

# ONLINE APPENDIX FOR “EXPLANATIONS”

Thomas Graeber

Christopher Roth

Constantin Schesch

September 15, 2025

This Online Appendix contains a full list of identified arguments (Section F) and additional results on confidence (Section G) and respondent characteristics (Section H) in our main experiment. Furthermore, it provides additional results on our Richness (Section I) and Balls-and-urns (Section J) experiments. Finally, it contains survey screens (Section K) and full survey instructions (Sections L, M, N, O, P, Q, R).

## F List of identified arguments

Table F1 shows all naturally-occurring arguments in the 15 financial reasoning tasks we study in our main experiment, as identified by our annotation scheme detailed in Appendix B.

**Online Appendix Table F1:** Arguments Table

Argument	Description	Category
<b><i>Task: Actively managed funds</i></b>		
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

Continued on next page

Argument	Description	Category
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 track 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
<b>Task: Bid ask spread</b>		
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
<b>Task: Crypto mining</b>		
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 large 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
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
<b>Task: Disposition effect</b>		

Continued on next page

Argument	Description	Category
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
<b>Task: Diversification</b>		
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
<b>Task: Exponential growth bias</b>		
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
<b>Task: Good company heuristic</b>		
Higher growth brings higher returns	Investing in the firm with higher growth prospects will yield higher returns due to its potential for growth.	Fallacious

Continued on next page

Argument	Description	Category
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
Irrelevant argument	Argument unrelated to the question; or no answer is implied by the argument.	Irrelevant
<b><i>Task: Herding</i></b>		
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
<b><i>Task: Historical stock returns</i></b>		
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

Continued on next page

Argument	Description	Category
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
<b>Task: Home bias</b>		
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
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
<b>Task: Interest rates and bond prices</b>		
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
<b>Task: Interest rates and stock prices</b>		
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

Continued on next page

Argument	Description	Category
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
<b>Task: Nominal illusion</b>		
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
<b>Task: Stock picking</b>		
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

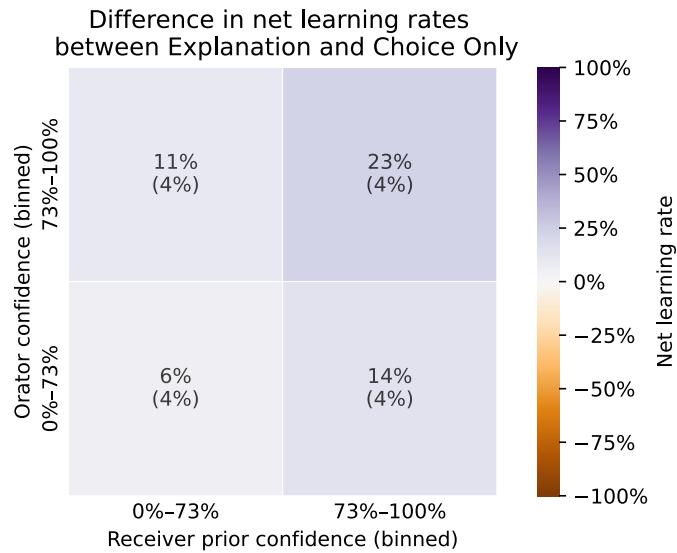
Continued on next page

Argument	Description	Category
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
<b>Task: Value of call option</b>		
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

## G Additional results on confidence

### G.1 Differences in net learning rates

Figure G1 shows how imitation by receiver and orator confidence combine by showing the difference in net learning rate between *Explanation* and *Choice Only* by orator and receiver prior confidence. The net learning rate is defined as the difference in imitation rates between learning and unlearning situations. We split both variables by the aggregate median confidence of 74%. We find that confidence effects largely compound: the explanation effect on net learning is not significant when both orator and receiver are below the median; it is significant and of similar magnitude when either one is above the median; and it is highest when both are confident.



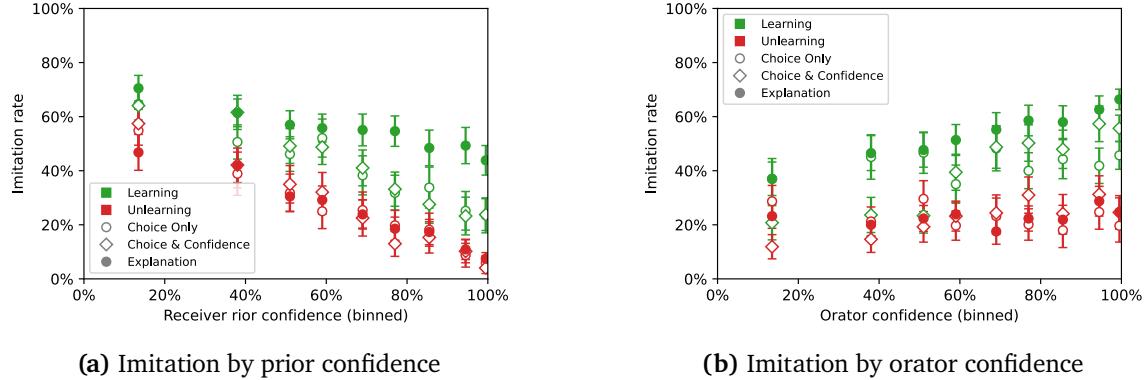
**Online Appendix Figure G1:** Effect of prior and orator confidence on imitation. *Notes:* Parentheses show SEs. *Explanation* sample is the main Receiver experiment (1,103 receivers, 13,111 obs.), *Choice Only* is pooled from all collections (2,733 receivers, 8,232 obs.).

### G.2 Confidence and the content of explanations

We shed light on the mechanism behind these patterns by analyzing the interplay of confidence on the content of explanations.

We first repeat our previous analysis while also including the *Choice & Confidence* treatment. In Figure G2a, the effect of *Choice & Confidence* and *Choice Only* by receiver prior appears identical. Confidence signals do not help confident but wrong receivers

adjust their answer. Although Figure G2b may appear to show a slightly larger effect of *Choice & Confidence* relative to *Choice Only* for confident orators in learning situations, the difference is not significant above ( $p = 0.37$ ) or below ( $p = 0.08$ ) median confidence. Numerical certainty statements therefore do not explain the effect of confidence, highlighting the importance of explanation content.<sup>33</sup>



**Online Appendix Figure G2:** Effect of prior and orator confidence on imitation in *Choice & Confidence* treatment. *Notes:* See notes for Figure G1. Sample for *Choice & Confidence* is the corresponding receiver experiment (713 receivers, 8,522 obs.). Whiskers show 95% CIs.

We therefore turn to the role of orator confidence on content differences in explanations. Figure G3a plots the average number of high certainty markers (e.g., “I am certain that”, “I am sure that”) by orator confidence. The number of markers increases with confidence, but there is no clear difference between correct and incorrect explanations. Low certainty markers (e.g., “It might”, “I’m not sure but”), shown in Figure G3b, conversely decrease with confidence, though right and wrong explanations do not significantly differ over most of the confidence range.

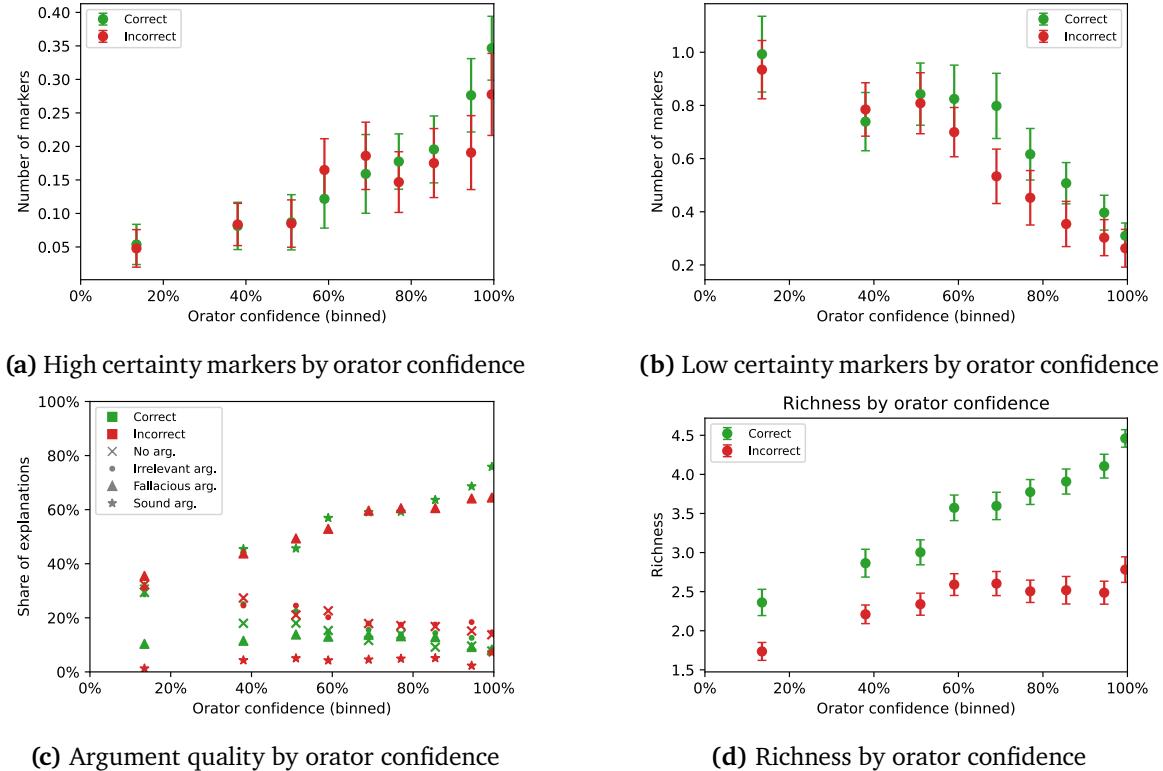
We then examine the effect of confidence on intellectual content by computing the shares of argument qualities in Figure G3c. The presence of sound arguments for correct explanations and of fallacious arguments for incorrect explanations clearly trend up in orator confidence. Other categories, and in particular no argument, in turn trend down, though there is no significant differentiating pattern between situations or confidence.

Finally, we focus on our main predictive variable by displaying richness by orator confidence in Figure G3c. We document three patterns. First, richness slopes up in orator confidence. Second, there is a richness gap between correct and incorrect explanations at all levels of confidence. Third, this gap becomes larger when orators become more

<sup>33</sup>Interestingly, even in the *Choice & Confidence* treatment which makes the latter very salient, there is no effect of orator confidence on imitation in unlearning situations.

confident. More precisely, the richness gap between correct and incorrect explanations is 0.48 SD ( $p < 0.01$ ) for orators with a confidence below the median against 0.91 SD ( $p < 0.01$ ) with a confidence above it, a 0.42 SD difference ( $p < 0.01$ ) which is slightly more than half the size of the unconditional richness gap.

In conclusion, even within learning and unlearning situations, we uncover important differences in the content of explanations given by orators with differing certainties.



**Online Appendix Figure G3:** Explanation content differences by orator confidence in learning and unlearning situations. *Notes:* Sample is the Orator experiment (466 orators, 6,910 obs.). For (c), see argument categorization in Section 5.2.1. Whiskers show 95% CIs. In Figure (c), CIs are not shown for legibility, but all SEs are smaller than 4.4%.

### G.3 Richness imitation and confidence

Just as there are substantial content differences driven by the orator side, receivers could differentially imitate explanations depending on their own and the orator's confidence. In particular, they could react differently to richness, which summarizes a broad range of explanation features.

Table G2 repeats our main decomposition, from Columns (1) and (7) in Table 1, separately for confident and unconfident receivers. Columns (1) and (2) reflect our

earlier finding that explanations have a learning effect for both, but that it is about two times larger for confident receivers. Columns (2) and (4) show that the imitation of (standardized) richness is approximately the same for both, around 10 p.p., with an insignificant difference ( $p = 0.09$ ). Since the richness gap between correct and incorrect explanations experienced by receivers is the same in both groups, richness therefore drives the same amount of learning asymmetry, of approximately  $0.1 \times 0.744 = 7.4$  p.p.. This means it effectively explains the entire learning asymmetry for unconfident receivers, but only half of it for confident receivers. Put differently, confident receivers receive a benefit from explanations that is not explained by differential imitation of richness.

Turning to orators, we repeat the same decomposition separately for confident and unconfident orators in Table G3. Columns (1) and (2) again reflect our finding that explanations have a roughly two times larger learning effect when given by confident orators. Columns (2) and (4) establish that the imitation of richness is approximately the same in both samples in this case too, around 10 p.p., also with an insignificant difference ( $p = 0.31$ ). Since we have found that the richness gap is 0.48 SD among unconfident orators but two times larger at 0.91 SD among confident ones, richness drives a two times larger learning effect among the latter. In turn, the residual not explained by richness is also roughly two times larger. In conclusion, the difference in explanation richness accounts for about half of the higher imitation of confident correct orators.

In conclusion, we have found that explanations have a learning benefit for confident and unconfident receivers, that this benefit is larger for more confident receivers, but that this effect is not driven by differential imitation of richness. It therefore reflects either content differences other than richness or a greater propensity to change their answer based on the arguments contained in explanations. In parallel, we have documented that explanations also have a larger learning benefit when they are delivered by confident orators, and that about half of this effect is driven by the fact that the richness gap between correct and incorrect orators is larger when they are more confident, which leads listeners to imitate correct and confident orators more.

**Online Appendix Table G2:** Richness imitation by prior confidence

	Dependent variable: Imitation			
	Receiver prior confidence $\leq 73\%$	Receiver prior confidence $> 73\%$	(1)	(2)
Intercept	0.347*** (0.017)	0.339*** (0.019)	0.118*** (0.011)	0.122*** (0.012)
Explanation	-0.008 (0.022)	0.055** (0.023)	0.007 (0.014)	0.048*** (0.016)
Learning	0.165*** (0.023)	0.175*** (0.025)	0.169*** (0.024)	0.163*** (0.025)
Explanation $\times$ Learning	0.098*** (0.030)	0.013 (0.031)	0.196*** (0.030)	0.126*** (0.031)
Richness		-0.016 (0.013)		0.008 (0.012)
Explanation $\times$ Richness		0.122*** (0.016)		0.085*** (0.017)
Observations	4849	4849	3940	3940
R <sup>2</sup>	0.054	0.077	0.133	0.155

*Notes:* Sample is the main Receiver experiment 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, *Richness* is standardized explanation richness. 73% is the median pooled orator and receiver prior confidence. We drop the 0.6% of observations with missing receiver prior confidence from all regressions.

**Online Appendix Table G3:** Richness imitation by orator confidence

	Dependent variable: Imitation			
	Orator confidence $\leq 73\%$	Orator confidence $> 73\%$	(1)	(2)
Intercept	0.244*** (0.013)	0.226*** (0.016)	0.208*** (0.016)	0.210*** (0.017)
Explanation	-0.029* (0.017)	0.035* (0.020)	0.036* (0.022)	0.072*** (0.022)
Learning	0.176*** (0.024)	0.192*** (0.024)	0.225*** (0.023)	0.221*** (0.025)
Explanation $\times$ Learning	0.086*** (0.029)	0.030 (0.030)	0.156*** (0.030)	0.077** (0.032)
Richness		-0.033** (0.014)		0.005 (0.013)
Explanation $\times$ Richness		0.113*** (0.018)		0.089*** (0.016)
Observations	4583	4583	4206	4206
R <sup>2</sup>	0.061	0.073	0.124	0.144

*Notes:* See notes for Table G2.

## H Additional results on orator & receiver characteristics

We examine how orator and receiver characteristics across learning and unlearning situations, as well as differential imitation based, e.g., on sociodemographics, may contribute to the asymmetric treatment effect of explanations.

We first document the heterogeneity of participants on observable characteristics. We then study which *orator* characteristics predict imitation, and whether this helps explain the asymmetric effect above and beyond the language differences associated with different groups of orators. We then test whether the characteristics of *receivers* predict their responsiveness to explanations, and to what extent this contributes to the asymmetric effect in the aggregate.

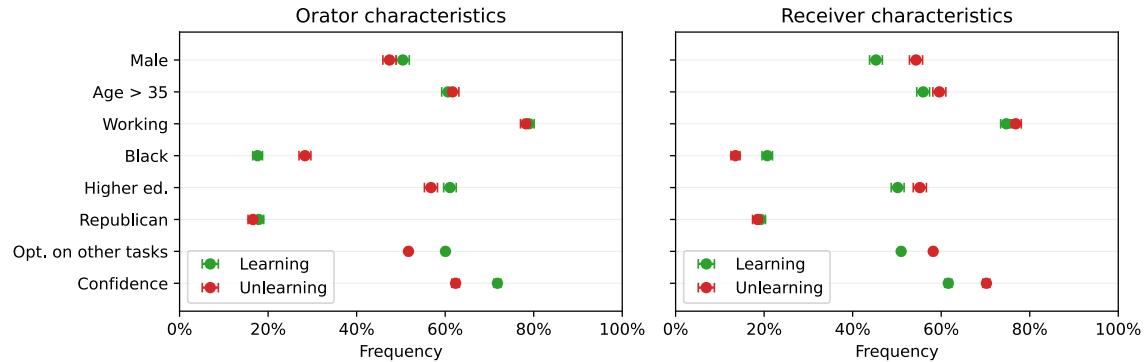
### H.1 Participant Characteristics in Learning and Unlearning

Because receivers and orators are drawn from the same population and because learning and unlearning are determined by the same endogenous variable—receiver prior accuracy—, orators in learning and receivers in unlearning situations *should* have the same characteristics on average; similarly, orators in unlearning and receivers in learning situations should be similar. Figure H4 shows the full set of observable characteristics we elicit in the study, for orators in the left panel and receivers in the right panel, separately for learning and unlearning situations.

Among orators, we find significant differences across seven of the eight characteristics we examine between those in learning and unlearning situations. The first six characteristics capture participant sociodemographics. Orators with a correct 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 our study: orators with correct answers have a substantially higher prior accuracy rate in the 14 remaining tasks (61% vs. 52%) and a higher confidence on the present task (72% vs. 62%). Among receivers, we find very similar patterns for those with correct and incorrect priors.

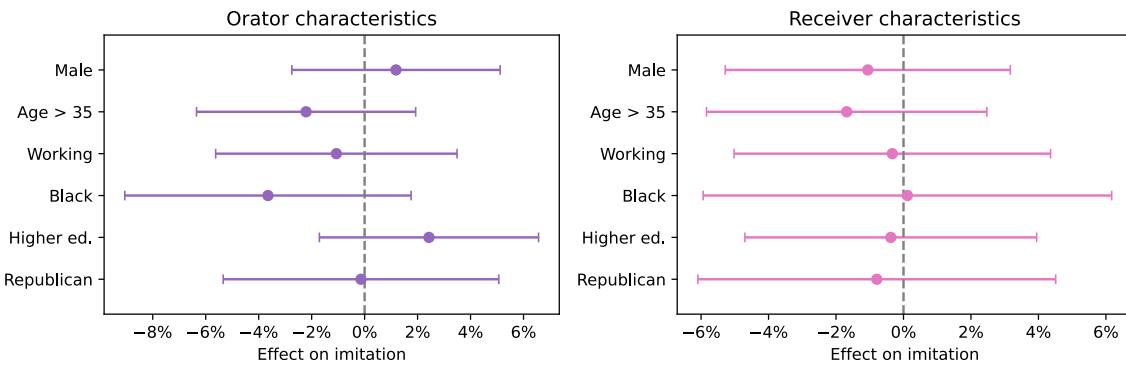
### H.2 The Role of Orator Characteristics

**Heterogeneous treatment effects by orator characteristics.** We ask whether orator characteristics drive imitation patterns. We begin by examining the raw relationship between orator characteristics and the treatment effect of explanations based on the *Choice Only* and *Explanation* conditions. In the left panel of Figure H5, we show the



**Online Appendix Figure H4:** 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. Sample for both panels is the main Receiver experiment, matched with the Orator experiment for the left panel. Whiskers show 95% CIs.

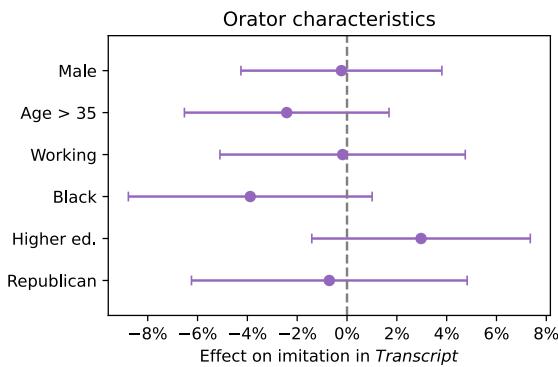
effect of orator characteristics on imitation, controlling for the orator's and receiver's accuracy. For example, the imitation rate is 1.2 p.p. higher if the orator is male than when they are female. Point estimates suggest that that explanations from male and highly educated speakers tend to increase imitation, while explanations from Black or older speakers result in less imitation, but none are significant.



**Online Appendix Figure H5:** Effect of orator and 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 the same results for receiver characteristics. Sample for *Explanation* is the main Receiver experiment, *Choice Only* is pooled from all collections. Whiskers show 95% CIs.

**Do orator characteristics operate via content differences?** Orator characteristics could drive imitation via their content, via their oral delivery, or because the orator's voice sig-

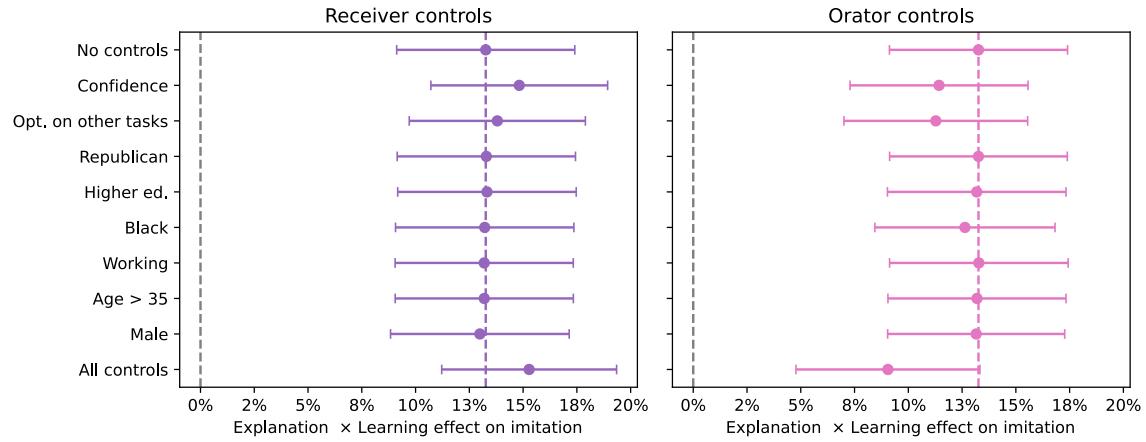
nals sociodemographic features that could influence imitation. To separate the role of content difference, we can focus on the transcript of each explanation and repeat the previous analysis while contrasting *Choice Only* with the *Transcript* treatment. In Figure H6, we can see that the effects of orator characteristics in *Transcript* are very close to those in *Explanation*, and that none of the differences between effects are statistically significant. This suggests that the effect of orator characteristics arises not from the orator's identity as revealed by speech, but from the different kinds of explanations different demographics use.



**Online Appendix Figure H6:** Effect of orator characteristics on imitation in *Transcript* treatment. *Notes:* See notes in Figure H5. *Transcript* sample is the corresponding Receiver experiment, *Choice Only* is pooled from all collections. Whiskers show 95% CIs.

**Do orator characteristics explain away the asymmetric effect?** Intuitively, orators in learning situations may signal characteristics through their voice, e.g., gender, which could make them more or less likely to be imitated. However, they also deliver different explanatory content. The preceding estimates should be interpreted as encapsulating every feature correlated with orator characteristics, both those contained in words and in their oral delivery.

Column 3 of Table 1 shows that orator characteristics explain away 32% of the differential effects, far below the 59% explained by richness. Moreover, columns 6 and 7 show that, once richness is controlled for, adding controls for orator characteristics does not further decrease the size of the unexplained learning effect. Figure H7 examines which orator characteristics explain away the differential effect. It shows that prior confidence and accuracy in the other tasks are most predictive, consistent with our finding that the effects of orator characteristics primarily operate through the content of explanations, though these effects are again very weak.



**Online Appendix Figure H7:** Learning asymmetry after controlling for orator or receiver characteristics. Notes: Coefficient on *Explanation*  $\times$  *Learning* in a regression of imitation on *Explanation*, *Learning*, *Explanation*  $\times$  *Learning*, *Control* and *Explanation*  $\times$  *Control*. *Control* is a receiver control in the left panel and an orator control on the right. All regressions except 'No controls' and 'All controls' contain a single control. *Explanation* sample is the main Receiver experiment, *Choice Only* is pooled from all collections, both restricted to learning and unlearning situations. *Explanation* sample is the corresponding Receiver experiment, *Choice Only* is pooled from all collections. Whiskers show 95% CIs.

### H.3 The Role of Receiver Characteristics

**Heterogeneous treatment effects by receiver characteristics.** We plot the effect of receiver characteristics on imitation in the left panel of Figure H5. For example, after controlling for orator and receiver accuracy, a male receiver is 1.1 p.p. less likely to imitate. Again, it should be stressed that these effects are relatively weak since none of them are significant. 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, who are randomly assigned to receivers.

**Do receiver characteristics explain away the asymmetric effect?** From column 4 of Table 1, we can see that receiver characteristics, if anything, *increase* the asymmetric treatment effect between learning and unlearning. 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. Appendix Figure H7 studies the effect of controlling for each characteristic separately, finding that accounting for prior confidence and accuracy in the other tasks slightly increases this gap.

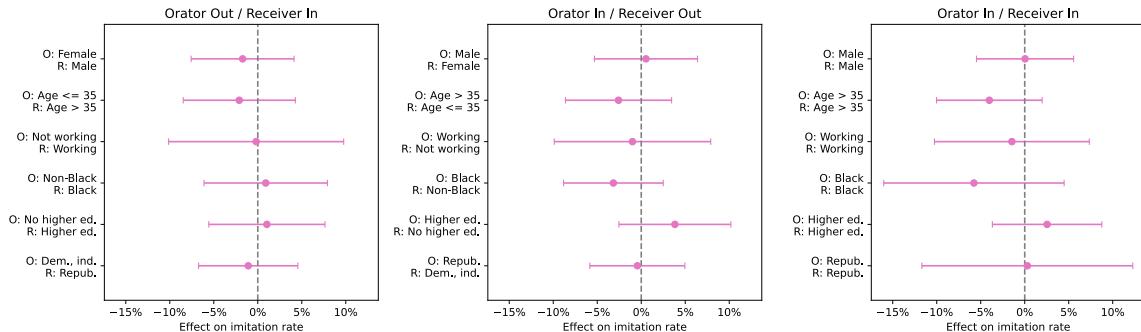
## H.4 The Role of Interacted Orator and Receiver Characteristics

We have studied orator and receiver characteristics separately, and found relatively mild effects, but it could be that their interaction produces strong effects. For example, men might be more likely to imitate other men, women might be less likely to imitate men, etc. This example shows that, for binary characteristics like ours, there are four relevant configurations, depending on whether the orator or receiver are in the in- or out-group.

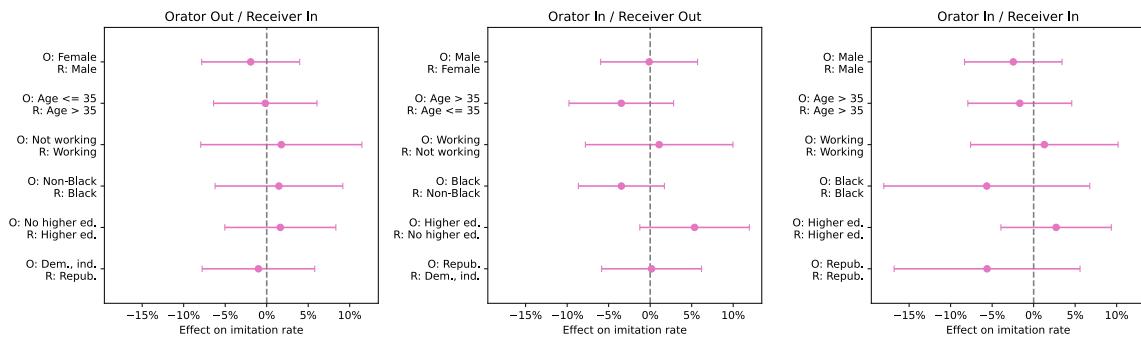
For example, when a male receiver is 2.5 p.p. less likely to imitate a female orator, while a woman is 1.9 p.p. more likely to imitate a man and a man is 0.1 p.p. more likely to imitate a man. All of these effects are relative to a woman listening to a woman, contrast *Explanation* with *Choice Only*, and control for orator and receiver accuracy. However, much like before, none of these effects appear strong enough to be significant.

Figure H9 repeats the same analysis but contrasts *Transcript* with *Choice Only*. Estimates are very similar and not statistically different from the previous one. This suggests that the effects of orator-receiver characteristics, if at all, seem to operate mainly through the content of explanations rather than pure signaling of socioeconomic characteristics.

In conclusion, although online speech-based surveys are promising to study bias in social learning or even discrimination, we do not find strong effects of orator or receiver characteristics when discussing financial reasoning tasks.



**Online Appendix Figure H8:** 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 experiment, *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% CIs.



**Online Appendix Figure H9:** Effect of orator-receiver characteristics on imitation in *Transcript* treatment. Notes: See notes in Figure H8. *Transcript* sample is the corresponding Receiver experiment, *Choice Only* is pooled from all collections. Whiskers show 95% CIs.

# I Additional results from Richness experiment

This Section summarizes additional details about the experiment exogenously manipulating linguistic richness.

## I.1 Details on the generation of rich & sparse versions

Our experimental design aims to take each explanation given by a respondent and generate a rich and a sparse variant that (i) preserve the informational content, and (ii) remain as close as possible to the natural tone of respondents.

We address (i) by heavily emphasizing in our prompt that the informational content, and in particular “all arguments, facts or references” should be exactly preserved. An inspection of original explanations and generated versions shows this works quite well. The ex-post argument identification exercise described below confirms this finding.

Achieving (ii) is more challenging because LLMs are pre-trained on corpora of written text and fine-tuned to be maximally intelligent and helpful assistants. We therefore give the model examples, for each question, of 20 explanations randomly drawn from the 50 with the lowest richness rating (to avoid having only extremely sparse explanations), and the 5 with the highest richness rating (since these are longer).

We then prompt it to generate a maximally sparse and a maximally rich version by imitating the style of the examples while keeping the informational content constant. We instruct the LLM to pay close attention to specific features when imitating the sparse and rich examples. These features are informed by our analysis of the content of explanations and of the determinants of richness (cf. Section 5.2).

To generate the versions, we used GPT-4.1 (more specifically, gpt-4.1-2025-04-14), OpenAI’s “flagship model for complex tasks” as of June 2025. Below, we reproduce the full prompt used to generate both versions at once:

Survey respondents were asked the following financial literacy question: {question}

They also recorded oral responses to explain their choice. A rich explanation is detailed, comprehensive, logically structured, nuanced, and tailored to the context. A sparse explanation is basic, narrow, unclear or disorganized, presents only surface-level understanding, lacks depth or specific details, and fails to relate clearly to the context. Richness is measured as an integer between 0 and 10, both inclusive. If the explanation provides no argument, it is 0. Richness should not depend on whether the respondent’s answer or their premises are factually correct—only the quality of explanation matters.

To give some examples, here are twenty maximally sparse and five maximally rich explana-

tions given by respondents:

{20 sparse examples}

{5 rich examples}

We now focus on a specific respondent, who chose the following answer: {answer}

The respondent also gave the following oral response to explain their choice: {speech}

Please produce two versions of the above explanation, one that is maximally sparse (richness score: 1/10) and one that is maximally rich (richness score: 9/10). Each version should exactly preserve the informational content of the original explanation. In particular, it should preserve all arguments, facts or references of the original explanation, and not add to them. Instead, your re-writings should focus on structure, coherence, presentation and wording of the explanation, and very closely imitate the corresponding examples given above.

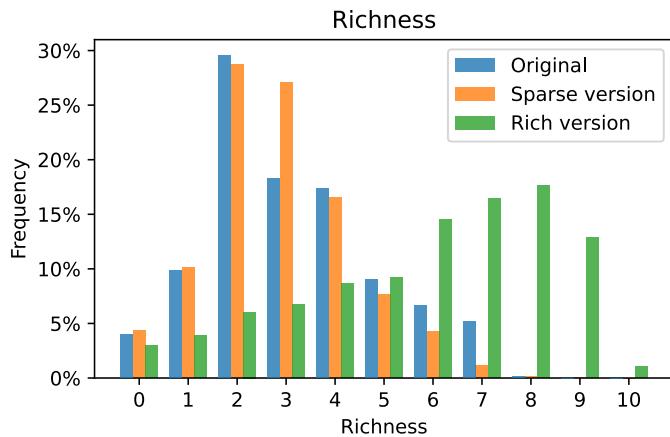
Both versions should closely imitate the oral style of corresponding explanations. In particular, you should pay close attention to filler words ("uh", "um", "huh"), false starts ("So, uh, I think, maybe..."), repetitions and restarts ("I believe that, um, I believe that..."), rhetorical questions ("Why do they outperform?") and first-person framing ("I believe...") and meta-commentary ("This is an easy/difficult question", "To answer this question", "To maximize your chances of getting it right" etc.).

The maximally sparse version should not be a pure summarization of the explanation. When imitating the given examples, you should pay close attention to contradictions or self-corrections mid-sentence ("I think it's one—no wait, maybe two..."), vague or circular language, replacement of specific vocabulary with informal terms ("stuff", "thing"), fragmented or run-on sentences ("So I say one, because, like, I think..."), clauses connected without logical links ("I think, it has to be that, I mean clearly"), abrupt cutoffs or trailing thoughts ("So yeah, number one, because they..."), disconnected or illogical flow, introduction of arguments without relating to the overall context and over-reliance on the prompt.

The maximally rich version should not be a mere rewording of the explanation. When imitating the given examples, you should pay close attention to their clear logical sequence and explicit connectives ("because", "therefore", "for instance"), clarification of underlying assumptions and concepts, specific vocabulary and avoidance of vague words, illustrative examples and real-world analogies, contextual framing of arguments ("Let's say the interest rate is 2%..."), and structure designed to help someone else choose the correct answer rather than just restating their own.

## I.2 Richness of rich & sparse versions

We display the distribution of richness scores in Figure I10, which shows that the sparse versions are indeed sparser and the rich versions richer than the originals. Moreover, all distributions appear relatively well-behaved, with no extreme clustering at either end.<sup>34</sup>



**Online Appendix Figure I10:** Distribution of richness of original, rich & sparse versions. *Notes:* Sample is the Orator experiment (466 orators, 6,910 obs.) and variants generated with an LLM.

We can characterize the effect of the richness manipulation depending on the accuracy of the orator’s answer. When the orator is correct, original richness is 3.74 on average, against 3.38 for *Sparse version* ( $p < 0.01$ ) and 6.74 for *Rich version* ( $p < 0.01$ ). On the other hand, when he is incorrect, average richness is 2.40, 2.24 ( $p < 0.01$ ) and 4.73 ( $p < 0.01$ ) respectively.

## I.3 Features of rich & sparse versions

To shed more light on what is driving this exogenous linguistic richness variation, we analyze the features of the generated versions in more depth. We therefore repeat the feature annotation exercise described in Section 5.2 separately for each version, and display results in Figure I11.

Beginning with certainty markers, we find that both richer and sparse versions contain more low-confidence markers than original explanations, meaning that even in rich versions the LLM communicates more extensively about the uncertainty contained in the initial explanation. Sparse explanations contain fewer high confidence markers

---

<sup>34</sup>This well-behavedness is somewhat at odds with our prompt, which instructed the LLM to aim for a richness of 1/10 and 9/10 respectively. Experimentation shows LLMs are not very good at obtaining very specific richness scores, so that these indications were only meant to broadly increase richness differences.

**Online Appendix Table I4:** Effect of original richness on sparse and rich versions' richness

	Sparse version richness		Rich version richness	
	(1)	(2)	(3)	(4)
Intercept	0.657*** (0.019)	0.649*** (0.027)	2.212*** (0.042)	1.487*** (0.055)
Orator correct		0.157*** (0.041)		1.571*** (0.092)
Original richness	0.705*** (0.006)	0.662*** (0.012)	1.157*** (0.011)	1.353*** (0.020)
Original richness × Orator correct		0.026* (0.015)		-0.370*** (0.025)
Observations	6910	6910	6910	6910
R <sup>2</sup>	0.701	0.706	0.650	0.671

*Notes:* All richness measures are standardized using the mean and standard deviation of original explanations. Sample is the Orator experiment (466 orators, 6,910 obs.) and variants generated with an LLM.

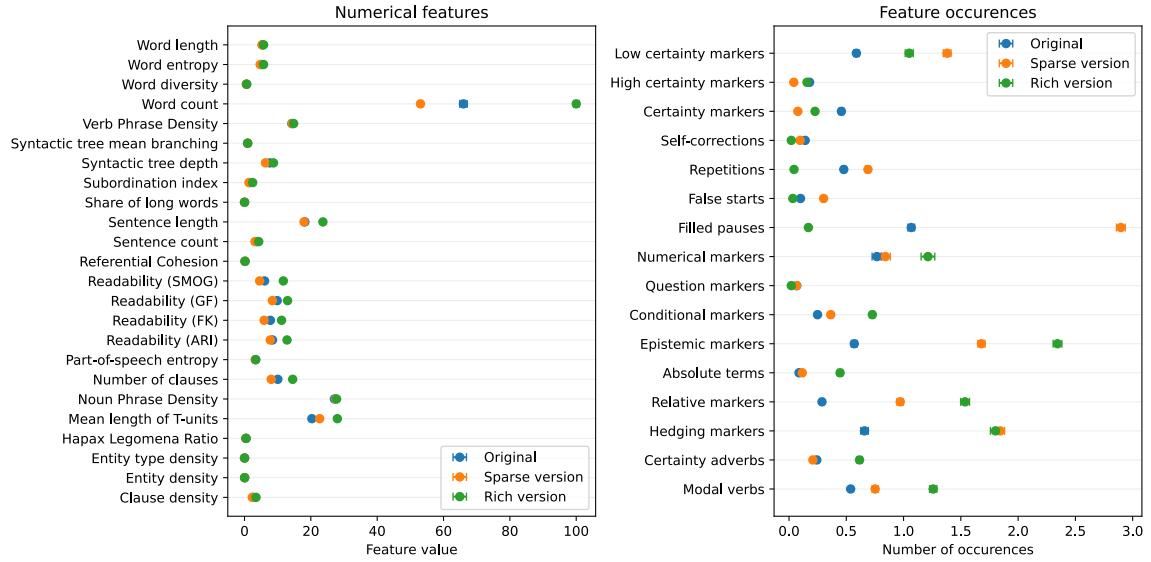
and even rich versions contain slightly fewer than the originals, perhaps because the LLM is slightly overcautious in wanting to preserve informational content.

Disfluencies are more common in sparse versions than in the original and rarer in rich versions, a pattern that applies to self-corrections, repetitions, false starts and especially strongly to filled pauses. Indeed, originals contain 1.06 filled pauses (e.g., “uh” or “uhm”) against only 0.16 in rich versions and 2.89 in sparse versions. Clearly, the LLM identifies these as prominent features of maximally rich or sparse explanations. In most other features, we also obtain that they are less frequent in sparse versions and more frequent in rich ones.

Taken together, these results show that the linguistic richness manipulation affects the presence of numerical correlates of richness and of language markers that contribute to it. We therefore conclude that it achieved its aim in an effective, broad and balanced way.

#### I.4 Arguments of rich & sparse versions

Finally, we examine how each versions' richness depends on the original's richness in Table I4. Column 1 shows that the sparse version dampens original richness but has a slight fixed boost, when one might have expected a negative coefficient, while column 3 shows the rich version amplifies original richness differences and also has a much more sizeable fixed richness boost. Columns 2 and 4 show that there is no strong effect of orator accuracy on this pattern for sparse versions, but that rich versions amplify richness heterogeneity less strongly for correct orators. Perhaps this is because correct explanations are often already comparatively rich, so that there is less room to increase it further.



**Online Appendix Figure I11:** Features and markers of original, rich & sparse versions. *Notes:* Sample is the Orator experiment (466 orators, 6,910 obs.) and variants generated with an LLM.

To verify that each version indeed preserves the informational content of explanations, as is stressed in the LLM prompt, we repeat the argument-tagging exercise described in Section 5.1.1 independently for each version. More precisely, we use GPT-4 again to tag which of our manually identified arguments are present in each edited version, separately for each version of each explanation.

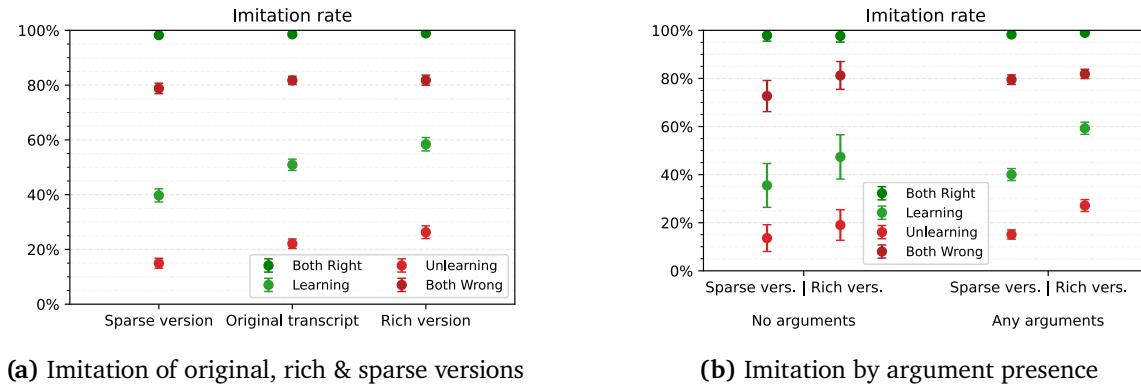
Across all rounds and all arguments (including “Other” and “Irrelevant”), the correlation between arguments identified in the original explanation and in the sparse version is 0.68, with a Cohen’s  $\kappa$  of 0.67 indicating “substantial agreement”. Similarly, the correlation between arguments in the original and the rich version is 0.66, and a Cohen’s  $\kappa$  of 0.66 again reflecting “substantial agreement”. Agreement between the two versions, the condition which ensures the validity of our experimental design, is even higher with a correlation of 0.76, and a Cohen’s  $\kappa$  at 0.76 showing we have “substantial agreement” and are very close to “almost perfect agreement”.

When an argument is not identified in the rich version, it is also not identified in the sparse version in 98% of cases. However, when an argument is identified in the rich version, it is also identified in the sparse version 73% of the time. Indeed, sparse versions contain 1.09 arguments on average, slightly less than the rich versions which contain 1.32, a comparatively small but significant difference ( $p < 0.01$ ).

In conclusion, the richness manipulation appears very effective at preserving the informational content of explanations, though it is not perfect. It should be noted that

assessing the presence of specific arguments is inherently noisy, so that even preserving informational content perfectly is likely to lead to some disagreement in the ex-post argument identification. Moreover, since rich versions contain more words and LLMs work on tokenized words, there is a higher chance that the LLM picks up on an instance of an argument for richer transcripts, so that the small differences we do observe could reflect this algorithmic bias.

Nonetheless, as a conservative robustness check, we repeat our main analysis on the explanations for which there is complete agreement on the informational content of both versions. For 63% of the sample, GPT-4 exactly agrees on the presence or absence of each of the arguments, including “Other” and “Irrelevant”, in both versions. This fraction is quite high given the noisy nature of argument tagging, with another substantial fraction of explanations agreeing on all but one or two arguments. Among this sample with perfect agreement, imitation is 8.1 p.p. higher in *Rich version* than in *Sparse version* ( $p < 0.01$ ). Moreover, it is 20.2 p.p. higher in learning situations ( $p < 0.01$ ), 10.7 p.p. higher in unlearning situations ( $p < 0.01$ ), 1.6 p.p. higher when both are wrong ( $p = 0.42$ ) and 0.7 p.p. higher when both are right ( $p = 0.37$ ). Since these effects are significant and very close to those in the full sample, we conclude that our experimental findings are not driven by changes in the informational content of explanations.



**Online Appendix Figure I12:** Effect of richness manipulation on imitation, relative to *Transcript* and by presence of arguments. *Notes:* Sample is the Richness receiver survey (972 receivers, 14,580 obs.). Whiskers show 95% CIs.

## I.5 Heterogeneity by presence of arguments

Another check we can implement is to test whether the richness manipulation has effects that differ based on the presence of any argument. Large differences in the effect of the treatment between explanations with and without arguments suggest that variation

in informational content plays some role in the results. For explanation without any arguments, imitation is 57.0% and 62.1% in *Sparse version* and *Rich version* respectively, against 65.0% and 72.8% for explanations with arguments. This 2.7 p.p. difference in effect sizes is not very large and not statistically significant ( $p = 0.36$ ). Figure I12b plots the imitation rates for each condition separately for each configuration, which also all have small and insignificant differences in effect sizes. However, it should be noted that this test is not very powerful because explanations without any arguments are quite rare, making up only 7.9% of the sample.

## I.6 Comparison with original *Transcript* treatment

Although results should be interpreted more carefully since they are not part of the same experimental design, it is nonetheless interesting to compare imitation of the edited versions with that of the original explanations in the *Transcript* condition. Overall, it sits between the sparse and rich versions, since its imitation rate is 69.5%, above the 64.3% of *Sparse version* ( $p < 0.01$ ) and below the 71.9% of *Rich version* ( $p < 0.01$ ). Figure I12a documents that, also within each situation, original explanations are in-between the sparse and rich versions. Since the richness of originals also sits in-between that of the versions, these findings further point toward an effect of richness on imitation.

## J Additional results from Balls-and-urns experiment

### J.1 Logistics

Respondents in both studies received a base reward of \$2 for completing the study. Participants could win a bonus of \$10 following the same incentive structure as in the main experiment, defining an answer as correct if it is within  $\pm 3$  p.p. of the Bayesian posterior. Beliefs are elicited as integer percentages between 0% and 100% using a text box. Median completion times were 13 minutes in both experiments. Both experiments were run in April 2025 using a U.S. sample on Prolific. The Orator experiment was run for a total of 554 respondents, of whom 464 gave valid data. The Receiver experiment was run for 1,975 respondents, of whom 1,856 gave valid data.

In line with our pre-registration, we drop respondents who self-reported using searching for answers online or using external tools. To curb the use of AI assistants,<sup>35</sup> we asked receivers whether the orator used an automatic voice generator or seemed to be reading a script. As specified in our pre-registration, we then exclude all orators for whom a majority of receivers agree: these cases represents 2.2% of valid recordings.

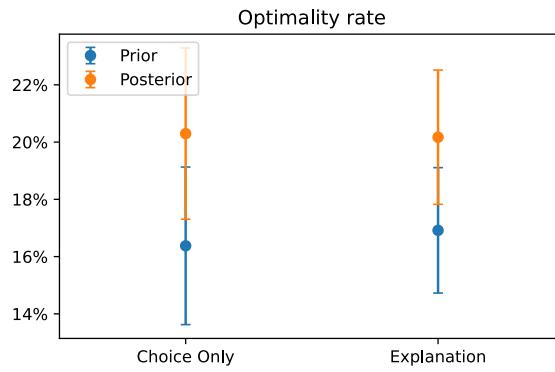
In light of prior research, we expected the share of respondents selecting a correct response to be low (Bordalo et al., 2025). This implies learning and unlearning situations should be very rare in the sample under random matching. To increase power to detect learning effects, we therefore stratify the orator assignment based on the receiver's prior answer. If a receiver's prior answer is incorrect, we randomly assign them to a correct or incorrect orator with equal probability. If it is correct, we assigned them an orator without stratification. In our raw sample, receiver and orator are both wrong in 40.4% and both right in 2.9% of observations, while learning and unlearning situations occur 40.4% and 13.4% of the time. We then de-stratify our sample using expected frequencies under random matching with an optimality rate of 16.7%, estimated by pooling orators and receiver priors. All summary statistics, regressions and figures presented in the main body and in the following sections are re-weighted to the population frequency.

---

<sup>35</sup>In all experiments, the question text is not selectable and therefore difficult to copy-and-paste into a search engine or AI assistant. However, between the main experiment in December 2023 and Balls-and-urns experiment in April 2025, multimodal AI assistants that can analyze screenshots have become much more prevalent. We believe this explains the noticeable rise in orators that appear to read a script from an AI assistant.

## J.2 Optimality

In line with the instructions and incentives given to survey participants, we define a correct answer a belief within  $\pm 3$ p.p. of the Bayesian posterior. Figure J13 displays optimality rates for the balls-and-urns task. Prior optimality stands at 16.4% and 16.9% in *Choice Only* and *Explanation* respectively ( $p = 0.76$ ), which, as expected, is considerably lower than the optimality rates around 50% seen in the main experiment. As before, both treatments help listeners select correct answers and significantly increase their optimality rates by 3.9 p.p. ( $p = 0.03$ ) and 3.3 p.p. ( $p = 0.05$ ) respectively. However, the difference in posterior optimality between the two conditions is very small, with *Choice Only* 0.1 p.p. higher than *Explanation*, and not significant ( $p = 0.95$ ).



**Online Appendix Figure J13:** Effect of *Explanation* on optimality in Balls-and-urns experiment.  
*Notes:* Share of correct receivers before and after exposure to the orator's choice or explanation. An answer is defined as correct if it is within  $\pm 3$  p.p. of the Bayesian posterior. Sample is the Balls-and-urns receiver experiment (1,856 receivers and obs.). Whiskers show 95% CIs.

## J.3 Imitation

In this abstract balls-and-urns setting, choices are continuous probabilities between 0% and 100%, so that imitation is harder to define and quantify. For completeness, we have therefore pre-registered four relevant measures of imitation:

1. *No imitation*: the receiver reports the same posterior belief as their prior belief
2. *Full imitation*: the receiver reports the same posterior belief as the orator
3. *Continuous imitation in  $[-1,2]$* : to measure imitation continuously, we leverage a Grether (1980) decomposition to represent beliefs as log-odds in  $\mathbb{R}$ ,<sup>36</sup> and then

---

<sup>36</sup>Observations with beliefs of 0% or 100% are dropped from all analyses using continuous imitation.

define continuous imitation as the interpolation weight placed on the orator's belief:

$$\text{Continuous imitation} = \frac{\text{logit}(\text{Receiver posterior belief}) - \text{logit}(\text{Receiver prior belief})}{\text{logit}(\text{Orator belief}) - \text{logit}(\text{Receiver prior belief})}$$

No imitation yields a continuous imitation of 0, and full imitation yields 1. It can be negative if the receiver updates in the opposite direction, and larger than 1 if they go beyond the orator's belief. To limit the influence of extreme outliers due to beliefs close to 0% or 100%, we winsorize continuous imitation to  $[-1, 2]$ .

4. *Continuous imitation in  $[0, 1]$ :* for completeness, we also report results for continuous imitation winsorized to  $[0, 1]$ .

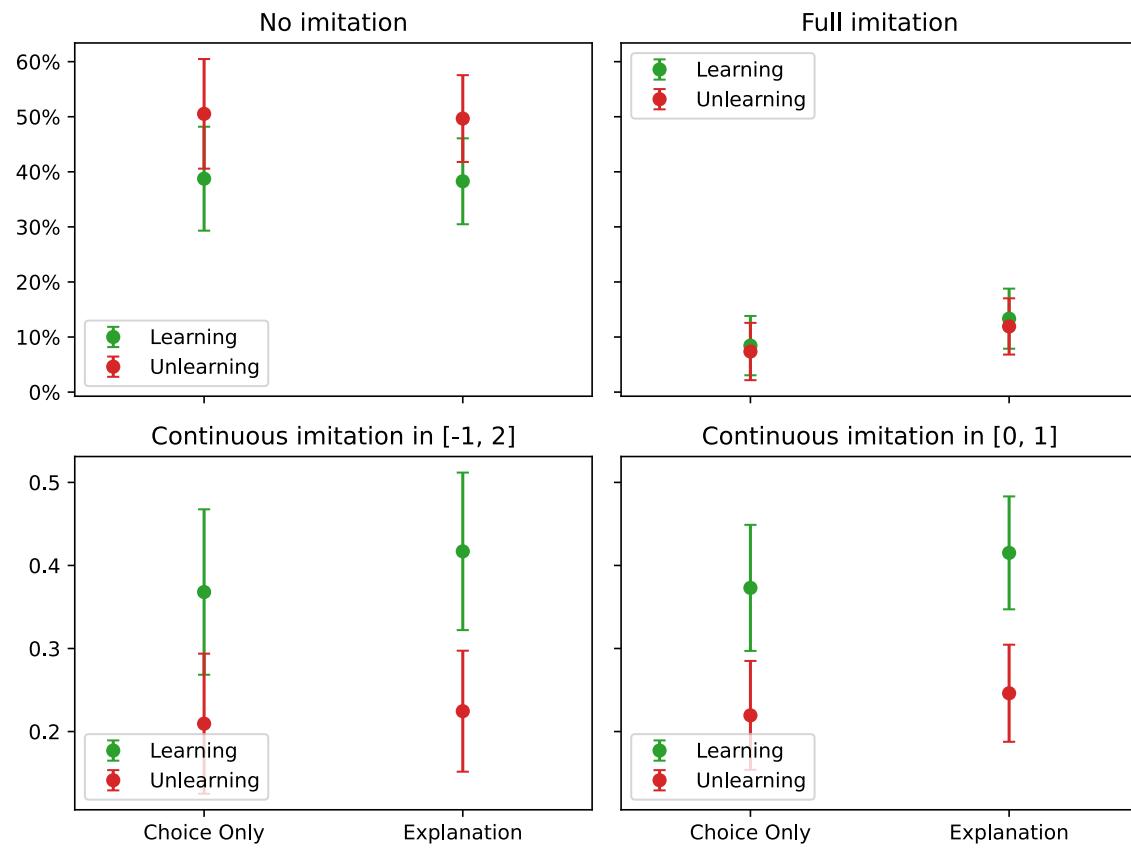
Across all situations, the share of receivers who do not imitate at all is 50.8%, while 25.3% imitate fully. These two sets have a small overlap, as in 16.4% of cases the receiver has the exact same prior belief as the orator. These confirming signals are typically one of the salient modes of the problem, e.g., the prior or diagnosticity. On the other hand, 38.5% of receivers pick a posterior belief that is neither their own prior belief nor the orator's, with two thirds of these picking answers that are strictly between the two.

Figure J14 shows imitation in learning and unlearning situations. The share of receivers not imitating is close to 40% in learning situations for both treatments ( $p = 0.89$ ) and closer to 50% for learning situations ( $p = 0.94$ ). Full imitation rates are much lower, hovering around 10% for both situations, and again without a significant difference between treatments ( $p = 0.21$  in learning and  $p = 0.22$  in unlearning). Continuous imitation winsorized to  $[-1, 2]$  is about 40% in learning and 20% in unlearning situations, with insignificant differences between treatments ( $p = 0.49$ ,  $p = 0.79$ ). Findings are qualitatively similar for winsorization to  $[0, 1]$  ( $p = 0.60$ ,  $p = 0.48$ ).

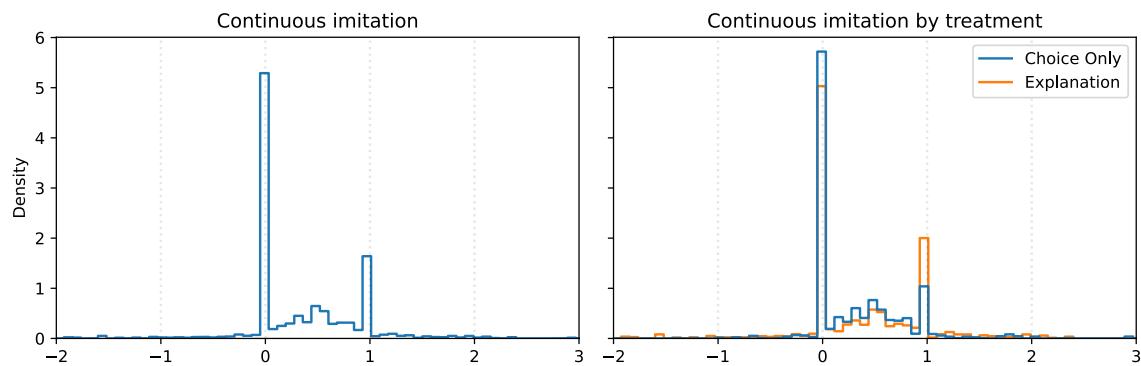
We further examine the full distribution of continuous imitation, as is done in Figure J15. The left panel shows the distribution pooling treatments, which displays strong modes at 0 and 1, corresponding to no and full imitation. There is a weaker mode around 0.5, suggesting that many receivers update somewhere roughly in the middle of their own prior belief and the receiver's belief.<sup>37</sup> Imitation outside of these bounds is rare, as 12.6% of observations are outside of  $[0, 1]$ , and only 4.5% outside of  $[-1, 2]$ . Splitting by treatments, the right panel reveals that the difference between *Choice Only* and *Explanation* appears small and unlikely to be statistically meaningful.

---

<sup>37</sup>Note that, with integer percentage beliefs, it is impossible to achieve an imitation of exactly 0.5 or a similar round value in log-odds space.



**Online Appendix Figure J14:** Effect of *Explanation* on imitation in Balls-and-urns experiment.  
*Notes:* See main text for definitions of imitation measures. Sample is the Balls-and-urns receiver experiment (1,856 receivers and obs.). Whiskers show 95% CIs.

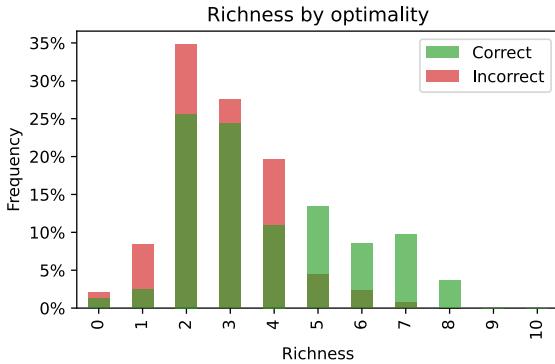


**Online Appendix Figure J15:** Distribution of continuous imitation in Balls-and-urns experiment.  
*Notes:* Left panel shows histogram of continuous imitation across treatments. Right panel shows histogram of continuous imitation in *Choice Only* and *Explanation*. The  $x$ -axis is cut to  $[-2, 3]$ , leaving 2.3% of observations outside of axis limits. Sample is the Balls-and-urns receiver experiment. Whiskers show 95% CIs.

In conclusion, we fail to find a treatment effect of *Explanation* relative to *Choice Only* in the Balls-and-urns task. However, despite our oversampling strategy designed to maximize power, standard errors remain relatively large, so that this is not a very precise null. The difficulty of the task, the complexity of social learning in this context, and the difficulty in defining optimality and imitation make aggregate improvement and learning hard to measure.

#### J.4 Imitation of richness

We follow the same procedure as in the main experiment to annotate the richness of explanations for the Balls-and-urns experiment. Figure J16 shows the distribution of richness rating by optimality. There is a statistically significant, 0.733 SD difference in richness between correct and incorrect explanations, which is very close to the gap found among financial reasoning tasks.



**Online Appendix Figure J16:** Richness gap by orator optimality in Balls-and-urns experiment.  
Notes: Sample is the Balls-and-urns orator survey.

We therefore examine whether richness predicts imitation in the Balls-and-urns experiment. Table J5 replicates Column (7) of Table 1, which examines whether richness predicts imitation within learning and unlearning situations. We find that richness has a negative but statistically insignificant effect on no imitation ( $p = 0.23$ ). However, it has a statistically significant effect on full imitation, at +3.6 p.p. ( $p = 0.05$ ), and on continuous imitation winsorized to  $[-1, 2]$  and  $[0, 1]$ , at +0.056 ( $p = 0.06$ ) and +0.043 ( $p = 0.05$ ) respectively.

Although we fail to find effects of *Explanation* on optimality and learning in this ball-and-urns task, we therefore do find evidence that our main hypothesized mechanism, the effect of richness on imitation, holds here. Crucially however, the magnitude of this effect is considerably smaller than for financial reasoning tasks.

**Online Appendix Table J5:** Imitation of richness in Balls-and-urns experiment

	No imitation (1)	Full imitation (2)	Continuous imitation (3)	Cont. in [0,1] (4)
Explanation	-0.013 (0.067)	0.050 (0.039)	0.022 (0.054)	0.032 (0.046)
Learning	-0.110* (0.062)	0.013 (0.033)	0.172*** (0.048)	0.155*** (0.039)
Explanation × Learning	0.032 (0.078)	-0.024 (0.047)	-0.009 (0.070)	-0.017 (0.054)
Richness	-0.011 (0.023)	-0.003 (0.010)	-0.019 (0.020)	-0.003 (0.014)
Explanation × Richness	-0.035 (0.028)	0.036* (0.018)	0.056* (0.031)	0.043* (0.023)
Observations	1031	1031	1026	1026
R <sup>2</sup>	0.020	0.014	0.036	0.053

Notes: Sample is the Balls-and-urns receiver experiment, restricted to learning and unlearning situations. *Explanation* is a dummy for the *Explanation* treatment, *Learning* is a dummy for learning situations. *Continuous imitation* is winsorized to [-1,2], *Cont. in [0,1]* to [0,1]. See main text for definition of imitation measures.

This finding can help us shed light on our failure to find aggregate effects due to explanations in our Balls-and-urns experiment. In learning and unlearning situations, a one-standard deviation increase in richness raises full imitation, the margin that matters for optimality, by 3.6 p.p. In turn, since the optimality rate stands at 16.7%, learning and unlearning situations occur in  $2 \times 0.167 \times (1 - 0.167) = 27.8\%$  of matches. A back-of-the-envelope calculation then shows that the effect of explanations on optimality via richness is  $0.73 \times 0.036 \times 0.278 = 0.7$  p.p.. This is considerably smaller than the  $\pm 2.4$  p.p. confidence interval for posterior optimality in Figure J13, suggesting that despite our large sample this abstract task would have made it difficult to detect any effect of explanations on optimality via richness.

To understand why this potential effect is considerably smaller than for financial reasoning tasks, we perform the same calculation for our main experiment. There, the richness gap between correct and incorrect explanations is very similar at 0.77 SD. However, the effect of richness on imitation is three times larger at 10.7 p.p. (cf. Table 1, Column (7)). Moreover, since the optimality rate across orators and receiver priors is 55.4%, learning or unlearning occurs in  $2 \times 0.554 \times (1 - 0.554) = 49.4\%$  of matches, approximately two times more often. The total predicted effect of explanations on optimality via richness is then  $0.77 \times 0.107 \times 0.494 = 4.1$  p.p.,<sup>38</sup> six times larger than for the

<sup>38</sup>This is not far from the 3.2 p.p. effect observed in the data, with the difference being due to error in

balls-and-urns. On the other hand, because we can pool the 15 financial reasoning tasks and because optimality rates are higher, the confidence interval for posterior optimality in Figure 1 is  $\pm 0.7$  p.p., considerably smaller than the richness effect.

In conclusion, we find some evidence that receivers are more likely to imitate rich explanations in the balls-and-urns task, though the effect is substantially weaker than for financial reasoning tasks. Combined with the difficulty of the task, which makes learning situations rare, this shows the potential effect of explanations is small, and well below our detection threshold. While the rarity of learning situations has an obvious effect, it is less certain why respondents are less likely to imitate richness here. Since the task is very difficult, one could expect more confused respondents to focus more strongly on richness over content. One possible explanation could therefore be that, given the complexity of the task, respondents could have very diffuse priors for the richness of explanations about it, and therefore only use it very weakly to update.

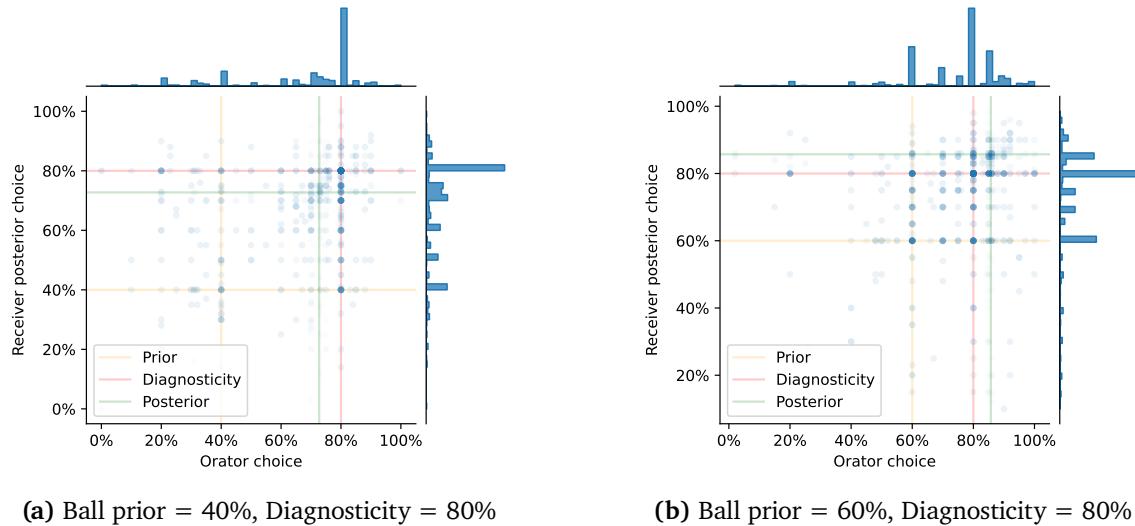
## J.5 Imitation of salient modes

Previous research has shown that, when faced with abstract numerical problems like the balls-and-urns task, respondents rely on salient features to form an answer (Bordalo et al., 2025). In particular, experiments show that salience causes neglect of other relevant features, leading to biased answers that cluster around specific modes. For a balls-and-urns inference task similar to ours, Bordalo et al. (2025) find that respondents very often answer the prior or 100%-prior, the diagnosticity or 100%-diagnosticity, or fall back to an ignorance default of 50%-50%. We obtain very similar modes in our data.

Our experiment allows us to study whether social learning about numerical problems also relies on salient features. We can examine whether other modes, besides the posterior, are more likely to be imitated. We confirm this approach is sensible by plotting orator beliefs and receiver posterior beliefs in Figure J17. To increase power, we have grouped symmetric configurations based on the color of the ball draw, so that there are only two groups: configurations with a prior of 40% on the color of the ball draw, and configurations with a prior of 60%. The left panel shows the first case, where the posterior of 72.73% is below the diagnosticity of 80%, and the right panel shows the second case, where the posterior of 85.71% is above. The marginal distributions of beliefs exhibit strong modes at the prior and the diagnosticity, and a weaker, more diffuse mode around the Bayesian posterior. The modes around 100%-prior, 100%-diagnosticity and 50%-50% are comparatively weaker.

The scatter plots, where the opacity of a dot reflects the number of overlapping  
the estimates used for this calculation.

observations, make clear that receiver posterior beliefs also strongly cluster on modes. Social learning is therefore largely based on switching from one mode to another, rather than smooth updating between likelihoods. For example, the dark blue dots along the “Diagnosticity” vertical line indicate a tendency by receivers to switch to this mode when faced with an orator that supports it. Salient features of a numerical problem therefore determine how subjects form responses to it, but also how they learn about it from others.



**Online Appendix Figure J17:** Effect of *Explanation* on imitation in Balls-and-urns experiment. Notes: Symmetric configurations are grouped based on the color of the ball draw. Left panel shows configurations with a corresponding prior of 40%, right panel shows configurations with a corresponding prior of 60%. Histograms show marginal distribution, scatter plots show joint distribution, with opacity reflecting the number of observations. Sample is the Balls-and-urns receiver experiment, pooling *Choice Only* and *Explanation* treatments.

In line with our pre-registration, we identify orator’s and receiver’s reasoning errors directly from their estimated likelihood. Among orators, 12.5% respond with the prior, 36.7% answer the diagnosticity inverted according to the signal (e.g., if the diagnosticity is 80% and the ball draw is blue, the belief that the ball is red would be 20%), 16.7% give an answer within  $\pm 3$ p.p. of the Bayesian posterior, and 34.1% state some other belief. Orators responding 50%, the non-signal-flipped diagnosticity or the flipped prior are relatively rare in our data. Due to our oversampling scheme designed to increase the frequency of learning situations, we have limited power to study these rarer modes, and therefore group them into the “Other” category.

Receiver prior belief modes are very similar. Importantly, receiver posterior belief modes are also very similar. In particular, the share of answers not falling in one of the

**Online Appendix Table J6:** Imitation of modes in Balls-and-urns experiment

	No imitation (1)	Full imitation (2)	Continuous imitation (3)	Cont. in [0,1] (4)
Intercept	0.450*** (0.038)	0.205*** (0.036)	0.410*** (0.041)	0.382*** (0.032)
Diagnosticity · Signal	0.151*** (0.045)	0.264*** (0.043)	-0.047 (0.052)	-0.006 (0.041)
Posterior $\pm 3$ p.p.	-0.043 (0.042)	-0.055 (0.039)	-0.006 (0.047)	0.012 (0.036)
Other	0.027 (0.046)	-0.119*** (0.039)	-0.136*** (0.050)	-0.096** (0.038)
Observations	1822	1822	1619	1619
R <sup>2</sup>	0.023	0.152	0.011	0.014

Notes: Variables indicate orator belief mode: *Diagnosticity · Signal* is the diagnosticity inverted according to the signal draw, *Posterior  $\pm 3$ p.p.* are belief within  $\pm 3$ p.p. of the Bayesian posterior, *Other* are other answers. *Prior* is left-out. *Continuous imitation* is winsorized to [-1,2], *Cont. in [0,1]* to [0,1]. See main text for definition of imitation measures. Sample is the Balls-and-urns receiver experiment, pooling *Choice Only* and *Explanation* treatments.

three modes is 35.3% and 35.1% respectively ( $p = 0.89$ ), confirming that social learning heavily relies on modes.

Table J6 regresses imitation on the orator's choice mode, with "Prior" being the held out mode as it represents the simplest heuristic. It shows that "Diagnosticity · Signal" is significantly more likely to be fully imitated, with a 26.4p.p. increase in full imitation, but also that it is more likely not to be imitated at all, with a somewhat smaller 15.1p.p. increase in no imitation. Correspondingly, there is no effect on continuous imitation. Presumably, this is because this mode reflects an (erroneous) mathematical understanding of the problem, and therefore calls for complete agreement or (slightly more rarely) disagreement, but not smooth updating somewhere in-between, so that the full- and no-imitation effects cancel out.

We can further observe that "Posterior  $\pm 3$ p.p." has no significant effect on imitation, indicating it is not much more or much less convincing than "Prior", which is an interesting finding. Finally, "Other" has a strong negative effect on full and continuous imitation, indicating receivers are less prone to update fully or even partially based on orators that are not clearly on one of the dominant modes.

So far, we have pooled *Choice Only* and *Explanation* to maximize power. For completeness, Table J7 repeats the same analysis but distinguishes between *Choice Only* and *Explanation* by adding an interaction term for the latter. It fails to find significant effects

of *Explanation* on the imitation of modes, in line with our previous findings for correct answer. However, it should be noted that our experiment is not very well-powered to study imitation of modes other than “Posterior  $\pm 3\text{p.p.}$ ” because of our stratification strategy, so that we can only detect relatively large effects.

**Online Appendix Table J7:** Effect of *Explanation* on imitation of salient modes in Balls-and-urns experiment

	No imitation (1)	Full imitation (2)	Continuous imitation (3)	Cont. in [0,1] (4)
Intercept	0.571*** (0.070)	0.231*** (0.057)	0.336*** (0.071)	0.313*** (0.059)
Explanation	-0.193** (0.094)	-0.040 (0.072)	0.114 (0.102)	0.106 (0.083)
Diagnosticity · Signal	0.047 (0.080)	0.217*** (0.069)	0.002 (0.083)	0.017 (0.071)
Explanation $\times$ Diagnosticity · Signal	0.165 (0.106)	0.077 (0.089)	-0.072 (0.118)	-0.033 (0.097)
Posterior $\pm 3\text{p.p.}$	-0.141* (0.075)	-0.087 (0.062)	0.024 (0.076)	0.052 (0.063)
Explanation $\times$ Posterior $\pm 3\text{p.p.}$	0.154 (0.101)	0.050 (0.080)	-0.040 (0.112)	-0.057 (0.089)
Other	-0.054 (0.081)	-0.204*** (0.058)	-0.075 (0.081)	-0.069 (0.066)
Explanation $\times$ Other	0.128 (0.108)	0.136* (0.077)	-0.092 (0.118)	-0.038 (0.093)
Observations	1822	1822	1619	1619
R <sup>2</sup>	0.029	0.157	0.013	0.021

*Notes:* See notes for Table J6. *Explanation* is a dummy for the *Explanation* treatment. Sample is the Balls-and-urns receiver experiment.

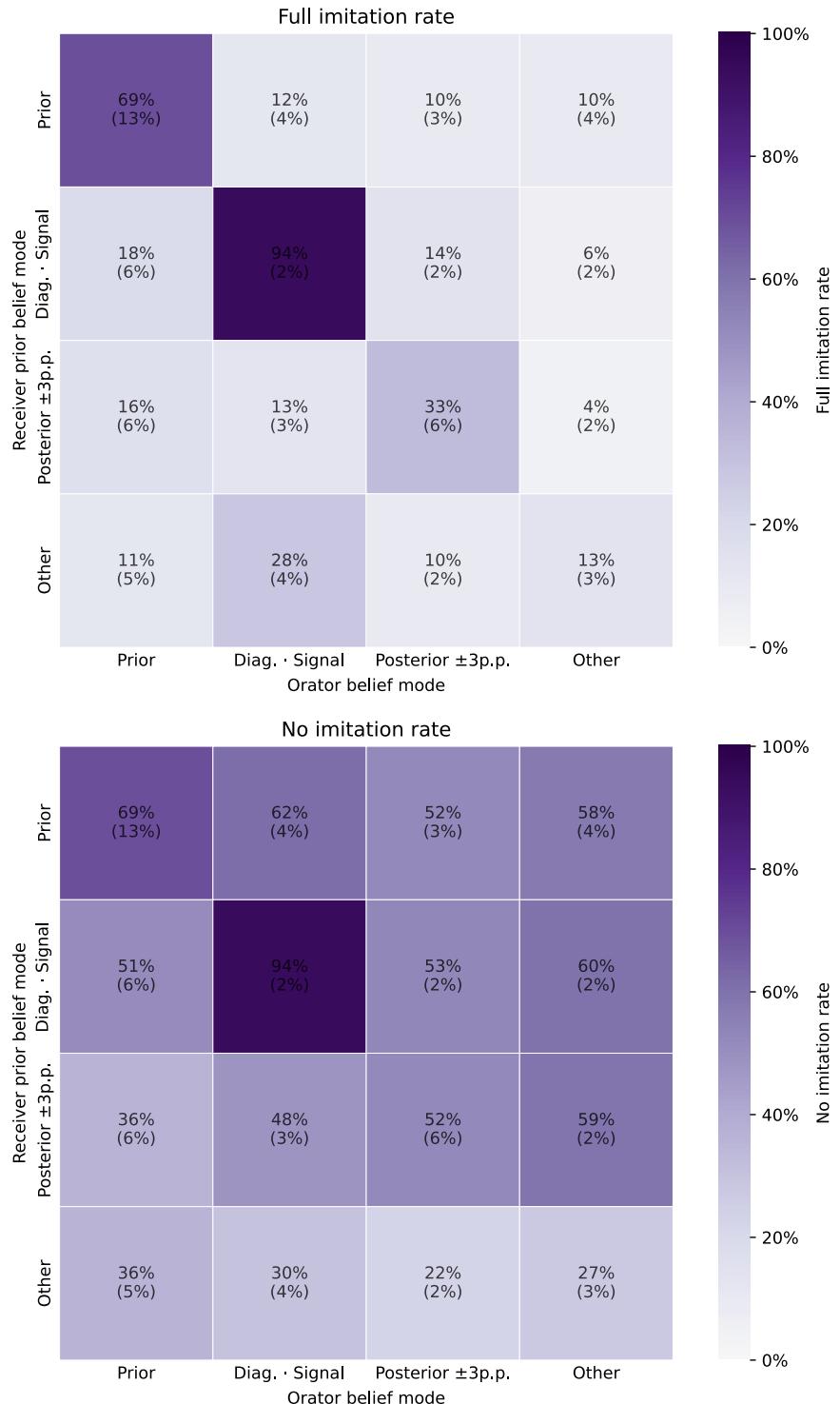
Finally, we can examine the effect of the receiver’s prior mode on imitation. Figure J18 plots the full and no-imitation rate by orator and receiver prior belief modes. Recall that two thirds of receivers either imitate fully or not at all, so that this Figure reflects transitions for an important fraction of the sample. Receivers who answered the prior and saw an orator who answered the prior maintained their answer in 69% of cases, a figure that goes up to 94% for diagnosticity signal, indicating that these are socially stable modes. On the other hand only 33% of receivers who gave a correct answer fully imitate the orator, though this effect is partly mechanical because this mode covers a range of  $\pm 3\text{p.p.}$ . Indeed, a receiver who holds their belief while matched with an orator whose belief also falls within this range—but differs numerically—will not be classified as fully imitating, even if no belief updating occurs. The corresponding no-imitation rate is relatively high at 53%.

Transitions between modes are lower but still relatively high, typically around 10%-15%, and these smaller differences are unlikely to be significant. No-imitation rates for these tasks are correspondingly lower, around 50%.

On the other hand, orators giving an answer in the “Other” category see full-imitation rates closer to 5-10%, lower than the other modes, and no-imitation rates around 60%, much higher than the others. Moreover, receivers whose prior belief was “Other” have no-imitation rates around 20%-30%, again considerably lower than for the other modes. They also appear more likely to imitate fully, in particular they fully imitate the “Diagnosticity · Signal” mode in 28% of cases. In conclusion, non-modal answers are more likely to be deviated from, and less likely to be imitated.

Our experimental setting therefore empowers us to study social learning for abstract numerical tasks. We find that salient numerical features, which are important in explaining respondents’ beliefs, also play an important role in the diffusion of beliefs. However, we do not find that social learning, and explanations in particular, attenuate modes that reflect errors in reasoning.

This suggests that, at least for complex, abstract problems like this balls-and-urns inference task, explanations may only lead to weak aggregate improvements in optimality. Bordalo et al. (2025) find that manipulating the framing of tasks, for example adopting the more naturalistic taxicab hit-and-run format (Kahneman and Tversky, 1972), changes the distribution of modes. It is therefore an open question, which we leave to further research, whether more natural tasks also lead to more beneficial social learning.



**Online Appendix Figure J18:** Imitation by receiver prior belief mode and orator belief mode in Balls-and-urns experiment. Notes: Top panel shows share of receivers stating the same belief as the orator, bottom panel shows share of receivers maintaining the same belief as their prior belief. Sample is the Balls-and-urns receiver experiment.

## K Survey screens

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

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.

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

**Start Recording**

**Online Appendix Figure K19:** Recording screen from the Orator 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%**

**Online Appendix Figure K20:** *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:

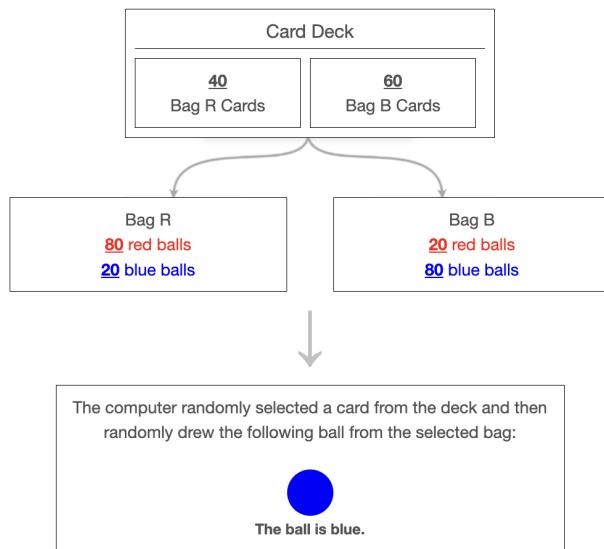
**Actively managed funds outperform passively managed ones.**

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.

**Online Appendix Figure K21:** *Transcript* treatment screen from the Receiver experiment in Section 4.3.

Listen to the other respondent's answer



Given that this ball was drawn, **how likely** do you think it is that Bag R (as opposed to Bag B) has been selected?

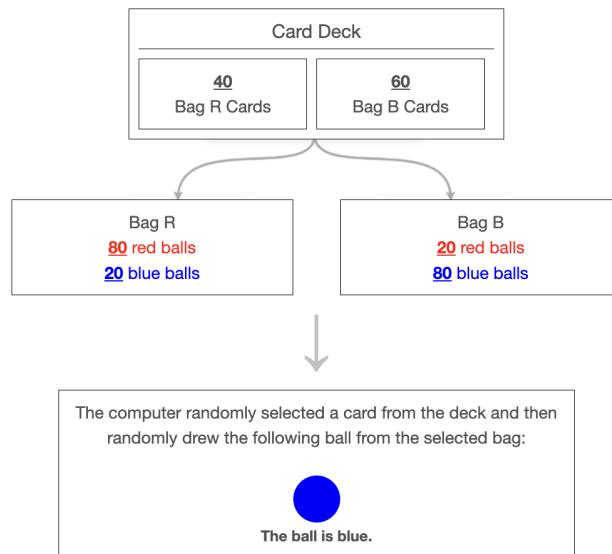
Other person's answer:

85%



Online Appendix Figure K22: Recording screen from the Balls-and-urns orator experiment.

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



Given that this ball was drawn, **how likely** do you think it is that Bag R (as opposed to Bag B) has been selected?

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

**Start recording**

**Online Appendix Figure K23:** *Explanation* screen from the Balls-and-urns receiver experiment.

## L Survey: Orator experiment

### Welcome

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?

[Recording box, activated manually]

## **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.

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.]

# M Survey: Explanation receiver experiment

## Welcome

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% (Fully 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.

[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.]

## **N Survey: Transcript receiver experiment**

### **Welcome**

In this survey, you will be asked questions on various topics. This study is designed for computer (PC or Mac) users only (desktop, laptop, etc.). If you are accessing this study on a smartphone, a tablet or any other non-PC devices, please switch to PC and enter the study again, or return the submission on Prolific.

Please write at least 15 words describing your opinion about daylight savings in the United States. Whether you are in favor or against daylight savings does not affect your eligibility to participate in this study. However, we ask that you write at least 15 words on your thoughts about this topic.

[Text box]

## **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 see another participant's answer to the question and additionally read the transcript of the explanation they recorded.
  - For other questions, you will only see the participant's answer.
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 read a number of explanations on different questions. Which one of the following statements is true?

- Before each of the 15 questions I'll be asked, I will read the transcript of an explanation recorded by a previous respondent.
- Before each of the 15 questions I'll be asked, I will see what a previous respondent answered.
- On some of the 15 questions I'll be asked, I might read the transcript of an explanation recorded by a previous respondent. On other questions, I will see what the previous respondent answered.

Which one of the following statements is true?

- Previous respondents typed their explanations into a text box.
- Previous respondents recorded their explanations as voice messages which were then automatically transcribed into text.

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% (Fully certain)]

**[Transcript Treatment:]**

Read the other respondent's explanation

[Box with question text]

Other person's answer:

[Answer of other respondent]

Other person's explanation:

[Transcript of explanation, displayed progressively]

**[Choice Only Treatment:]**

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.

[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 read 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.]

## **O Survey: Choice & Confidence receiver experiment**

### **Welcome**

In this survey, you will be asked questions on various topics. This study is designed for computer (PC or Mac) users only (desktop, laptop, etc.). If you are accessing this study on a smartphone, a tablet or any other non-PC devices, please switch to PC and enter the study again, or return the submission on Prolific.

Please write at least 15 words describing your opinion about daylight savings in the United States. Whether you are in favor or against daylight savings does not affect your eligibility to participate in this study. However, we ask that you write at least 15 words on your thoughts about this topic.

[Text box]

### **General Instructions**

This study will take approximately 15 minutes to complete. You will earn a reward of \$3.00 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 three 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 see another participant's answer to the question and the participant's confidence in their answer, on a scale from 0% to 100%. 0% means the respondent was not at all certain, 100% means the respondent was fully certain.
  - For other questions, you will only see the participant's answer.
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).

The confidence question previous respondents were asked was exactly like the confidence question you will be asked after giving your answers in each round.

## **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 gave the previous answers 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 read a number of explanations on different questions. Which one of the following statements is true?

- Before each of the 15 questions I'll be asked, I will see what a previous respondent answered and how confident they were.
- Before each of the 15 questions I'll be asked, I will see what a previous respondent answered.
- On some of the 15 questions I'll be asked, I might see what a previous respondent answered and how confident they were. On other questions, I will see what the previous respondent answered.

Which one of the following statements is true?

- Confidence is expressed as a sentence such as “I am not very confident” or “I’m extremely confident.”
- Confidence is expressed as a percentage from 0% to 100%.

Which one of the following statements is true?

- The answers I choose 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% (Fully certain)]

**[Choice & Confidence Treatment:]**

Read the other respondent's answer

[Box with question text]

Other person's answer:

[Answer of other respondent]

Other person's confidence:

[Confidence of other respondent, from 0% to 100%]

**[Choice Only Treatment:]**

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.

[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

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

# P Survey: Richness manipulation receiver experiment

## Welcome

In this survey, you will be asked questions on various topics. This study is designed for computer (PC or Mac) users only (desktop, laptop, etc.). If you are accessing this study on a smartphone, a tablet or any other non-PC devices, please switch to PC and enter the study again, or return the submission on Prolific.

Please write at least 15 words describing your opinion about daylight savings in the United States. Whether you are in favor or against daylight savings does not affect your eligibility to participate in this study. However, we ask that you write at least 15 words on your thoughts about this topic.

[Text box]

Please also complete the captcha below.

[Captcha]

## Attention Check

This 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 opinion about daylight savings in the United States. Whether you are in favor or against daylight savings does not affect your eligibility to participate in this study. However, we ask that you record a message that lasts at least 20 seconds.

[Recording box, activated manually]

## General Instructions

This study will take approximately 25 minutes to complete. You will earn a reward of \$4.50 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 three steps:

1. You provide your best answer to the question.
2. You will see another participant's answer to the question and additionally read an edited version of the explanation they recorded.
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 an explanation, the explanation will be displayed automatically. You will only be able to read it once.

We ask you not to search the answers on the internet or to use external tools:

- To confirm that you do not search for answers, the survey will monitor whether the survey window remains active.
- If you use the internet or external tools, you will not be eligible for the \$4.50 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 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 read a number of explanations on different questions. Which one of the following statements is true?

- Before each of the 15 questions I'll be asked, I will read an explanation written by the researchers behind this survey.
- Before each of the 15 questions I'll be asked, I will only see what a previous respondent answered.
- Before each of the 15 questions I'll be asked, I will read an edited version of an explanation recorded by a previous respondent.

Which one of the following statements is true?

- Previous respondents typed their explanations into a text box.
- Previous respondents recorded their explanations as voice messages which were then automatically transcribed and edited.

Which one of the following statements is true?

- The answers I choose 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% (Fully certain)]

**[Rich or Sparse Version Treatment:]**

Read the other respondent's explanation

[Box with question text]

Other person's answer:

[Answer of other respondent]

Other person's explanation:

[Transcript of explanation, displayed progressively]

**[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.

[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

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

## **Q Survey: Balls-and-urns orator experiment**

### **Welcome**

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 2025 will the average temperature recorded across all of the

U.S. be below 65°F?

[Recording box, activated manually]

## General Instructions

Thanks for recording your first voice message! This study will take approximately 10 minutes to complete. You will earn a reward of \$2.00 for completing the survey. To complete the study, you will need to read all instructions carefully and correctly answer the comprehension questions.

### Survey Structure

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.

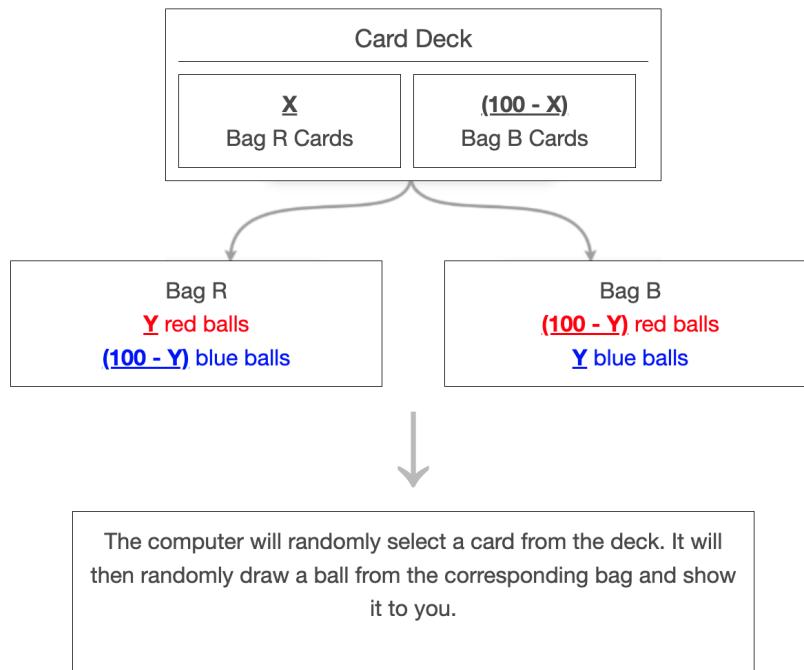
We ask you not to search the answers on the internet:

- 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 \$2.00 reward.
- You should remain focused on the survey window and answer questions as best you can using your previous knowledge.

## The Task

In this study, your task is to guess which of two bags was secretly selected by a computer. As a hint, you will see a colored ball that was drawn from the secretly-selected bag.

The setup is as follows:



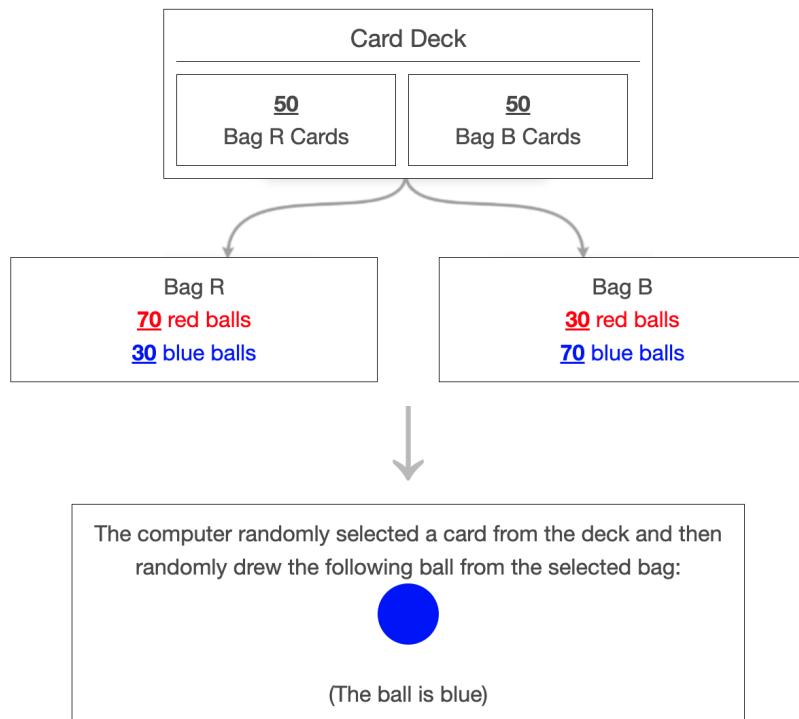
- There is a deck of 100 cards, some of which have “Bag R” written on them and some of which have “Bag B” written on them. We denote the number of cards with Bag R on them by “ $X$ ” (the number of cards with Bag B on them is  $100 - X$ ).
- There are two bags, Bag R and Bag B. Both bags contain 100 balls each, some of which are red and some of which are blue. Bag R always contains at least as many red balls as blue balls, and Bag B always contains at least as many blue balls as red balls. It is always the case that the number of red balls in Bag R is equal to the number of blue balls in Bag B. We denote this number by “ $Y$ ”.
- For example, Bag R might contain 95 red balls and 5 blue balls, and so Bag B would contain 5 red balls and 95 blue balls.

The task proceeds as follows:

1. You will be told how many cards have "Bag R" vs. "Bag B" written on them.

2. You will be told how many balls in Bag R and Bag B are red or blue.
3. The computer will randomly select one of the 100 cards. If the card has "Bag R" written on it, the computer selects Bag R. If the card has "Bag B" written on it, the computer selects Bag B. You will not observe which card was drawn, so you will not know for sure which bag was selected.
4. The computer will then randomly draw one ball from the selected bag and show it to you.
5. You will then be asked to provide a percentage chance to indicate how likely you think it is that the computer selected Bag R or B.

### Example



- In the example above, there are 50 cards labeled "Bag R" and 50 cards labeled "Bag B". Bag R contains 70 red and 30 blue balls and bag B contains 30 red and 70 blue balls.
- A blue ball was drawn.

- You would then tell us the percentage chance you think the computer selected Bag R, based on the information provided.

### **Your certainty**

During the task:

- You will tell us the percentage chance you think the computer selected Bag R.
- We will ask you how certain you are about your answer. Specifically, we are interested in how likely you think it is (in percentage terms) that your answer is within +/- 3 percentage points of the statistically correct answer.

### **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 participants who listened to your answer will be randomly chosen.
- You will receive \$10 if the participant's answer is within +/- 3 percentage points of the statistically correct answer.
- 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 their answer is within +/- 3 percentage points of the statistically correct answer. Moreover, participants listening to your recordings will be informed that you will receive a bonus if they select an answer within +/- 3 percentage points of the statistically 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 record a voice message on a question. 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 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.

Which one of the following statements is true?

- When I'm asked to indicate my certainty about my decision, the people running this study are interested in how certain I am that my estimate of the percentage chance that Bag R was selected is statistically correct.
- When I'm asked to indicate my certainty about my decision, the people running this study are interested in how certain I am that Bag R was selected.

Which one of the following statements is true?

- If the computer draws a card with “R” on it from the deck of cards, it will draw a ball from Bag R and show it to me.
- If the computer draws a card with “R” on it from the deck of cards, it will draw a ball from a randomly selected bag and show it to me.

Which one of the following statements is NOT true?

- Bags R and B always have the same proportion of red and blue balls.
- Bag R always has at least as many red balls as blue balls.
- Bag B always has at least as many blue balls as red balls.

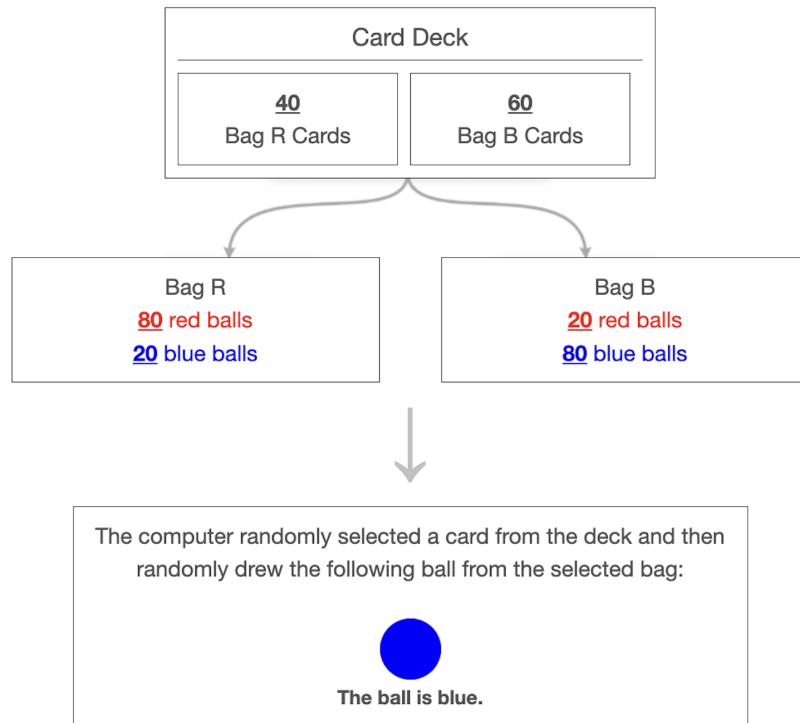
## Remember!

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

## Main Part

On the next page, the main task 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 explanation, you will select your own answer to the question.

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



Given that this ball was drawn, how likely do you think it is that Bag R (as opposed to Bag B) has been selected?

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.

Given that this ball was drawn, how likely do you think it is that Bag R (as opposed to Bag B) has been selected?

[Input X]% likely that Bag R was selected.

[ $100 - X$ ]% likely that Bag B was selected.

How certain are you that your above answer is within +/- 3 percentage points of the statistically correct answer?

[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 the question or use external tools, like an AI assistant, 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.]

## **R Survey: Balls-and-urns receiver experiment**

### **Welcome**

In this survey, you will be asked to make an estimate. Besides providing your estimate, you might listen to a voice message of someone else's thoughts on the correct estimate. 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**

Before providing your estimate, you may be asked to first listen to a voice message from another participant. In a previous survey, we asked respondents to record their thoughts on the same estimation task.

## **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.

[Play button screenshot here]

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 10 minutes to complete. You will earn a reward of \$2 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 complete an estimation task. In each round, there are four steps:

1. You provide your best answer to the task.
2. You get information about a previous respondent’s answer:
  - You might listen to a voice message of another participant once.
  - You might see the answer of another participant.
3. You have a second chance to provide your best answer to the task. 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).

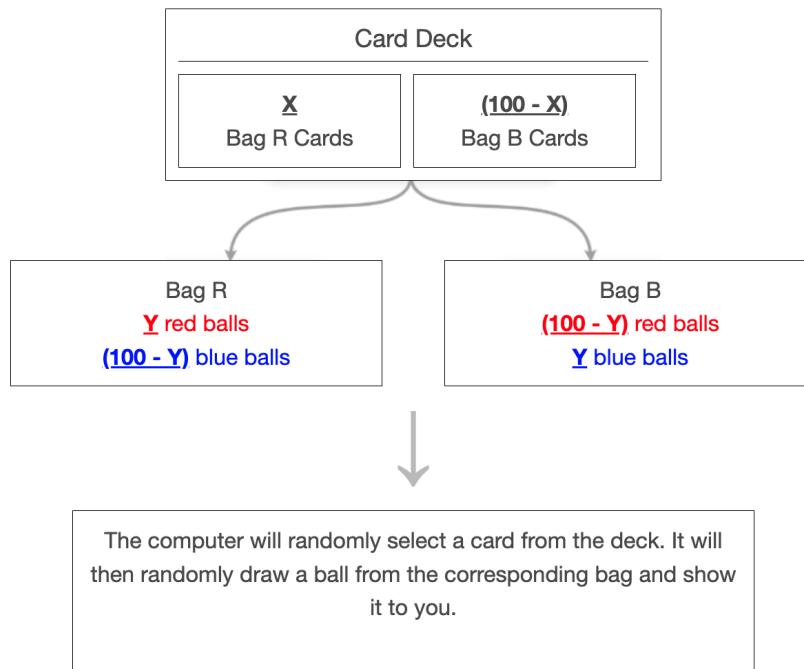
When you enter a page with a recording, the recording will play automatically. You will only be able to listen to it once.

We ask you not to search the answers on the internet or to use external tools:

- 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 \$2 reward.
- You should remain focused on the survey window and answer questions as best you can using your previous knowledge.

### The Task

In this study, your task is to guess which of two bags was secretly selected by a computer. As a hint, you will see a colored ball that was drawn from the secretly-selected bag.



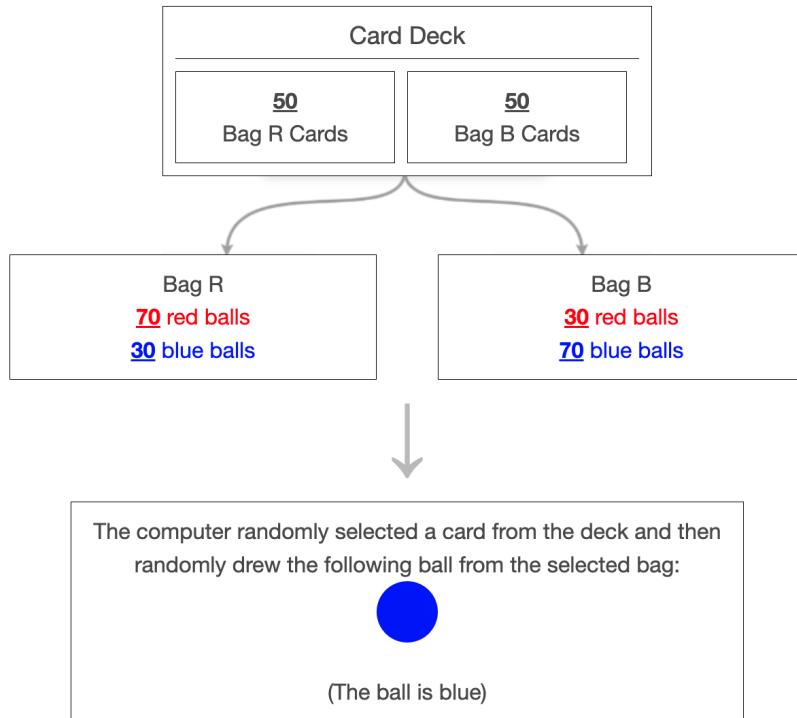
- There is a deck of 100 cards, some of which have “Bag R” written on them and some of which have “Bag B” written on them. We denote the number of cards with Bag R on them by “ $X$ ” (the number of cards with Bag B on them is  $100 - X$ ).
- There are two bags, Bag R and Bag B. Both bags contain 100 balls each, some of which are red and some of which are blue. Bag R always contains at least as many red balls as blue balls, and Bag B always contains at least as many blue balls as red balls. It is always the case that the number of red balls in Bag R is equal to the number of blue balls in Bag B. We denote this number by “ $Y$ ”.

- For example, Bag R might contain 95 red balls and 5 blue balls, and so Bag B would contain 5 red balls and 95 blue balls.

The task proceeds as follows:

1. You will be told how many cards have "Bag R" vs. "Bag B" written on them.
2. You will be told how many balls in Bag R and Bag B are red or blue.
3. The computer will randomly select one of the 100 cards. If the card has "Bag R" written on it, the computer selects Bag R. If the card has "Bag B" written on it, the computer selects Bag B. You will not observe which card was drawn, so you will not know for sure which bag was selected.
4. The computer will then randomly draw one ball from the selected bag and show it to you.
5. You will then be asked to provide a percentage chance to indicate how likely you think it is that the computer selected Bag R or B.

### **Example**



- In the example above, there are 50 cards labeled "Bag R" and 50 cards labeled "Bag B". Bag R contains 70 red and 30 blue balls and Bag B contains 30 red and 70 blue balls.
- A blue ball was drawn.
- Step 1: You enter the percentage chance you think the computer selected Bag R, based on the information provided.
- Step 2: You see a randomly drawn other participant's answer, and you may listen to their explanation. Note: that participant saw exactly the same problem, including the same draw of the ball.
- Step 3: You enter the percentage chance you think the computer selected Bag R, based on the information provided and the other participant's answer.

### Your certainty

During the task:

- You will tell us the percentage chance you think the computer selected Bag R.

- We will ask you how certain you are about your answer. Specifically, we are interested in how likely you think it is (in percentage terms) that your answer is within +/- 3 percentage points of the statistically correct answer.

### **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:

- Either your answer from step (1) or your answer from step (3) will be randomly chosen.
- You will receive \$10 if your answer is within +/- 3 percentage points of the statistically correct answer.

Participants who made the recordings were informed they had a chance to win a bonus of \$10 if you selected an answer within +/- 3 percentage points of the statistically correct answer. They were also informed that you had a chance to win a bonus of \$10 if you selected an answer within +/- 3 percentage points of the statistically 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.

1. In this study, you will listen to a number of voice messages on different questions. Which one of the following statements is true?
  - I will listen to a voice message from another respondent about the task.
  - I will see what another respondent answered on the task.
  - I might listen to a voice message from another respondent about the task, or see what another respondent answered on the task.
2. Which one of the following statements is true?
  - On a page with a voice recording, the recording only starts playing once I click “Play”.
  - On a page with a voice recording, the recording starts playing automatically when I enter, so I should read the question and pay attention to the recording.
3. Which one of the following statements is true?

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

4. Which one of the following statements is true?

- When I'm asked to indicate my certainty about my decision, the people running this study are interested in how certain I am that my estimate of the percentage chance that Bag R was selected is statistically correct.
- When I'm asked to indicate my certainty about my decision, the people running this study are interested in how certain I am that Bag R was selected.

5. Which one of the following statements is true?

- If the computer draws a card with "R" on it from the deck of cards, it will draw a ball from Bag R and show it to me.
- If the computer draws a card with "R" on it from the deck of cards, it will draw a ball from a randomly selected bag and show it to me.

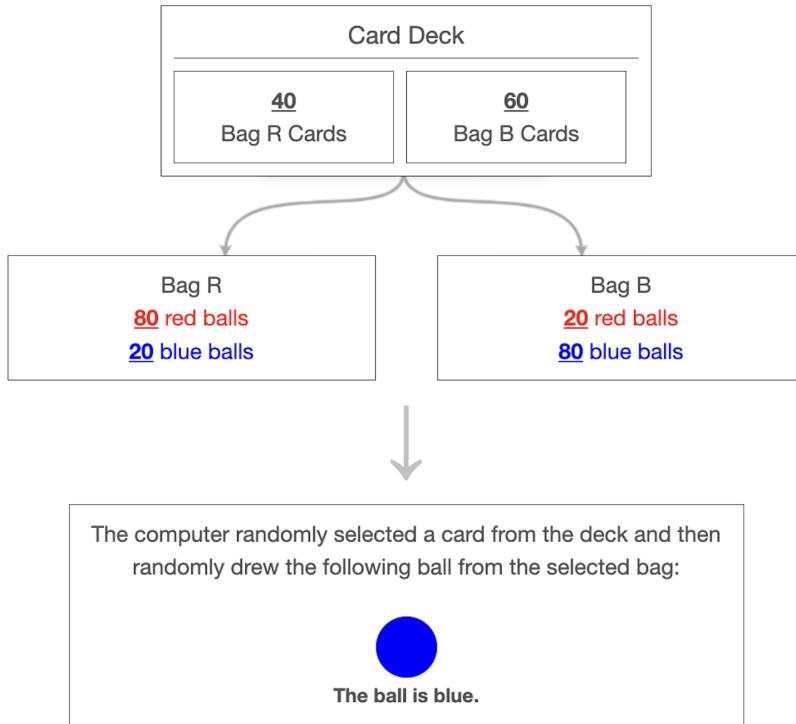
6. Which one of the following statements is NOT true?

- Bags R and B always have the same proportion of red and blue balls.
- Bag R always has at least as many red balls as blue balls.
- Bag B always has at least as many blue balls as red balls.

## Main Part

On the next page, the main task will be displayed.

**[Prior Choice:] Provide your best answer**



Given that this ball was drawn, how likely do you think it is that Bag R (as opposed to Bag B) has been selected?

[Input X]% likely that Bag R was selected.

[100 – X]% likely that Bag B was selected.

How certain are you that your above answer is within +/- 3 percentage points of the statistically correct answer?

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

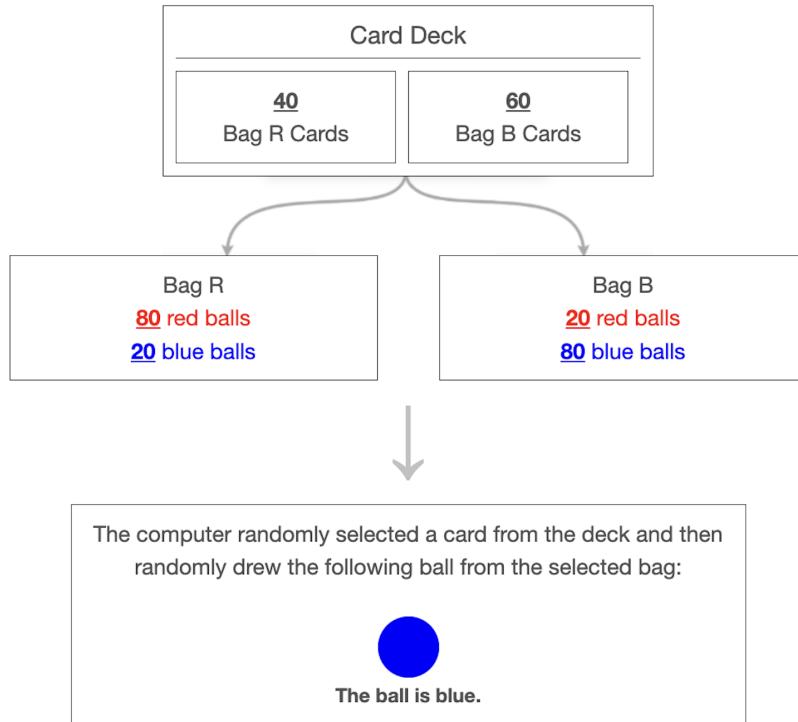
#### [Explanation Treatment:]

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

Other person's answer:

[Answer of other respondent]

[Recording of other respondent, on auto-play]

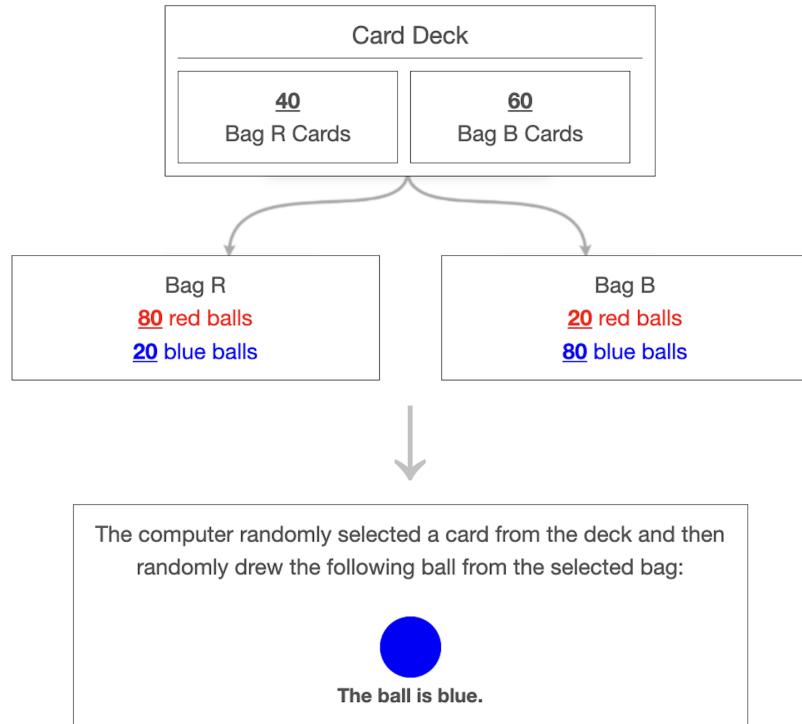
#### **[Choice Only Treatment:]**

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



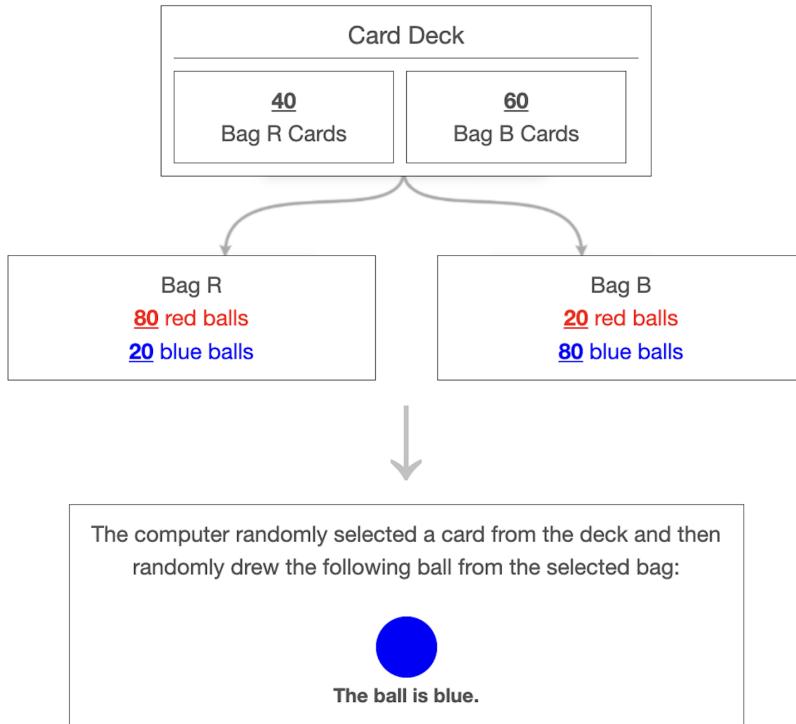
Other person's answer:

[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.



Given that this ball was drawn, how likely do you think it is that Bag R (as opposed to Bag B) has been selected?

[Input X]% likely that Bag R was selected.

[100 – X]% likely that Bag B was selected.

How certain are you that your above answer is within +/- 3 percentage points of the statistically correct answer?

[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 the question or use external tools, like an AI assistant, 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.]