loss. Moreover, differential information loss does not seem to be an artifact of a specific incentive scheme, and appears to reflect a more general mechanism.

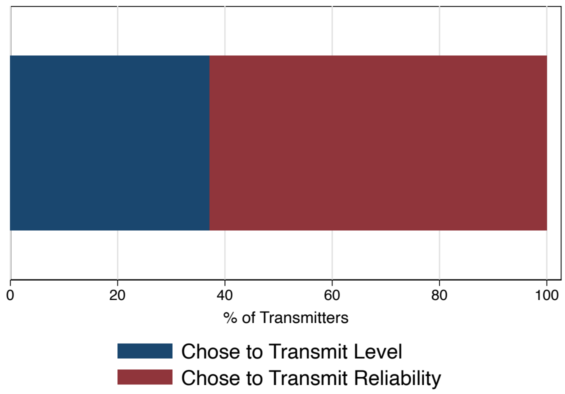
4.2 Perceived Costs of Transmitting Level and Reliability

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 as more costly or difficult to transmit before doing so; alternatively, they might try to transmit reliability information but then *experience* it as being very difficult to properly transmit. Our analysis in Section 3.3, 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.

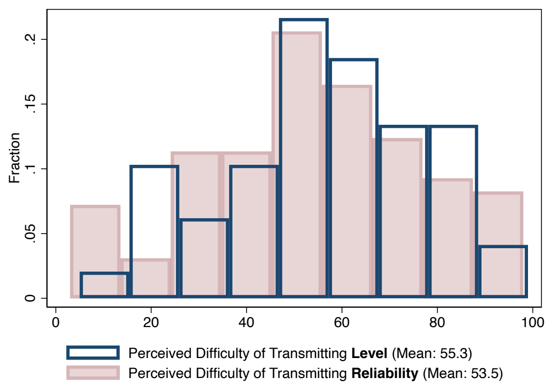
As a direct test of the initially *anticipated* costs of transmitting level versus reliability information, we study whether transmitters prefer to be paid for their performance in transmitting information about (i) the level of the original prediction or (ii) the reliability of the original prediction. By "performance," we mean an external evaluator's assessment of how well the transmitter's message passed on the level or reliability, 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 results described below show that, if anything, people believe reliability information is *easier* to pass on, and this does not change once they experience the transmission task.

Design. This 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.

a) Transmitters' Choices of Incentives



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



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

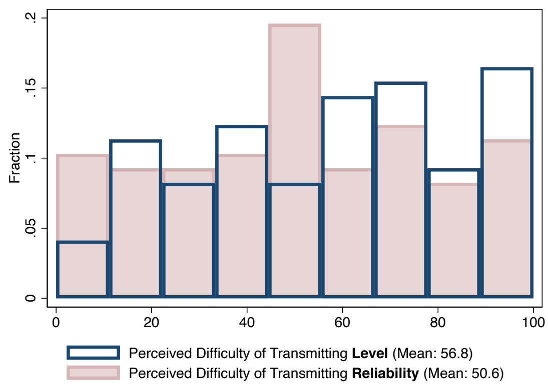


Figure 6: This figure presents data from the Choice of Incentive Experiment. Panel (a) shows the share of people choosing to be incentivized based on their transmission of level information versus reliability information. Panel (b) shows the distribution of respondents' beliefs about the difficulty of 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.

Logistics. We conducted this experiment with 97 respondents on Prolific in November 2023. This collection was also pre-registered at `https://www.socialscienceregistry.org/trials/12119`.

Results. Panel (a) of Figure 6 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. This additional experiment provides strong evidence that differences in the anticipated or experienced costs of transmitting level versus reliability information cannot account for the differential information loss.

Result 3. *Mechanism experiments suggest that differential transmission loss of reliability information 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.

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 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.

4.3.1 Memory Constraints in Cued Recall

We analyze the beliefs of transmitters in the baseline experiment, measured after they complete their recordings.[25] 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).[26]

Appendix Figure A10 demonstrates that there is virtually no memory loss among transmitters about the original message's reliability: several minutes after hearing the original recording and after performing the cognitively demanding task of recording their own voice message in the interim, transmitters are *just as sensitive to variations in reliability* as listeners whose beliefs are elicited immediately after hearing the original recordings. If anything, there is more memory 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. This hence leaves open the possibility that reliability information simply does not come to transmitters' minds when

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

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

recording their voice message.

4.3.2 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, during the recording process, about the level-reliability distinction, which effectively turns the free recall setup of the recording into a cued recall situation.

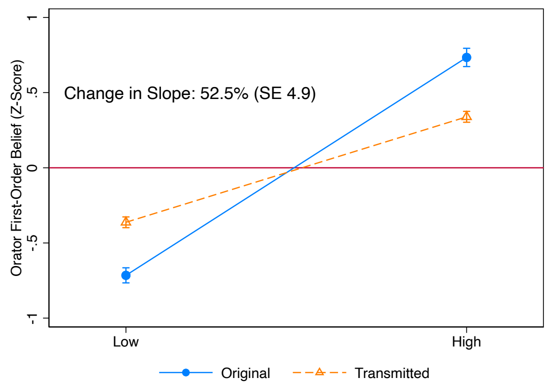
Design. The design closely follows the *explicit incentives* treatment described in Section 4.1.2, in which transmitters were explicitly incentivized to transmit both the level and reliability of the original message's prediction. It adds three features to increase the salience of the level-reliability 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. This collection was also pre-registered at
`https://www.socialscienceregistry.org/trials/12119`.

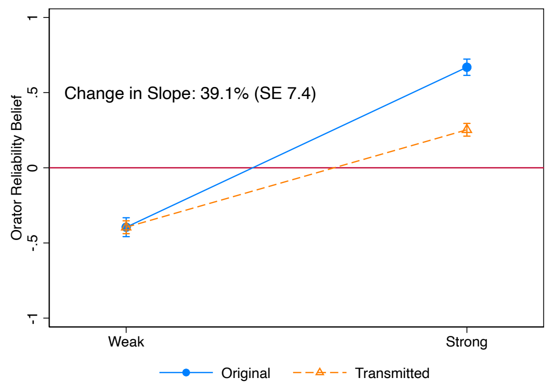
Results. Figure 7 visualizes the results of the *high salience* experiment. In line with our hypothesis that reliability information comes to mind more easily under the added cues, we document a strong reduction in reliability information loss together with a complete disappearance of differential information loss. Analyzing message beliefs, Panels (a) and (b) of Figure 7 show that reliability information loss decreases to 39% (from 65.1% in the *explicit incentives* treatment), while level information loss increases slightly from 39.8% to 53%, possibly reflecting crowding-out of level information as transmitters talk more about reliability. Interestingly, Panel (b) 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 (c) shows that transmitters indeed 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 slightly, from 90-95% to 80-90%.

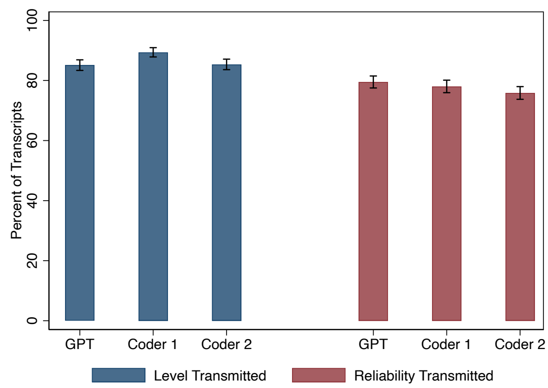
a) Salience Experiment: Level Info Loss



b) Salience Experiment: Reliability Info Loss



c) Fraction of Scripts Containing Level/Reliability Statements



d) Belief Updates About the Economic State

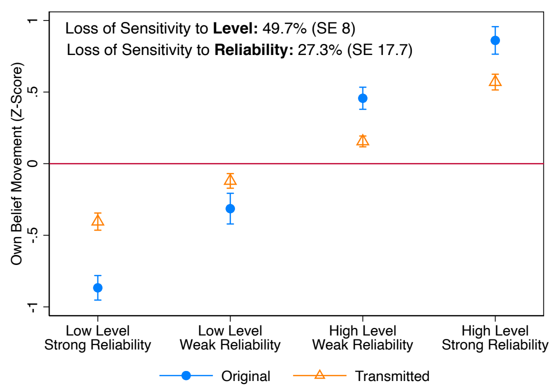


Figure 7: This figure presents data from the High Salience Experiment. Panels (a) and (b) replicate Figure 1, showing beliefs about the original message's level and reliability, separately by whether level is low/high, reliability is weak/strong, and the listener is hearing an original or transmitted message. Panel (c) replicates Panel (a) of Figure 3, showing which fraction of transmitted scripts contain statements about the level or reliability of the original prediction. Panel (d) replicates Panel (c) of 2, showing listeners' average belief updates about the economic variable. Bars are standard error bars. $N = 1,010$ listeners and 244 transmitters.

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 3.2). 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.

The resulting pattern of state belief updates illustrates a tradeoff arising from our salience intervention: on the one hand, the intervention restores the gap in the influence of weak- and strong-reliability messages relative to our baseline results. Put differently, transmission no longer renders weak- and strong-reliability messages similarly influential. On the other hand, the intervention further weakens absolute belief updates from transmitted messages, both because it slightly exacerbates level information loss and because it dilutes the partially offsetting force of reliability information loss. 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. Of course, such a slowdown may be desirable if this convergence would otherwise happen on the basis of unreliable information.

Result 4. *Reliability information is lost in transmission largely because it fails to come to mind during transmission. We show that differential information loss can be eliminated through simple interventions that remind people at the time of transmission to also consider the reliability of information. These interventions have the side effect of exacerbating the attenuation of absolute belief movements through transmission.*

Potential origins of differential memory failures. Our final mechanism finding raises the question of *why* reliability information is less likely to come to mind than level information, absent explicit reminders. One possibility is that the differential ease of retrieval is related to the hierarchical relationship between level and reliability information. Level information, or the *signal realization*, comprises substantive statements or reasoning about the subject matter, whereas reliability is meta-information *tagged to* the level information. There are various potential implications of this hierarchical relationship. First, level information can be interpreted and used to update beliefs even in the absence of specific reliability information (the learner can simply use their default or prior reliability). The reverse is not true: reliability information is not interpretable—in the precise sense that it cannot shift beliefs—in the absence of an accompanying signal realization. Second, this hierarchical relationship may cause the corresponding memory associations to be directed (as is often shown in memory research, e.g. Kahana (2012)), so that retrieval of level information serves as a memory cue for reliability information, but not vice versa. Either of these facts could cause level information to naturally come to mind first, and reliability information may fail to subsequently come to mind due to processing constraints.

While we are not aware of existing evidence from memory research that directly speaks to the pattern we have documented, our findings seem at least compatible with the previously documented result that individuals have difficulty accurately attributing a particular piece of memory to a source, also referred to as "source monitoring errors in memory" (Johnson et al., 1993; Johnson, 1997). The reliability of a message often has a lot to do with the message's source (in our naturalistic reliability manipulations, the credentials and characteristics of the speaker are a key reliability indicator), so difficulties retrieving a source could lead to difficulties retrieving reliability.

5 Discussion

We now discuss issues related to the interpretation of our experimental findings, their external validity, and additional implications.

Transmission incentives. Our baseline incentive schemes attempt to mimic the circumstances transmitters face in many real-world settings, where the incentive to transmit reliability information is embedded implicitly in a broader goal of conveying all relevant information to the listener. Incentives to faithfully transmit all relevant information appear in many contexts. For example, information diffusion is critical to the productivity of workers and the functioning of organisations (Sandvik et al., 2020). When people work in teams or organisations to achieve a goal, e.g., developing a new product, predicting future revenues or meeting a sales target, they usually have incentives for faithful information transmission. Problems often arise from transmission loss in these contexts, for example in shift-to-shift handoffs in hospitals (e.g. Riesenberg, 2012). Another example concerns the supply of advice: when people give advice to friends, their incentive is to only convey reliable advice and to flag unreliable pieces of information as such. Similar remarks apply to the provision of financial advice to clients or strategic advice within companies.

At the same time, we acknowledge that people face a variety of other transmission incentives in practice. First, people often pass on information with the goal of *entertaining*, rather than simply informing the listener. Such incentives are likely to, if anything, exacerbate differential information loss: caveats and meta-commentary about reliability are boring. Second, people often have incentives to *persuade* others, which may conflict with the full transmission of information. This also incentivizes the dropping of caveats or nuance, but may encourage the speaker to play up indicators of strong reliability. 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.

Our paper abstracts from entertainment or persuasion incentives to focus on the distortions emerging in attempts to faithfully transmit information. But beyond the incentives we focus on, the mechanism supported by our findings—that reliability information fails to come to mind—likely reflects a general cognitive mechanism that plausibly extends across incentive schemes and circumstances. That only targeted reminders eliminate differential loss suggests that differential loss likely emerges in many real-world contexts, because this form of specific cue is typically absent. A widespread argument holds that as information gets passed on and stories are told and re-told, *nuance* is lost. This "loss of nuance" hypothesis is frequently discussed in the media (e.g., Herrman, 2018) and research (e.g., Hirshleifer, 2020), but concrete evidence remains scant. The reliability loss we document may provide a concrete instance of this broader hypothesis.

Related field evidence on information sharing. Pennycook et al. (2021) conduct survey and field experiments with Facebook and Twitter focusing on users' decisions to share news stories. The authors find that social media users are equally likely to share true and false headlines they are exposed to. However, *when prompted,* users are able to distinguish between true and false headlines, and claim that it is important to only share true headlines. When their *attention is nudged towards a source's accuracy*, users share more accurate stories. Although these findings come from a very different setting and focus on the extensive margin of information transmission, they are remarkably consistent with our own: we show that, *when prompted,* people remember whether a forecast was reliable or unreliable and claim that passing on reliability information is important. We also find that when people's *attention is nudged towards reliability information,* they pass on reliability information at higher rates. This suggests that the phenomenon and mechanism we have identified is likely to apply in the field—in more realistic information-sharing environments and under real-world incentive schemes.

Moreover, the finding that people are equally likely to share true and false headlines (despite being able to distinguish them when prompted) bolsters the importance of one of our key implications: that transmission causes reliable and unreliable messages to converge in influence on downstream beliefs. One might object that this is an artifact of the fact that we *force* people to pass on unreliable messages; perhaps, in practice, when people encounter an unreliable forecast they simply do not pass it on, rather than passing it on but dropping information about its reliability. The results from Pennycook et al. (2021) suggest that people *do indeed* pass on unreliable information without mentioning its reliability.

Differential information loss in the real world. How common is the omission of reliability information in the real world? There are many examples of contexts where differential information loss seems, anecdotally, to be very important. One example is science journalism. Coverage of scientific results tends to omit caveats or expressions of uncertainty written in the original paper (Jensen, 2008). A second example is the oft-observed gradual disappearance of caveats as a report moves through an organization. A famous and consequential instance is the 2002 National Intelligence Estimate on WMDs in Iraq, which was widely interpreted as a justification for the Iraq War. A RAND retrospective from 2014 concluded that when the report was transformed into an executive summary for higher-ups, "[it] contained several qualifiers that were dropped" and "as the draft NIE went up the intelligence chain of command, the conclusions were treated increasingly definitively" (Gompert et al., 2014). As a third example, doctors often complain that clinical guideline committees often compile lists of recommendations without mentioning the (highly variable) quality of the evidence underlying each recommendation (Spellberg et al., 2022).

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 as accurately as possible through voice messages. Other participants listen to either the original recorded voice messages or transmitted versions of those messages, 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.

We show two main implications of the findings we document. First, transmission strongly increases the relative influence of weak-reliability messages on beliefs. Second, transmission shrinks average belief updates, meaning belief polarization can be sustained even in the presence of new information. We experimentally document a trade-off between these two effects: interventions to restore the distinction between weak- and strong-reliability messages can further weaken average belief updates.

References

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, and Johannes Wohlfart, "Narratives about the Macroeconomy," 2022.

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," *arXiv preprint arXiv:2109.09871*, 2024.

Ba, Cuimin, J Aislinn Bohren, and Alex Imas, "Over-and underreaction to information," *Available at SSRN*, 2022.

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 2023.

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.

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

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

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

__ , 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.

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.

Braghieri, Luca, "Biased Decoding and the Foundations of Communication," *Available at SSRN 4366492*, 2023.

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.

__ , Georgy Egorov, Ingar Haaland, Aakaash Rao, and Christopher Roth, "Justifying Dissent," *The Quarterly Journal of Economics*, 2023, *138* (3), 1403–1451.

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.

Chandrasekhar, Arun G, Esther Duflo, Michael Kremer, João F Pugliese, Jonathan Robinson, and Frank Schilbach, "Blue Spoons: Sparking Communication About Appropriate Technology Use," Technical Report, National Bureau of Economic Research 2022.

Conlon, John J, Malavika Mani, Gautam Rao, Matthew W Ridley, and Frank Schilbach, "Not Learning from Others," Technical Report, National Bureau of Economic Research 2022.

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

— **and Cassidy Shubatt**, "Quantifying lottery choice complexity," Technical Report, National Bureau of Economic Research 2023.

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

— **,** — **, and Ryan Oprea**, "Complexity and Time," 2023.

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.

Fehr, Dietmar, Johanna Mollerstrom, and Ricardo Perez-Truglia, "Listen to her: Gender differences in information diffusion within the household," Technical Report, National Bureau of Economic Research 2022.

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

Gabaix, Xavier, "Behavioral inattention," in "Handbook of behavioral economics: Applications and foundations 1," Vol. 2, Elsevier, 2019, pp. 261–343.

Galeotti, Andrea, Sanjeev Goyal, Matthew O Jackson, Fernando Vega-Redondo, and Leeat Yariv, "Network Games," *The Review of Economic Studies*, 2010, *77* (1), 218–244.

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.

Gorodnichenko, Yuriy, Tho Pham, and Oleksandr Talavera, "The voice of monetary policy," *American Economic Review*, 2023, *113* (2), 548–584.

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.

Granovetter, Mark S, "The Strength of Weak Ties," *American journal of sociology*, 1973, *78* (6), 1360–1380.

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.

Herrman, John, "Science in the Media: A Game of Telephone," 2018.

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

Jackson, Matthew O and Leeat Yariv, "Diffusion of behavior and equilibrium properties in network games," *American Economic Review*, 2007, *97* (2), 92–98.

Jensen, Jakob D, "Scientific uncertainty in news coverage of cancer research: Effects of hedging on scientists' and journalists' credibility," *Human communication research*, 2008, *34* (3), 347–369.

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.

Kendall, Chad W and Constantin Charles, "Causal Narratives," Technical Report, National Bureau of Economic Research 2022.

Lazarsfeld, Paul F, Bernard Berelson, and Hazel Gaudet, *The people's choice: How the voter makes up his mind in a presidential campaign*, Columbia University Press, 1968.

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.

Oprea, Ryan, "What makes a rule complex?," *American Economic Review*, 2020, *110* (12), 3913–51.

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.

Riesenberg, Lee Ann, "Shift-to-shift handoff research: where do we go from here?," 2012.

Sandvik, Jason J, Richard E Saouma, Nathan T Seegert, and Christopher T Stanton, "Workplace knowledge flows," *The Quarterly Journal of Economics*, 2020, *135* (3), 1635–1680.

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

— , *Narrative economics*, Princeton University Press, 2020.

Spellberg, Brad, Gloria Aggrey, Meghan B Brennan, Brent Footer, Graeme Forrest, Fergus Hamilton, Emi Minejima, Jessica Moore, Jaimo Ahn, Michael Angarone et al., "Use of novel strategies to develop guidelines for management of pyogenic osteomyelitis in adults: a WikiGuidelines group consensus statement," *JAMA network open*, 2022, *5* (5), e2211321–e2211321.

Thaler, Michael, "The supply of motivated beliefs," *arXiv preprint arXiv:2111.06062*, 2021.

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.

Vespa, Emanuel and Georg Weizsäcker, "Do we talk too much?," Technical Report, CRC TRR 190 Rationality and Competition 2023.

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

A A Model of Noisy Transmission

This Appendix briefly lays out a fully structural model giving rise to the reduced-form equations laid out in Section 3.1 under appropriate conditions. It presents the inference and transmission problem as a coherent, global noisy inference problem.

All participants believe that the true state ℓ is drawn from some prior distribution $\ell \sim \mathcal{N}(\ell^d, v)$, where ℓ^d stands for the prior or *default* level of the state and v for the state's variance. The original message provides a noisy signal about the state to the transmitter. This noisy signal, ℓ^o, has a specific reliability r^o:

$$\ell^o = \ell + \varepsilon \quad \text{with} \quad \varepsilon \sim \mathcal{N}(0, e^{-r^o}) \tag{8}$$

The transmitter conveys their noisy signal, but the process of transmission adds additional *level transmission noise,* η. As a result, the receiver of the transmitted message gets a different noisy signal about the level of state:

$$\ell^t = \ell^o + \eta = \ell + \varepsilon + \eta \quad \text{with} \quad \eta \sim \mathcal{N}(0, v_\eta) \tag{9}$$

The listener of the transmitted message does not know the original signal reliability r^o, only that it is drawn from the prior $r^o \sim \mathcal{N}(r^d, v_r)$. Along with the signal about the state, the transmitter sends a signal about the original message's reliability to the receiver, which in turn is subject to *reliability transmission noise,* χ:

$$r^t = r^o + \chi \quad \text{with} \quad \chi \sim \mathcal{N}(0, v_\chi) \tag{10}$$

Under this interpretation, the defaults l^d and r^d from Equations 2 and 3 correspond to prior beliefs about the level and reliability of the original message, respectively. The transmission process adds zero-mean noise to the level and reliability expressed in the original message. The listener to a transmitted message, aware of this introduction of noise, combines their prior with the transmitted signal realization to form a Bayesian posterior about the level and reliability of the original message.

Specifically, we assume that listeners of the transmitted message behave as if they treat the signal extraction problem sequentially, by forming *message beliefs* that serve as the input for their *state belief*. They first infer a posterior estimate for the reliability as:

$$\hat{r} = r^d + \frac{v_r}{v_r + v_\chi}(r^t - r^d) \tag{11}$$

Listeners then infer a posterior estimate for the level information in the original message as:

$$\widehat{\ell^o} = \ell^d + \frac{v + e^{-\hat{r}}}{v + v_\eta + e^{-\hat{r}}}(\ell^t - \ell^d) \tag{12}$$

Given these message beliefs, they form a posterior estimate for the true state as:

$$\hat{\ell} = \ell^d + \frac{v}{v + v_\eta + e^{-\hat{r}}}(\ell^t - \ell^d) \tag{13}$$

These reduced forms represent a modest deviation from full Bayesian inference. In fact, (11) is fully Bayesian. (12) and (13) are only fully Bayesian conditional on r^o and on setting aside non-linearities that have no first-order effect. This approach also assumes that agents do not make cross-inference from the extremity of the level signal, ℓ^t, about the reliability of the original signal r^o.

Importantly, these inference equations straightforwardly map to the reduced-form equations (2) and (3). In particular, reliability attenuation is given $\lambda_r := \frac{v_r}{v_r + v_\chi}$, while level attenuation is given by $\lambda_\ell := \frac{v + e^{-\hat{r}}}{v + v_\eta + e^{-\hat{r}}}$.

Moreover, the model's key comparative statics, presented in Section 3.2, can also be found here. Transmission always attenuates message beliefs towards the prior. Indeed, (11) shows that reliability beliefs are more compressed when reliability transmission noise v_χ increases. Similarly, (12) shows that level message beliefs are more compressed when message transmission noise v_η increases. On the other hand, transmission has a more intricate effect on state belief movement. (13) shows that level transmission noise, which is higher when v_η is higher, always shrinks absolute belief movements. On the other hand, (11) and (13) show that reliability transmission noise, which is higher when v_χ is higher, has an effect on belief movement that depends on the level of reliability r^o: it reduces belief movement when reliability is high, i.e. $r^o > r^d$, but increases belief movement when reliability is low, i.e. $r^o < r^d$.

Under some simplifications, these comparative static statements can be turned into a more precise statement about sample means. Conditional on a signal ℓ^o and a reliability r^o, orators form the belief:

$$\hat{\ell}^o = \ell^d + \frac{v}{v + e^{-r^o}}(\ell^o - \ell^d) \tag{14}$$

In turn, listeners form the average belief:

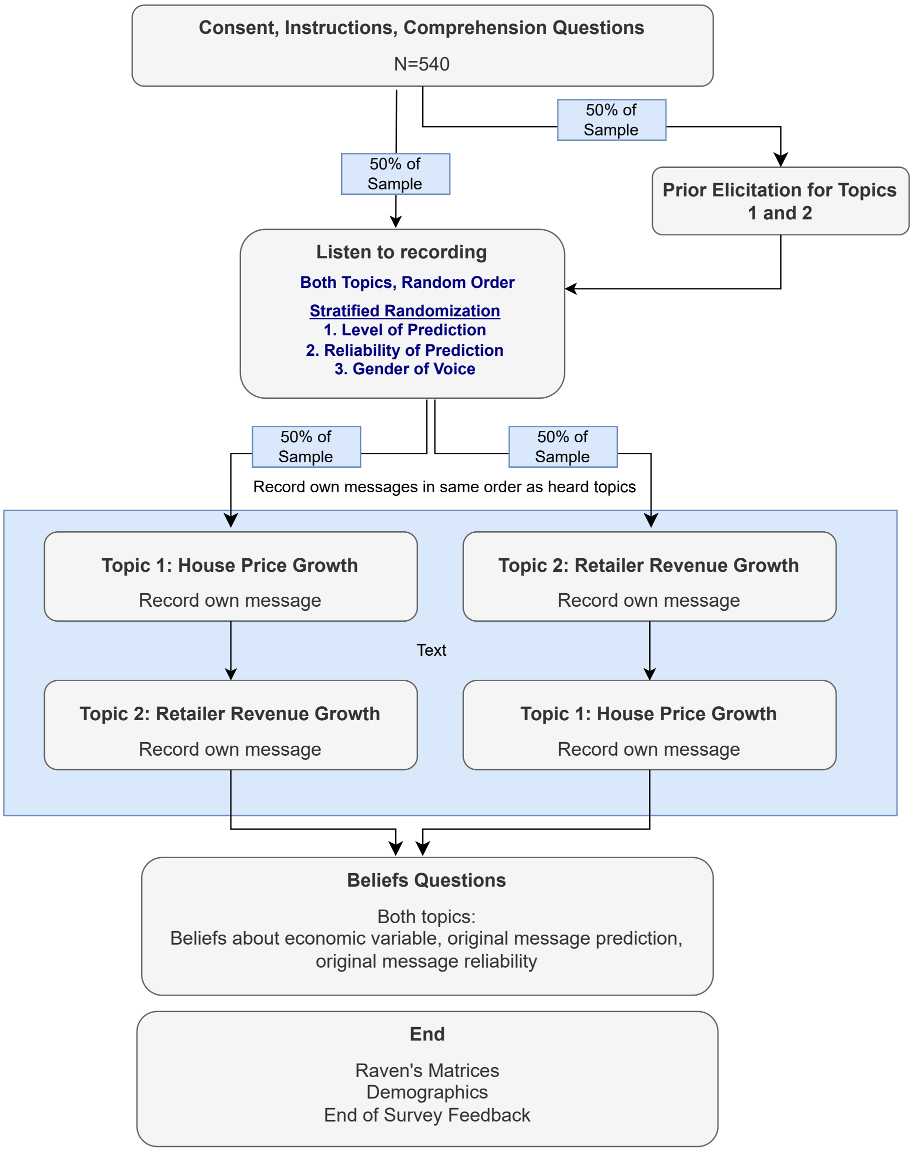
$$E(\hat{\ell}|\ell^o, r^o) = \ell^d + E\left(\frac{v}{v + v_\eta + e^{-\hat{r}}}\right)(\ell^o - \ell^d) \approx \ell^d + \frac{v}{v + v_\eta + e^{-r^d - \frac{v_r}{v_r + v_\chi}(r^o - r^d))}}(\ell^o - \ell^d) \tag{15}$$

The first-order approximation used above applies in the small reliability transmission noise limit ($v_\chi \to 0$). Since we are faced with the expectation of a logit-normal variable, for which there is no analytic formula, little can be said in full generality. Moreover, the sigmoid function of $\hat{r}^t$ will feature concavity or convexity depending on the value of r^o, so that even the second-order effect of transmission noise v_χ cannot be signed without parametrizing the model.

This shows that the contrasting effects of transmission noise on belief movement, applied above, also apply to sample means, given appropriate simplifying assumptions.

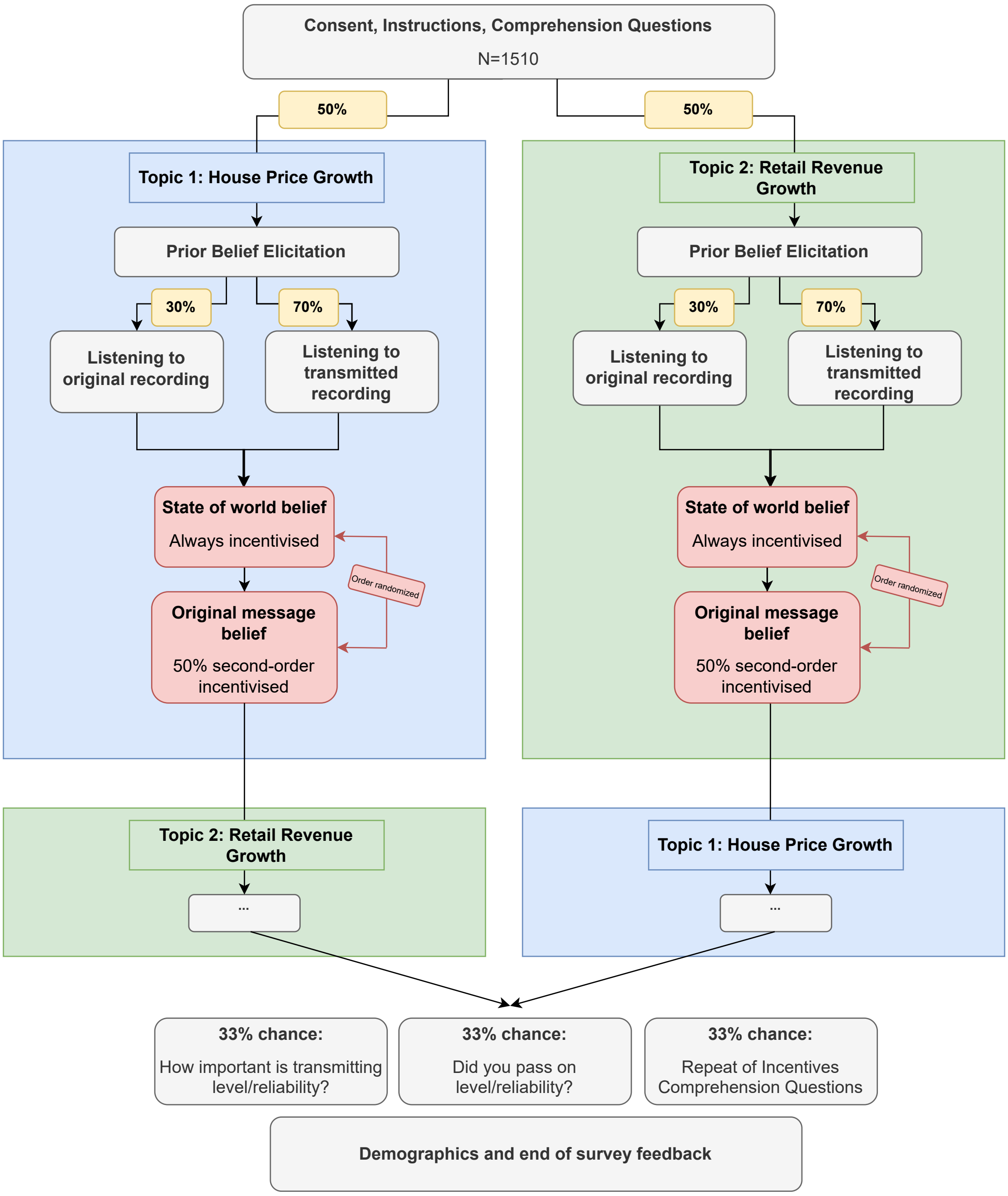
B Additional Exhibits

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

Collection	Sample	Content Treatments	Additional Features/Treatments	Main outcomes
Baseline experiments				
Transmitter Experiment: **Belief Movement Incentives**	Prolific (540 respondents)	High level, high reliab.; High level, low reliab.; Low level, high reliab.; Low level, low reliab.	None	Speech recordings, beliefs about originator level prediction and reliab..
Listener Experiment: **Belief Movement Incentives**	Prolific (1,510 respondents)	High level, high reliab.; High level, low reliab.; Low level, high reliab.; Low level, low reliab.	Original versus transmitted recording	Own beliefs about state, beliefs about originator level prediction and reliab..
Robustness experiment				
Transmitter Experiment with quantitative information: **Belief Movement Incentives**	Prolific (834 respondents)	High level, high reliab.; High level, low reliab.; Low level, high reliab.; Low level, low reliab.	Original scripts contain quantitative information about both the level and reliability.	Speech recordings, beliefs about originator level prediction and reliab..
Listener Experiment with quantitative information: **Belief Movement Incentives**	Prolific (181 respondents)	High level, high reliab.; High level, low reliab.; Low level, high reliab.; Low level, low reliab.	Original versus transmitted recording	Own beliefs about state, beliefs about originator level prediction and reliab.
Mechanism experiments				
Transmitter Experiment: **Content Transmission Incentives**	Prolific (501 respondents)	High level, high reliab.; High level, low reliab.; Low level, high reliab.; Low level, low reliab.	Explicit versus implicit incentives for transmission of reliab. information	Speech recordings, own beliefs about state, beliefs about originator.
Listener Experiment: **Content Transmission Incentives**	Prolific (1,509 respondents)	High level, high reliab.; High level, low reliab.; Low level, high reliab.; Low level, low reliab.	Original versus transmitted recording	Own beliefs about state, beliefs about originator level prediction and reliab..
Transmitter Experiment: **Choice of Incentives**	Prolific (97 respondents)	High level, high reliab.; High level, low reliab.; Low level, high reliab.; Low level, low reliab.	Respondents choose which type of information they need to transmit	Choice of incentives, perceived difficulty of transmitting level and reliab. information.
Transmitter Experiment: **High Salience**	Prolific (244 respondents)	High level, high reliab.; High level, low reliab.; Low level, high reliab.; Low level, low reliab.	Salient reminders of incentives to transmit reliab.	Speech recordings, beliefs about originator level prediction and reliab..
Listener Experiment: **High Salience**	Prolific (1,010 respondents)	High level, high reliab.; High level, low reliab.; Low level, high reliab.; Low level, low reliab.	Original versus transmitted recording	Own beliefs about state, Beliefs about originator level prediction and reliab..

This Table provides an overview of the different data collections. The sample sizes refer to the final sample of respondents that completed the survey and satisfied the pre-specified inclusion criteria for each of our collections. All of the data collections were pre-registered on the AEA RCT registry: `https://www.socialscienceregistry.org/trials/12119`.

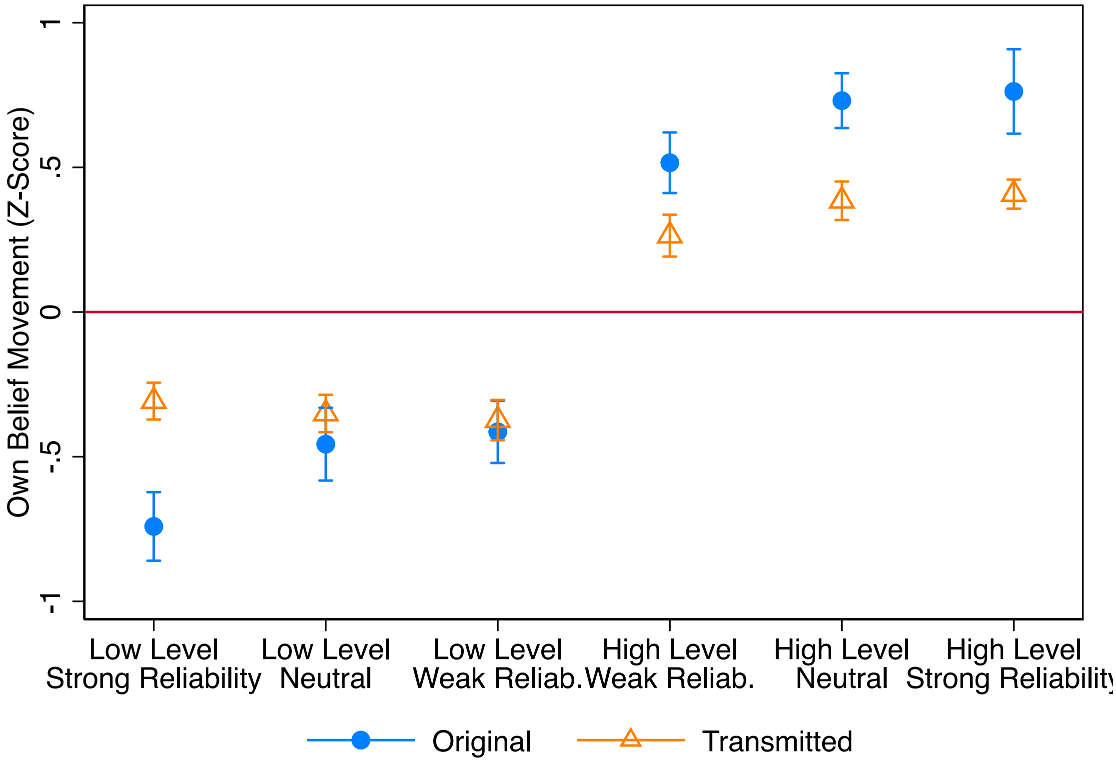
Appendix Table A2: Example Transmitted Messages

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 US retail company and specifically this US 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 US 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.

B.1 Additional Figures

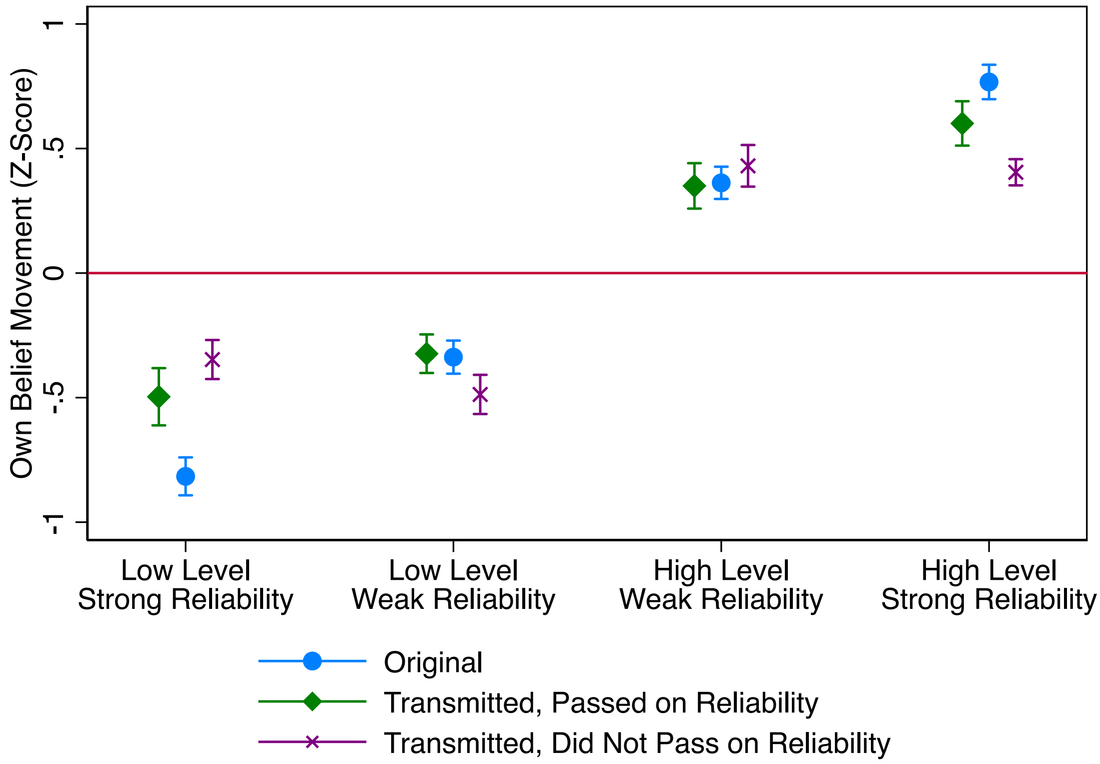
B.1.1 Baseline Experiment: Belief Movement Incentives

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



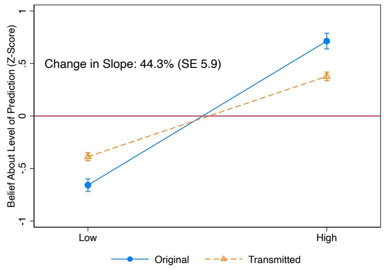
Appendix Figure A3: This figure presents data from our baseline experiment (Belief Movement Incentives). It is an alternative version of Panel (c) of Figure 2. It shows the average belief updates of listeners, restricting to the Modular reliability manipulation, which has a weak-reliability, strong-reliability, 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

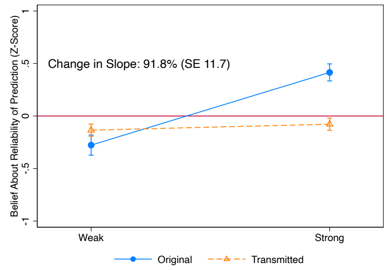


Appendix Figure A4: This figure presents data from our baseline experiment (Belief Movement Incentives). It is an alternative version of Panel (c) of Figure 2. It shows the average belief updates of listeners, splitting listeners who hear transmitted recordings by whether the transmitted recording is 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).

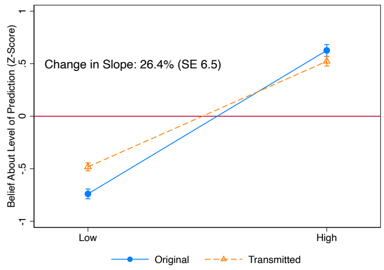
a) Level Info Loss: Modular Manipulation



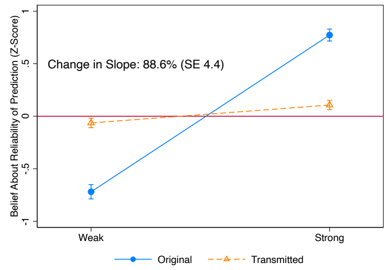
b) Reliability Info Loss: Modular Manipulation



c) Level Info Loss: Naturalistic Manipulation

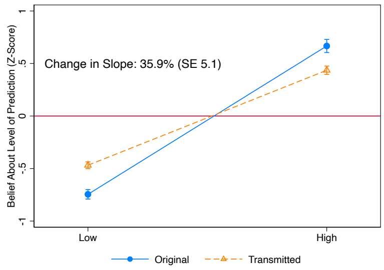


d) Reliability Info Loss: Naturalistic Manipulation

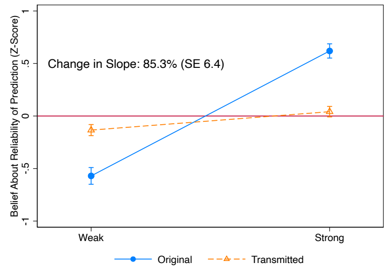


Appendix Figure A5: This figure presents data from our baseline experiment (Belief Movement Incentives). It replicates Figure 1, showing beliefs about the level and reliability of the original prediction, separately by respondents in our *modular* versus *naturalistic* reliability manipulations, described in Section 2.

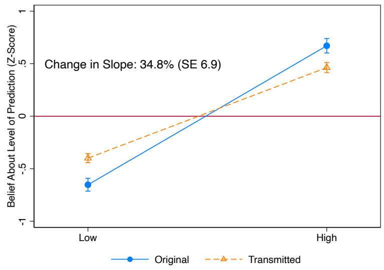
a) Level Info Loss: No Incentives



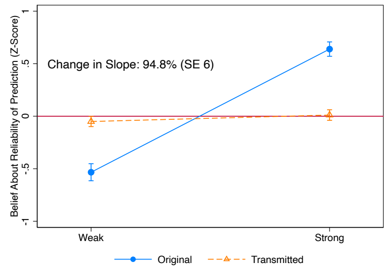
b) Reliability Info Loss: No Incentives



c) Level Info Loss: Second-Order Incentives

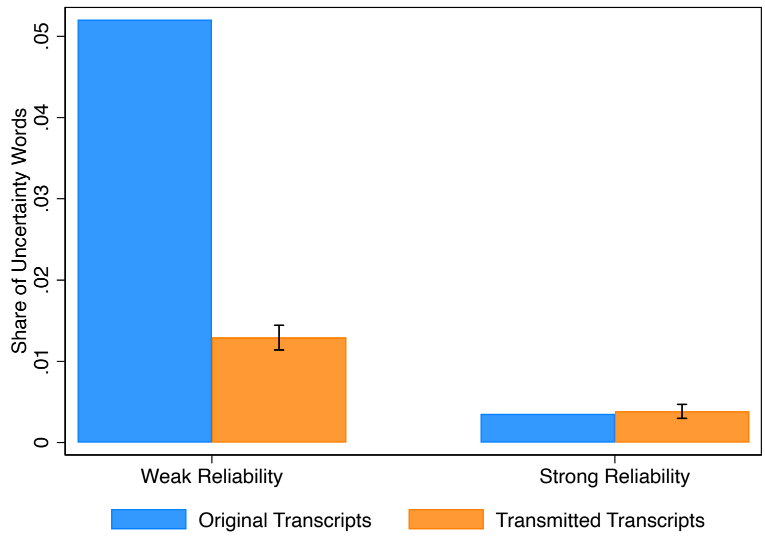


d) Reliability Info Loss: Second-Order Incentives

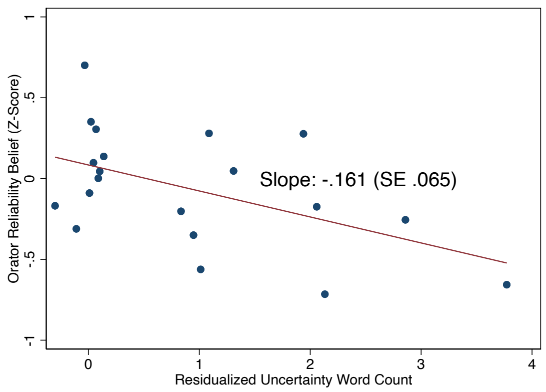


Appendix Figure A6: This figure presents data from our baseline experiment (Belief Movement Incentives). It replicates Figure 1, showing beliefs about the level and reliability of the original prediction, separately by respondents who are asked these questions directly and not incentives, compared to 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, as described in Section 2.

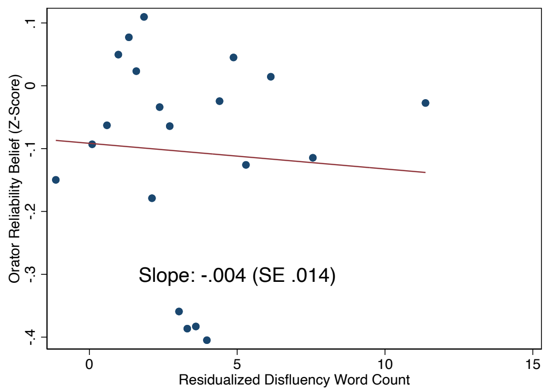
a) Uncertainty Words are Lost in Transmission



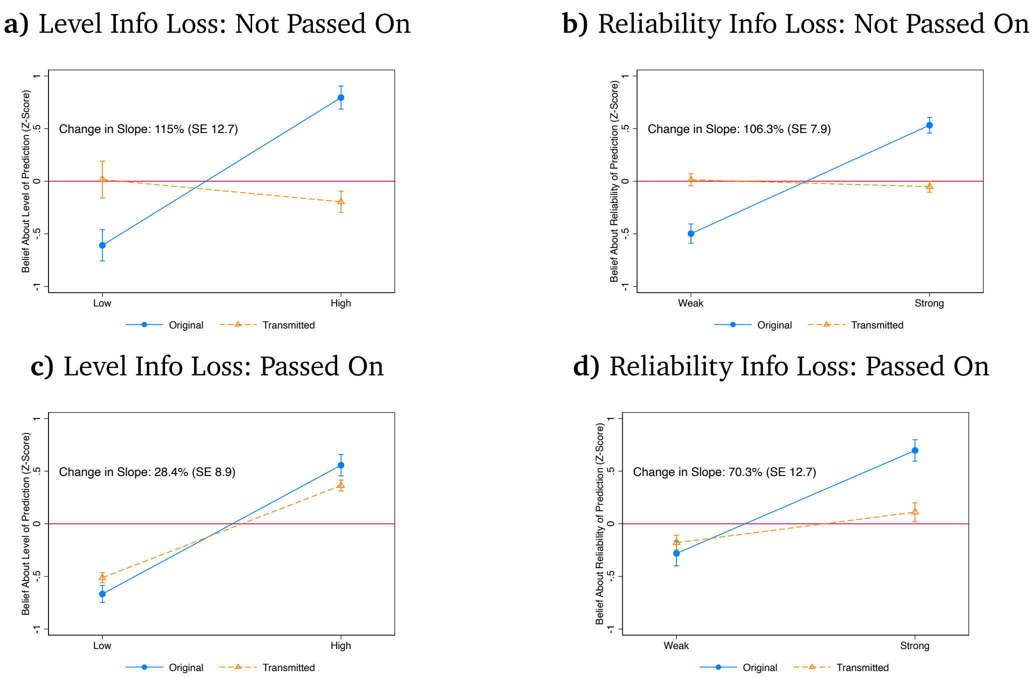
b) Uncertainty Words in Transmitted Scripts Affect Listener Beliefs



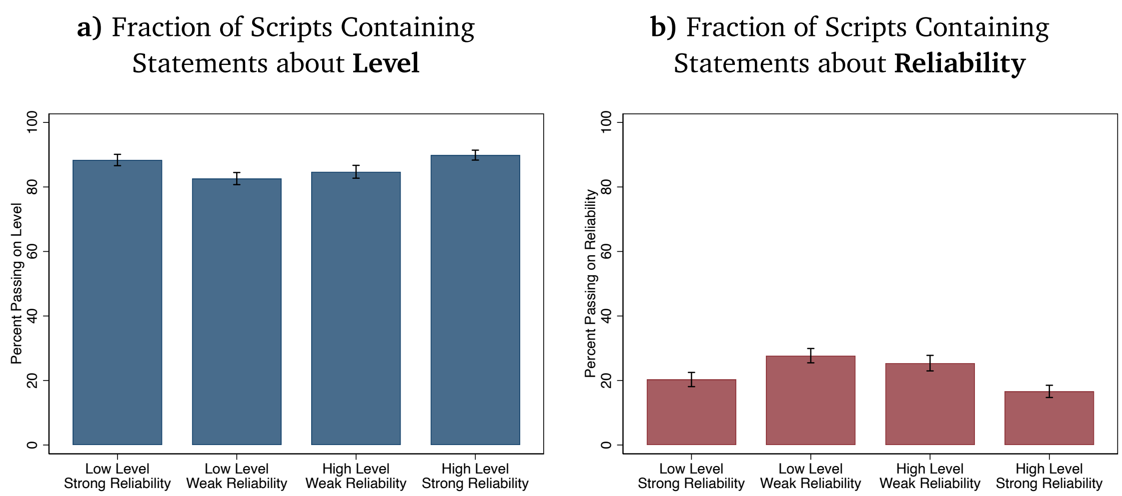
c) Disfluencies in Transmitted Scripts Do Not Affect Listener Beliefs



Appendix Figure A7: 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. Panel (c) does the same for disfluencies, automatically counted by GPT-4 and encompassing various kinds of disruptions in the flow of the original transcript.

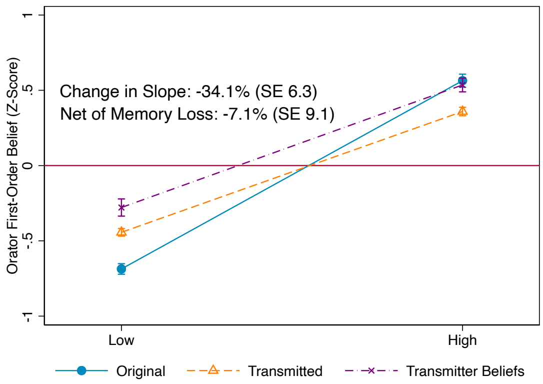


Appendix Figure A8: This figure presents data from our baseline experiment (Belief Movement Incentives). It replicates Figure 1, showing beliefs about the level and reliability of the original prediction. Panels (a) and (b) restrict to recordings that both human coders and GPT-4 unanimously agree do *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.

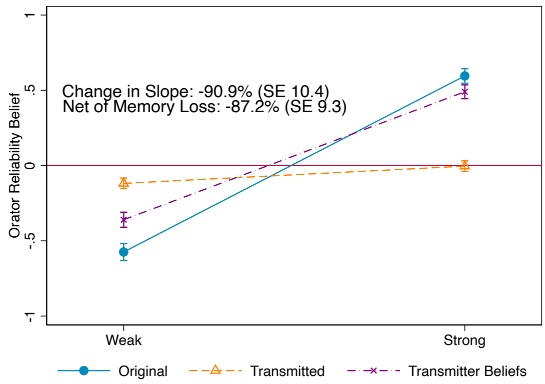


Appendix Figure A9: This figure disaggregates Figure 3 by the four conditions in our level × reliability manipulation. It shows the percent of transmitted messages that are unanimously classified by our coders as containing statements about level or reliability.

a) Transmission of **Level**



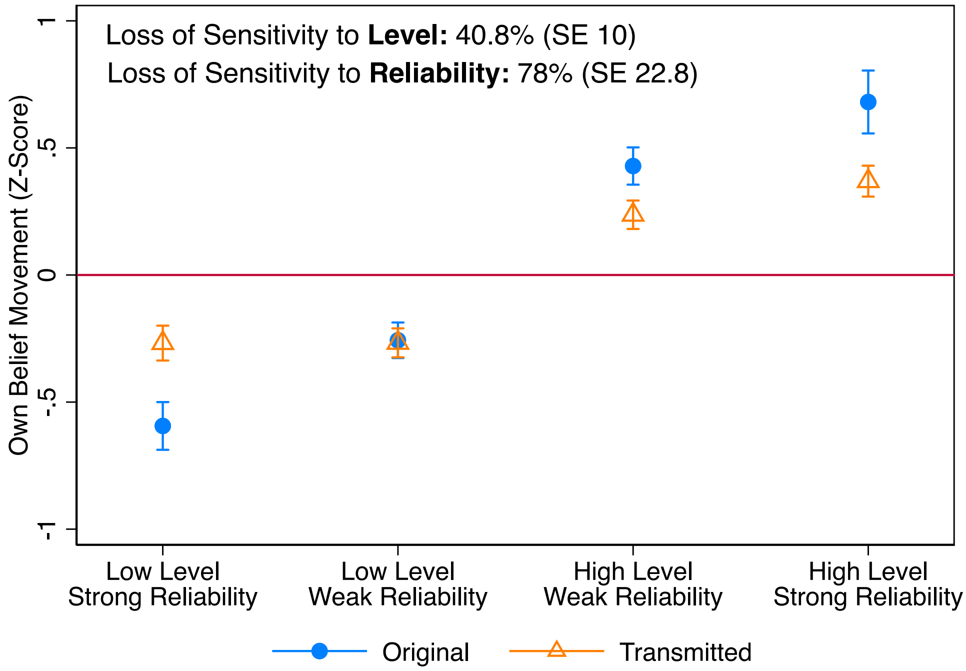
b) Transmission of **Reliability**



Appendix Figure A10: This figure presents data from our baseline experiment (Belief Movement Incentives). It replicates Figure 1 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.

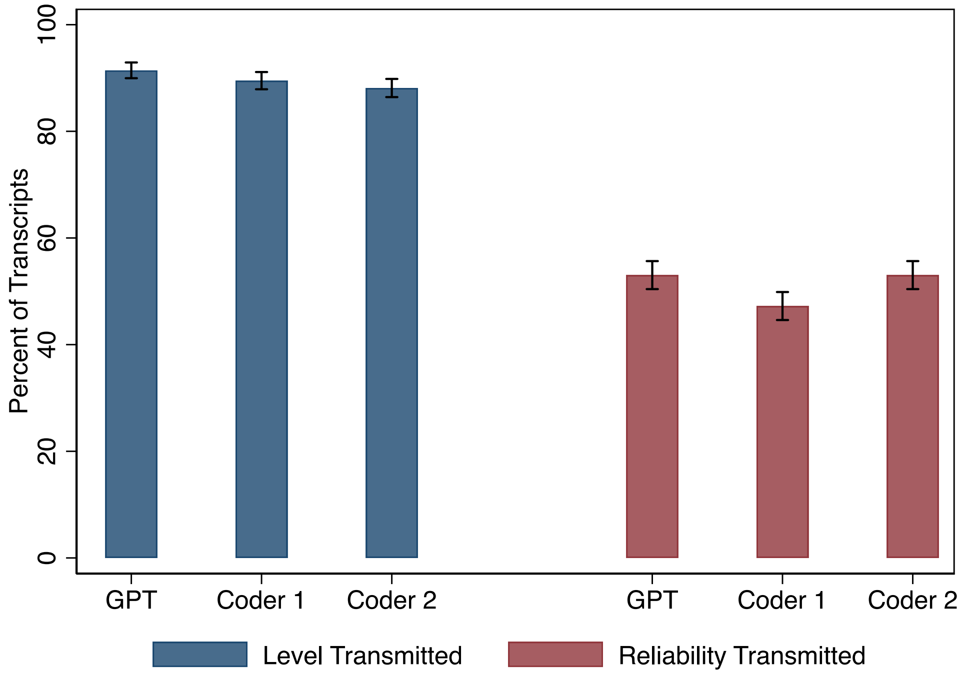
B.1.2 Robustness Experiment: Quantitative Communication

a) Belief Movements About the Economic Variable (Quantitative Scripts)



Appendix Figure A11: This figure presents data from our Quantitative Scripts experiment. It is an alternative version of Panel (c) of Figure 2. It shows the average belief updates of listeners, by quadrant of our level/reliability manipulation and whether they listened to a transmitted recording.

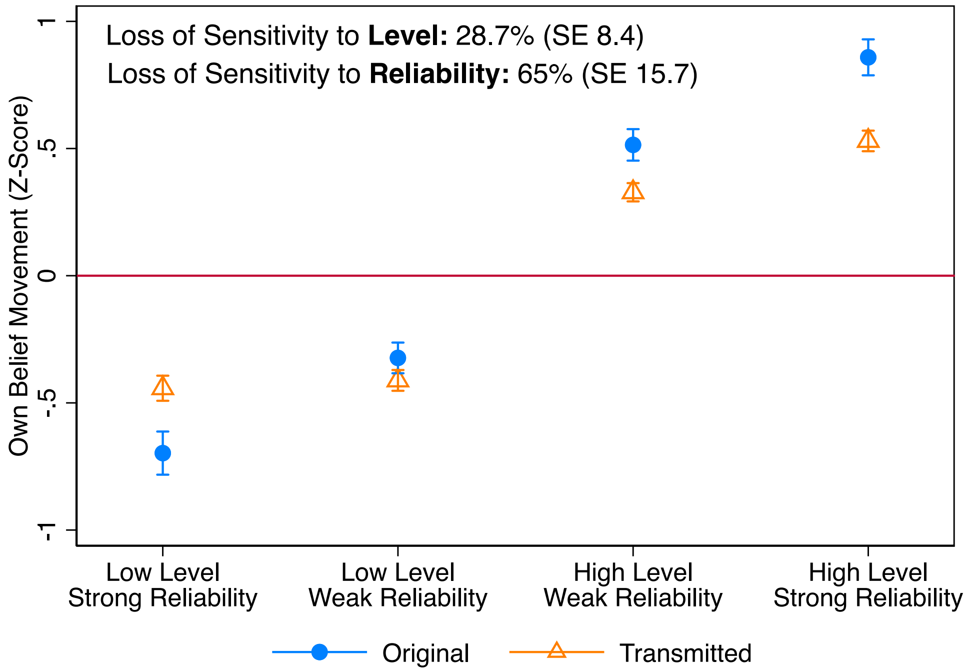
a) Fraction of Scripts Containing Level/Reliability Statements (Quantitative Scripts)



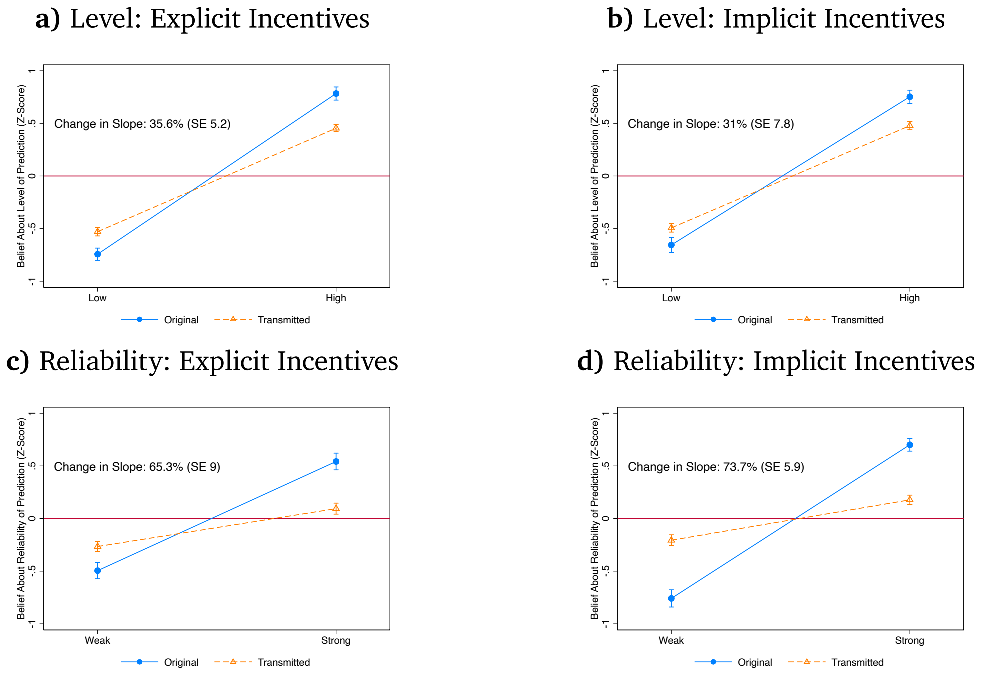
Appendix Figure A12: This figure replicates Figure 3 Panel (a) for our Quantitative Scripts Experiment. It shows the fraction of transmitted messages classified by GPT-4 and our two human coders as containing statements about the level or reliability of the original forecast.

B.1.3 Supplementary Experiment: Content Transmission Incentives

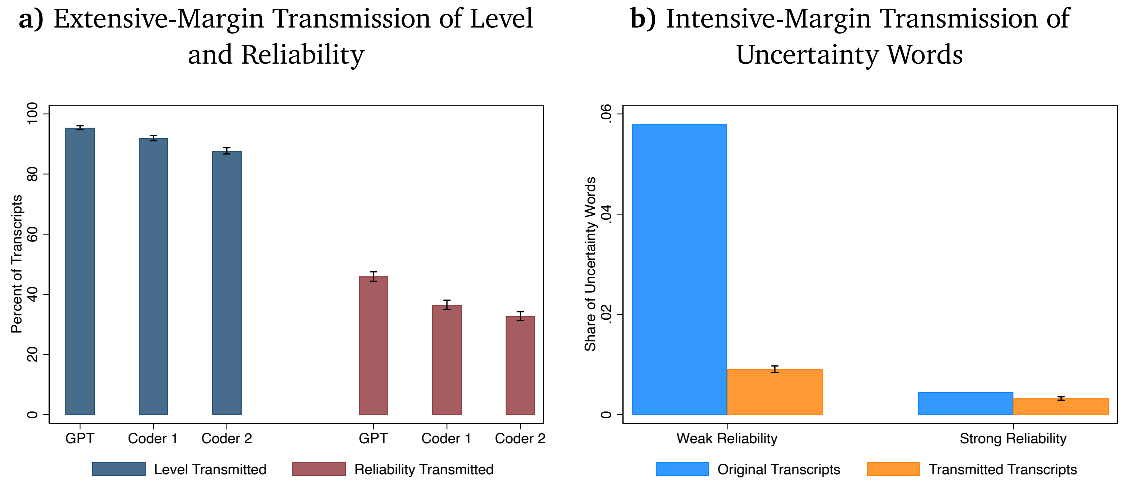
a) Belief Movements About the Economic Variable (Content Transmission Incentives)



Appendix Figure A13: This figure presents data on belief movement about the true state of the world from the Content Transmission Incentive Experiments. Panel (a) shows belief movement about the true state of the world in response to original and transmitted recordings across the four different main recording conditions. Panel (b) shows the transmission of information about the level, pooling across the weak and strong reliability conditions. Panel (c) displays the transmission of reliability information about the level, pooling across the low and high level conditions. Error bars represent 1 SE in either direction.



Appendix Figure A14: This figure replicates Figure 1 in the Content Transmission Incentives data, separately by respondents randomized into the explicit and implicit transmission incentives.



Appendix Figure A15: This figure presents data from the Content Transmission Incentives focusing on the scripts of transmitted recordings. Panel (a) shows data on the fraction of transcripts that convey any information about level and reliability information, using the same GPT-4 and human coder methods. Panel (b) shows data on the share of uncertainty words in original versus transmitted scripts.

B.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

B.3 Regression Tables

B.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 Belief Movement Incentive experiment. 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 Belief Movement Incentives Experiment. 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 Belief Movement Incentives Experiment. 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.3.2 Quantitative Scripts

Appendix Table A7: 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	-0.596***	-0.316**	-0.729***	-0.536***	-0.680***
	(0.125)	(0.150)	(0.051)	(0.142)	(0.098)
Trans. × Low Level × Strong Reliability	0.328**	-0.004	0.499***	0.322*	0.344**
	(0.143)	(0.167)	(0.116)	(0.168)	(0.143)
Low Level × Weak Reliability	-0.256***	-0.398***	-0.111*	-0.138*	-0.409***
	(0.090)	(0.086)	(0.061)	(0.079)	(0.109)
Trans. × Low Level × Weak Reliability	0.000	0.087	-0.087	-0.159	0.201
	(0.109)	(0.124)	(0.108)	(0.108)	(0.151)
High Level × Weak Reliability	0.431***	0.398***	0.468***	0.412***	0.464***
	(0.073)	(0.113)	(0.069)	(0.128)	(0.096)
Trans. × High Level × Weak Reliability	-0.198**	-0.152	-0.250***	-0.210	-0.181
	(0.090)	(0.138)	(0.093)	(0.140)	(0.140)
High Level × Strong Reliability	0.684***	0.618***	0.765***	0.704***	0.660***
	(0.049)	(0.036)	(0.072)	(0.079)	(0.082)
Trans × High Level × Strong Reliability	-0.319***	-0.259**	-0.396***	-0.304***	-0.347**
	(0.086)	(0.119)	(0.112)	(0.110)	(0.155)
Nb. obs	1,253	615	638	718	535

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

Note: This table presents data from the baseline Quantitative Scripts experiment. 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 A8: 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	0.698***	0.696***	0.712***	0.741***	0.638***
	(0.095)	(0.132)	(0.128)	(0.108)	(0.144)
High Level × Transmitted	-0.097	-0.016	-0.183	-0.068	-0.129
	(0.118)	(0.163)	(0.163)	(0.138)	(0.180)
Transmitted	0.066	0.026	0.106	0.059	0.067
	(0.097)	(0.129)	(0.137)	(0.116)	(0.150)
Constant	-0.364***	-0.392***	-0.345***	-0.413***	-0.298**
	(0.083)	(0.109)	(0.117)	(0.097)	(0.128)
Nb. obs	1,253	615	638	719	534

Standard errors in parentheses

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

Note: This table presents data from the Quantitative Scripts experiment. 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 A9: 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	1.511***	1.422***	1.603***	1.531***	1.478***
	(0.100)	(0.176)	(0.090)	(0.116)	(0.133)
Strong Reliability × Transmitted	-1.342***	-1.317***	-1.368***	-1.193***	-1.536***
	(0.124)	(0.207)	(0.137)	(0.142)	(0.181)
Transmitted	0.562***	0.511***	0.611***	0.609***	0.494***
	(0.080)	(0.121)	(0.081)	(0.097)	(0.129)
Constant	-0.673***	-0.583***	-0.765***	-0.713***	-0.613***
	(0.064)	(0.099)	(0.045)	(0.081)	(0.092)
Nb. obs	1,252	615	637	719	533

Standard errors in parentheses

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

Note: This table presents data from the Quantitative Scripts experiment. 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.3.3 Content Transmission Incentives

Appendix Table A10: Belief Updates About State of the World

	(1) Pooled	(2) Modular Only	(3) Naturalistic Only	(4) Explicit Incentives	(5) Implicit Incentives
Low Level × Strong Reliability	-0.697***	-0.431***	-0.944***	-0.696***	-0.699***
	(0.142)	(0.092)	(0.012)	(0.165)	(0.127)
Trans. × Low Level × Strong Reliability	0.255*	0.052	0.439***	0.268	0.246*
	(0.152)	(0.116)	(0.078)	(0.186)	(0.144)
Low Level × Weak Reliability	-0.323***	-0.411***	-0.270***	-0.334***	-0.308***
	(0.043)	(0.036)	(0.045)	(0.039)	(0.059)
Trans. × Low Level × Weak Reliability	-0.088	-0.019	-0.131*	-0.102	-0.077
	(0.061)	(0.079)	(0.070)	(0.070)	(0.085)
High Level × Weak Reliability	0.514***	0.578***	0.474***	0.567***	0.453***
	(0.047)	(0.065)	(0.057)	(0.067)	(0.131)
Trans. × High Level × Weak Reliability	-0.186***	-0.289***	-0.119	-0.275***	-0.085
	(0.060)	(0.088)	(0.074)	(0.086)	(0.139)
High Level × Strong Reliability	0.859***	0.922***	0.815***	1.012***	0.718***
	(0.056)	(0.089)	(0.059)	(0.112)	(0.022)
Trans × High Level × Strong Reliability	-0.329***	-0.399***	-0.281***	-0.541***	-0.126*
	(0.070)	(0.118)	(0.076)	(0.123)	(0.067)
Nb. obs	2,500	1,288	1,212	1,245	1,255

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

Note: This table presents data from the Content Transmission Incentives Experiment. 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 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, Column (3) for the naturalistic reliability manipulation, Column (4) for the explicit-incentives group (where incentives to transmit level and reliability are explicitly separated), and Column (5) for the implicit-incentives group (where transmitters are generically incentivized to pass on all information).

Appendix Table A11: Beliefs About Level of Original Message's Prediction

	(1) Pooled	(2) Modular Only	(3) Naturalistic Only	(4) Explicit Incentives	(5) Implicit Incentives
High Level	1.468***	1.325***	1.628***	1.527***	1.408***
	(0.065)	(0.091)	(0.089)	(0.086)	(0.097)
High Level × Transmitted	-0.492***	-0.406***	-0.588***	-0.543***	-0.437***
	(0.073)	(0.108)	(0.104)	(0.096)	(0.114)
Transmitted	0.190***	0.109	0.279***	0.214***	0.162*
	(0.054)	(0.077)	(0.079)	(0.070)	(0.085)
Constant	-0.700***	-0.578***	-0.838***	-0.743***	-0.655***
	(0.047)	(0.065)	(0.065)	(0.059)	(0.074)
Nb. obs	2,500	1,288	1,212	1,245	1,255

Standard errors in parentheses

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

Note: This table presents data from the Content Transmission Incentives Experiment. 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, Column (3) for the naturalistic reliability manipulation, Column (4) for the explicit-incentives group (where incentives to transmit level and reliability are explicitly separated), and Column (5) for the implicit-incentives group (where transmitters are generically incentivized to pass on all information).

Appendix Table A12: Beliefs About Reliability of Original Message's Prediction

	(1) Pooled	(2) Modular Only	(3) Naturalistic Only	(4) Explicit Incentives	(5) Implicit Incentives
Strong Reliability	1.236***	0.751***	1.567***	1.037***	1.460***
	(0.077)	(0.120)	(0.096)	(0.112)	(0.105)
Strong Reliability × Transmitted	-0.860***	-0.433***	-1.130***	-0.677***	-1.076***
	(0.098)	(0.156)	(0.119)	(0.138)	(0.131)
Transmitted	0.375***	-0.032	0.628***	0.230**	0.553***
	(0.073)	(0.118)	(0.088)	(0.092)	(0.105)
Constant	-0.612***	-0.304***	-0.802***	-0.495***	-0.759***
	(0.058)	(0.093)	(0.072)	(0.079)	(0.084)
Nb. obs	2,082	870	1,212	1,052	1,030

Standard errors in parentheses

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

Note: This table presents data from the Content Transmission Incentives Experiment. 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 transmitter level. Column (1) does this for our full pooled sample, Column (2) for our subsample hearing the modular reliability manipulation, Column (3) for the naturalistic reliability manipulation, Column (4) for the explicit-incentives group (where incentives to transmit level and reliability are explicitly separated), and Column (5) for the implicit-incentives group (where transmitters are generically incentivized to pass on all information).

C Original Recordings: Transcripts and Links to Recordings

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

Revenue growth of a retail company

Modular

Introduction

This prediction is about the annual revenue growth of a large US 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 US 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 US 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 US 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 US 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.