# Explanations*

Thomas Graeber Christopher Roth Constantin Schesch

July 28, 2024

**Abstract**

When people exchange knowledge, both truths and falsehoods can proliferate. We study the role of explanations for the spread of truths and falsehoods in 15 financial decision tasks. Participants record the reasoning behind each of their answers with incentives for accuracy of their listeners' responses, providing over 6,900 unique verbal explanations in total. A separate group of participants either only observe one orator's choice or additionally listen to the corresponding explanation before making their own choice. Listening to explanations strongly improves aggregate accuracy. This effect is asymmetric: explanations enable the spread of truths, but do not curb the contagion of falsehoods in knowledge exchange. To study mechanisms, we extract every single argument provided in the explanations, alongside a large collection of speech features, revealing the nature of financial reasoning on each topic. Explanations for truths exhibit a significantly richer message space and higher argument quality than explanations for falsehoods. These content differences in the supply of explanations for truths versus falsehoods account for 60% of their asymmetric benefit, whereas orator and receiver characteristics play a minor role.

*Keywords:* Explanations, Social Learning, Speech Data, Financial Knowledge

* We thank Simon Cordes, Pietro Ducco, Maximilian Fell, Paul Grass, Jindi Huang, Milena Jessen, Julian König, Malte Kornemann, Maximilian Müller, Nicolas Röver, Gabriel Saliby, and Georg Schneider for outstanding research assistance. We thank seminar audiences at the MPI Bonn, Duke, CESifo behavioral, the MiddExLab,Innsbruck, the University of Pittsburgh, the University of Zurich and Harvard Business School for useful feedback. The research described in this article was approved by the Institutional Review Board at Harvard Business School and the ethics committee of the University of Cologne. Roth: Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2126/1-390838866. The data collections were pre-registered on AsPredicted (#155323; #157147; #159277). Graeber: Harvard Business School, tgraeber@hbs.edu. Roth: University of Cologne, IZA, ECONtribute, CEPR, NHH, and Max-Planck Institute for Research on Collective Goods, roth@wiso.uni-koeln.de. Schesch: Harvard Business School, c.schesch@gmail.com.

# 1 Introduction

We obtain most ideas, news and knowledge from listening to others (Hirshleifer, 2020; Schotter, 2023). Some of the information we receive is accurate, while some of it is flawed. Whether learning from others improves or impairs our decisions therefore critically depends on our ability to discern what is right and what is wrong. The epitome of a welfare-improving social aggregation of information is the *marketplace of ideas*: the truth will emerge and prevail in an environment where thoughts, insights and knowledge are freely exchanged.[1] Yet, misbeliefs can spread rapidly, too, whenever individuals systematically fail to identify falsehoods (Lazer et al., 2018; Pennycook and Rand, 2021). Indeed, some argue that recent technological advances catalyze the spread of falsehoods, marking the onset of a "post-truth era."

At the center of any knowledge exchange are explanations: people share justifications for and reasoning behind their beliefs and choices, often conveyed by word of mouth from peer to peer (Shiller, 2017). Unlike in canonical economic models of social learning, people do not only learn from observing *what* someone else does or believes, but also from *why* they do so. To examine how explanations affect the contagion of truths and falsehoods, we conduct large-scale experiments in which respondents solve canonical financial decision tasks, receive one of over 6,900 explanations recorded by other respondents and are then allowed to update their answer. We focus on financial decisions because they are known to be shaped by information that circulates through social networks (Duflo and Saez, 2003; Brown et al., 2008) and include topics where false narratives appear to be widespread, such as cryptocurrencies or stock-picking.

We run a series of pre-registered experiments using a design comprising two separate stages. We start with an Orator experiment that characterizes the *supply of explanations*. We focus on the spread of knowledge: respondents complete 15 canonical financial decision problems with an objectively correct answer, allowing us to characterize mistakes. These include questions on nominal illusion, the net returns of active and passive investing, the relationship between interest rates and bond prices, compounding of interest and other topics. In each task, respondents first indicate their choice, with incentives for accuracy. Then, they record a voice message in which they provide an explanation for their answer to randomly matched participants in a separate study. Orators' incentives induce aligned interests with the listeners of their recording: an orator's likelihood of receiving a bonus payment increases with the accuracy of the listener's subsequent

[1] The marketplace of ideas is a foundational rationale for freedom of speech and open discourse, frequently attributed to U.S. Supreme Court Justice Oliver Wendell Holmes Jr.

response to the question.[2] The faithful exchange of knowledge pervades real-world interactions in education (e.g., classroom teaching), business (e.g., collaboration across business functions), healthcare (e.g., patient-doctor consultations, public health campaigns), the legal system (e.g., advice on contracts) and finance, including peer-to-peer information sharing.

To study the *interpretation of explanations* and its consequences for social learning, we then conduct a Receiver experiment, in which respondents face the same 15 tasks. They first make their own incentivized choice. For each task, we randomly assign respondents to the *Choice Only* or the *Explanation* condition. In *Choice Only*, they learn about the choice of a randomly chosen respondent in the Orator experiment. In the *Explanation* treatment arm, participants additionally listen to the respondent's explanation. In both conditions, respondents then again select their own best choice, which may now differ from their initial choice. The comparison between *Explanation* and *Choice Only* allows us to identify the specific effect of listening to a verbal explanation on imitation, above and beyond the mere observation of another respondent's choice. The *Choice Only* condition provides a natural benchmark that captures learning from mere observation in the absence of an explanation, and further allows us to control for the direct effects of the respondents' confidence in their prior answers, measurement error in priors and other factors, such as experimenter demand.

***Do explanations matter?*** We begin by analyzing how explanations shape aggregate optimality rates. The prior optimality rate across tasks is 55%, meaning participants perform better than the 50% rate implied by choosing randomly in a two-option task. In our control, just observing someone else's answer increases the frequency of optimal choices to 59.5% (treatment *Choice Only*, $p < 0.01$). Explanations boost the aggregate improvement from exposure to others: across all tasks, explanations raise the accuracy rate to 62.7% (*Explanation*, $p < 0.01$). This treatment effect of *Explanation* is large: it is 82% higher than that of *Choice Only* ($p < 0.01$).

To understand the drivers of the aggregate effect on optimality, we focus on receivers who are confronted with an answer that *conflicts* with their own prior answer. First, those with incorrect priors may learn from correct choices, creating learning opportunities. Second, those with correct priors encounter incorrect choices, leading to unlearning opportunities. We find that the aggregate benefit of explanations over merely observing someone's answer is entirely driven by learning opportunities: in those cases, the imitation rate is 55.8% in *Explanation* but only 42.8% in *Choice Only* ($p < 0.01$). This corresponds to a 30.4% increase in seizing learning opportuni-

[2] Our *explanations* thus differ from *persuasive messages*, which can reflect misaligned incentives.

ties. The *Explanation* treatment does not, however, decrease the frequency with which receivers switch to a wrong answer in unlearning opportunities. Receivers switch from accurate to inaccurate answers in 23.1% and 22.9% of the unlearning conditions in *Choice Only* and *Explanations*, respectively ($p = 0.87$). Asymmetric learning emerges in 14 out of 15 tasks, among listeners with weakly and strongly held priors, and is robust to various sample restrictions.

***Why do explanations help?*** To develop a better understanding of the economic interpretation of this treatment effect, we cast our treatments in a standard belief formation model. The model illustrates that the treatment effect can only arise from how explanations shape the perceived accuracy of the orator's answer. This raises the question of whether the impact of explanations is akin to observing the orator's numerically expressed confidence, which could systematically differ across learning and unlearning opportunities. To test this hypothesis, we conduct an additional Receiver experiment, in which respondents observe both the orator's choice and their stated confidence. Compared to only observing the orator's choice, additionally learning about their confidence does not significantly shift imitation in learning and unlearning opportunities. Therefore, an explanation is a signal of accuracy distinct from the orator's numerical confidence.

Which features of oral explanations convey the additional information? Both the content—*what* is said—and its delivery through the speaker's voice—*how* it is said—may shape the receivers' behavior. To distinguish between these two channels, we design an additional Receiver study, in which respondents read the transcript of the orator's explanation instead of listening to the corresponding recording. This preserves the exact verbal content conveyed by the explanation while eliminating the role of the speaker's delivery and voice. A strongly asymmetric effect of explanations on learning and unlearning persists in the *Transcript* treatment, suggesting that it is primarily driven by the supply and interpretation of the *content* of explanations. While we exactly replicate the average null effect of explanations in unlearning situations of the *Transcript* treatment, there is a 8.1 p.p. increase of the imitation rate in learning opportunities (relative to *Choice Only*), which equals 62% of the effect size induced by the *Explanation* treatment.[3]

Having established that the impact of explanations is distinct from confidence statements and primarily operates through content, the question remains as to why the asymmetric effect

[3]One interpretation of this lower treatment effect might be that people are less attentive in the *Transcript* treatment than the *Explanation* treatment. Yet, our decomposition exercise displayed in Table A3 shows that 75% of the differential effect of explanations in the *Transcript* treatment is explained away by our content-based measure of explanation richness. This, in turn, suggests that respondents are highly attentive to differences in the content of the scripts.

of explanations emerges. To answer it, we conduct a mechanism analysis that systematically explores the two margins of variation that could explain the asymmetry: first, it may arise from content differences in the supply of explanations. Second, it might result from differences in the characteristics of participants in learning versus unlearning situations.

***Mechanisms: The content of explanations.*** In the first step of our mechanisms analysis, we characterize the content of the more than 6,900 provided explanations and study its effects on imitation. Analyzing the content of speech recordings is non-trivial due to the high-dimensional nature of language data: each sentence has innumerable features and interpretations. We pursue a two-pronged approach based on the following distinction. On the one hand, explanations are characterized by the substantive content that rationalizes the answer to a question: specifically, they provide *arguments*. Arguments tend to be domain-specific; they directly relate to a specific question and answer. We identify and code every argument provided across the universe of our explanations, delivering the first dataset of its kind for studying their effect on social learning and unlearning in a controlled setting. On the other hand, explanations are characterized by a large number of text features: they exhibit various classes of expressions (e.g., certainty phrases, hedges or questions), many linguistic and rhetorical attributes, and can be described by speech and text metrics, among others. We code a large collection of text features culled from the existing literature. These features are domain-general in the sense that they similarly apply to explanations for different questions and answers. We employ a combination of dual human coding, a large language model and machine learning methods to analyze the universe of our explanation transcripts in a robust and replicable manner.

Our argument annotation delivers, for each of the 15 tasks, the collection of all provided arguments, their frequency in explanations for correct versus incorrect choices, and estimates of their idiosyncratic effects on imitation rates (Figure 3, pp. 25-27). To illustrate our results at the task level, consider the question of whether actively managed investment funds systematically outperform passively managed ones in terms of expected net returns. The most frequent argument is that active funds—unlike passively managed ones—can quickly adapt to changes in the market, which is present in 57.2% of the explanations for incorrect answers but almost absent (2.9%) from explanations for correct answers. The second most common one argues that active funds charge higher fees, which is prominent among explanations for correct answers (22.7%) and nearly absent from those for incorrect answers (3.2%). In total, we identify 11 unique arguments. Among the most effective ones are the one on active funds being able to react to the market (in unlearning situations) and the argument that passive funds target long-term growth (in learning situations), both of which raise imitation rates by more than 20 p.p. relative to *Choice Only*. We discuss patterns of heterogeneity in the number of arguments, the degree of consensus and variation in the effects on imitation across tasks. This angle on the data delivers unprecedented access to the nature of people's reasoning about a given topic and its likelihood of influencing others, but does not yet shed light on the drivers of the asymmetric treatment effect, since the substantive content of arguments is naturally difficult to compare across tasks.

To draw broader conclusions about differences between explanations for correct and incorrect choices based on our argument data, we measure the prevalence of four classes of arguments with increasing "quality:" (i) the absence of any argument, (ii) irrelevant arguments, i.e., off-topic reasoning, (iii) fallacious arguments in which the premises are false or do not establish the conclusion and (iv) sound arguments, which have true premises and a valid conclusion. This quantification of argument classes is important as respondents could in theory reach the right conclusion even based on fallacious or irrelevant arguments. We document large differences in the presence of argument classes between learning and unlearning opportunities. While respondents in unlearning situations are more likely to encounter no (21.8% vs. 16.0% in learning), irrelevant (22.6% vs. 17.7%), or fallacious arguments (51.0% vs. 10.5%), respondents in learning situations are more likely confronted with sound arguments (55.8% vs. 4.7% in unlearning).

Are these different argument classes associated with different effects on imitation rates? Reassuringly, higher-quality arguments are associated with higher effects on imitation rates. While the absence of an argument or the presence of an irrelevant argument – if anything – is associated with a decrease in imitation rates, even fallacious arguments increase imitation rates. Yet, even conditional on encountering an argument from the same class, the impact on imitation rates remains far higher in learning than in unlearning situations. In fact, we find that the argument gap in explanations at most accounts for 25% of the asymmetric effect.

What content features other than the quality of arguments drive imitation decisions? We turn to an analysis of the second component of our content annotation approach, domain-general speech and text features, such as certainty markers. While this analysis reliably confirms intuitions about the prevalence of specific features, the core insight is that explanations for correct answers reflect a significantly *richer* message space: they contain more occurrences of most features, even conditional on length. Indeed, 24 out of 31 features more frequently occur in explanations for learning opportunities. This finding is corroborated by a quantitative, pre-registered measure of richness annotated using a large language model.The average richness score of explanations for correct choices is 0.76 SD higher than for incorrect choices.[4] We investigate the key drivers of the richness effect and find that *lexical* richness—the complexity, variety, and uniqueness of vocabulary—plays a pivotal role.

Can these differences in the features and richness of explanation content, then, account for the asymmetric treatment effect? First, we test to which extent different explanation features are predictive of imitation per se, revealing that the richness of a message is by far the strongest predictor. A one-standard deviation increase in the richness score is associated with an 11.8 p.p. ($p < 0.01$) higher imitation rate, relative to not receiving an explanation, after controlling for all other features, including the length of the script. The strong benefit of richer explanations diverges from the principle of *Occam's Razor*, which posits that the simplest explanation is usually the preferred one. While simplicity may be valued, our findings show that comprehensiveness and detail in an explanation can enhance its social influence in the case of financial reasoning. Second, we ask to what extent differences in the average richness of message spaces account for the differential effect in learning and unlearning opportunities. Our analyses suggest that the richness gap explains approximately 60% of the differential treatment effect of explanations.

***Mechanisms: Orator and receiver characteristics.*** In the final step of our mechanisms analysis, we study the role of participant characteristics. The asymmetric effect may, in part, stem from the unique characteristics of participants in learning situations. They may have traits that, as orators, make them more influential or, as listeners, lead them to interpret information in a manner that induces less imitation. Our analyses show that some orator characteristics indeed predict imitation: for example, male and more educated orators induce more imitation through their explanations, whereas Black speakers are imitated less. These effects may reflect inferences made from both content and voice, but note that receivers were not explicitly informed about the orators' sociodemographics. While orator characteristics are somewhat predictive of imitation in general, they contribute virtually nothing to the asymmetric treatment effect of explanations above and beyond content. Similarly, differences in receiver characteristics do not explain away the asymmetric effects. If anything, accounting for differences in receiver characteristics somewhat widens the differential treatment effects in learning and unlearning.

[4]Note that we document this richness-truth correlation in a setting with aligned incentives between sender and receiver, and factual questions on which *motivated* beliefs are unlikely. The association between richness and truth may differ—and perhaps even reverse—in domains that involve incentives for persuasion and motivated beliefs. For instance, conspiracist explanations in politics can often be very rich.

In conclusion, our mechanism evidence suggests that content differences rather than personal characteristics of orators or receivers are the key determinant of imitation decisions in our setting. We leverage novel methods that were unavailable to the earlier literature on advice and social learning (see, e.g., Schotter, 2023), providing direct access to *what* people communicate in unconstrained natural language and how it affects imitation. Our data allow us to compare and emphasize the role of content over that of the identity of speakers on imitation, the focus in much of the previous literature (Cialdini, 2001, 2007; Cialdini and Goldstein, 2004). The overall beneficial effect of explanations on optimality directly hinges on the positive association between richness and truth. This relationship may be specific to settings that have, like ours, aligned incentives between speakers and listeners to identify the truth in the exchange of knowledge rather than opinions. At the same time, our portable paradigm provides a blueprint for studying analogous dynamics in settings that lack these features. For example, in the case of persuasion, orators' motives typically deviate from receivers'. While it is likely that some findings about explanations generalize to persuasive messages—e.g., richer persuasive message may turn out to be more effective—other features may not: for example, the positive relationship between richness and truth likely breaks in such settings, which is a fruitful avenue for future work.[5]

***Literature.*** Our paper contributes to an emerging literature on learning from qualitative information, e.g., in the form of stories (Graeber et al., 2023; Aina, 2023) and narratives (Andre et al., 2022; Kendall and Charles, 2022; Barron and Fries, 2023; Hüning et al., 2022; Shiller, 2017; Ambuehl and Thyssen, 2024; Schwartzstein and Sunderam, 2021; Andre et al., 2023; Bursztyn et al., 2023; Han et al., 2024; Eliaz and Spiegler, 2020, 2024). Barron and Fries (2023) study strategic communication of model parameters as a persuasive tool when financial advisors hold incentives that differ from those of the individuals they are advising. Graeber et al. (2024) examine how verbal transmission distorts the supply of qualitative economic information and show that information about signal reliability gets lost in transmission more than information about signal values. We differ from existing work in our focus on characterizing the supply and interpretation of explanations for people's choices in canonical financial decision problems.

We relate to an interdisciplinary literature on explanations (Lombrozo, 2006; Rozenblit and Keil, 2002) and arguments (Sloman et al., 1998; Gick and Holyoak, 1980). Our contribution lies

[5]While our paper already covers a relatively large set of 15 questions, our experimental approach could be applied to other topics, such as beliefs about macroeconomic quantities or politics, allowing us to compare and test how broadly our findings generalize.

in providing a characterization of the supply of qualitative explanations and in estimating their consequences for economic decisions in a controlled setting. While early work by Langer et al. (1978) shows that people are more likely to comply with a request if it is justified, irrespective of whether the reason is good or bad, we document that people are more likely to imitate choices justified by rich and sound arguments and less likely to imitate irrelevant arguments.

We further contribute to a literature that studies how social learning among non-experts—labeled "naive advisors" by Schotter (2003)—affects the prevalence of biases and misinformation (Hirshleifer, 2020).[6] In a setting that—unlike ours—features incentives for deception, Serra-Garcia and Gneezy (2021) show that individuals fail to detect others' lies when they are shown a video message of another respondent paid to invent a news story. A series of papers has examined social learning in the context of motivated beliefs (Oprea and Yuksel, 2022; Thaler, 2021). Grunewald et al. (2024) study whether biases are contagious in a setting with motivated beliefs. They find that communication of personal opinions via a written text message amplifies belief biases relative to a setting of observational learning. Conlon et al. (2022) show that in the context of a balls-and-urns updating task, people are less sensitive to information others discover than to equally relevant information they receive themselves.

Finally, by characterizing the spread of truths versus falsehoods through social learning, we contribute to a long-standing literature on whether individual-level biases matter for aggregate market-level outcomes (Russell and Thaler, 1985; List, 2003; Fehr and Tyran, 2005). Enke et al. (2023) show that awareness about biases reduces the impact of individual-level biases on aggregate outcomes through institutions that rely on self-selection, while Amelio (2023) studies how meta-cognition shapes social learning. Unlike those findings, ours cannot, by design, be due to meta-cognition. Instead, we examine how explanations affect perceptions of others' accuracy and thereby the proliferation of truths and falsehoods.

[6]The literature on social learning (typically defined as observational learning from others' actions, our control condition) is vast (Weizsäcker, 2010; Conlon et al., 2021; Eyster and Rabin, 2014; Mobius and Rosenblat, 2014; Jackson and Yariv, 2007; Galeotti et al., 2010), especially in the context of financial decisions (Ambuehl et al., 2022; Hvide and Östberg, 2015; Bursztyn et al., 2014; Akçay and Hirshleifer, 2021; Han et al., forthcoming; Hirshleifer et al., 2023). The present paper builds on an earlier literature on advice-giving (Schotter and Sopher, 2003; Çelen et al., 2010; Schotter, 2023), which has focused on pre-structured messages and not yet studied the nature and causal effect of explanations in natural language.

# 2 Experimental Design

## 2.1 Overview

Our experimental design studies 15 canonical financial decision problems and consists of two stages. In the *Orator* experiment, respondents record an explanation for their answer for each of the tasks. In the subsequent *Receiver* experiment, respondents first provide their choice. Then, they either only see another respondent's choice (from the Orator experiment) or additionally listen to that respondent's explanation, before providing their answer to the same task again.

## 2.2 Financial Decision Problems

We select 15 financial decision tasks based on three criteria. First, we aim for a collection that is broadly representative of the reasoning and decision biases studied in the finance literature. This spans behavioral phenomena like exponential growth bias and nominal illusion but also more specific knowledge about different asset classes and investment decisions, for example expected returns under active versus passive investing. Many of the problems we study are tightly linked to common high-stakes financial decisions, such as whether to invest in active or passive funds. Second, we restrict our attention to questions with an objectively correct answer or ones where a broad consensus exists in the financial economics literature. This means we exclude questions that rely on tastes. Moreover, participants in our studies are made aware that a "correct" solution exists.[7] Third, the questions should be reasonably short.

Some tasks have a logically correct answer, for example the following question about the concept of inflation, with the correct answer underlined:

> *Imagine that the interest rate on your savings account was 2.5% per year and inflation was 3% per year. After 1 year, how much would you be able to buy with the money in this account?*
>
> 1. *More than today*
> 2. *Exactly the same as today*
> 3. <u>*Less than today*</u>

Other questions relate to a broad consensus in financial economics, like the following:

[7]This is motivated both by our focus on the exchange of knowledge—rather than opinions—and a precondition for incentivizing answers. The effect of explanations on imitation choices might be different in settings where people think that no correct answer exists.

*Do actively managed investment funds systematically outperform passively managed investment funds in terms of expected net returns, i.e., after accounting for investment fees?*

1. *Actively managed funds outperform passively managed ones.*
2. *Actively managed funds do not outperform passively managed ones.*

We embrace that the differences across these tasks will likely evoke structurally different explanations. A participant might give a wrong answer because they have not heard of the concept of a call option—in a sense, they may not really know what the question is about—or they fully understand the question but still do not know its answer. This difference captures two separate important features of bias in practice and may be reflected in explanations as we discuss in the following sections. Table A1 outlines the motivation and origin of the different tasks, as well as the exact wording of all questions. Two tasks have two options, while the others have three. In most analyses, we are only interested in whether the correct option is chosen.[8]

## 2.3 Part 1: Orator Experiment

The main objective of the Orator experiment is to obtain recordings of people's verbal explanations for each of the financial decision tasks.[9] In the beginning, participants are told that we are interested in how they would give advice in an informal conversation. They are informed that they should share an explanation behind their response and that their recording will be played to a few other participants who will have to respond to the same question. We ask respondents not to search for answers on the internet.[10]

In practice, people typically have (at least some) time to think about an explanation they are asked to give. Correspondingly, rather than forcing respondents to talk immediately upon reading the question, we show them the question first and they decide when to start their recording. An example screen is shown in Appendix Figure A12. After recording their explanation, respondents first select their preferred answer and then state their confidence in its accuracy by answering "*How certain are you that your above answer is optimal?*" on a scale from 0 "*Not at all certain*" to 100 "*Fully certain*".

[8]Due to these differences, one should expect different frequencies of correct responses, because randomizing would create an optimality rate of 50% in a two-option and 33% in a three-option task. However, this is constant across conditions and thus cannot affect treatment comparisons.

[9]The full set of instructions is reproduced in Appendix C.4.

[10]We ask participants at the end of the study whether they searched for any answers, stressing there is no penalty for indicating that they did. We then exclude the 7.0% of participants who indicated that they searched for answers from our data.

**Incentives.** With a 10% chance, a respondent is eligible for a bonus payment of $10. Whether a selected respondent receives a bonus is based on one randomly drawn task. The orator is matched with another randomly selected participant in the Receiver experiment, who either only sees that orator's answer or additionally listens to their voice recording. The bonus is paid if the matched receiver gives the correct answer after exposure to the orator's answer. Our experiment thus creates aligned incentives between the orator and the receiver: the orator is incentivized not to be imitated *per se*, but to induce the receiver to make the right choice. The orators' instructions emphasize that their incentives will be known to listeners ("Participants listening to your recordings will be informed that you will receive a bonus if they select the correct answer."). We confirm that orators understand the aligned incentives scheme using a control question.

**Speech recordings.** The Orator experiment relies on speech recordings of people's explanations. Relative to written text, speech recordings have a series of advantages for our purposes. A voluminous literature outside of economics has characterized the 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 and cognitively taxing to produce (e.g., Bourdin and Fayol, 2002). Because writing text is typically more exhausting than speaking, the transcripts of orally provided explanations are often substantially longer. Much of social learning follows from oral conversations, making speech recordings an ideal testing ground to study the effects of explanations. Second, speech data include features of natural language that plausibly affect social learning but are mostly absent from written texts, including tone, emphasis, and disfluencies such as pauses, repetitions, revisions, hesitations, or filler words. Third, writing text as opposed to spontaneously talking about one's thoughts adds another filter that may distort measured explanations compared to explanations people give spontaneously in the real world.

## 2.4 Part 2: Receiver Experiment

To characterize the effect of explanations on social learning and unlearning, we conduct a Receiver experiment that leverages the choices and recordings from the Orator experiment.[11]

As in the Orator experiment, respondents iterate through the 15 decision tasks. To measure imitation rates at the individual level, we use a within-design with five steps in each round. First, respondents read the financial decision task and are incentivized to indicate their preferred choice, which provides our measure of their prior belief. Second, they indicate their confidence in

[11] We provide the full set of instructions in Appendix D.

the accuracy of their response in the same format as respondents in the Orator experiment. Third, they either only learn about the choice of another randomly selected respondent in the Orator experiment (*Choice Only* condition) or additionally listen to the recording of their explanation (*Explanation* condition). Fourth, the receiver again has an opportunity to select their preferred choice with incentives for accuracy. Fifth, they indicate their confidence in their posterior answer.

**Treatments.** In the *Choice Only* treatment, receivers may infer and adjust their belief about the optimal answer from learning what someone else chose, even absent an explanation. This same source of learning is present in the *Explanation* treatment, but the explanation provides an additional source of information. We randomize treatments between participants, at the task level. For each task, 80% of receivers are sampled into the *Explanation* condition, while the remaining 20% are assigned to *Choice Only*.[12]

The comparison between *Explanation* and *Choice Only* allows us to identify the specific effect of listening to a recorded explanation on learning and unlearning, above and beyond the mere observation of another respondent's choice. The *Choice Only* condition is critical to control for (i) the effects of confidence, (ii) measurement error in priors, and (iii) other confounders, such as experimenter demand effects. At the same time, this comparison captures various potential channels of learning which we disentangle through additional treatments.[13]

**Incentives.** Receivers have a 10% chance of being eligible for an additional $10 bonus payment. Whether they receive the bonus is determined by the accuracy of their answer in a randomly selected reasoning task. For every task, we randomly select whether their first answer or their second answer is the decision that counts for the bonus.

## 2.5 Logistics

Respondents in both studies received a base reward of $6 for completing the study. Median completion times were 25 minutes in the Orator experiment and 26 minutes in the Receiver experiment. All experiments were conducted on the online platform Prolific, which is widely used for experiments in the social sciences (Eyal et al., 2021). The Orator experiment was run for a total of 505 U.S. respondents in December 2023, of whom 466 provided valid responses. Participants were required to have a working microphone to record their voice message. The Orator

[12]We oversample the *Explanation* to obtain the statistical power needed to examine heterogeneous effects by features of the explanations.

[13]We compare the effect of hearing an oral explanation to simply observing the orator's choice and confidence score (Section 4.2) and simply reading a transcript (Section 4.3).

experiment yields a total of 6,910 valid recordings obtained by integrating speech recordings with Phonic into Qualtrics surveys. We use an Amazon Web Services backend to stratify and distribute recordings into our Receiver experiment. The Receiver experiment was run with 1,385 U.S. respondents in December 2023, of whom 1103 provided valid responses.[14] Appendix Table A2 provides an overview of all data collections and corresponding pre-registrations.

# 3 The Effects of Explanations on Social Learning

We start by providing basic descriptives about our respondents' explanations. We then turn to the effects of explanations on imitation and optimality and break them down by learning and unlearning situations. We conclude with additional results on heterogeneity and robustness.

## 3.1 Basic Characteristics of Explanations

Our Orator experiment generated 6,910 audio recordings with a median duration of 26 seconds. Appendix Figure A1 shows substantial variation in length: the 10th percentile is 11 seconds and the 90th percentile 55 seconds. The audio quality of recordings obtained through our online experiments is high. Analyses of the audio files show that only 1.1% of recordings are unusable, typically because of a technical microphone problem or because respondents accidentally submitted it too early.[15] We use transcripts of the recordings that preserve details and nuances of spoken language, notably filler words such as "um" and "eh." The median length of the resulting transcripts is 55 words.

To parse basic features of the scripts, human coders classify them using a simple coding scheme (see Appendix C.1). We find that 13.1% of the explanations are pure restatements of the question and/or the answer given, without adding any content matter. These may both be due to people not trying to give or not having any explanation for their answer. We characterize a negligible minority of 2.6% of recordings as nonsensical. Looking at the cases that reflect some form of actual explanation, we find that 8.6% of all explanations contain *no substantive arguments* while 74.3% of all explanations contain *substantive arguments*. Finally, a substantial 13.7% of recordings contain explicit expressions of the speaker's confidence.

This first look at basic descriptives suggests the Orator experiment provides a rich and varied

[14]For both experiments we exclude participants who indicated that they looked up answers to the financial decision questions online, in accordance with our pre-registration.

[15]Incomprehensible messages or high background noise appear very rarely and are therefore not relevant concerns for our study.

database of heterogeneous explanations for our set of tasks. In Section 5, we go much further in examining their characteristics.

## 3.2 Explanations and Optimality Rates

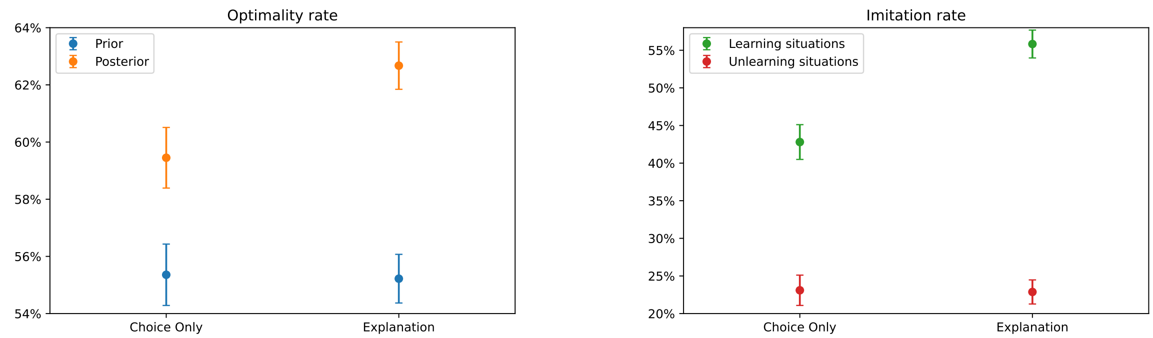
We start by analyzing the effects of our treatments on the frequency of correct choices, which we refer to as the *optimality rate*. For comparability with additional treatments and to maximize statistical power, we pool observations for *Choice Only* obtained from different between-subject collections that use this exact same control condition (see Sections 4.2 and 4.3). The prior optimality rate reflects receivers' knowledge about a task before learning from another respondent. The posterior optimality rate captures accuracy after receivers observe another respondent's answer only (*Choice Only*) or additionally listen to their verbal explanation (*Explanation*).

Figure 1a shows these optimality rates pooled across all 15 tasks. Prior to exposure, 55.4% and 55.2% of the respondents provided correct answers in the *Choice Only* and *Explanation* conditions, respectively ($p = 0.85$). We document two main findings on posterior optimality rates. First, just observing another's choice increases optimality rates by 4.1 p.p. ($p < 0.01$), creating an aggregate improvement. Second, additionally listening to another person's explanation significantly raises the size of the improvement to 7.5 p.p. (0.15 SD, $p < 0.01$). This difference in improvement rates across the *Choice Only* and *Explanation* condition is statistically significant ($p < 0.01$) and quantitatively large: explanations induce a 82% larger improvement than mere observations of another's choice. The analysis of optimality rates establishes that explanations, on average, have a strongly positive effect on social learning. At the same time, this average effect masks variation across initially correct and incorrect listeners, different tasks and explanations.

## 3.3 Explanations and Imitation Rates

Our sample is composed of structurally distinct sets of orator-receiver matches. Specifically, pairs vary along two margins relevant for social learning: the initial accuracy of the listener and the accuracy of the orator. This creates four distinct groups, characterized by whether a receiver was initially correct, and whether they were subsequently exposed to a *confirming* or *conflicting* signal. Intuitively, the two groups of receivers matched with an orator who gave the same answer should not change their initial response and thus have little effect on aggregate improvements.[16] Instead, changes of prior choices should be driven by receivers who are confronted with

[16] We confirm in our data that switching after affirming advice is rare and does not differentially occur among initially correct and incorrect receivers; see Appendix B.5.2 for further results.




**(a)** Prior and posterior optimality by treatment

**(b)** Imitation rate in learning and unlearning

**Figure 1:** Effect of *Explanation* on optimality and imitation. *Notes:* Left panel shows share of correct receivers before and after exposure to the orator's choice or explanation. Right panel shows share of receivers picking the same answer as the orator in learning and unlearning. *Explanation* sample is the main Receiver survey (1,103 receivers, 13,111 observations), *Choice Only* is pooled from all collections (2,733 receivers, 8,232 observations). Whiskers show 95% CI.

a conflicting choice. We therefore focus on the two groups that drive treatment effects on social learning: receivers with incorrect prior choices exposed to correct choices on the one hand, and receivers with correct choices exposed to incorrect choices on the other. We refer to these as *learning opportunities* and *unlearning opportunities*, respectively.[17] These two types of matches are equally frequent in our sample (due to random matching of receivers and orators): there are 21.2% learning opportunities and 20.3% unlearning opportunities.

Figure 1b displays the frequency of imitation in learning and unlearning opportunities. This figure illustrates two key findings. First, the unlearning rate does not differ significantly between *Choice Only* and *Explanation*, at 23.1% vs. 22.9% ($p = 0.87$). About one of every four receivers with a correct prior confronted with another respondent's wrong answer switches away from the correct one. Thus, participants in unlearning opportunities do not, on average, infer information from explanations that systematically helps them identify the answer as wrong.

Second, we do find a quantitatively large treatment effect on the learning rate. Learning opportunities are far more likely to be seized in *Explanation*, where people imitate in 55.8% of cases, than in *Choice Only*, with 42.8% ($p < 0.01$). This 30.4% increase in the learning rate shows that, on average, explanations are highly beneficial for identifying a correct answer.

**Result 1.** *Listening to another respondent's explanation strongly increases the optimality rate, rela-*

[17]We will correspondingly refer to explanations associated with the correct answer to a question as *learning explanations* and to explanations provided for incorrect answers as *unlearning explanations*.

*tive to just observing their answer. This benefit of explanations is asymmetric: explanations increase imitation in learning opportunities but do not decrease imitation in unlearning opportunities.*

### 3.4 Heterogeneity and Robustness

**Cross-task variation.** We examine cross-task variation in our main findings (see Appendix Figure A2). We estimate a positive treatment effect of *Explanation* on optimality rates in 14 out of the 15 tasks. This improvement rate is not significantly correlated with the prior optimality rate in a given task ($r = -0.06$, $p = 0.82$). The asymmetric benefit of explanations, defined as the difference in learning rates between *Explanation* and *Choice Only* subtracted from the corresponding difference in unlearning rates, is similarly pervasive: we obtain a positive estimate in 14 tasks. At the same time, there is substantial heterogeneity, with estimates ranging from a maximum of +25.9 p.p. in the "Exponential growth bias" task to a median of +12.1 p.p. in the "Interest rates and stock prices" task to a minimum of -3.1 p.p. in the "Value of call option" task.

**Robustness.** In Appendix B.5, we find that the asymmetric effects of explanations are robust to various sample restrictions, and we confirm that our main effect on optimality rates is indeed almost entirely driven by receivers confronted with a conflicting response. In additional analyses, we find that asymmetric learning emerges both among listeners with strong and weak priors, and is more pronounced when priors are strong (see Appendix Figure A11).

## 4 Interpreting the Treatment Effect of Explanations

Having established a treatment effect of explanations, we attempt to analyze the drivers of this treatment effect. To structure our analyses, we first illustrate the role of explanations in a canonical belief formation framework. We then present additional experiments that further characterize our treatment effect in light of this framework.

### 4.1 Summary of Conceptual Framework

In Appendix A, we discuss a simple model of imitation based on Bayesian updating from a signal with subjectively perceived precision (or diagnosticity). Imitation is shaped by two forces: first, an individual's confidence in their prior answer, often referred to as *meta-cognition*; and second, the subjective perception of whether the orator is accurate. The latter is central: to learn from the "signal" that the decision-maker receives through the treatment—seeing another's answer or additionally hearing their explanation—, they need to assign a diagnosticity to it. The subjective diagnosticity is the perceived likelihood that the observed answer matches the truth. This belief about the orator's accuracy is the channel through which the explanations affect social learning.

First, this shows that an explanation can theoretically be thought of as conveying a signal of the *reliability* of the orator's choice. This, in turn, raises the question of whether the verbal explanation an orator provides is equivalent to their self-assessed *confidence*, i.e., their quantitative belief in their accuracy. We empirically test this in subsection 4.2.

Second, the framework clarifies the relationship between optimality and imitation. Under random matching of orators and receivers, there is an equal share of learning and unlearning opportunities. As a consequence, assuming that receivers confronted with confirming answers do not switch systematically, the sign of the difference between learning and unlearning rates is a sufficient statistic for whether there are aggregate improvements. In our data, we find this assumption to be overwhelmingly borne out, although there is a small difference between *Explanation* and *Choice Only* when orator and receiver are both wrong. Appendix B.5.2 provides additional details on this finding, explains why its aggregate impacts are small and shows our results are robust to keeping all situations by distinguishing only by prior accuracy. Moreover, as predicted by the model, the sign of the difference between learning and unlearning rates perfectly predicts whether there is improvement in all of our tasks.

Third, the model illustrates that our main reduced-form finding of treatment differences between *Choice Only* and *Explanation* cannot be explained by meta-cognition (unlike in, e.g., Enke et al., 2023), because the distribution of prior answers and confidence is—by virtue of treatment randomization—the same across conditions. The treatment effect has to arise, instead, from the effects of explanations on the perceived accuracy of orators' explanations. This implication of the model is important because it establishes why our design allows us to abstract from the role of prior confidence in examining the mechanisms underlying the treatment effect.

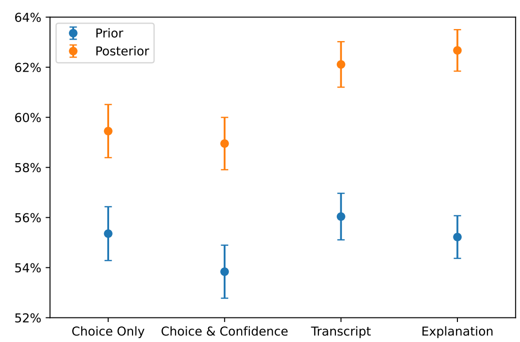
In conclusion, a simple but central insight from the model is that, in economic terms, an explanation can be productively thought of as a signal of an answer's reliability. In the following, we will shed light on two central aspects of this signal: (i) does learning about the orator's stated confidence produce similar effects as does receiving an explanation? (ii) Do explanations mainly convey information through their content or through their oral delivery?

## 4.2 Are Explanations Equivalent to Numerical Confidence Statements?

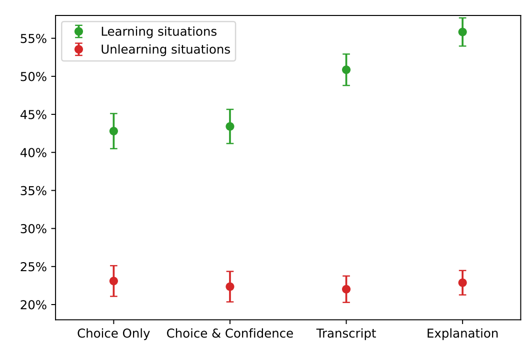
To assess this question empirically, we conduct an additional Receiver experiment that allows us to benchmark the effect of explanations against directly observing the orator's confidence.

**Design.** This additional experiment closely follows the baseline Receiver experiment and also uses on the orator data collected in the baseline Orator experiment. Condition *Choice Only* is identical to the baseline. Condition *Confidence* is identical to *Choice Only* except that the listener also sees the level of the orator's stated posterior confidence, a number between 0% and 100%. Example screens from this experiment are provided in Appendix B.6. In each task, we randomly assign respondents to the *Choice Only* (20%) treatment or the *Confidence* treatment (80%).

**Logistics.** This experiment was run with 860 U.S. respondents in January 2024, of whom 713 provided valid responses.




**(a)** Prior and posterior optimality by treatment




**(b)** Imitation rate in learning and unlearning

**Figure 2:** Effect of *Choice & Confidence* and *Transcript* on optimality and imitation rates. *Notes:* See notes for Table 1. Samples are the corresponding Receiver surveys for *Choice & Confidence* (713 receivers) and *Transcript* (917 receivers).

**Results.** We compare the treatment effect of the *Confidence* treatment on optimality and imitation rates to the treatment effect of *Explanation*. The results are visualized in Figure 2. Panel (a) shows that the *Confidence* treatment also induces a substantial, 5.1 p.p. ($p < 0.01$) improvement of the average optimality rate. Yet, adding confidence does *not* create a significant treatment effect on the posterior optimality rate, which is at 59.5% in *Confidence* compared to 59.0% in *Choice Only* ($p = 0.51$).[18] Turning to the imitation patterns, we find that *Confidence* has virtually

[18] The prior optimality rate in *Confidence*, at 53.8%, lies marginally below that in *Choice Only* at 55.4% ($p = 0.05$) and *Explanation* at 55.2% ($p = 0.05$). We attribute these small differences to sampling noise

no treatment effect on both the learning and unlearning rate. At 22.4% and 23.1%, respectively, the unlearning rates of *Confidence* and the control *Choice Only* are virtually identical ($p = 0.61$). Learning rates are similarly close at 43.4% and 42.8% ($p = 0.70$), especially in comparison to the 55.8% learning rate in *Explanation*.

From the absence of treatment effects in *Confidence*, we conclude that explanations operate differently from merely conveying a quantitative signal of the orator's confidence. There are many possible reasons. For example, explanations can convey information above and beyond a confidence level: they may provide objective justifications for an answer that the listener can evaluate independently. This result motivates our mechanism analyses in Sections 5 and 6.

**Result 2.** *The effect of explanations on social learning differs from that of merely observing a sender's confidence. Unlike explanations, confidence observations (i) do not have a treatment effect on the optimality rate, and (ii) do not affect learning and unlearning rates.*

## 4.3 The Effect of Explanations: What Is Said Versus How It Is Said

The drivers of the treatment effect can be broken down into two factors: the content of the explanation (*what* is said) and its delivery through the speaker's voice (*how* it is said). This distinction is important because it allows us to understand whether the benefits of explanations for social learning are likely to be limited to oral conversations or may occur similarly, for example, in written exchanges.[19] To distinguish the effects of content and delivery, we run an additional experiment in which receivers read the transcript of an explanation, instead of listening to it. This effectively shuts down the effect of oral delivery, while keeping the content channel constant.

**Design.** We transcribe the explanations from our Orator experiment in a way that preserves the nuances of the spoken text, including, for example, filler words such as *um* and *eh*. The design is then identical to our baseline Receiver experiment except that *Explanation* is replaced with a *Transcript* treatment, in which participants read the transcript of a recording rather than listening to it. In each task, we randomly assign respondents to the *Choice Only* (20%) treatment or the *Transcript* treatment (80%). To keep the voice message and transcript treatments as comparable as possible, the text is displayed progressively. Example screens are provided in Appendix B.6.

across data collections. Our conclusion of a non-significant treatment effect in *Confidence* also holds when analyzing the improvement (difference between posterior and prior, thereby accounting for the variation in priors), instead of the posterior optimality rate ($p = 0.09$).

[19]Ever since Mehrabian (1971), it has been evident that non-content features may play a crucial role in the effectiveness of communication.

**Logistics.** This experiment was run with 1,266 U.S. respondents in January 2024, of whom 917 provided valid responses.

**Results.** Panel (a) of Figure 2 shows that explanation transcripts also strongly increase optimality rates relative to the *Choice Only* condition, with nearly similar effect sizes as the corresponding voice recordings. *Transcript* induces a posterior optimality rate of 62.1%, significantly above *Choice Only* (59.5%, $p < 0.01$) and not significantly different from *Explanation* (62.7%, $p = 0.37$). Looking at improvements, which net out the minor across-treatment differences in the prior optimality rate, *Transcript* produces a 2.0 p.p. ($p < 0.01$) larger increase from prior to posterior than *Choice Only*. This corresponds to approximately 58.9% of the size of the additional improvement in *Explanation* (at 3.4 p.p., $p < 0.01$). Panel (b) shows that a strong asymmetric effect between learning and unlearning also emerges in the transcript treatment. While transcripts have a strong treatment effect of 8.1 p.p. on the learning rate (at 50.9%, $p < 0.01$), the unlearning rate at is virtually unaffected relative to *Choice Only*, at 22.0% ($p = 0.43$). The size of the treatment effect of *Transcript* on the learning rate corresponds to 61.9% of the treatment effect in *Explanation*. This evidence shows that listening to a spoken explanation leads to somewhat more imitation than just reading the same explanation in learning opportunities, though the asymmetric treatment effects qualitatively emerge in both *Transcript* and *Explanation*.[20]

We conclude from this experiment that both substantive content features and the oral delivery of explanations matter for social learning, with content driving the majority of the effect.

**Result 3.** *The treatment effect of explanations on social learning is largely due to their content, with a smaller role played by non-content features conveyed orally.*

# 5 The Supply and Interpretation of Explanations

All findings about the effects of explanations up to this point were identified through experimental manipulations and did not require any assumptions about or analysis of the actual content of the explanations. The objective of Sections 5 and 6 is to shed light on the mechanisms underlying our main finding: the asymmetric treatment effect of explanations in learning versus unlearning opportunities. To that end, we go beyond the reduced-form analysis of treatment effects and directly study the rich data on explanations.

[20]Note that the *Transcript* treatment still relies on text that was originally produced in *spoken* format. In Section 2 we reviewed the systematic differences between written and spoken text production.

To structure the mechanism analysis, we systematically explore two margins of empirically measurable variation that could drive the asymmetric treatment effects: first, Section 5 asks whether differences in imitation rates might derive from content differences in the supply of explanations. Intuitively, explanations for correct choices may differ from those for incorrect choices in ways that increase the likelihood of imitation. For example, individuals who know the right answer might give "better," more compelling explanations, inducing the asymmetric treatment effect. Second, Section 6 asks whether different imitation rates could be the result of differences between the characteristics of participants in learning versus unlearning situations. Since learning and unlearning opportunities are determined by the accuracy of a respondent's answer, participants in learning situations are likely to differ systematically from those in unlearning situations. This sample selection is not a confound or flaw of the design, but a basic feature affecting *any* exchange of knowledge: different explanations are shared by people with systematically different characteristics.

## 5.1 Dissecting Explanations

To study the role of explanations' content, we examine the language used in the recordings. Such data is hard to analyze comprehensively because language is high-dimensional: each sentence and its oral delivery have innumerable features and interpretations (Batista et al., 2024).

Our analysis follows a two-pronged approach based on the following distinction: on the one hand, explanations are characterized by the substantive content of the answer to a question. Specifically, explanations frequently invoke *arguments*, defined as a series of statements with the purpose of establishing a conclusion. This content tends to be domain-specific, i.e., it directly relates to a specific question and the chosen answer. On the other hand, an explanation is characterized by (an infinite number of) text features: it exhibits speech and text metrics, markers of certainty, linguistic and rhetorical features, etc. We think of such features as domain-general, i.e., similarly pertaining to explanations for different questions and answers. These two ways of analyzing the data provide complementary perspectives that may shed light on central questions, such as whether imitation is affected more by the substantive content of an explanation or by specific speech figures, expressions or structural elements, such as length.

### 5.1.1 Argument Annotation

**Identifying arguments.** To identify arguments from the unstructured text data, we develop a coding scheme detailed in Appendix C. First, we provide a state-of-the-art Large Language Model

(LLM), OpenAI's GPT-4, with all explanations for a given task and make it identify all distinct arguments. This extraction encompasses any type of argument: not only valid or sound ones, but also fallacious and irrelevant ones. Second, based on the initial list of arguments identified by the LLM, we manually fine-tune the categories, e.g., to avoid duplicates or distinguish between variants. Third, a team of six graduate level research assistants annotates 100 responses in each of the different tasks; whenever they encounter arguments not captured by our scheme, we add them to it. This yields the final list of arguments.

**Annotating explanations.** The team of research assistants then annotates the presence of all arguments in the scheme across all 6,910 explanations. To assess the quality of the main annotation, the manual annotation is then performed again, with each task allocated to a new research assistant who is blind to the previous results. Inter-rater reliability is high. If one coder identifies a specific argument, there is a 72% chance the coder does so as well. If one coder does not identify an argument, there is a 95% chance the other coder does not either. Cohen's $\kappa$ is 0.67, indicating "substantial agreement" (Cohen, 1960; Landis and Koch, 1977). Inter-rater reliability is even higher for argument categories. When one coder identifies "any argument", there is a 100% chance that the other coder does so as well; when one coder does not identify "any argument", there is a 0% chance the other coder does so as well. These chances are 82% and 89% for "any fallacious argument" and 83% and 89%. As a second test, we performed the annotation again using GPT-4. When a human coder identifies an argument, there is a 79% chance GPT-4 does so as well. If a human coder does not identify an argument, there is a 90% chance GPT-4 does not either. Cohen's $\kappa$ is 0.61, again indicating "substantial agreement". Agreement on categories is similarly high. We view these benchmarks as validating our annotation approach.

### 5.1.2 Feature Annotation

The second part of our approach leverages a collection of 31 text features, all of which are domain-general, i.e., they similarly apply to all explanations independent of the different questions and answer options. We first annotate 25 text features using GPT-4. They are culled from the vast literature on text analysis in communications research and previous work that uses natural language data in economics and related fields. They include explicit markers of uncertainty such as modal verbs ("could", "might"), epistemic stance markers ("I believe"), hedges ("probably", "perhaps"), relative language ("almost", "nearly"), absolute language ("always") and references to certainty ("definitely", "certainly"). They also comprise implicit markers of uncertainty, such as hesitations ("um", "uh"), filled pauses ("you know", "like"), repetitions and self-corrections. We further annotate mentions of sources, references to personal experiences or authority as well as directive addresses to the receiver and apologetic phrases. We additionally compute six simple textual and speech metrics such as word count, speed of delivery and the Flesch-Kincaid language complexity score. Appendix Table A5 provides an overview of all features.

**Discussion.** We see our classifications of substantive arguments and text features as complementary to one another. For example, the substance of an argument may be unrelated to other text features, such as the speed of talking or implicit markers of uncertainty. As such, we start by examining the data on arguments and text features as independent sources of variation in explanations and study their relationship at the end of this section.

## 5.2 The Content of Explanations

We first discuss the descriptive results from our annotation approach. We then estimate how content differences affect imitation rates among receivers. We begin with task-specific arguments before turning to the domain-general text features of explanations.

### 5.2.1 Arguments: The Substance of Explanations

Figure 3 displays the results of our final argument annotation across all 15 tasks. For each task, the left-hand panel shows the frequency of a given argument separately for the sample of explanations for correct and incorrect answers, ordered by total frequency in our data. The bottom of the panel also shows the frequency of irrelevant arguments, pure restatements of the answer, any expressions of the speaker's certainty, and any non-argument reasoning. The right-hand panel displays the estimated effect of a given argument on the likelihood of imitation, separately for learning and unlearning situations. To ensure sufficient statistical power, we only show effects for arguments occurring in at least 5% of explanations in the corresponding situation.

**Illustration of results: Actively managed funds task.** To illustrate these results at the level of an individual task, we take the question on actively managed funds (top row panels).[21] The most frequent argument is that active funds—unlike passively managed ones—can quickly adapt to changes in the market, which is present in 57.2% of the explanations for incorrect answers but almost absent (2.9%) from explanations for correct answers. The second most common one argues that active funds charge higher fees, which is prominent (22.7%) among explana-

[21]The wording of the question and answer options can be found in Appendix Table A1.

tions for correct answers and nearly absent (3.2%) from those for incorrect answers. The third most frequent argument suggests that active funds are managed by experts—more common in the incorrect category—and the fourth most frequent states that it is impossible to predict stock markets—present in just less than 20% of explanations in correct and absent from incorrect. We identified eight additional unique arguments (pertaining to historical performances, differences in levels of diversification and risk levels, among other topics), all of which individually occur in fewer than 5% of the explanations. Irrelevant arguments, defined as those that have premises unrelated to the questions or answer options, are common, and slightly more so in explanations for correct (41.5%) than for incorrect answers (29.5%). Pure restatements are similarly frequent at about 10%. We identify indications of certainty in roughly 20% of all explanations.

Turning to the right-hand panel on the effects of imitation, we find that the two most common arguments have sizable effects on imitation rates, whereas most others do not. The argument that active funds can quickly adapt to changes is associated with an increase of 32.1 p.p. in the imitation rate in unlearning situations, and the argument that active funds charge higher fees increases the imitation rates by 20.7 p.p. in learning situations. We find that irrelevant arguments decrease the likelihood of imitation in both situations. Pure restatements increase imitation in learning but decrease it in unlearning. Expressions of uncertainty do not have significant systematic effects in this task, nor does non-argumentative reasoning.

**Variation across tasks: Stylized facts.** The previous perspective on a single task illustrates the richness of insights on how people reason about a topic of interest emerging from our analysis of the substance of verbal explanations. Synthesizing these findings across tasks, we document the following stylized facts.

First, we see pronounced heterogeneity in the number of distinct arguments circulating for a given topic. It ranges from a maximum of eleven arguments in the case of actively managed funds, as discussed above, to eight arguments about the relationship between interest rates and stock prices and all the way down to just two distinct arguments in the case of an exponential growth calculation. The median number of arguments across tasks is five.

Second, we see varying degrees of consensus on specific arguments across tasks. For example, none of the nine arguments identified in the task on historical stock returns exceeds a frequency of 20%, whereas we see two arguments occurring with more than 60% and 80% likelihood in the bid-ask spread question. Figure 3 also illustrates the degree of consensus across tasks.

Finally, we document large heterogeneity in the effects of different substantive arguments

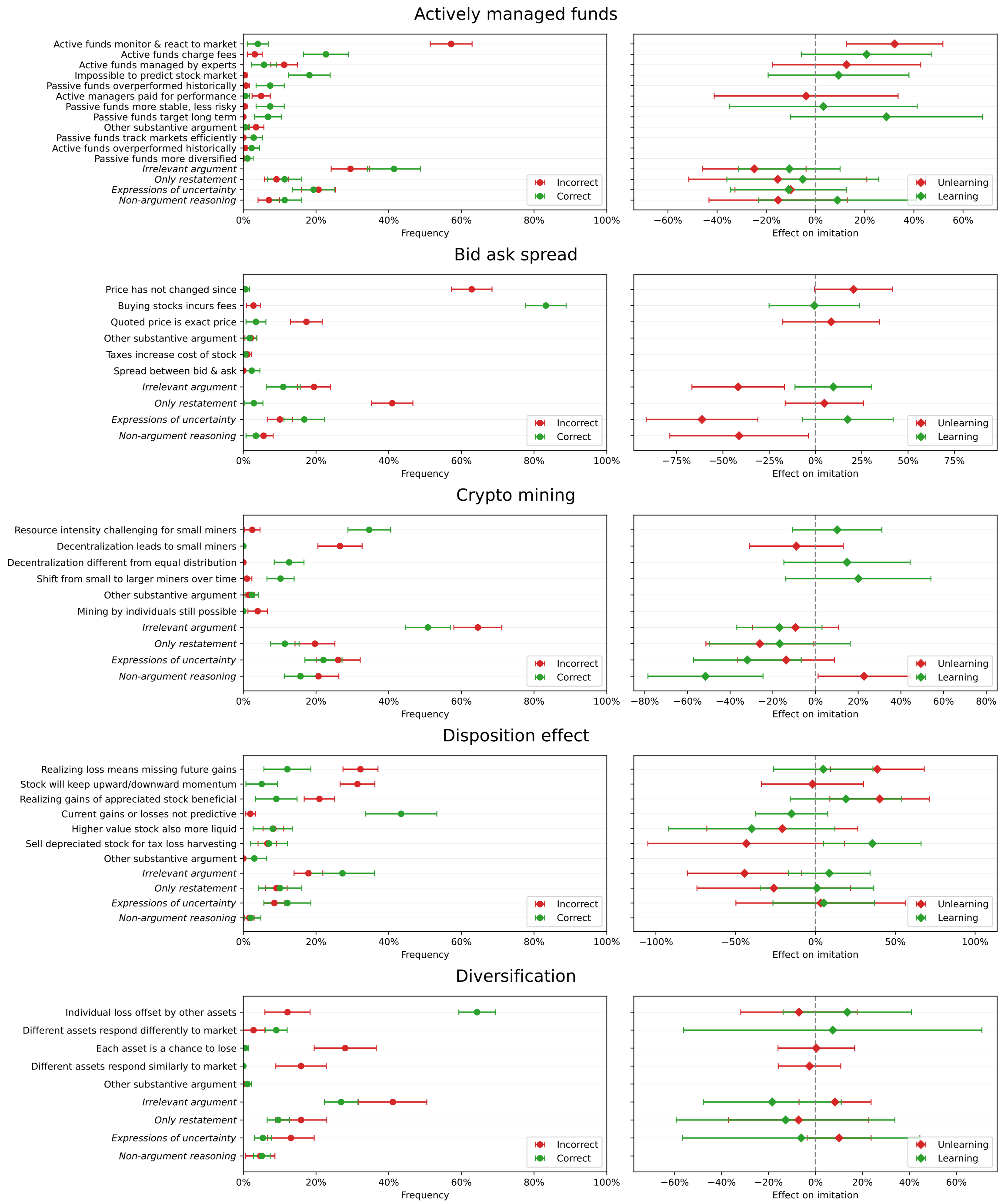



**Figure 3:** Arguments by task I/III. *Notes:* Left panel shows the frequency of an argument in correct and incorrect explanations. Right panel shows the difference-in-differences of the imitation rate between *Explanation* and *Choice Only*, between explanations that contain the argument and those that do not, in learning and unlearning situations. Only arguments appearing in more than 5% of corresponding explanations are shown in the right panel. Whiskers show 95% CI.

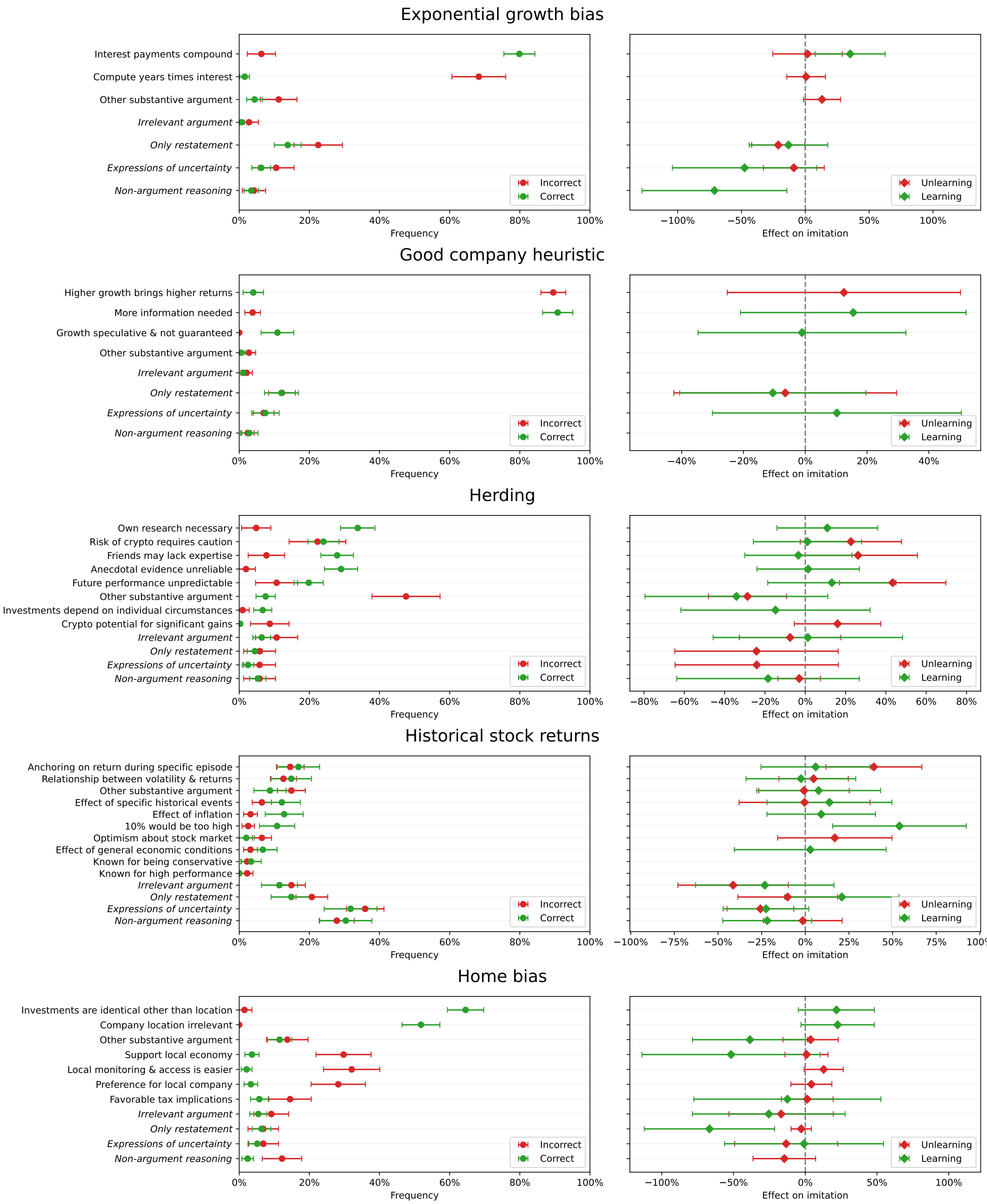



**Figure 3:** Arguments by task II/III. *Notes:* See above.

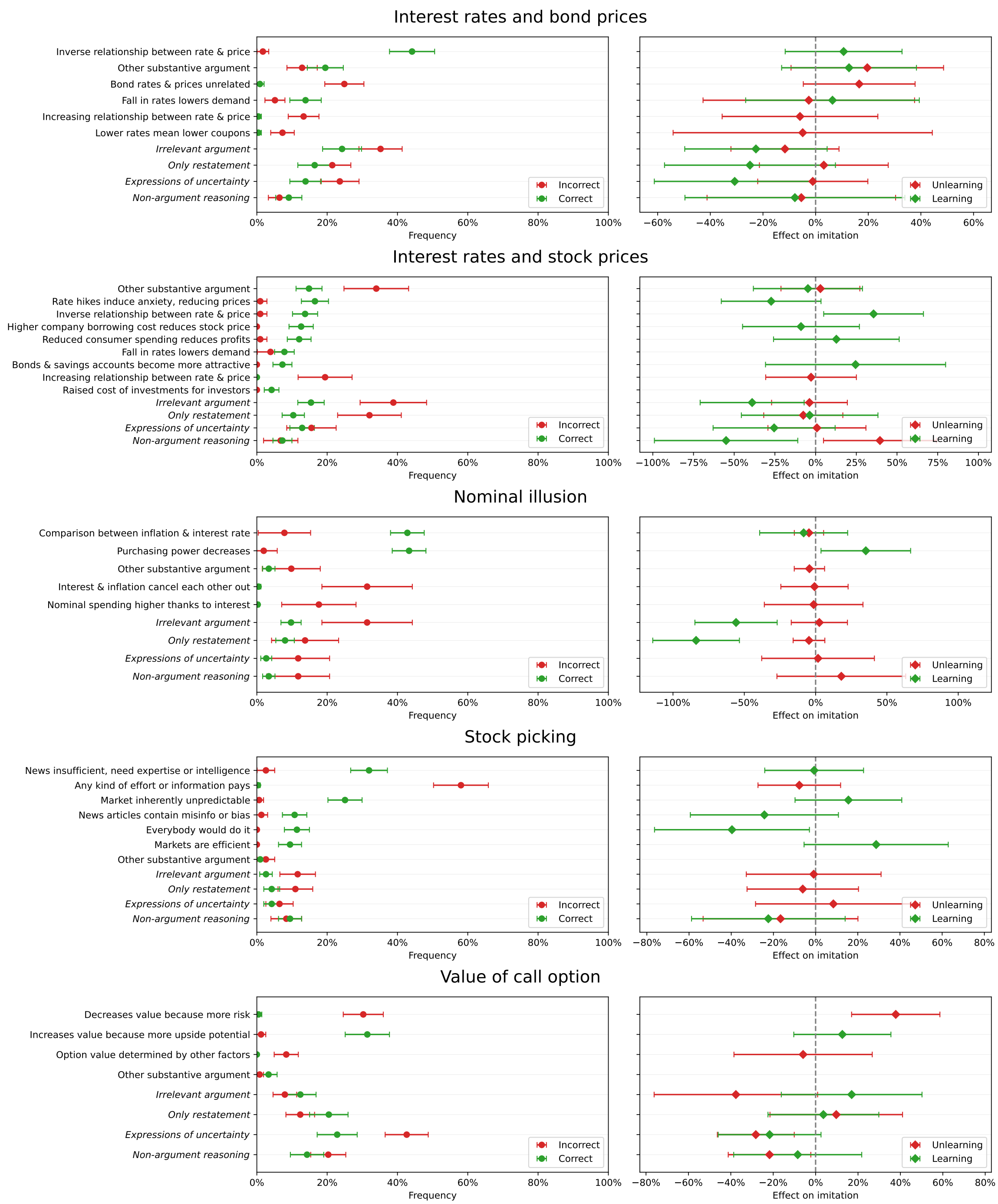



**Figure 3:** Arguments by task III/III. *Notes:* See above.

on imitation rates. In some tasks, arguments for wrong choices increase imitation rates by up to 40 p.p., while other arguments for wrong choices decrease imitation rates by up to 40 p.p. We observe similar heterogeneity for arguments in explanations for correct choices. While some arguments for correct choices increase imitation rates by more than 50 p.p., others decrease the likelihood to imitate by close to 60 p.p..

While this perspective delivers valuable descriptives for each topic as well as on the general nature of financial explanations, it remains difficult to compare substantive content across questions and this perspective does not yet shed light on the drivers of the asymmetric treatment effect. As a next step, we attempt to put more structure on the nature of different arguments that permits an aggregation of the different tasks.

**Categorizing arguments across tasks.** To draw more general conclusions about differences between explanations for correct and incorrect choices, we define four domain-general argument categories. First, we code the absence of any argument. Second, we define an argument as irrelevant if the premises are unrelated to the question or its answer, i.e., an argument might be entirely off-topic. Third, leveraging standard notions from the discipline of logic, an argument might be relevant but fallacious: one or more of the premises are false, or the conclusion is not valid given the premises. Finally, we classify as a sound argument one that has correct premises and where the conclusion follows from the premises. In classifying explanations, we embrace the fact that there are often several explanations for the correct answer that are sound. We also include "weakly sound" arguments in the "sound" category where, from a strict logician's perspective, the premises might not quite be sufficient for the conclusion. Appendix Table A7 lists the categorization of arguments in each task.

Analyzing the prevalence of different arguments classes is important as it is *ex ante* unclear whether right or wrong answers are supported by different kinds of arguments. For example, it is in principle possible to give the right answer based on fallacious or irrelevant arguments or to give a wrong answer even if some part of the explanation relies on a weakly sound argument.

To classify each explanation, we treat these four categories as hierarchical: conditional on having any argument, an explanation will be coded based on the category of the "highest-quality" argument it contains. For example, if an explanation contains both a sound and a fallacious argument, we will assign this explanation to the sound bucket. In practice, 91.8% of explanations contain at most one argument type.

**The argument gap.** The left panel of Figure 4 shows the frequency of different classes of arguments encountered in learning versus unlearning opportunities. A large fraction of explanations in unlearning opportunities contain no (21.8%), irrelevant (22.6%) or, most often, fallacious arguments (51.0%). All three types of arguments are significantly less common in learning opportunities, at 16.0% for none, 17.7% for irrelevant and 10.5% for fallacious arguments, respectively. This means that the three categories of "lower quality" explanations are more frequent in unlearning situations, with the most pronounced gap in the case of fallacious arguments. Sound arguments, by contrast, are practically absent in unlearning explanations (4.7%),[22] yet they constitute the dominant category in learning explanations (55.8%). We refer to this stark imbalance in the distribution of argument types as the *argument gap*: learning explanations contain "better" types of arguments than unlearning explanations according to this taxonomy.

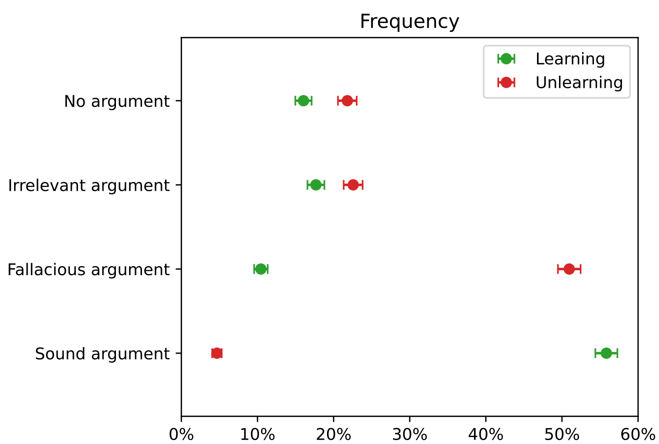



**Figure 4:** The Argument Gap: argument quality and effect on imitation. *Notes:* Left panel shows frequency of explanations by argument quality. Right panel shows differences in imitation rates between *Explanation* and *Choice Only* by argument quality. Whiskers show 95% CI.

**Argument types and the asymmetric treatment effect.** We next examine whether these different classes of arguments are associated with different effects on imitation rates. The right panel of Figure 4 displays the treatment effects associated with each category of argument separately for learning and unlearning situations. Recall that the treatment effect is calculated as the difference in the imitation rates between the *Explanation* treatment and the corresponding matches in the *Choice Only* treatment. We make the following observations.

[22]This small occurrence stems from our definition of soundness that includes arguments in which the premises might not strictly be true under all circumstance or only weakly establish the conclusion.

First, consider the treatment effects in unlearning situations. A perhaps striking feature of our original findings in Figure 1b was that there is a precisely estimated null effect of unlearning explanations in the aggregate. This could mean that listeners in these situations tend to not respond to explanations, or that the average null effect simply masks heterogeneity across different types of explanations. Figure 4 shows that the treatment effect in unlearning situations is significantly different from zero in two out of four categories. Most importantly, it is strongly positive at 4.6 p.p. ($p = 0.01$) for fallacious arguments, which account for more than half of explanations in unlearning. This implies that even a fallacious argument makes initially correct listeners significantly more likely to switch to a wrong answer. By contrast, we see a large and significant negative effect of irrelevant arguments: those make initially wrong participants 10.1 p.p. ($p < 0.01$) less likely to imitate. We further see a slightly negative effect of no argument (-2.3 p.p., $p = 0.41$). This provides a clear conclusion: the null effect of explanations in unlearning situations, on average, masks the differential effect of different types of explanations.

Second, turning to learning situations, we find a positive treatment effect on imitation across all categories. It is strongest in the most common category of sound arguments (18.8 p.p., $p < 0.01$), less pronounced for fallacious arguments (12.1 p.p., $p = 0.01$), and about the same for explanations with no (3.5 p.p., $p = 0.35$) or irrelevant arguments (4.5 p.p., $p = 0.21$). This suggests that, if it supports the right answer, *any* explanation tends to help, even if it contains no or low-quality arguments. That said, we reassuringly see good arguments are more persuasive.

Third, we analyze the difference in treatment effects between learning and unlearning situations. We see a level shift in the treatment effects of learning versus unlearning in *each* category of arguments. The magnitude of the asymmetric effect varies across categories: it ranges from 14.9 p.p. ($p = 0.02$) for sound and 14.7 p.p. ($p < 0.01$) for irrelevant arguments down to 5.8 p.p. ($p = 0.21$) for none and 7.4 p.p. ($p = 0.14$) in the case of fallacious arguments.

Finally, we can combine the results on the heterogeneity of asymmetric treatment effects with those on the argument gap—i.e., the different frequencies of the argument types between learning and unlearning situations—to ask which fraction of the overall asymmetric effect can be *explained away* by the argument classification. Comparing the coefficient estimates for the asymmetric effect in columns 1 vs. 2 in Table 1, we show that as much as 25% of the asymmetric effect is accounted for by the different composition of argument categories across learning and unlearning opportunities.

**Table 1:** Decomposition of differential learning effects

| | *Dependent variable: Imitation* | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| Explanation | -0.002 | 0.042** | -0.143*** | -0.109*** | -0.167*** | -0.126*** | -0.159*** | -0.227*** | -0.266*** |
| | (0.014) | (0.019) | (0.020) | (0.028) | (0.046) | (0.046) | (0.050) | (0.051) | (0.064) |
| Learning | 0.198*** | 0.197*** | 0.205*** | 0.199*** | 0.199*** | 0.127*** | 0.197*** | 0.118*** | 0.113*** |
| | (0.016) | (0.020) | (0.017) | (0.020) | (0.016) | (0.016) | (0.020) | (0.020) | (0.020) |
| Explanation × Learning | 0.134*** | 0.101*** | 0.055** | 0.056** | 0.091*** | 0.153*** | 0.051** | 0.075*** | 0.071*** |
| | (0.020) | (0.025) | (0.022) | (0.026) | (0.021) | (0.020) | (0.026) | (0.025) | (0.025) |
| Argument controls | | ✓ | | ✓ | | | ✓ | ✓ | ✓ |
| Richness controls | | | ✓ | ✓ | | | ✓ | ✓ | ✓ |
| Orator controls | | | | | ✓ | | ✓ | | ✓ |
| Receiver controls | | | | | | ✓ | | ✓ | ✓ |
| Observations | 8800 | 8800 | 8800 | 8800 | 8800 | 8800 | 8800 | 8800 | 8800 |
| $R^2$ | 0.092 | 0.099 | 0.112 | 0.113 | 0.107 | 0.166 | 0.118 | 0.190 | 0.195 |

*Notes:* Sample is the main Receiver survey for *Explanation* and all collections for *Choice Only*, both restricted to learning and unlearning situations. *Explanation* is a dummy for the *Explanation* treatment, *Learning* is a dummy for learning situations. All controls contain the variable itself and an interaction with *Explanation*. *Argument controls* denotes dummies for *No argument*, *Irrelevant argument*, *Fallacious argument* and *Sound argument*. *Richness* denotes richness, cf. Section C.3. Orator and receiver controls are: *Republican*, *Higher education*, *Black*, *Working*, *Age above 35*, *Male*, *(Prior) Confidence*, *Optimality on all others tasks*. We drop the 0.6% of observations with missing receiver prior confidence from all regressions.

**Discussion.** Explanations induce more imitation in learning than unlearning opportunities across the four different argument classes we examine. Why does an asymmetric effect persist across the whole range of argument types? Explanations in learning and unlearning opportunities may differ beyond the arguments they contain, which we will study in the following.

### 5.2.2 The Features of Explanations

We now turn to our second perspective on the content of explanations. Figure 5 summarizes the results from our annotation of domain-general characteristics of explanations, separately for learning and unlearning situations. The left-hand panel displays the frequency with which each feature occurs in learning and unlearning explanations. Given the variety of features, a range of distinct insights emerge. First, the results confirm a number of intuitions about how explanations for correct and incorrect explanations might compare. For example, low certainty markers—indicating low confidence—are more common among unlearning explanations but high certainty markers are more common among learning explanations. Low certainty markers appear more than twice as often overall as high certainty markers, plausibly reflecting that people understand the absence of confidence statements as indicating high confidence. Many features that are plausibly associated with a higher quality of explanations, such as empirical statements or indications of sources, are indeed more common in learning explanations. Second, we find that for the vast majority of features (24 out of 31, or 77.4%), explanations in learning situations exhibit more occurrences. Moreover, learning explanations feature higher scores in all of the quantitative text metrics, such as language complexity scores or sentence length. The central insight coming out of our feature analysis is, therefore, that explanations for correct answers reflect a *richer* message space. We will explore this insight more systematically in the following.

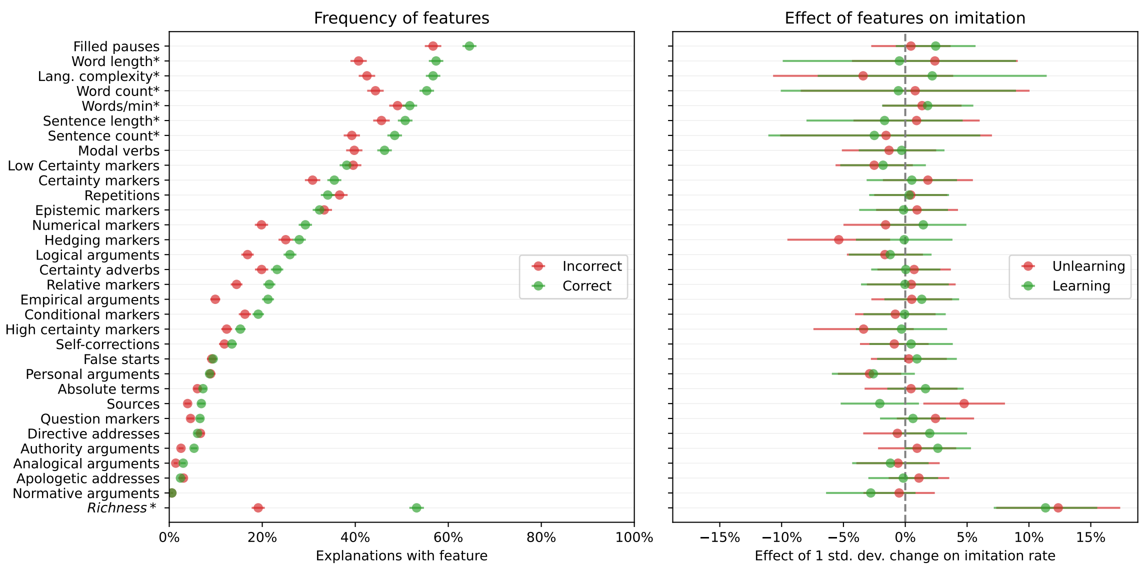



**Figure 5:** Frequency of explanation features and effects on imitation. *Notes:* Left panel shows share of explanations with features, split by orator optimality. Features with a * instead show the fraction above the overall median. Right panel shows the coefficients on *Explanation* × *Feature* in a multiple regression of imitation on *Explanation*, *Feature* and *Explanation* × *Feature*, for all listed features, applied separately to learning and unlearning situations. Whiskers show 95% CI.

**The richness gap.** We next explore the richness of explanations in learning versus unlearning situations in a more systematic way. The motivation is that (i) if explanations in learning opportunities are indeed richer and (ii) richness is associated with imitation, the richness gap may account for the asymmetric treatment effect of explanations.

The concept of richness is used in different fields across economics, most commonly to characterize the space of numerical messages in models of communication.[23] To characterize the richness of a message from natural language, we apply the following pre-registered definition in our coding instructions: "A rich explanation is detailed, comprehensive, logically structured, nuanced, and tailors the argument to fit the context. A sparse explanation is basic, narrow, unclear or disorganized, presents only surface-level understanding, lacks depth or specific details and fails to clearly relate to the context."Our coding approach relies on both human and machine coding, and follows similar procedures as our main annotation (Section 5.1). We obtain richness scores on an 11-point Likert scale, ranging from 0 to 10 (both inclusive).

We document a richness gap in explanations: the average richness score is 0.76 SD ($p < 0.01$) higher in learning than in unlearning explanations. The richness gap remains at 0.65 SD ($p < 0.01$) after controlling for differences in the length of transcripts. Appendix Figure A3 shows the distribution of richness scores in each situation as well as the richness gap separately for each of the four categories of arguments. We document a pronounced richness gap across the board: the magnitude of the gap is 0.28 SD ($p < 0.01$) for none, 0.46 SD ($p < 0.01$) for irrelevant, 0.73 SD ($p < 0.01$) for fallacious and 0.28 SD ($p < 0.01$) for sound arguments.

**The effect of text features and richness on imitation.** Are the differences in the features of explanations we document predictive of imitation? The right Panel of Figure 5 estimates the treatment effect of each feature on the imitation rates in learning and unlearning situations. These estimates are obtained from multiple regressions for learning and unlearning situations that include all features, relative to *Choice Only*. Like for our analysis of the effects of arguments, we only report estimates for features present in at least 5% of the corresponding explanations.

We document three key findings. We begin by analyzing the effects of the features annotated in our original coding approach, i.e., excluding the richness score. First, we find substantial heterogeneity in the degree to which specific features are associated with increases or decreases in imitation rates. A higher speaking pace and markers for questions are consistently associated with stronger imitation. Conversely, low certainty markers are correlated with less imitation. Indicating sources strongly positively predicts imitation in unlearning but not learning situations. Mentioning numerical markers correlates with more imitation in learning but less imitation in unlearning situations. A substantial number of features do not significantly predict imitation.

[23]While the literature uses various definitions of richness, they are related to the cardinality and/or granularity of the message space as well as the mapping between messages and states. Here, we attempt to transfer a variant of the construct to the case of messages in natural language.

Second, again abstracting from the richness score, we do not find that the raw features are jointly associated with more or less imitation in learning versus unlearning situations (in a joint F-test, $F = 0.33$ and $p = 0.56$). We only find significant (albeit small) differences in the feature coefficients for learning versus unlearning situations in 1 of the 32 variables. This suggests that differences in the degree to which these specific features lead to differential imitation are rather marginal and may not contribute significantly to the asymmetric effect in the aggregate.

Third, by far the most potent predictor of imitation is the richness score of an explanation. A 1SD increase in richness is associated with 11.4 p.p. ($p < 0.01$) and 9.5 p.p. ($p < 0.01$) increases in imitation in learning and unlearning situations, respectively, after controlling for all other features, including the length of the recording.

What does the richness score capture? In Appendix Figure A4, we systematically investigate which attributes of a text are most predictive of its richness score. Our analysis accommodates a wide array of metrics, spanning lexical features, part-of-speech and phrase structure, readability metrics, syntactic features, complexity and cohesion measures, as well as entity metrics. While many of these features influence richness, we document that *lexical* richness—the complexity, variety, and uniqueness of vocabulary—is a key determinant.

**Does the richness gap explain the asymmetric treatment effect of explanations?** Given that learning explanations are richer across the board and that richness is a strong determinant of imitation, a larger treatment effect of explanations might naturally emerge in learning situations. We examine which fraction of the asymmetric treatment effect is explained by differences in the richness of explanations encountered in learning versus unlearning situations.

Regression analyses in columns 1 to 4 of Table 1 show that across various specifications, a very significant portion of up to 65% of the asymmetric effect is explained by the richness gap. Richness remains a powerful determinant of the asymmetric effect once we also account for the role of argument categories. In fact, after accounting for richness, the estimated asymmetric treatment effect remains unchanged when additionally controlling for different argument types.

**Result 4.** *We document pronounced differences in the content of learning and unlearning explanations: learning explanations are richer and exhibit higher-quality arguments. Richness is the strongest predictor of imitation and accounts for approximately 60% of the asymmetric treatment effect of explanations.*

# 6 The Role of Orator and Receiver Characteristics

In this final step of our mechanism analysis, we turn to examining how the variation in orator and receiver characteristics across learning and unlearning situations may contribute to the asymmetric treatment effect of explanations.

This analysis is motivated once again by the fact that the asymmetric treatment effect reflects heterogeneity across an endogenous variable, because orators in learning and receivers in unlearning situations are those with a correct prior answer, while orators in unlearning and receivers in learning situations have an incorrect prior. Intuitively, the asymmetric effect might arise from characteristics of those with correct priors that, as orators, makes them more likely to be imitated or, as receivers, makes them infer systematically less from explanations.

We begin by documenting the heterogeneity of participants on observable characteristics (Section 6.1). In Section 6.2 we study which *orator* characteristics predict imitation, and whether this helps explain the asymmetric effect above and beyond the content differences associated with different groups of orators. In Section 6.3 we then test whether the characteristics of *receivers* are predictive of the responsiveness to explanations, and to what extent this contributes to the asymmetric effect in the aggregate.

## 6.1 Participant Characteristics in Learning and Unlearning

Note that because the receiver and orator samples are drawn from the same population and because the learning and unlearning samples are determined by the same endogenous variable—prior accuracy—, the orators in learning and the receivers in unlearning situations *should* have the same characteristics on average; similarly, the orators in unlearning and the receivers in learning situations should be similar. Figure 6 shows the full set of observable characteristics we elicit in the study. The left-hand panel summarizes the features of the orator sample, separately for learning and unlearning situations. The right-hand panel presents the characteristics of the receiver sample, again separately for learning and unlearning situations.

Among orators, we indeed find significant and pronounced differences across seven of the eight characteristics we examine between those with incorrect and correct answers. The first six characteristics capture participant sociodemographics. Orators with a correct prior answer have more education, are more likely to be male, less likely to be Black, and similarly likely to be Republican, to be older than 35 (the approximate median in our dataset) and to be working. The remaining two features characterize participants within the context of their answers in our study:

orators with correct answers have a substantially higher prior accuracy rate in the 14 remaining tasks (60.1% vs. 51.7%) and a higher confidence on the present task (71.8% vs. 62.4%). Among receivers, we find very similar patterns for those with correct and incorrect priors.

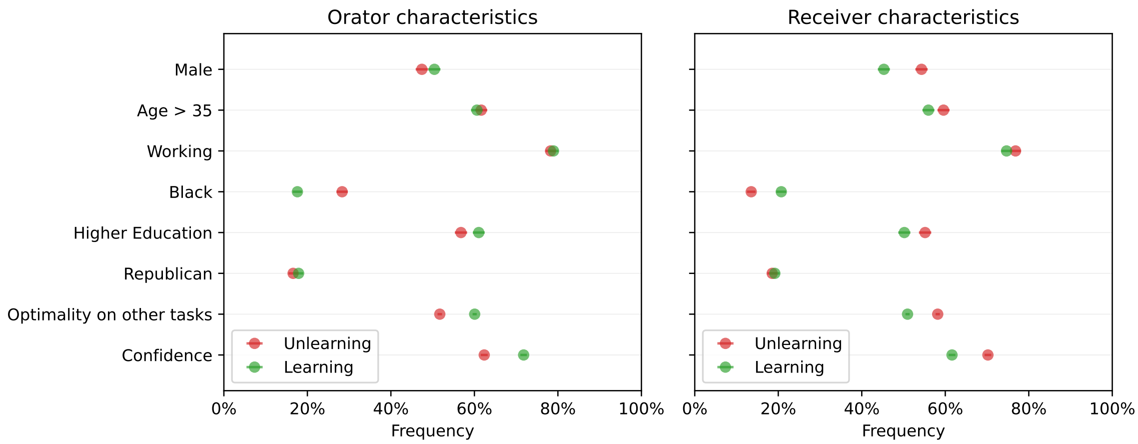



**Figure 6:** Characteristics of orators and receivers in learning and unlearning situations. *Notes: Optimality on other tasks* is the average optimality in the 14 other tasks. Confidence is rescaled from $[0, 100]$ to $[0, 1]$. 35 is the approximate median age in our data. Whiskers show 95% CI.

## 6.2 The Role of Orator Characteristics

**Heterogeneous treatment effects by orator characteristics.** In the left panel of Figure 7, we report results that examine to what extent different observables moderate the treatment effect of explanations on imitation, controlling for the accuracy of the orator's answer. Our findings show that explanations from male and highly educated speakers tend to increase imitation. In contrast, explanations from Black or older speakers result in less imitation.

The right panel investigates how the similarity in observable characteristics between the orator and the listener affects imitation rates. It reveals no significant effects of similarity, suggesting that *homophily* does not play a significant role in our setting.

In Appendix Figure A8, we report a complete breakdown of the effect of orators' and receivers' in- and out-group memberships on imitation. Similar effects of orator characteristics are also observed in the *Transcript* treatment (see Appendix Figures A6 and A9). This indicates that variations in *content* associated with different demographic groups play a crucial role.

**Do orator characteristics explain away the asymmetric effect?** We now turn to the question of whether differences between orators in learning and unlearning situations contribute

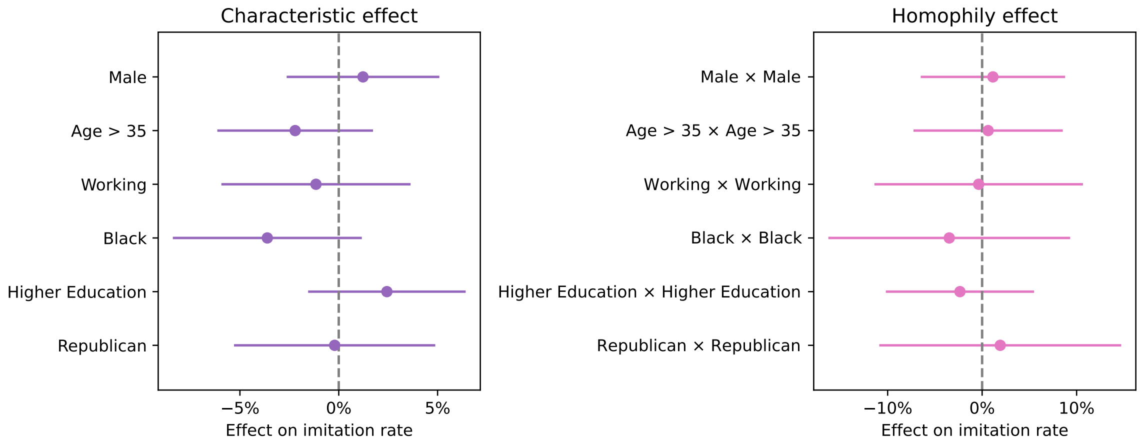



**Figure 7:** Effect of orator and orator-receiver characteristics on imitation in *Explanation*. *Notes:* Left panel shows coefficients on *Explanation* interacted with orator characteristics, in a linear regression of imitation on orator and receiver optimality, *Explanation*, orator characteristics, and *Explanation* interacted with orator characteristics. Right panel shows coefficient on *Explanation* interacted with orator-receiver characteristics, in a regression like the left panel with additional controls for receiver characteristics, orator-receiver characteristics, and *Explanation* interacted with *orator-receiver* characteristics. Orator-receiver characteristics are a dummy equal to 1 if the characteristic is shared, e.g., *Male* × *Male* is a male listening to a male. Whiskers show 95% CI.

to the asymmetric treatment effect of explanations. Intuitively, orators in learning situations may signal specific characteristics through their voice, making them more likely to be imitated. However, they also deliver different explanatory content. The preceding estimates should be interpreted as encapsulating *every* feature in the experiment that is correlated with orator characteristics, both those conveyed through transcript differences and differences in the oral delivery.

To decompose the combined effect, we examine whether observable orator characteristics account for some of the asymmetric effect above and beyond the most predictive content features identified in Section 5, richness and argument types. Column 4 of Table 1 shows that orator characteristics explain away 30% of the differential effects.[24] Yet, Column 7 of Table 1 reveals that these characteristics account for barely any of the differential effects once we account for content differences in learning and unlearning situations.

[24]Appendix Figure A7 examines which orator characteristics explain away the differential effect. It shows that prior confidence and accuracy in the other tasks are most predictive. This is consistent with our finding that the effects of orator characteristics primarily operate through the content of explanations.

## 6.3 The Role of Receiver Characteristics

We continue with the receiver side and ask how much of the asymmetric treatment effect is explained by differences in their observable characteristics. To illustrate, consider the following thought experiment. Given that the samples of listeners in learning and unlearning situations are determined endogenously, the observed differences in responsiveness to explanations may be attributed to various factors correlated with each situation, as shown in Figure 6. For example, listeners in learning situations have lower prior confidence. This not only makes them more likely to imitate in *Choice Only* scenarios but also may increase their responsiveness to explanations. Our analysis controls for these observable differences by comparing two hypothetical listeners: one in a learning situation and the other in an unlearning situation, both with otherwise identical characteristics. Do we still observe an asymmetric treatment effect in that case? Columns 5 to 9 of Table 1 study whether the observable sample differences between listeners documented above can account for the differential effect. If anything, we find that these observable differences *increase* the differential treatment effect. This is consistent with the idea that receivers with higher prior accuracy, i.e., those with unlearning opportunities, actually have a *better* assessment of whether another respondent's choice is right or wrong after listening to their explanation.[25]

These estimates may potentially underestimate the true role of sample differences between receivers in learning and unlearning situations, due to measurement errors in characteristics and unobservable features not accounted for in our analysis.

Unlike in our analogous study of the orator side, the receiver analysis is not affected by content differences, since the content is controlled by orators and, conditional on a learning or unlearning situation, randomly assigned to receivers.

**Result 5.** *Differences in observable orator characteristics in learning and unlearning situations account for as much as 30% of the asymmetric treatment effect, though these differences almost exclusively operate through differences in content supplied by different groups of orators. Corresponding differences in receiver characteristics, if anything, somewhat widen the differential treatment effect.*

# 7 Discussion and Conclusion

We examine how explanations influence the propagation of truths and falsehoods in the context of 15 financial decision-making problems. In our first experiment, one group of participants

[25]Appendix Figure A7 studies the effect of controlling for each individual characteristic separately. It shows that accounting for prior confidence and accuracy in the other tasks somewhat increases the gap.

record an explanation for each of their answers with incentives for the accuracy of their listeners' responses. In a second experiment, a separate set of respondents either only observes an orator's choice in a question or also hears one of the over 6,900 verbal explanations before potentially updating their own decisions. Our main finding on the effect of explanations on aggregate optimality rates is an optimistic one: when people talk to each other instead of just observing each other's choices, aggregate optimality rises. Notably, this improvement is entirely driven by the greater spread of truths, whereas falsehoods do not become less contagious. A comprehensive analysis of underlying mechanisms reveals that explanations for truths contain fewer fallacious and more sound arguments, and a far richer message space than explanations for falsehoods. These content differences account for approximately 60% of the differential treatment effect.

**Limitations and future directions.** The evidence in this paper may be extended in various directions. We find that explanation richness is correlated with truth, which is a central relationship underlying the overall beneficial effect of explanations. This relationship may be specific to explanations in settings with aligned incentives. Our setup might fruitfully serve as a blueprint for studying analogous patterns in the case of *persuasive messages*, where the orator wants the receiver to take a specific action. In the case of persuasion, it is conceivable that the richness-truth association in the supply of arguments weakens, or, in some situations, even reverses.

In many contexts, social interactions with others are not determined by random matching. This implies that learning and unlearning opportunities will not be equally frequent, and hence the sign of the difference between learning and unlearning rates does not serve as a sufficient statistic for whether there is aggregate improvement anymore. Moreover, in many situations, people not only listen to but also see each other. This broadens the scope of potential cues that can be used to infer the accuracy of another's advice. Furthermore, interactions are often repeated rather than one-shot, both in the dyadic back-and-forth within a conversation, and across different contexts. All of these considerations are specifically associated with interactions that occur with people that are not strangers, unlike in our experiments. This suggests a productive extension of our work to study the contagion of truths and falsehoods in real social networks.

Finally, our study focuses on the spread of truths and falsehoods in financial decision-making. While this domain is highly relevant for economists, ecologically valid to study social learning and often involves large stakes, it is possible that our findings do not carry over to other domains, such as politics. Future research should examine whether and how the supply and interpretation of explanations differ in settings where identity and motivated cognition play a central role.

## References


**Aina, Chiara**, "Tailored Stories," *Working paper*, 2023.

**Akçay, Erol and David Hirshleifer**, "Social finance as cultural evolution, transmission bias, and market dynamics," *Proceedings of the National Academy of Sciences*, 2021, *118* (26), e2015568118.

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

**Ambuehl, Sandro and Heidi Christina Thyssen**, "Competing Causal Interpretations: An Experimental Study," *Working paper*, 2024.

**__ , B Douglas Bernheim, Fulya Ersoy, and Donna Harris**, "Peer Advice on Financial Decisions: A case of the blind leading the blind?," *Review of Economics and Statistics*, 2022, pp. 1–45.

**Amelio, Andrea**, "Social Learning, Behavioral Biases and Group Outcomes," *Working paper*, 2023.

**Andre, Peter, Ingar Haaland, Christopher Roth, and Johannes Wohlfart**, "Narratives about the Macroeconomy," *Working paper*, 2022.

**__ , Philipp Schirmer, and Johannes Wohlfart**, "Mental Models of the Stock Market," 2023.

**Atkinson, Adele and Flore-Anne Messy**, "Measuring Financial Literacy: Results of the OECD/International Network on Financial Education (INFE) Pilot Study," *OECD Working papers on Finance, Insurance and Private Pensions*, 2012, *15*, 1.

**Barron, Kai and Tilman Fries**, "Narrative persuasion," *Working paper*, 2023.

**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.

**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.

**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.

**Brown, Jeffrey R, Zoran Ivković, Paul A Smith, and Scott Weisbenner**, "Neighbors matter: Causal community effects and stock market participation," *Journal of Finance*, 2008, *63* (3), 1509–1531.

**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," *Quarterly Journal of Economics*, 2023.

**Çelen, Boğaçhan, Shachar Kariv, and Andrew Schotter**, "An experimental test of advice and social learning," *Management Science*, 2010, *56* (10), 1687–1701.

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

**Cialdini, Robert B**, "The science of persuasion," *Scientific American*, 2001, *284* (2), 76–81.

_ , *Influence: The psychology of persuasion*, Vol. 55, Collins, 2007.

**_ and Noah J Goldstein**, "Social influence: Compliance and conformity," *Annual Review of Psychology*, 2004, *55*, 591–621.

**Cohen, Jacob**, "A coefficient of agreement for nominal scales," *Educational and Psychological Measurement*, 1960, *20* (1), 37–46.

**Conlon, John J, Malavika Mani, Gautam Rao, Matthew W Ridley, and Frank Schilbach**, "Learning in the Household," *NBER Working Paper No. 28844*, 2021.

**_ , _ , _ , _ , and _** , "Not Learning from Others," *NBER Working Paper No. 30378*, 2022.

**Duflo, Esther and Emmanuel Saez**, "The role of information and social interactions in retirement plan decisions: Evidence from a randomized experiment," *Quarterly Journal of Economics*, 2003, *118* (3), 815–842.

**Eliaz, Kfir and Ran Spiegler**, "A model of competing narratives," *American Economic Review*, 2020, *110* (12), 3786–3816.

**_ and _** , "News Media as Suppliers of Narratives (and Information)," *arXiv preprint arXiv:2403.09155*, 2024.

**Enke, Benjamin, Thomas Graeber, and Ryan Oprea**, "Confidence, Self-selection and Bias in the Aggregate," *American Economic Review*, 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.

**Eyster, Erik and Matthew Rabin**, "Extensive imitation is irrational and harmful," *Quarterly Journal of Economics*, 2014, *129* (4), 1861–1898.

**Fehr, Ernst and Jean-Robert Tyran**, "Individual irrationality and aggregate outcomes," *Journal of Economic Perspectives*, 2005, *19* (4), 43–66.

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

**Gick, Mary L and Keith J Holyoak**, "Analogical problem solving," *Cognitive Psychology*, 1980, *12* (3), 306–355.

**Graeber, Thomas, Christopher Roth, and Florian Zimmermann**, "Stories, Statistics, and Memory," *Quarterly Journal of Economics*, 2023.

**_, Shakked Noy, and Christopher Roth**, "Lost in Transmission," *Working paper*, 2024.

**Grunewald, Andreas, Victor Klockmann, Alicia von Schenk, and Ferdinand A. von Siemens**, "Are Biases Contagious? The Influence of Communication on Motivated Beliefs," *Working paper*, 2024.

**Haaland, Ingar and Ole-Andreas Elvik Næss**, "Misperceived Returns to Active Investing," *Working paper*, 2023.

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

**Han, Yi, David Huffman, and Yiming Liu**, "Bounded Rationality and Strategic Competition: Evidence from Gas Station Managers," *Working paper*, 2024.

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

**_, Lin Peng, and Qiguang Wang**, "News diffusion in social networks and stock market reactions," *NBER Working Paper No. 30860*, 2023.

**Hüning, Hendrik, Lydia Mechtenberg, and Stephanie Wang**, "Using Arguments to Persuade: Experimental Evidence," *Available at SSRN 4244989*, 2022.

**Hvide, Hans K and Per Östberg**, "Social interaction at work," *Journal of Financial Economics*, 2015, *117* (3), 628–652.

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

**Kendall, Chad W and Constantin Charles**, "Causal narratives," *NBER Working Paper No. 30346*, 2022.

**Landis, J Richard and Gary G Koch**, "The measurement of observer agreement for categorical data," *Biometrics*, 1977, pp. 159–174.

**Langer, Ellen J, Arthur Blank, and Benzion Chanowitz**, "The mindlessness of ostensibly thoughtful action: The role of placebic information in interpersonal interaction.," *Journal of Personality and Social Psychology*, 1978, *36* (6), 635.

**Lazer, David MJ, Matthew A Baum, Yochai Benkler, Adam J Berinsky, Kelly M Greenhill, Filippo Menczer, Miriam J Metzger, Brendan Nyhan, Gordon Pennycook, David Rothschild et al.**, "The science of fake news," *Science*, 2018, *359* (6380), 1094–1096.

**List, John A**, "Does market experience eliminate market anomalies?," *Quarterly Journal of Economics*, 2003, *118* (1), 41–71.

**Lombrozo, Tania**, "The structure and function of explanations," *Trends in Cognitive Sciences*, 2006, *10* (10), 464–470.

**Lusardi, Annamaria and Olivia S Mitchell**, "Financial literacy and retirement planning: New evidence from the Rand American Life Panel," *Michigan Retirement Research Center Research Paper*, 2007, *157.*

**Mehrabian, Albert**, *Silent Messages*, Belmont, CA: Wadsworth, 1971.

**Mobius, Markus and Tanya Rosenblat**, "Social learning in economics," *Annual Review of Economics*, 2014, *6* (1), 827–847.

**Oprea, Ryan and Sevgi Yuksel**, "Social exchange of motivated beliefs," *Journal of the European Economic Association*, 2022, *20* (2), 667–699.

**Pennycook, Gordon and David G Rand**, "The psychology of fake news," *Trends in Cognitive Sciences*, 2021, *25* (5), 388–402.

**Rozenblit, Leonid and Frank Keil**, "The misunderstood limits of folk science: An illusion of explanatory depth," *Cognitive Science*, 2002, *26* (5), 521–562.

**Russell, Thomas and Richard Thaler**, "The relevance of quasi rationality in competitive markets," *American Economic Review*, 1985, *75* (5), 1071–1082.

**Schotter, Andrew**, "Decision making with naive advice," *American Economic Review*, 2003, *93* (2), 196–201.

_ , *Advice, Social Learning and the Evolution of Conventions*, Cambridge University Press, 2023.

_ **and Barry Sopher**, "Social learning and coordination conventions in intergenerational games: An experimental study," *Journal of Political Economy*, 2003, *111* (3), 498–529.

**Schwartzstein, Joshua and Adi Sunderam**, "Using models to persuade," *American Economic Review*, 2021, *111* (1), 276–323.

**Serra-Garcia, Marta and Uri Gneezy**, "Mistakes, overconfidence, and the effect of sharing on detecting lies," *American Economic Review*, 2021, *111* (10), 3160–3183.

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

**Sloman, Steven A, Bradley C Love, and Woo-Kyoung Ahn**, "Feature centrality and conceptual coherence," *Cognitive Science*, 1998, *22* (2), 189–228.

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

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

# Online Appendix for "Explanations"

Thomas Graeber Christopher Roth Constantin Schesch

July 28, 2024

## A Conceptual Framework

We briefly set out a framework that casts our experimental setup as a standard belief formation model that speaks to the existing economics literature. It conceptualizes our reduced-form findings and provides a structure for our mechanism analyses. At the same time, it is not meant to be a micro-foundation of the structure of explanations in natural language and their interpretation.

**Prior beliefs and choice.** Consider a binary question with answers $\{0, 1\}$, where the correct answer $\omega$ is always 1.[26] A decision-maker (DM) enters with a prior belief $p = P(\omega = 1)$ that the correct answer is 1, and gives the prior answer $x = \mathbb{1}_{p>0.5}$. For simplicity, we assume the agent's prior belief relates to the prior answer according to:

$$p = \alpha_0 + \alpha_1\, x + \varepsilon \tag{1}$$

Here, $\varepsilon$ is a noise term, and we assume that $p \in [0, 1]$ for all its realizations. To build intuition, suppose that $\alpha_0 = 0$ and $\alpha_1 = 1$, and notice that correct and incorrect receivers would then be perfectly confident in their answers and could never be convinced to change their mind. If instead $\alpha_0 = \alpha_1 = 0.5$, a DM taking the correct decision is perfectly confident, whereas an incorrect DM is not confident at all but rather perfectly indifferent between answers at $p = 0.5$.

**Signal, posterior beliefs and choice.** The DM then observes the answer of another respondent, which we model as a signal $s \in \{0, 1\}$. To learn the signal, the DM needs to assign a perceived *diagnosticity* $d = \hat{\mathbb{P}}(\omega = s|s)$ to it, i.e., a belief about the probability that the observed

[26] In settings with multiple options, we simply map all incorrect options to 0.

answer matches the true state. We assume it can be represented by the functional form:

$$d = \beta_0(1-s) + \beta_1\, s + \gamma_0(1-s)\mathbb{1}_{Explanation} + \gamma_1\, s\, \mathbb{1}_{Explanation} + \delta \tag{2}$$

The parameters $\beta_0$ and $\beta_1$ reflect the baseline perceived diagnosticity of a correct and incorrect answer in the *Choice Only* treatment. $\mathbb{1}_{Explanation}$ is an indicator for the *Explanation* treatment, so that $\gamma_0$ and $\gamma_1$ encode the effect of *Explanation* on the perceived reliability of a a correct and incorrect answer. $\delta$ is a noise term, and we assume that $d \in [0,1]$ for all its realizations. More importantly, we assume that the DM never interprets an answer as evidence for its opposite:

**Assumption 1.** *For all realizations of $\delta$, the perceived diagnosticity $d$ is greater or equal to* 0.5.

The DM then forms a posterior belief $\pi = P(\omega = 1|s)$ according to Bayes' rule:

$$\pi(s=1) = \frac{pd}{pd + (1-p)(1-d)} \qquad \pi(s=0) = \frac{p(1-d)}{p(1-d) + (1-p)d} \tag{3}$$

Finally, he makes a posterior choice $y = \mathbb{1}_{\pi > 0.5}$.

**Irrelevance of confirming advice.** Taken together, Assumption 1 and Bayes' rule (3) imply that a DM receiving confirming advice $s = x$ forms a more extreme belief, i.e., $|\pi - 0.5| > |p - 0.5|$. In particular, signals confirming the prior choice do lead to an identical posterior choice:

**Proposition 1.** *If the signal coincides with the prior answer, the answer does not change, $y = s = x$.*

**Learning and unlearning situations.** This establishes that only learning and unlearning situations can lead to a change in answers. We call the learning and unlearning rates:

$$l = \mathbb{P}[\pi(s=1) > 0.5 \mid p < 0.5] \qquad u = \mathbb{P}[\pi(s=1) < 0.5 \mid p > 0.5]. \tag{4}$$

The following establishes how they depend on the parameters of the functional forms.

**Proposition 2.** *The learning rate $l$ always rises in $\alpha_0$ and $\beta_1$, and rises in $\gamma_1$ in the Explanation treatment. The unlearning rate $u$ always falls in $\alpha_0$ and $\alpha_1$, always rises in $\beta_0$, and rises in $\gamma_0$ in the Explanation treatment.*

Indeed, higher $\alpha_0$ means that an initially incorrect DM has a belief closer to 0.5, and a less diagnostic signal can therefore move them over the threshold of 0.5. $\beta_1$ captures the baseline perceived diagnosticity of a correct answer, so that these are more likely to be imitated. $\gamma_1$ encodes the additional contribution of *Explanation* to this mechanism. On the other hand, $\beta_0$ and $\gamma_0$ have no effect in learning situations.

Symmetrically, in unlearning situations, lower $\alpha_0$ and $\alpha_1$ mean an initially correct receiver is less confident. Moreover, the perceived persuasiveness of an incorrect signal increases in $\beta_0$, and additionally in $\gamma_0$ in *Explanation*, while $\beta_1$ and $\gamma_1$ have no effect.

**Optimality rates.** Next, we turn to the expected rate of correct choices across the receiver population. We denote the prior optimality rate $\theta^{prior} = \mathbb{P}(p > 0.5)$. Given the random matching mechanism of our experimental design, each DM's signal is uniformly drawn from the pool of choices. Therefore, the expected fraction of participants with a correct answer observing an incorrect signal equals $\theta^{prior}(1-\theta^{prior})$. The fraction of participants with an incorrect answer observing a correct signal equals $(1-\theta^{prior})\theta^{prior}$, which is identical. In expectation, there will always be an identical fraction of learning and unlearning opportunities. This simple but crucial insight implies that posterior optimality $\theta^{posterior} = \mathbb{P}(\pi(s) > 0.5)$ follows:

$$\theta^{posterior} = \theta^{prior} + \theta^{prior}(1-\theta^{prior}) \times (l-u) \tag{5}$$

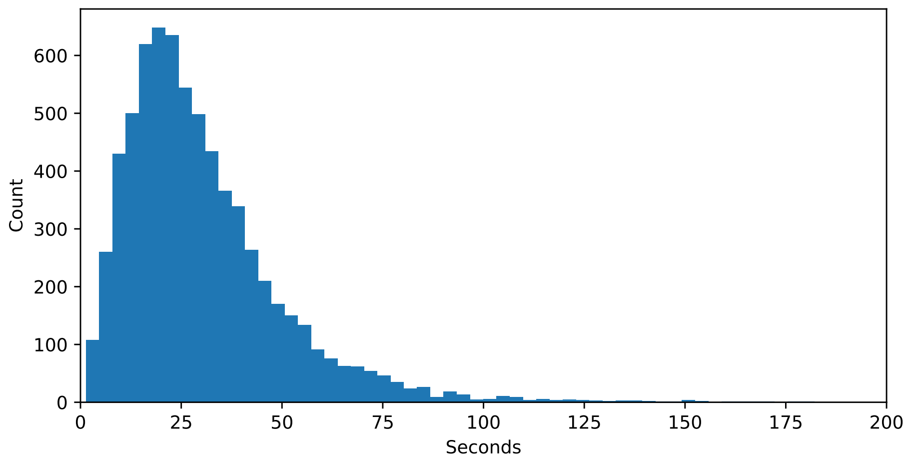
**Proposition 3.** *The posterior optimality rate exceeds the prior optimality rate if and only if the learning rate exceeds the unlearning rate. The posterior optimality rate rises with the learning rate and falls with the unlearning rate.*

This Proposition provides a simple formal justification to study learning and unlearning rates as the drivers of aggregate improvements, as in Section 3.

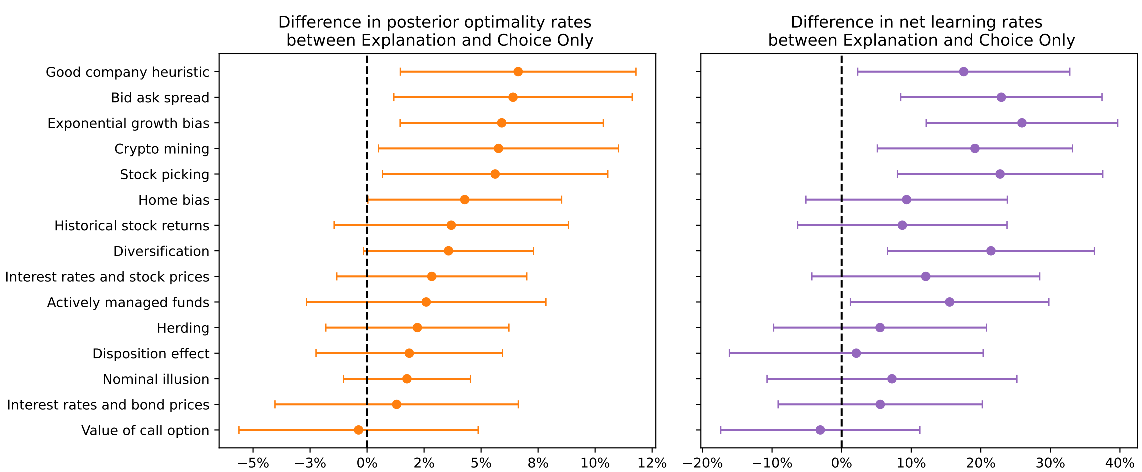
**Linking model to data.** Our main reduced-form finding in stated in Result 1 is that *Explanation* increase imitation rates compared to *Choice Only* in learning situations, but not in unlearning situations. In terms of our model, this first implies $\gamma_0 \approx 0$, i.e. that the perceived diagnosticity of incorrect answers is not higher when listening to an explanation. Second, it means that $\gamma > 1$, i.e. that explanations provide an additional way of assessing the diagnosticity of correct answers.

# B Additional figures and tables

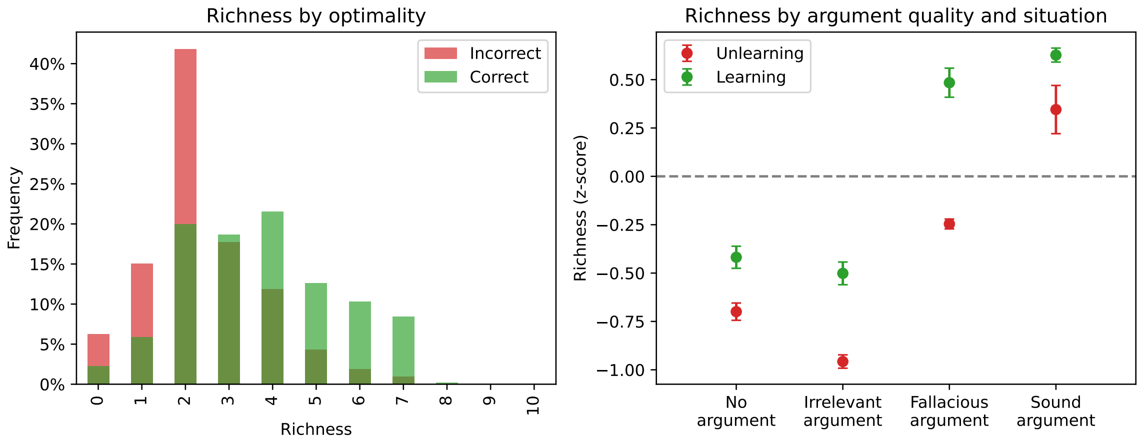
## B.1 Additional figures



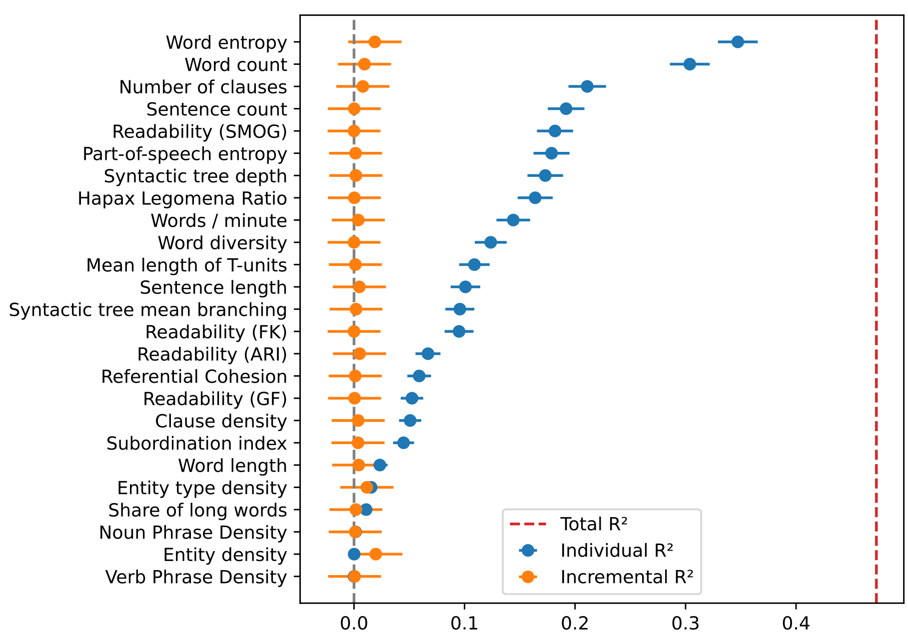

**Appendix Figure A1:** Histogram of recording lengths. *Notes:* Sample is the Orator survey with 466 orators and 6,910 valid explanations.




**Appendix Figure A2:** Difference in posterior optimality rates between *Explanation* and *Choice Only* by task, and difference in net learning rates between *Explanation* and *Choice Only* by task. *Notes:* The net learning rate is defined as the difference in imitation rates between learning and unlearning situations. *Explanation* sample is the main Receiver survey, *Choice Only* is pooled from all collections. Whiskers show 95% CI.




**Appendix Figure A3:** Richness gap by orator optimality, and by argument quality and situation. *Notes: Explanation* sample is the main Receiver survey, *Choice Only* is pooled from all collections. See Appendix C.3 for details on richness ratings. Whiskers show 95% CI.




**Appendix Figure A4:** Explanatory power of explanation features for richness. *Notes: Individual* $R^2$ is the $R^2$ in a regression of richness on a given feature and a constant. *Incremental* $R^2$ shows the decrease in $R^2$ when removing a given feature from a regression of richness on all features and a constant. *Total* $R^2$ shows the $R^2$ in a a regression of richness on all features and a constant. Whiskers show 95% CI.

## B.2 Additional tables

**Appendix Table A1:** Financial Decision Questions

| Task | Motivation | Question |
|---|---|---|
| Actively managed funds | Overestimating the return (after fees) of actively vs. passively managed funds. Adapted from Haaland and Næss (2023). | Do actively managed investment funds systematically outperform passively managed investment funds in terms of expected net returns, i.e., after accounting for investment fees? (i) Actively managed funds outperform passively managed ones. (ii) Actively managed funds do not outperform passively managed ones. |
| Bid ask spread | Assessing knowledge about features of financial transactions. | You look up live stock prices on the internet and see that the current trading price of a stock you're interested in buying is $30. You go to your online broker and buy that stock. Assuming the trading price hasn't changed in the meantime, how much do you have to pay for the stock? (i) Less than $30 (ii) Exactly $30 (iii) More than $30 |
| Crypto mining | Testing knowledge of the structure of the Bitcoin network. | Since the blockchain is decentralized, most Bitcoin mining is done by many small miners. (i) True (ii) False |
| Disposition effect | Failing to account for the random walk of stock prices. Investors have a stronger tendency to sell assets at a profit than to sell at a loss. | You have two stocks in your portfolio: one went up a lot in value since you bought it whereas the other one lost value. You need to sell one to raise cash. Is it optimal to sell the one that has lost value since you bought it? (i) Yes (ii) No (iii) This does not make a difference |
| Diversification | Assessing how investing in several different asset classes affects risk. Taken from Atkinson and Messy (2012). | When an investor spreads his money among different assets, does the risk of losing money: (i) Increase (ii) Decrease (iii) Stay the same |
| Exponential growth bias | Underestimating the exponential effects of compounding. Taken from Lusardi and Mitchell (2007). | Suppose you had $100 in a savings account and the interest rate was 2 percent per year. After 5 years, how much do you think you would have in the account if you left the money to grow: (i) More than $110 (ii) Exactly $110 (iii) Less than $110 |
| Good company heuristic | Failing to consider that market prices reflect available information, including growth prospects. | Imagine two hypothetical firms from the same industry, Firm A and Firm B, which have equal risk. However, Firm A has much higher growth prospects than Firm B. Imagine investing into one of the two firms. Which investment yields higher returns? (i) Firm A (ii) Firm B (iii) Need to know more information |
| Herding | Being influenced by "old news" from others, e.g., stories of friends, when investing. | Some of your friends with no prior experience or expert knowledge in financial markets tell you that they bought cryptocurrencies and made a lot of money with those cryptocurrencies; they mention that they bought after they came across an interesting newspaper article which describes the past price movements of cryptocurrencies. For your long-run investment strategy, how should the experience and information received from your friends influence your decision to invest (more) into cryptocurrencies? (i) Should invest more (ii) Should invest less (iii) Should not affect my decision |
| Historical stock returns | Estimating average historical returns of the S&P500. | What is the average annual return of the S&P 500 stock market index over the past 20 years? (i) Less than 10% (ii) Between 10% and 15% (iii) More than 15% |
| Home bias | Believing that firms headquartered close to home outperform better investments. | Imagine two hypothetical companies that are identical in every possible way except that one is headquartered in your home state, whereas the other one is not. Assume you're deciding between investing in one firm or the other. Which one is the better investment? (i) The firm headquartered in my home state. (ii) The firm headquartered outside of my home state. (iii) Given the assumptions, both are equally good investments. |
| Interest rates & bond prices | Assessing the interaction between interest rates and bond prices. Taken from Lusardi and Mitchell (2007). | If the interest rate falls, what should generally happen to bond prices? (i) Rise (ii) Fall (iii) Bond prices are not affected |
| Interest rates & stock prices | Assessing the interaction between interest rates and stock prices. Adaptation from Lusardi and Mitchell (2007). | When the Fed increases interest rates more aggressively than expected by markets, what should happen to stock prices on average? (i) Stock prices will rise (ii) Stock prices will fall (iii) Stock prices will stay the same |
| Nominal illusion | Failing to assess purchasing power in real terms. Taken from Lusardi and Mitchell (2007). | Imagine that the interest rate on your savings account was 1% per year and inflation was 2% per year. After 1 year, would you be able to buy: (i) More than today (ii) Exactly the same as today (iii) Less than today |
| Stock picking | Overconfidence in the value of free online news to "beat the market". Many investors actively pick stocks despite evidence that this leads to underperformance for most participants. | Most people could systematically outperform the stock market by carefully reading free online news articles about how recent events will affect different companies and picking the right stocks based on those readings. (i) True (ii) False |
| Value of a call option | Inferring how uncertainty affects the value of financial derivatives. | Holding everything else constant, how is the value of a call option for a stock generally affected by a higher volatility of that stock? (i) Higher volatility increases the value of a call. (ii) Higher volatility decreases the value of a call. (iii) Higher volatility has no effect on the value of a call. |

*Notes:* Correct answers are underlined.

**Appendix Table A2:** Overview of data collections

| Collection | Sample | Respondents | Treatments | Main outcomes | Pre-analysis plan |
|---|---|---|---|---|---|
| ***Baseline experiments*** | | | | | |
| Orator Experiment | Prolific | 466 | None | Choices in 15 financial decision tasks and voice recordings of explanations for choices | `https://aspredicted.org/56V_NLR` |
| Explanation Speech Receiver Experiment | Prolific | 1,103 | *Choice Only*<br>*Explanation* | Choices in 15 financial decision tasks | `https://aspredicted.org/56V_NLR` |
| ***Additional experiments*** | | | | | |
| Confidence Receiver Experiment | Prolific | 713 | *Choice Only*<br>*Choice & Confidence* | Choices in 15 financial decision tasks | `https://aspredicted.org/RH4_375` |
| Transcript Receiver Experiment | Prolific | 917 | *Choice Only*<br>*Transcript* | Choices in 15 financial decision tasks | `https://aspredicted.org/VPC_5NH` |

*Notes:* Sample sizes refer to the final sample of respondents that satisfied the pre-specified inclusion criteria for each collection.

**Appendix Table A3:** Decomposition of differential learning effect in *Transcript* treatment

| | *Dependent variable: Imitation* | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| Explanation | -0.009 | 0.022 | -0.134*** | -0.109*** | -0.096** | -0.032 | -0.093* | -0.124** | -0.115* |
| | (0.015) | (0.019) | (0.021) | (0.028) | (0.048) | (0.048) | (0.053) | (0.053) | (0.067) |
| Learning | 0.198*** | 0.197*** | 0.205*** | 0.199*** | 0.199*** | 0.127*** | 0.197*** | 0.118*** | 0.113*** |
| | (0.016) | (0.020) | (0.017) | (0.020) | (0.016) | (0.015) | (0.020) | (0.019) | (0.020) |
| Explanation × Learning | 0.091*** | 0.045* | 0.024 | 0.056** | 0.068*** | 0.092*** | 0.011 | 0.016 | 0.018 |
| | (0.021) | (0.026) | (0.022) | (0.026) | (0.022) | (0.021) | (0.027) | (0.026) | (0.026) |
| Argument controls | | ✓ | | ✓ | | | ✓ | ✓ | ✓ |
| Richness controls | | | ✓ | ✓ | | | ✓ | ✓ | ✓ |
| Orator controls | | | | | ✓ | | ✓ | | ✓ |
| Receiver controls | | | | | | ✓ | | ✓ | ✓ |
| Observations | 7871 | 7871 | 7871 | 8800 | 7871 | 7871 | 7871 | 7871 | 7871 |
| $R^2$ | 0.072 | 0.079 | 0.086 | 0.113 | 0.078 | 0.161 | 0.089 | 0.178 | 0.180 |

*Notes:* See notes for Table 1. *Explanation* sample is the *Transcript* survey, *Choice Only* sample is pooled from all collections. We drop the 0.3% of observations with missing receiver prior confidence from all regressions.

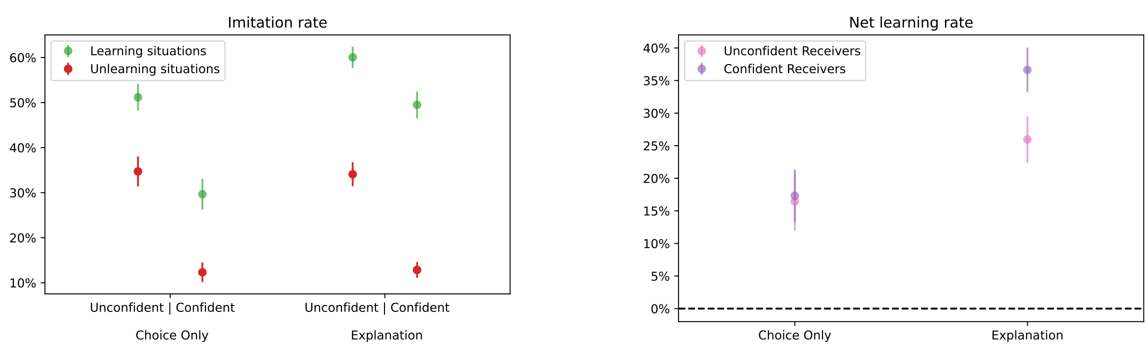
## B.3 Additional results on confidence

On average, receivers report a prior confidence of 68%, with a median of 71%. Prior confidence is 61% on average for incorrect receivers, against 72% for correct ones.[27]

We define receivers as confident or unconfident if they report a prior confidence above or below the sample median, respectively. Figure A5a then reports results from Figure 1b broken down by receiver prior confidence. Confident respondents are always less likely to imitate.

Figure A5b reports the net learning rate, which is the difference in imitation rates between learning and unlearning situations. It is positive in both treatments, but roughly two times larger in *Explanation* than in *Choice Only*. Since this difference is substantial and highly significant among confident ($p \leq 0.01$) and unconfident receivers ($p \leq 0.01$), we conclude that our reduced-form effects are not driven by a subsample of (un)confident receivers.

At the same time, the effect of explanations is stronger among confident receivers. The imitation rate stands at 15.7% for unconfident against 17.6% for confident receivers in *Choice Only*, an insignificant difference ($p = 0.53$). On the other hand, it is 25.3% for unconfident and 35.0% for confident receivers in *Explanation*, a substantial and highly significant gap ($p < 0.01$). These additional analyses show that the learning benefits of explanations show up at all confidence levels, but are especially strong among confident receivers. Put simply, confident but wrong receivers appear remarkably open to having their mind changed by a correct explanation.



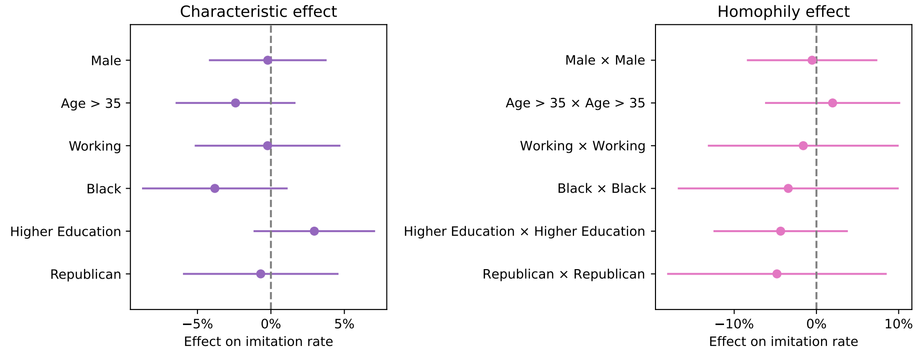

**(a)** Imitation rate by situation and confidence

**(b)** Net learning rate by confidence

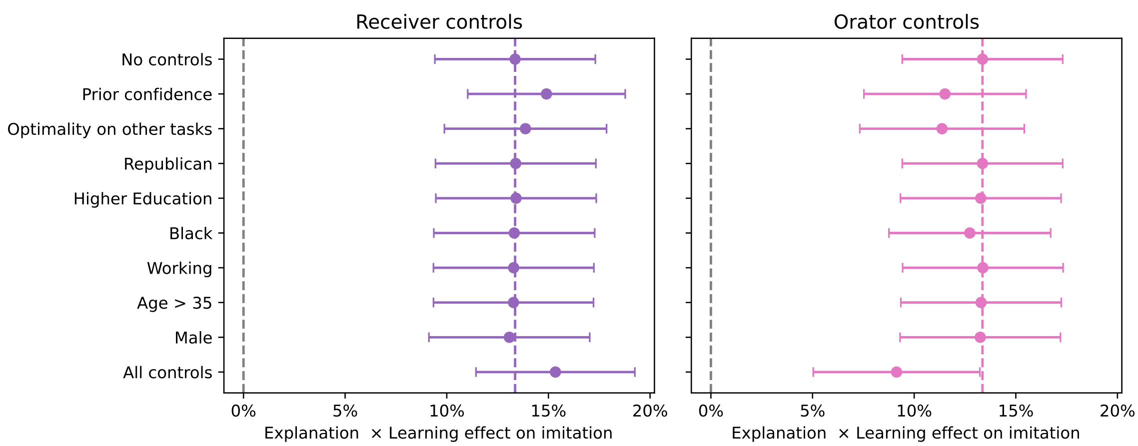
**Appendix Figure A5:** Effect of receiver prior confidence on imitation. *Notes:* Confident or unconfident receivers reported a prior confidence above or below the sample median (71%). The net learning rate is the imitation rate in learning minus unlearning. *Explanation* sample is the main Receiver survey, *Choice Only* is pooled from all collections. Whiskers show 95% CI.

[27]Because of an error in the survey which allowed respondents to skip the question, prior confidence is missing for 0.6% of the sample. We drop these observations in this Section.

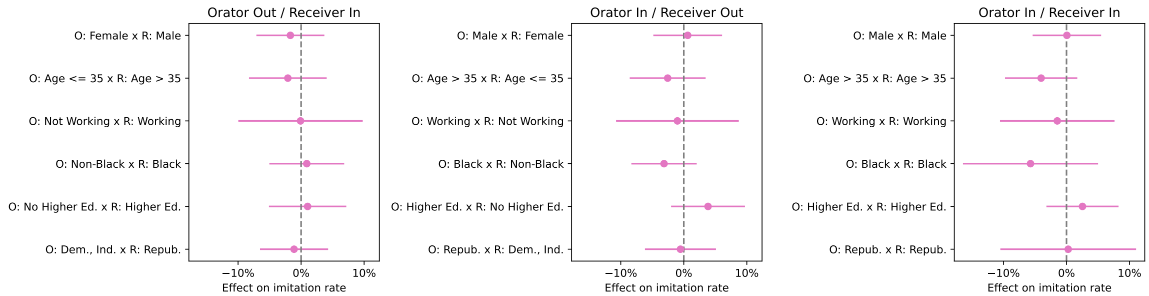
## B.4 Additional results on orator & receiver characteristics



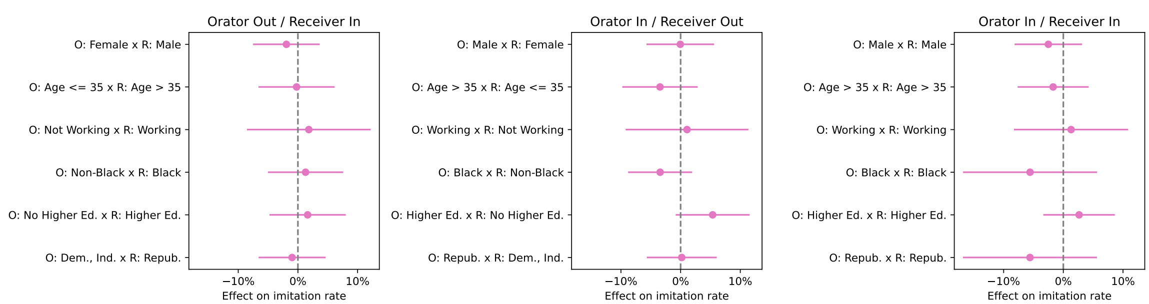

**Appendix Figure A6:** Effect of orator and orator-receiver characteristics on imitation in *Transcript* treatment. *Notes:* See notes in Figure 7. *Transcript* sample is the corresponding Receiver survey, *Choice Only* is pooled from all collections. Whiskers show 95% CI.




**Appendix Figure A7:** Learning asymmetry after controlling for orator or receiver characteristics. *Notes:* Coefficient on Explanation × Learning in a regression of imitation on Explanation, Learning, Explanation × Learning, Control and Explanation × Control. In the left panel, Controls are receiver controls. In the right panel, Controls are orator controls. All regression except 'No controls' and 'All controls' contain a single control. *Explanation* sample is the main Receiver survey, *Choice Only* is pooled from all collections, both restricted to learning and unlearning situations. See also Table 1. Whiskers show 95% CI.




**Appendix Figure A8:** Effect of orator-receiver characteristics on imitation in *Explanation* treatment. *Notes:* Coefficients on *Explanation* interacted with orator-receiver characteristics, in a linear regression of imitation on orator and receiver optimality, *Explanation*, orator-receiver characteristics and *Explanation* interacted with orator-receiver characteristics. *Explanation* sample is the main Receiver survey, *Choice Only* is pooled from all collections. Orator-receiver characteristics are an *Orator Out / Receiver In* dummy equal to 1 if the receiver has the characteristic but not the Orator; *Orator In / Receiver Out* and *Orator In / Receiver In* are analogously defined; *Orator Out / Receiver Out* is left out and serves as reference level. Whiskers show 95% CI.




**Appendix Figure A9:** Effect of orator-receiver characteristics on imitation in *Transcript* treatment. *Notes:* See notes in Appendix Figure A8. *Transcript* sample is the corresponding Receiver survey, *Choice Only* is pooled from all collections. Whiskers show 95% CI.

## B.5 Additional results from robustness checks

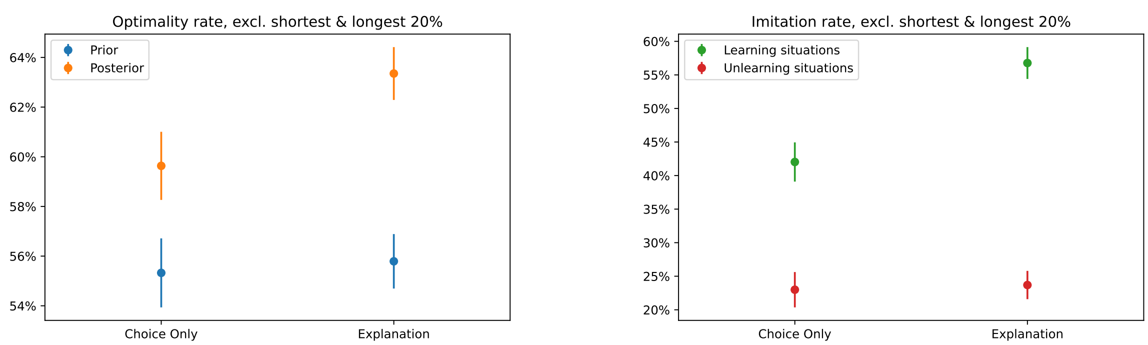
### B.5.1 Excluding the shortest and longest recordings

To ensure our reduced-form findings are not driven by a small subsample of explanations, e.g., extremely succinct or long-winded, we verify that they are robust to excluding the shortest and longest 20% of recordings. We filter recordings based on the total duration of the audio file.

Figure A10b shows the resulting optimality rates, mirroring Figure 1b. The share of receivers giving the correct answer before having been exposed to the orator's explanation is 55.3% in *Choice Only* and 55.8% in *Explanation* ($p = 0.61$). After being exposed to the orator's explanation, the share of receivers giving the correct answer is 59.6% in *Choice Only* and 63.4% in *Explanation* ($p < 0.01$). Being exposed to an orator's explanation instead of only their choice therefore increases the optimality rate by 3.3 p.p. ($p < 0.01$).

Figure A10b repeats the analysis by learning and unlearning situations as in Figure 1b. In unlearning situations, being exposed to an orator's explanation increases the likelihood of imitating their answer an insignificant 0.7 p.p. ($p = 0.68$) relative to only seeing their choice. On the other hand, in learning situations, being exposed to an orator's explanation in addition to their choice increases the likelihood of taking over their answer by 14.7 p.p. ($p < 0.01$).

In conclusion, both findings confirm that our results from Section 3 are robust to excluding the shortest and longest 20% of explanations from the sample.




**(a)** Prior and posterior optimality

**(b)** Imitation by situation

**Appendix Figure A10:** Robustness of main findings to dropping shortest and longest recordings. *Notes:* Analyses and starting samples are the same as in Figures 1a and 1b. We additionally drop the 20% of shortest and 20% of longest explanations. Whiskers show 95% CI.

### B.5.2 Heterogeneity by prior accuracy

Our analyses in Section 3 focus on the difference between learning and unlearning situations. This is motivated by our conceptual framework in Appendix A, which shows that, under the assumption that receivers faced with confirming advice do not change their answers, the difference between learning and unlearning rates is a sufficient statistic for aggregate improvements.

This assumption is largely borne out in the data. In situations where the orator and receiver are both correct, the imitation rate is 99.1% for *Choice Only* and 99.2% for *Explanation* ($p = 0.77$). The corresponding posterior optimality rates are 99.1% for *Choice Only* and 99.2% for *Explanation* ($p = 0.77$).

When the orator and receiver are both incorrect, the imitation rate stands at 82.6% and 82.2% respectively ($p = 0.73$). This is lower than when both are correct mainly because we define imitation as the receiver picking the same option as the orator, so that in the 13 tasks with three options, receivers who gave a different wrong answer from the orator's and maintained it count as non-imitators. However, the posterior optimality rates when both are incorrect stand at 2.0% and 3.8% ($p < 0.01$) respectively in *Choice Only* and *Explanation* respectively.

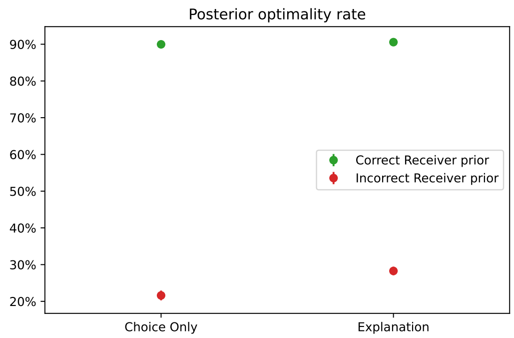
This means a small but significant number of receivers switches to the correct answer upon hearing a confirming incorrect explanation. However, this 1.8 p.p. effect is much smaller than the 13.0 p.p. effect in unlearning situations, while both situations occur similarly often (23.5% and 20.3% respectively). Decomposing the 3.2 p.p. increase in posterior optimality in *Explanation*, we find that 80% of it is driven by learning situations, 12% by situations where both are wrong and 8% by unlearning situations.

This justifies focusing on the decisive learning-unlearning margin in the rest of our analyses. We nonetheless show that our main results are robust to keeping all situations by distinguishing simply by prior accuracy, thereby analyzing the whole sample.

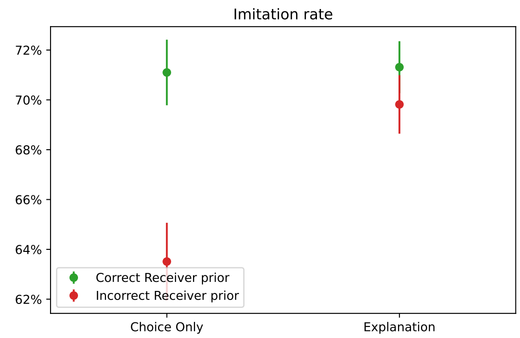
Figure A11a shows the share of receivers giving a correct answer after having been exposed to the orator's answer, split by treatment and receiver prior accuracy. Among receivers that were initially correct, posterior optimality is 90.0% for *Choice Only* and 90.6% for *Explanation* ($p = 0.29$). On the other hand, among receivers that were initially incorrect, posterior optimality rates are 21.6% for *Choice Only* and 28.3% for *Explanation*, a significant 6.7 p.p. increase ($p < 0.01$).

Figure A11b shows that initially correct receivers, imitation rates are very close at 71.1% in *Choice Only* and 71.3% in *Explanation* ($p = 0.80$), while among initially incorrect receivers, imitation rates increase by 6.3 p.p. ($p < 0.01$) from 63.5% to 69.8%.

Our main conclusion that hearing an orator's explanation has an effect *via* receivers that are initially wrong, but not *via* receivers that are already right, is therefore robust to keeping the whole sample instead of considering only learning and unlearning situations.



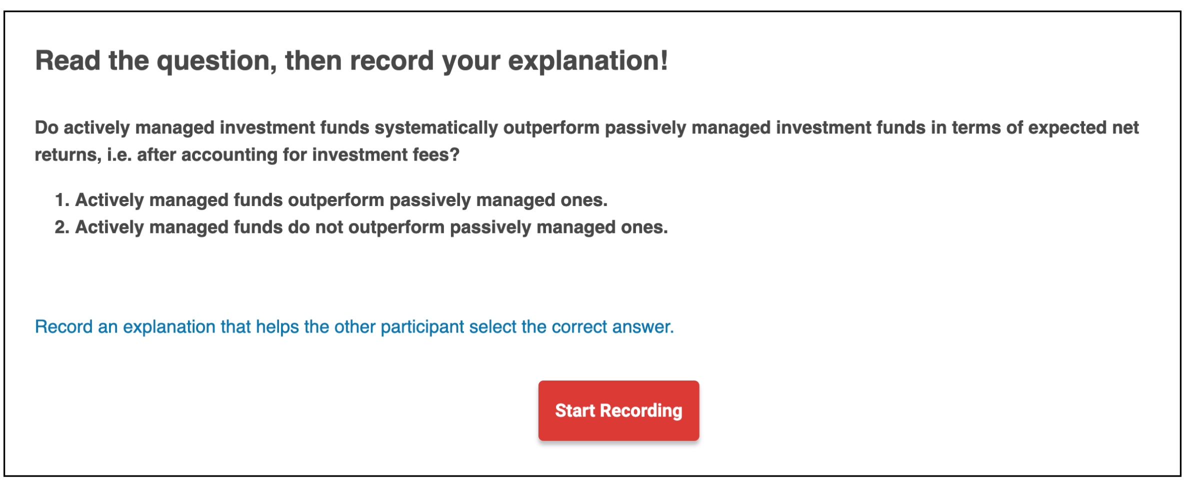

**(a)** Posterior optimality by prior optimality




**(b)** Imitation rate by prior optimality

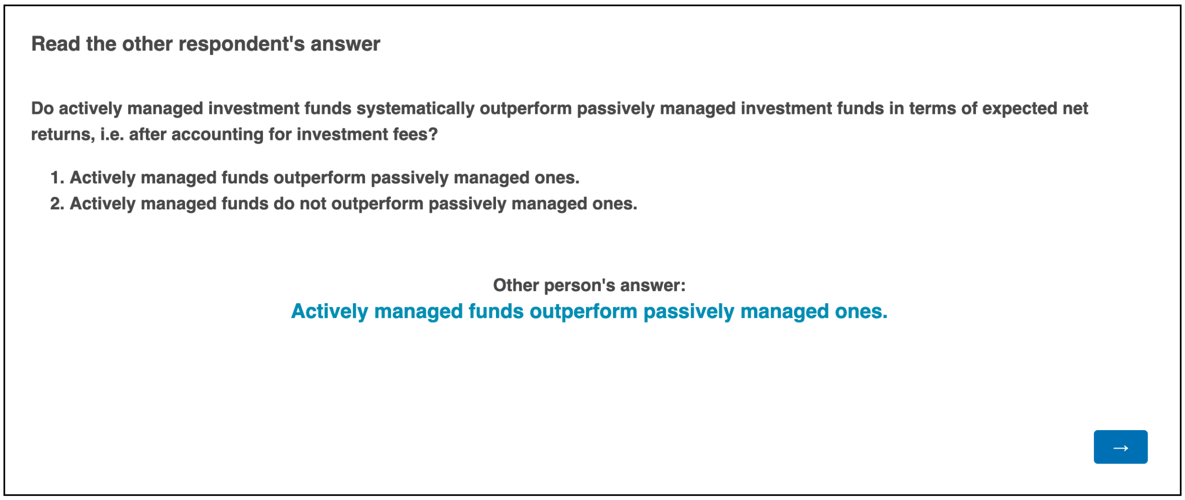
**Appendix Figure A11:** Robustness of main findings to considering receiver prior optimality. *Notes: Explanation* sample is the main Receiver survey, *Choice Only* is pooled from all collections. Whiskers show 95% CI.

## B.6 Survey screens

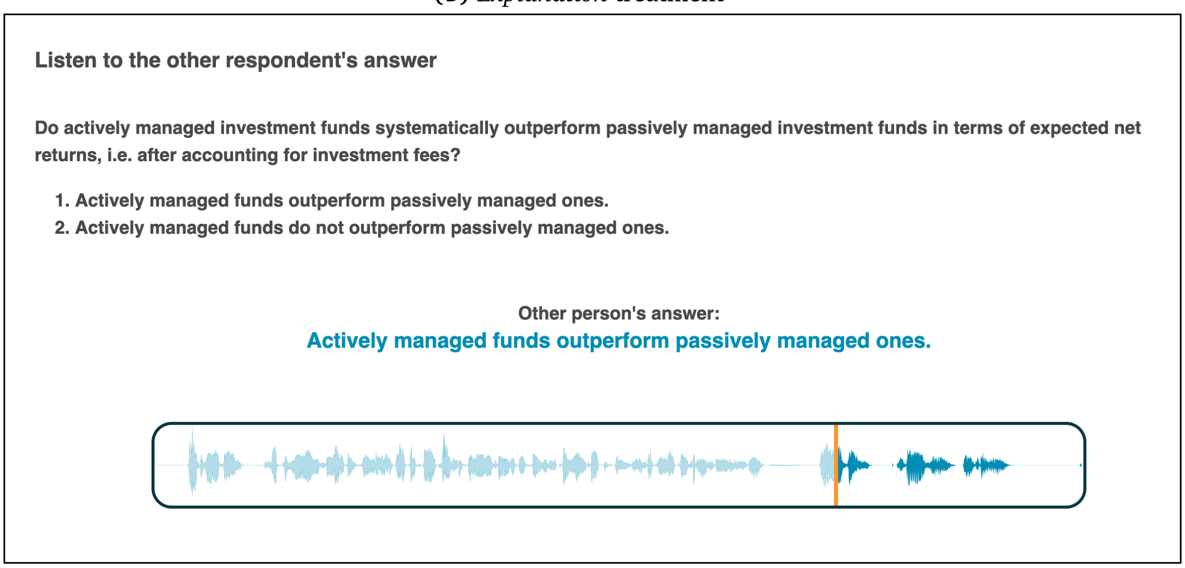



**Appendix Figure A12:** Recording screen from the Orator experiment.

**(a)** *Choice Only* treatment




**(b)** *Explanation* treatment




**Appendix Figure A13:** Observation screens from the Receiver experiment.

**Read the other respondent's answer**

**Do actively managed investment funds systematically outperform passively managed investment funds in terms of expected net returns, i.e. after accounting for investment fees?**

1. **Actively managed funds outperform passively managed ones.**
2. **Actively managed funds do not outperform passively managed ones.**

**Other person's answer:**

**Actively managed funds do not outperform passively managed ones.**

**Other person's confidence:**

**50%**

**Appendix Figure A14:** *Choice & Confidence* treatment screen from the Receiver experiment in Section 4.2.

**Do actively managed investment funds systematically outperform passively managed investment funds in terms of expected net returns, i.e. after accounting for investment fees?**

1. **Actively managed funds outperform passively managed ones.**
2. **Actively managed funds do not outperform passively managed ones.**

**Other person's answer:**

**Actively managed funds outperform passively managed ones.**

**Other person's explanation:**

All right. So I'm going to say that actively managed funds, um, actively managed funds do outperform, passively managed funds. And I'm going to say that which is an answer number one because I'm factoring in the level of risk management. So if there's risk management being actively applied to a, you know, to a, a fun then a lot of that risk that would just go on, you know, uncontrolled gets mitigated.

**Appendix Figure A15:** *Transcript* treatment screen from the Receiver experiment in Section 4.3.

# C Annotation of explanations

Our annotation starts from transcripts generated by Phonic using Amazon Transcribe. Notably, these transcripts preserve disfluencies or hesitation markers like “um” or “eh” that are typically removed by speech-to-text software. We then annotate these transcripts using a combination of human coding by a team of RA’s and machine coding by a Large Language Model (LLM). For the latter, we use the state-of-the-art OpenAI GPT-4, with a temperature set to 0 for reproducibility.

We annotate four different dimensions of explanations. First, we categorize explanations into broad categories, e.g. to distinguish restatements of the answer from non-substantive or substantive argumentation. Second, we identify a large set of 31 features in the explanations, e.g. the word count, the number of uncertainty markers or of analogical arguments. Third, we rate the general richness of explanations using a pre-registered definition. Fourth, we identify the different arguments appearing in each task and tag their presence in each explanation.

## C.1 Explanation categorization

We first categorize speeches into general categories to acquire a broad overview of the different type of explanations. For that, we asked a team of RA to identify wether an explanation fell into one of the following categories: *Only Restatement*, *Any Uncertainty*, *Non-Substantive Explanation*, *Substantive Explanation*, *Correct Explanation*, *Incorrect Explanation*, *Unclear Explanation*, *Invalid Explanation* (see Table A4 for a detailed overview). They are not necessarily mutually exclusive.

To benchmark our fully manual categorization, we then performed the same categorization with GPT-4. When the human coder identified one of the categories, GPT-4 did so too in 79% of cases; when the human coder did not identify one of the categories, GPT-4 did so too in 82% of cases. Cohen’s $\kappa$ is at 0.53, indicating ’moderate agreement’. These statistics are higher for the more specific categories we rely on in our analyses, e.g., they stand at 56%, 96% and 0.55 for the *Only Restatement* category. Aggregate frequencies also seem more stable, e.g., with human coder finding 13.1% of explanations to be Only Restatements while GPT-4 identifies a close 11.1%.

## C.2 Feature identification

We identify 31 text features in explanation, which are domain-general and were largely taken from the vast existing research on text analysis and natural language data. We extract 25 features in five categories: language markers, disfluencies, certainty markers, reasoning content and addresses to the Receiver. Some features potentially overlap, e.g. we simultaneously extract high confidence markers, low confidence markers and any confidence markers.

We instruct GPT-4 to identify all instances of each feature and return them as a JSON dictionary of lists. The annotation can then easily be audited, and appears sensible upon inspection. Instances are then counted, and counts are then standardized (intensive margin) or turned into dummies equal to 1 if any instance has been detected (extensive margin).

We generate 6 simple textual & speech features via direct computation. To investigate which features explain richness in Appendix Figure A4, we additionally generate 19 more advanced features, notably based on the distribution of words, part-of-speech tagging, named entity recognition and syntactic structure identification. Table A5 provides an overview of all features.

## C.3 Richness rating

To assess the richness of explanations, we provide GPT-4 with the following, pre-registered definition of richness: *A rich explanation is detailed, comprehensive, logically structured, nuanced, and tailors the argument to fit the context. A sparse explanation is basic, narrow, unclear or disorganized, presents only surface-level understanding, lacks depth or specific details and fails to clearly relate to the context.* We instruct GPT-4 to rate each speech's richness individually on a numerical scale from 0 to 10 (both inclusive).

## C.4 Argument identification

Section 5.1.1 describes the argument identification and annotation scheme. It also provides statistics on inter-rater-reliability, from a second blind human annotation and from an annotation via GPT-4, all showing substantial agreement. Table A7 shows all arguments appearing in the final scheme. Each has a title used to denote it in Figures and a detailed description used in the annotation.

Table A6 further shows the four types of argument we have identified. Section 5.2.1 describes how each speech is then associated with a specific argument category based on the strongest type of argument it contains.

**Appendix Table A4:** Overview of explanation categories

| **Category** | **Description** | **Example** |
|---|---|---|
| Only Restatement | The explanation is purely a restatement of the answer, without any arguments or elaborations. | "I think it's number one." |
| Any Uncertainty | The explanation contains any expressions of (un)certainty in the answer or arguments presented. | "Um, this one is more tricky. I think it's, um, I think it would be that they do not outperform passively managed ones. Um, I'm not really sure of an exact explanation because to be honest, I don't have any idea. Um, sorry" |
| Non-Substantive Explanation | The explanation only contains non-substantive justifications: appeals to authority, appeals to emotion, etc. | "I believe that passively managed funds perform better. And I'm gonna say that as uh uh as I remember Warren Buffett uh during an interview [...]" |
| Substantive Explanation | The explanation contains any substantive justification, e.g., any form of argument. | "If active funds outperformed, passive funds wouldn't exist."; "A fund is just like a plant, if you take more care of it, it will grow better." |
| Correct Explanation | The explanation is correct in meaning. | "I believe that actively managed funds do not outperform passively managed ones, the account for fees is too high when constantly monitoring an actively managed account." |
| Incorrect Explanation | The explanation is incorrect in meaning. | "Actively managed funds, do outperform, passive ones because you're actively making decisions about it and doing what makes you the most money." |
| Unclear Explanation | The explanation is very unclear or non-sensical. | "Passively managed funds, outperform, actively managed funds. And this is why hedge funds have a very short life spans. So question number two." |
| Invalid Explanation | The explanation is empty or entirely incomprehensible due to transcription errors. | 'Yes, I conquer, actively managed form. I perform passively managed forms. Every time, every time I really conquer, I good choice." |

**Appendix Table A5:** Explanation features annotated via GPT-4 or computed directly

| Feature | Description |
|---|---|
| ***Language Markers*** | |
| Modal verbs | Verbs indicating possibility, probability, or necessity. Example: “might”, “could”, “would”. |
| Certainty adverbs | Adverbs indicating certainty or doubt. Example: “possibly”, “probably”, “likely”. |
| Hedging language | Phrases indicating hedged claims. Example: “it seems”, “appears to be”, “to the best of our knowledge”. |
| Relative language | Words indicating qualifiers or comparisons. Example: “almost”, “nearly”, “more or less”. |
| Absolute language | Words indicating absolutes or superlatives. Example: “Always”, “Best”. |
| Epistemic stance markers | Phrases indicating subjective judgment. Example: “I believe”, “we assume”, “in my opinion”. |
| Conditional statements | Sentences indicating “If-Then” constructs. Example: “If we don’t act now, then”, “Assuming X, then Y”. |
| Interrogation markers | Words indicating questions or uncertainty. Example: “who”, “what”, “where”, “when”. |
| Numerical expressions | Phrases indicating quantitative or probabilistic information. Example: “more than 100 banks”, “95% chance that”. |
| ***Disfluencies*** | |
| Filled pauses | Instances of filled pauses. Example: “um”, “ah”, “er”. |
| False starts | Sentences starting but not completed. Example: “If you look at - I believe that”. |
| Repetitions | Instances of word or phrase repetition. Example: “I I mean”, “this is, this is wrong”. |
| Repairs | Instances where the speaker corrects themselves. Example: “I have two- three dogs”. |
| ***Certainty Markers*** | |
| Certainty markers | Statements indicating overall confidence. Example: “Without a doubt”, “I am certain that”. |
| High certainty markers | Statements indicating high confidence. Example: “I am certain that”, “I am sure that”. |
| Low certainty markers | Statements indicating low confidence. Example: “It might”, “I’m not sure but”. |
| ***Reasoning Content*** | |
| Indications of origin | Statements indicating information origin. Example: “According to”, “My grandmother has always said that”. |
| Personal experience args. | Arguments based on personal experience. Example: “I have often found that”. |
| External authority args. | Arguments based on external authority. Example: “My girlfriend works at a bank and said”. |
| Empirical args. | Arguments based on empirical facts. Example: “I remember reading a newspaper article saying”. |
| Analogical args. | Arguments based on analogies. Example: “Investments funds are like babies”. |
| Logical reasoning args. | Arguments based on logical reasoning. Example: “Since active managers put in more research”. |
| Normative args. | Arguments based on ethical considerations. Example: “It would not be fair if”. |
| ***Addresses to Receiver*** | |
| Directive addresses | Directives to the listener. Example: “You should definitely say that”. |
| Apologetic or humble addresses | Apologetic or humble addresses. Example: “I apologize for not knowing more”. |
| ***Simple Computed Features*** | |
| Word count | Total number of words. |
| Word length | Average length of words. |
| Words per minute | Average number of words per minute. |
| Sentence count | Total number of sentences. |
| Sentence length | Average length of sentences. |
| Language complexity | Flesch-Kincaid readability grade. |
| ***Lexical Metrics*** | |
| Lexical Diversity | Ratio of unique words to total number of words. |
| Entropy of Words | Entropy of distribution of words. |
| Hapax Legomena Ratio | Percentage of words that appear only once. |
| Share of long words | Percentage of words that have more than 10 letters. |
| ***Additional Readability Metrics*** | |
| Gunning Fog Index | Years of education required to understand the text, based on sentence length and percentage of complex words. |
| SMOG Index | Years of education required to understand the text, based on polysyllabic word counts. |
| Automated Readability Index | Years of education required to understand the text, based on characters per word and words per sentence. |
| ***Cohesion Metrics*** | |
| Referential Cohesion | Mean word overlap between sentence and following sentence. |
| ***Syntactic Complexity Metrics*** | |
| Part-of-Speech Tag Entropy | Entropy of parts of speech (e.g., nouns, verbs, adjectives) in the text. |
| Mean Length of T-units | Average length of T-units, i.e. a main clause plus any subordinate clauses. |
| Subordination Index | Ratio of subordinate clauses to main clauses. |
| Clause Density | Average number of clauses per sentence. |
| ***Named Entity Recognition*** | |
| Entity Count / Words | Ratio of number of named entities (e.g., people, organizations, locations) to total number of words. |
| Entity Type Count / Words | Number of different types of named entities (e.g., person, organization, location) to total number of words. |
| ***Sentence Structure and Syntax Metrics*** | |
| Number of Clauses | Total number of clauses. |
| Syntactic Tree Depth | Maximum depth of the syntactic dependency tree. |
| Syntactic Tree Branching | Average number of branches per node in the syntactic tree. |
| Noun Phrase Density | Ratio of number of noun phrases to total number of words. |
| Verb Phrase Density | Ratio of number of verb phrases to total number of words. |

**Appendix Table A6:** Overview of argument types

| Type | Description | Example |
|---|---|---|
| Sound Argument | An argument that has correct premises and where the conclusion follows from the premises. The premises might not quite be sufficient for the conclusion. | "I believe that actively managed funds do not outperform passively managed ones, the account for fees is too high when constantly monitoring an actively managed account." (Active funds charge fees) |
| Fallacious Argument | An argument that is relevant to the question or its answer, but where one or more of the premises are false, or the conclusion is not valid given the premises. | "Actively managed funds will outperform passively managed ones because actively managed funds make more strategic decisions. While passively managed ones are kind of just going with the flow of the market. But actively managed funds can predict what the market is gonna do and make a decision based on that. So the answer is actively managed funds outperform, passively managed ones." (Active funds managed by experts) |
| Irrelevant Argument | An argument whose premises are unrelated to the question or its answer. | "Actively managed funds, outperform, passively managed ones because they are being actively managed. Whereas passively managed ones are being managed passively and actively sounds better than passively." |
| No Argument | No argument given at all. | "Um, actively managed funds outperform passively managed ones most times probably." |

**Appendix Table A7:** Arguments Table

| Argument | Description | Category |
|---|---|---|
| ***Task: Actively managed funds*** | | |
| Active funds monitor & react to market | Actively managed funds can monitor and quickly adapt to market changes. | Fallacious |
| Impossible to predict stock market | Human inability to predict market movements, performance pressure, errors or overconfidence limit the effectiveness of active management. | Sound |
| Active funds managed by experts | Expertise in active management can lead to better investment decisions. | Fallacious |
| Active managers paid for performance | Active managers get paid because clients expect them to bring higher results than passive funds. | Fallacious |
| Active funds overperformed historically | References to historical data showing active management's performance. | Fallacious |
| Passive funds overperformed historically | References to historical data showing passive management's performance. | Fallacious |
| Passive funds more stable, less risky | Passively managed funds maintain stability by not frequently changing investments, while actively managed funds are risky investments. | Sound |
| Passive funds more diversified | Passive management benefits from diversification across a broad market index. | Sound |
| Active funds charge fees | Investment fees of actively managed funds are higher than for passive management. They reduce net returns and negate potential gains. | Sound |
| Passive funds target long term | Passively managed funds tack market trends over the longer term, so that they are better at delivering long-term growth. | Fallacious |
| Passive funds track markets efficiently | Passively managed funds can achieve long-term growth by following market trends. Passive management is efficient in tracking market performance with minimal intervention. | Sound |
| Other substantive argument | Any other substantive argument not part of the other categories. | Other |
| Irrelevant argument | Argument unrelated to the question; or no answer is implied by the argument. | Irrelevant |
| ***Task: Bid ask spread*** | | |
| Spread between bid & ask | Stocks have a bid and an ask price, and one can only buy the stock at the ask price which is always higher than the midpoint. | Sound |
| Buying stocks incurs fees | Buying stock through an online broker incurs additional fees, leading to a cost higher than the stock's listed price. | Sound |
| Quoted price is exact price | The cost of purchasing a stock is exactly the listed trading price if no fees are applied. | Fallacious |
| Taxes increase cost of stock | The cost of purchasing the stock is higher because of taxes. | Fallacious |
| Price has not changed since | Since the price hasn't changed since it was quoted, the stock can be bought at this exact price. | Fallacious |
| Other substantive argument | Any other substantive argument not part of the other categories. | Other |
| Irrelevant argument | Argument unrelated to the question; or no answer is implied by the argument. | Irrelevant |
| ***Task: Crypto mining*** | | |
| Resource intensity challenging for small miners | Bitcoin mining requires significant energy and resources, making it difficult for small miners. Large miners have an economic advantage in Bitcoin mining due to their scale and resources. Mining may not be profitable for small miners. | Sound |
| Mining by individuals still possible | Despite challenges, mining Bitcoin by individuals on a small scale is still possible, so that small miners dominate. | Fallacious |
| Decentralization different from equal distribution | Decentralization means that everyone can mine, but not that everyone mines equally, so that in practice larrge miners dominate. | Sound |
| Decentralization leads to small miners | Decentralization means there is no central planner, so that it leads to a diversity of miners, in which small miners dominate. | Fallacious |

Continued on next page

| Argument | Description | Category |
|---|---|---|
| Shift from small to larger miners over time | There has been a historical shift from small miners to large mining operations over time. | Sound |
| Other substantive argument | Any other substantive argument not part of the other categories. | Other |
| Irrelevant argument | Argument unrelated to the question; or no answer is implied by the argument. | Irrelevant |
| ***Task: Disposition effect*** | | |
| Sell depreciated stock for tax loss harvesting | Selling a stock that has lost value can be beneficial for tax purposes, allowing for tax loss harvesting. | Sound |
| Realizing loss means missing future gains | A stock that has lost value may have the potential to increase in value in the future, making it unwise to sell. Avoid selling stocks at a loss to prevent realizing the loss and potentially missing out on future gains. | Fallacious |
| Realizing gains of appreciated stock beneficial | Selling a stock that has gained value realizes the profit, ensuring a positive return on investment. | Fallacious |
| Stock will keep upward/downward momentum | One should keep the stock that has gone up and sell the stock that has gone down, because these trends can be expected to continue in the future. | Fallacious |
| Current gains or losses not predictive | Stock values fluctuate, so current losses or gains do not reflect future performance. | Sound |
| Higher value stock also more liquid | The stock with the highest value will also be more liquid, one should therefore sell that one. | Fallacious |
| Other substantive argument | Any other substantive argument not part of the other categories. | Other |
| Irrelevant argument | Argument unrelated to the question; or no answer is implied by the argument. | Irrelevant |
| ***Task: Diversification*** | | |
| Individual loss offset by other assets | Investing in multiple assets prevents total loss if one specific investment fails, akin to not putting all eggs in one basket. | Sound |
| Different assets respond differently to market | Different assets respond differently to market changes, so spreading investments can mitigate losses due to geopolitical or macroeconomic events. | Fallacious |
| Different assets respond similarly to market | Different assets usually respond similarly to market changes, so that it does not change much to invest in multiple assets instead of a single one. | Fallacious |
| Each asset is a chance to lose | Each asset is a chance to lose, so investing in multiple assets increases the chances of losing money. | Fallacious |
| Other substantive argument | Any other substantive argument not part of the other categories. | Other |
| Irrelevant argument | Argument unrelated to the question; or no answer is implied by the argument. | Irrelevant |
| ***Task: Exponential growth bias*** | | |
| Interest payments compound | The total amount in the savings account increases due to compound interest, where interest is earned on both the initial principal and the accumulated interest from previous periods. | Sound |
| Compute years times interest | A simple calculation of 2% interest per year on the initial 100, $leading to a total of$ 110 after five years without considering compound interest. | Fallacious |
| Other substantive argument | Any other substantive argument not part of the other categories. | Other |
| Irrelevant argument | Argument unrelated to the question; or no answer is implied by the argument. | Irrelevant |
| ***Task: Good company heuristic*** | | |
| Higher growth brings higher returns | Investing in the firm with higher growth prospects will yield higher returns due to its potential for growth. | Fallacious |
| Growth speculative & not guaranteed | Growth prospects are speculative and not a guaranteed indicator of future success, thus more information is needed. | Fallacious |
| More information needed | More information is needed to make a decision because the provided details are insufficient. | Sound |
| Other substantive argument | Any other substantive argument not part of the other categories. | Other |

Continued on next page

| Argument | Description | Category |
|---|---|---|
| Irrelevant argument | Argument unrelated to the question; or no answer is implied by the argument. | Irrelevant |
| ***Task: Herding*** | | |
| Future performance unpredictable | Past performance of cryptocurrencies does not guarantee future results. Timing the market correctly when investing in cryptocurrencies is not possible. | Sound |
| Own research necessary | It is important to conduct one's own research before investing in cryptocurrencies. | Fallacious |
| Risk of crypto requires caution | High volatility, risk of scams, lack of backing and other risks associated with cryptocurrencies are a reason for caution. | Fallacious |
| Friends may lack expertise | Friends providing advice may lack expertise in financial markets or cryptocurrencies. | Sound |
| Anecdotal evidence unreliable | Anecdotal evidence from friends is not a reliable basis for investment decisions, can be coincidence, luck etc. | Sound |
| Investments depend on individual circumstances | Investment decisions should be based on individual circumstances and not influenced by others. Cryptocurrencies may not be suitable for all investors. | Sound |
| Crypto potential for significant gains | Cryptocurrencies have the potential for significant gains from investing. | Fallacious |
| Other substantive argument | Any other substantive argument not part of the other categories. | Other |
| Irrelevant argument | Argument unrelated to the question; or no answer is implied by the argument. | Irrelevant |
| ***Task: Historical stock returns*** | | |
| Effect of inflation | Arguments that consider the impact of inflation on the average annual return. | Other |
| Relationship between volatility & returns | Arguments about how economic volatility affects the stock market's performance. | Other |
| Optimism about stock market | Arguments expressing a general optimism about the stock market's performance and long-term growth. | Other |
| Effect of general economic conditions | Arguments considering the general economic conditions and their impact on the stock market. | Other |
| Effect of specific historical events | Arguments considering the impact of specific historical economic events on the stock market, such as COVID-19 pandemic, recessions and subsequent recoveries etc. | Other |
| Anchoring on return during specific episode | Arguments where some remembrance of a specific or general stock returns is used as an anchor for the average return of the S&P 500. | Other |
| Known for high performance | The S&P500 is known for its high performance, which is why it has a historical average return above 10%. | Other |
| Known for being conservative | The S&P500 is known for being a popular, steady and conservative investment, which is why it has a historical return below 10%. | Other |
| 10% would be too high | Arguments based on the idea that a historical return above 10% seems too high. This can also involve the idea that, if that were true, everybody would be investing in the S&P500, which is not true and/or would reduce the return. | Other |
| Other substantive argument | Any other substantive argument not part of the other categories. | Other |
| Irrelevant argument | Argument unrelated to the question; or no answer is implied by the argument. | Irrelevant |
| ***Task: Home bias*** | | |
| Company location irrelevant | The location of a company's headquarters does not impact its investment value. | Sound |
| Support local economy | Investing in a company headquartered in one's home state supports the local economy and community. This can also happen via taxes being paid in one's home state. | Fallacious |
| Local monitoring & access is easier | Investing in a company in one's home state allows for easier monitoring and access to the company. | Fallacious |
| Favorable tax implications | The choice between investing in a home state or out-of-state company may be influenced by different tax implications. | Fallacious |
| Preference for local company | A preference or bias towards investing in companies headquartered in one's home state. | Fallacious |
| Investments are identical other than location | Both investment options are considered equally good due to the companies being identical except for location. | Sound |

Continued on next page

| Argument | Description | Category |
|---|---|---|
| Other substantive argument | Any other substantive argument not part of the other categories. | Other |
| Irrelevant argument | Argument unrelated to the question; or no answer is implied by the argument. | Irrelevant |
| ***Task: Interest rates and bond prices*** | | |
| Inverse relationship between rate & price | Since there is an inverse relationship between interest rates and bond prices, bond prices will increase when the interest rate falls. | Sound |
| Increasing relationship between rate & price | Since there is a relationship in the same direction between interest rates and bond prices, bond prices will fall when the interest rate falls. | Fallacious |
| Fall in rates lowers demand | A fall in the interest rate leads to less demand and therefore a higher price of bonds. | Fallacious |
| Bond rates & prices unrelated | Bond prices remain stable and are not influenced by fluctuations in interest rates. | Fallacious |
| Lower rates mean lower coupons | Since the interest rate determines the interest payment that bondholders get from holding the bond, the bond's value will go down if the interest rate goes down. | Fallacious |
| Other substantive argument | Any other substantive argument not part of the other categories. | Other |
| Irrelevant argument | Argument unrelated to the question; or no answer is implied by the argument. | Irrelevant |
| ***Task: Interest rates and stock prices*** | | |
| Inverse relationship between rate & price | Since there is an inverse relationship between interest rates and stock prices, stock prices will increase when the interest rate falls. | Sound |
| Increasing relationship between rate & price | Since there is a relationship in the same direction between interest rates and stock prices, stock prices will fall when the interest rate falls. | Fallacious |
| Fall in rates lowers demand | A fall in the interest rate leads to less demand and therefore a lower price of stocks. | Fallacious |
| Higher company borrowing cost reduces stock price | Higher interest rates increase borrowing costs for companies, reducing their profitability and negatively affecting stock prices. | Sound |
| Bonds & savings accounts become more attractive | Higher interest rates make bonds and savings accounts more attractive compared to stocks, leading investors to shift their investments. | Fallacious |
| Reduced consumer spending reduces profits | Higher interest rates reduce consumer spending (e.g. due to borrowing constraints), negatively affecting company profits and stock prices. | Sound |
| Raised cost of investments for investors | Interest rate increases raise the cost of investments, making it more expensive for investors and negatively affecting stock prices. | Fallacious |
| Rate hikes induce anxiety, reducing prices | Interest rate hikes make market participants uncertain and anxious, which reduces stock prices. | Fallacious |
| Other substantive argument | Any other substantive argument not part of the other categories. | Other |
| Irrelevant argument | Argument unrelated to the question; or no answer is implied by the argument. | Irrelevant |
| ***Task: Nominal illusion*** | | |
| Comparison between inflation & interest rate | Since the inflation rate is higher than the interest rate, one would be able to buy less tomorrow than today. This argument is distinct from PurchasingPowerDecrease because it displays no understanding of the mechanisms behind inflation and interest, and is based solely on a comparison of numbers. | Sound |
| Purchasing power decreases | Even though the amount of money in a savings account has increased thanks to the interest rate, the price at which one needs to buy goods and services will have increased more because of the comparatively higher inflation, so that the net effect on real purchasing power is higher. This argument is distinct from NumericalComparison because it displays an understanding of the mechanisms behind inflation and interest, not just a comparison of numbers. | Sound |
| Nominal spending higher thanks to interest | Because the amount of money in the savings account has increased thanks to the interest rate, one would be able to spend more than today. | Fallacious |
| Interest & inflation cancel each other out | Because the interest rate and inflation rate both cancel out, one would be able to buy exactly as much tomorrow as today. | Fallacious |
| Other substantive argument | Any other substantive argument not part of the other categories. | Other |
| Irrelevant argument | Argument unrelated to the question; or no answer is implied by the argument. | Irrelevant |

Continued on next page

| Argument | Description | Category |
|---|---|---|
| ***Task: Stock picking*** | | |
| Everybody would do it | If it was possible to outperform the stock market by reading free online news, everybody would be doing it. | Sound |
| Markets are efficient | All publicly available information is already factored into stock prices, so that markets will already have adjusted to stale news. | Sound |
| News articles contain misinfo or bias | News articles can contain misinformation or bias, leading to poor investment decisions. | Sound |
| Market inherently unpredictable | The stock market is inherently volatile and unpredictable, making systematic outperformance difficult. | Sound |
| News insufficient, need expertise or intelligence | News are not enough for everyday people to outperform the stock market, since, for example, they also need to be specially smart, to have financial expertise and/or to have access to other sources of information. | Fallacious |
| Any kind of effort or information pays | Any kind of effort, research or information will help to outperform the stock market. | Fallacious |
| Other substantive argument | Any other substantive argument not part of the other categories. | Other |
| Irrelevant argument | Argument unrelated to the question; or no answer is implied by the argument. | Irrelevant |
| ***Task: Value of call option*** | | |
| Increases value because more upside potential | Higher volatility in a stock increases the potential for larger price movements, which can be advantageous for call option holders seeking to profit from upward stock movements. | Sound |
| Decreases value because more risk | Higher volatility is seen as increasing risk, making the call option less attractive and decreasing its value due to the unpredictability of stock price movements. | Fallacious |
| Option value determined by other factors | The volatility of a stock has no direct effect on the value of a call option because the call option's value is determined by other factors, not just the stock's volatility. | Fallacious |
| Other substantive argument | Argument unrelated to the question; or no answer is implied by the argument. | Other |
| Irrelevant argument | Any other substantive argument not part of the other categories. | Irrelevant |

# Survey Instructions for "Explanations"

Thomas Graeber Christopher Roth Constantin Schesch

July 28, 2024

## Orator Experiment

### Survey Overview

In this survey, we will ask you to record voice messages to answer different questions. The study is designed for computer (PC or Mac) users only (desktop, laptop, etc.).

Please make sure you are in a quiet environment. You will only receive your completion payment if your voice is clearly recorded.

You are eligible for a bonus of $10 for one of your recordings! The details of how you might receive a bonus will be explained on the following pages.

For each question, your recording will actually be played to another respondent. The other respondent will have to answer the same question as you after listening to your recording.

#### Privacy & Anonymity

All voice messages are treated strictly anonymously. They will never be linked to your person and will never be published anywhere. This data will be used solely for academic research. You can therefore talk freely and informally in each voice message. The other respondent who will listen to your voice recording will not be given any more information about you: your anonymity will be preserved.

#### Test Your Microphone

Use the recorder below to test your microphone. Click "Record", say the sentence "The dog runs in the park.", then click "Stop Recording & Submit". You may have to give your browser permission to access the microphone after you click "Record". After a recording, it might take the website a few seconds to upload your recording: please be patient.

## Trial Question

This trial question is an attention check. To be eligible to participate in this study, we simply ask you to record a voice message that lasts at least 20 seconds on your thoughts about this topic. There is no correct answer, and it does not matter what you say, as long as you record a message that lasts at least 20 seconds.

On how many days in 2024 will the average temperature recorded across all of the U.S. be below 65°F?

## General Instructions

Thanks for recording your first voice message! This study will take approximately 30 minutes to complete. You will earn a reward of $6.00 for completing the survey. To complete the study, you will need to read all instructions carefully and correctly answer the comprehension questions.

#### Survey Structure

In this study, you will be asked to answer 15 questions on various topics. Questions will have two or three possible options. Exactly one of the options is the correct answer. For each question, you will be asked to record yourself once to give advice on the question and explain your reasoning.

We are interested in how you would give advice in an informal conversation:

- You should share an explanation behind your response.
- Your recording will be played to a few other participants who will have to respond to the same question.
- Other participants can win a bonus for selecting the correct answer.

**Importantly:**

- You should first read the question, think about your response, and then record your answer.
- The recording begins once you click “Start Recording”.
- After you click to submit a recording, it can take a little while to upload. We kindly ask you to be patient.

We ask you not to search the answers on the internet:

- We are interested in the explanations behind your answer.
- To confirm that you do not search for answers, the survey will monitor whether the survey window remains active.
- If you leave the browser tab of this survey, you will not be eligible for the $6.00 reward.
- You should remain focused on the survey window and answer questions as best you can using your previous knowledge.

**Bonus Payment**

At the end of the survey, one out of every ten participants is randomly selected to be eligible for an additional bonus of up to $10. If you are selected for the bonus payment:

- One of the 15 questions you have answered will be randomly chosen.
- You will receive the bonus of $10 if the participant selected the correct answer.
- After you click to submit a recording, it can take a little while to upload. We kindly ask you to be patient.

One of the participants who listened to your answer will be randomly chosen. You should therefore give your explanation in a way that makes the other respondent most likely to select the correct answer!

Much like you, participants listening to your recordings will have a chance to win a bonus of $10 if they select the right question in a randomly selected round. Moreover, participants listening to your recordings will be informed that you will receive a bonus if they select the correct answer.

This study will take approximately 30 minutes to complete. You will earn a reward of $6.00 for completing the survey. To complete the study, you will need to read all instructions carefully and correctly answer the comprehension questions.

## Comprehension Questions

Please answer the comprehension questions below. Note that if you fail them twice in a row, you will not be eligible for the completion payment.

In this study, you will record a number of voice messages on different questions. Which one of the following statements is true?

- I should answer as if I'm talking to myself because the recording will never be played to another person.
- I should give a well-rehearsed response as if I'm giving a speech to a large audience of people that I don't know.
- I should share an explanation behind my response as if I were to give advice in an informal conversation. My response will be shared with other participants who later have to respond to the same question.

Which one of the following statements is true?

- On questions where I will be recorded, each recording only starts once I click "Start Recording" on a page.
- On questions where I will be recorded, each recording starts as soon as I enter a page.

Which one of the following statements is true?

- I can leave the tab of the experiment to search for answers online without consequences on my payment.
- If I leave the tab of the experiment, I will not be paid.

## Remember!

Your chances of receiving the bonus payment are highest if the other participant chooses the correct answer.

## Main Part: Example Task (Inflation)

On the next page, a question will be displayed. You should first read the question, think about your response, and then record your answer. The recording begins once you click "Start Recording". After recording your advice, you will select your own answer to the question.

**Read the question, then record your explanation!**

Imagine that the interest rate on your savings account was 1% per year and inflation was 2% per year. After 1 year, would you be able to buy:

1. More than today
2. Exactly the same as today
3. Less than today

Record an explanation that helps the other participant select the correct answer.

[Recording box, activated manually]

**Provide your best answer**

Please answer what you think is the correct answer to the question.

[Question with multiple choice response]

How certain are you that your above answer is correct?

[Slider from 0% (Not at all certain) to 100% (Fully certain)]

## Additional Questions

Your answer to the following question will not affect your reward or bonus payment for this study, so please answer honestly. Did you search the answer to any of the 15 questions before providing your advice or your own answer?

- Yes
- No

[Followed, on a separate page, by elicitation of sex, age, ethnicity, education, employment and political affiliation.]

# D Receiver Experiment

## Survey Overview

In this survey, you will be asked questions on various topics. Besides giving your answer to a question, you will sometimes listen to a voice message of someone else's thoughts on the question. The study is designed for computer (PC or Mac) users only (desktop, laptop, etc.) and only works on Firefox and Chrome.

### Voice Messages from Previous Participants

For some of the questions, you will first listen to a voice message from another participant. In a previous survey, we asked respondents to record their thoughts on the same questions that you will be asked.

### Test Your Speaker

Use the play button below to test your speaker. Click "Play" to play back a voice message and select the sentence that you heard in the text box below.

[Player of a recording of someone saying "The dog runs in the park."]

### Attention Check

Please select the sentence that you listened to in the voice message above:

- The koala climbs up the tree.
- The dog runs in the park.
- The lion looks at the gazelle.
- The cat waits for the mouse to come back.
- The fox sneaks through the garden.
- The turtle swims in the sea.

## General Instructions

This study will take approximately 30 minutes to complete. You will earn a reward of $6.00 for completing the survey. To complete the study, you will need to read all instructions carefully and correctly answer the comprehension questions.

**Survey Structure**

In this study, you will be asked to answer 15 questions on various topics. Questions will have two or three possible options. Exactly one of the options is the correct answer. In each round, there are four steps:

1. You provide your best answer to the question.
2. You get information about a previous respondent's answer:
   - For some questions, you will listen to a voice message of another person once.
   - For other questions, you will see the answer of another participant to the question.
3. You have a second chance to provide your best answer to the question. Your answer may or may not be different from your response in (1), given what you learned about the other participant's answer in (2).

When you enter a page with a recording, the recording will play automatically. You will only be able to listen to it once.

**Bonus Payment**

At the end of the survey, one out of every ten participants is randomly selected to be eligible for an additional bonus of up to $10. If you are selected for the bonus payment:

- One of the 15 rounds you have answered will be randomly chosen.
- Either your answer from step (1) or your answer from step (3) will be randomly chosen.
- You will receive the bonus of $10 if you selected the correct answer.

Participants who made the recordings were informed they had a chance to win a bonus of $10 if you selected the correct answer. They were also informed that you had a chance to win a bonus of $10 if you selected the correct answer.

## Comprehension Questions

Please answer the comprehension questions below. Note that if you fail them twice in a row, you will not be eligible for the completion payment.

In this study, you will listen to a number of voice messages on different questions. Which one of the following statements is true?

- Before each of the 15 questions I'll be asked, I will listen to a voice message from another respondent.
- Before each of the 15 questions I'll be asked, I will see what another respondent answered.
- On some of the 15 questions I'll be asked, I might listen to a voice message from another respondent. On other questions, I will see what the respondent answered.

Which one of the following statements is true?

- On questions with a voice recording, each recording only starts playing once I click "Play" on a page.
- On questions with a voice recording, each recording starts playing automatically when I enter a page, so I should read the question and pay attention to the recording.

Which one of the following statements is true?

- The answers I chose have no effect on my expected bonus.
- I maximize my expected bonus by selecting my best answer to each question.

## Main Part: Example Question (Inflation)

**[Prior Choice:] Provide Your Best Answer**

Please answer what you think is the correct answer to the question.

Imagine that the interest rate on your savings account was 1% per year and inflation was 2% per year. After 1 year, would you be able to buy:

1. More than today
2. Exactly the same as today
3. Less than today

How certain are you that your above answer is correct?

[Slider from 0% (Not at all certain) to 100% (Extremely certain)]

**[Explanation Treatment:]**

Now, you will listen to a recording of a voice message from a previous respondent who shares the explanation behind their answer to the exact same question that you just answered. The voice message will automatically start playing.

Please listen closely to the recording.

You will be able to proceed to the next page once the recording has finished playing.

[PAGEBREAK]

Listen to the other respondent's answer

[Box with question text]

Other person's answer:

[Answer of other respondent]

[Recording of other respondent, on auto-play]

**[Choice Only Treatment:]**

Now, you will observe the answer from a previous respondent.

Please pay close attention to the other person's answer.

[PAGEBREAK]

Read the other respondent's answer

[Box with question text]

Other person's answer:

[Answer of other respondent]

**[Posterior Choice:] Provide Your Best Answer**

Your answers on this page may or may not be different from your previous response, given what you learned about the other participant's answer.

Please answer what you think is the correct answer to the question.

Your answer may or may not be different from your previous response, given what you learned about the other participant's answer.

[Question with multiple choice answer]

Your answer is correct if you selected the right answer.

How certain are you that your above answer is correct?

[Slider from 0% (Not at all certain) to 100% (Fully certain)]

## Additional Questions

Did you look up any answers on the internet? Your response to this question will not affect your payment. Please answer truthfully.

- Yes
- No

## Additional Information

The explanations you just listened to likely differed systematically in how rich or sparse they were. Rich explanations include substantial details on the reasoning and tend to be elaborate, while sparse explanations provide limited details.

Which statement do you most agree with? Over the course of this experiment, I learned about whether a given answer is correct...

- ...more from sparse explanations than from rich explanations.
- ...more from rich explanations than from sparse explanations.
- ...equally much from rich and sparse explanations.

Why do you think this is the case?

[Open text box]

[Followed, on a separate page, by elicitation of sex, age, ethnicity, education, employment and political affiliation.]