Online Appendix for:
The effect of incentives on motivated numeracy amidst COVID-19

Table of Contents

Survey questionnaire

General Instructions and Consent

This is a research survey on public attitudes. There are two parts to this survey. In the first part, you will be asked some questions related to numbers. All data questions are hypothetical. In the second part, you will be asked to answer basic questions on demographics.

All of your responses in this survey are anonymous and will be used for research purposes only. Your response will be of great help in policy research.

Taking part in this survey is voluntary, and all of your responses will be done through a computer or mobile device. The entire survey will take up to 15 minutes. Do not press the "back" button during the survey.

There are no known risks to this study. All of your responses will remain completely confidential. If you are uncomfortable with the survey, you may refuse to participate or stop at any time without consequences. Your personal information and answers are not disclosed. For questions, concerns or complaints about the study, or if you feel you have been harmed by taking part in the study, you may contact the research team at [researcher email]. Once you have read and understood the above, select the appropriate bubble below.

Select one option:
- Yes. I give my consent to participate in this study. (1)
- No. I do not give my consent to participate in this study. (2)

Skip To: End of Survey If Select one option: = No. I do not give my consent to participate in this study.

Start of Block: Treatment; Block randomized to appear in beginning of survey for roughly half of sample; Correct answer coded as 1, incorrect 0

Here is an arithmetic problem.

[**Incentive Treatment**:] We will give you a bonus of $1.00 for the correct answer.
[**No-Incentive Treatment**:] Your answer to this question will not affect the total amount you earn in this survey.

Researchers have measured COVID-19 rates across American cities to understand what characteristics of cities correspond to rising or declining COVID-19 cases. Whether a city adopted a mask-wearing mandate was one of the characteristics measured in this study.

In each group (cities with a mask mandate and cities without a mask mandate), researchers recorded the number of cities where COVID-19 cases increased/decreased since the mandate was implemented. These numbers are recorded in the table below. The exact number of cities in each group is not the same, but this does not prevent assessment of the results.

Looking at the results of the study summarized in the table below, do you think the cities with a mask-wearing mandate saw an increase or a decrease in COVID-19 cases? Look at the following table, calculate the correct answer, and write it below.

<Results: Either 1 or 2 is shown randomly>

Condition 1: COVID-19 cases decrease

	COVID-19-positive cases increased	COVID-19-positive cases decreased
Cities with a mask-wearing mandate	223	75
Cities without a mask-wearing mandate	107	21

Condition 2: COVID-19 cases increase

	COVID-19-positive cases decreased	COVID-19-positive cases increased
Cities with a mask-wearing mandate	223	75
Cities without a mask-wearing mandate	107	21

When you see only the results of this study, which of the following conclusions do you think this study supports? Please calculate based on the numbers given in the table above.

- Cities with a mask-wearing mandate saw an INCREASE in COVID-19 cases.
- Cities with a mask-wearing mandate saw a DECREASE in COVID-19 cases.

End of Block: Treatment - correct answer coded as 1=DECREASE in condition 1 and as 1= INCREASE in condition 2, incorrect as 0

Start of Block: Manipulation check

Please select one answer below.

The previous question asked me about whether cities with a mask-wearing mandate saw an increase or a decrease in COVID-19 cases.

- True (1)
- False (0)

End of Block: Manipulation check

Start of Block: Numeracy: Block randomized to appear in beginning of survey for roughly half of sample; correct answer coded as 1, all others 0

This section of the study contains questions related to numbers. Please answer the following questions to the best of your ability.

We will pay you for answering the questions in this section correctly. One of six numeracy questions will be randomly selected and if your answer is correct in that question, you will receive $1.

Imagine that we roll a fair, six-sided die 1,000 times. Out of 1,000 rolls, how many times do you think the die would come up as an even number?

- 250 times (0)
- 333 times (0)
- 500 times (1)
- 600 times (0)

In the Big Bucks Lottery, the chances of winning a $10.00 prize are 1%. What is your best guess about how many people would win a $10.00 prize if 1,000 people each buy a single ticket from Big Bucks?

- 1 person (0)
- 10 people (1)
- 100 people (0)
- 200 people (0)

In the Acme Publishing Sweepstakes, the chance of winning a car is 1 in 1,000. What percentage of tickets to Acme Publishing Sweepstakes win a car?

- 0.01% (0)
- 0.1% (1)
- 1% (0)
- 10% (0)

A baseball and bat cost $1.10 in total. The bat costs $1.00 MORE than the ball. How much does the ball cost?

- 5 cents (1)
- 10 cents (0)
- 50 cents (0)
- 55 cents (0)

Imagine that you are taking a class and your chances of being asked a question in class are 1% during the first week of class and double each week thereafter (i.e., you would have a 2% chance in Week 2, a 4% chance in Week 3, an 8% chance in Week 4). What is the probability that you will be asked a question in class during Week 7?

- 14% (0)
- 16% (0)
- 64% (1)
- 128% (0)

In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake?

- 24 days (0)
- 6 days (0)
- 7 days (0)
- 47 days (1)

End of Block: Numeracy: correct answer coded as 1, all others 0

Start of Block: Demographics

Thank you, welcome to the second part of the study.

Please answer the questions below.

What gender do you identify as?

- Female (1)
- Male (2)
- Other (3)
- Prefer not to say (4)

Please type your age, as a number, in the space below.

____

In what city / state do you live?

- City (1) ____
- State (2) ____

Which of the following best describes your racial or ethnic identification? Select all the apply.

- White European (1)
- African American (2)
- Jewish (3)
- Latino (4)
- Asian / Pacific Islander (5)
- Native American (6)
- Other (7)
- Prefer not to say (8)

Page Break

What is the final degree that you have obtained?

- Did not complete high school (1)
- High school graduate (2)
- Some college (3)
- College graduate (4)
- Postgraduate (5)
- Other (6)

Generally speaking, how politically liberal or conservative are you?

- Very Liberal (1)
- Liberal (2)
- Somewhat Liberal (3)
- Moderate (4)
- Somewhat Conservative (5)
- Conservative (6)
- Very Conservative (7)

With which party do you identify, and how strongly?

- Strong Democrat (1)
- Democrat (2)
- Democratic-Leaning Independent (3)
- Independent (4)
- Republican-Leaning Independent (5)
- Republican (6)
- Strong Republican (7)

Who did you vote for in the 2016 presidential elections?

- Donald Trump
- Hilary Clinton
- Other candidate
- Did not vote
- I'd rather not say
- Other

End of Block: Demographics

Start of Block: Post-Experimental Questionnaire for All

This is the last question in the survey. What do you think the survey was about?

____

Detailed description of simulations for pre-analysis power calculations

Simulations for power calculations

The power was calculated using Monte Carlo (MC) simulations with 1000 repetitions. In each repetition, a dataset of sample size N (where N is varied) was generated using information about the distribution of variables and the logistic regression model in Table A1 from Kahan et al. (2017), and the following three effect size assumptions for the incentivized condition. First, we assumed that the incentives would increase accuracy by 30% [1] on average for each numeracy/ideology level. Second, for each unit increase in partisanship, we assumed that incentives would produce a 15% decrease in bias.[2] Third, we assumed that among incentivized respondents, higher numeracy would increase the difference in answers across the uncongenial and congenial condition, but at a 30% lower rate than in the non-incentivized condition.[3]

Using this dataset, we employed the regression models specified in equations (1), (2), and (3) to test the null hypotheses at a 5% significance level by calculating the average number of rejections of the null hypotheses across the 1000 repetitions. This average number of rejections is the power associated with rejecting the null hypotheses at a 5% level with the sample size N. We find the N that satisfies the requirement of 80% power to reject the null at a 5% level for the proposed hypotheses. The obtained sample size of 3050 respondents allows for greater than 90% probability of rejecting each of the four null hypotheses at the 5% level of significance.

The following sections include the detailed pre-analysis of the power calculations for each hypothesis—these pre-analyses plans were peer-reviewed before data collection started.

Pre-analyses for hypotheses 1 and 2

$$\begin{aligned}\text{Correct}_i = \beta_0 &+ \beta_1 \text{Numeracy}_i \\ &+ \beta_2 \text{Congenial}_i \\ &+ \beta_3 \text{Numeracy}_i \times \text{Congenial}_i \\ &+ \beta_4 \text{Numeracy}_i^2 + \epsilon_i\end{aligned} \tag{1}$$

To test hypotheses 1 and 2, we will run a linear probability model to estimate the parameters in equation (1) for data from the non-incentivized conditions.

Hypothesis 1 posits that accuracy is higher when interpreting data more congenial with one's ideological beliefs, i.e., $\beta_2 > 0$. A two-sided z-test will verify whether to reject the null $\beta_2 =$

[1] Incentives generate an 18% (Study1 in Prior et al., 2015) and 55% increase in accuracy (Bullock et al., 2015). We choose a conservative effect size of 30% as the average finding in previous research.
[2] Prior et al. (2015) find a 60% reduction in partisan bias. Since our measure of conservatism is on a 5-point scale, the reduction in bias with one-unit increase is assumed to be 60/4, i.e., 15%.
[3] We don't have previous literature to guide us in this assumption of effect size but we will be able to detect a meaningful effect size of 30% or more.

0. If β_2 is not statistically significant at 5% level, we will interpret the result as the lack of evidence to reject the null. If β_2 is positive and statistically significant at 5% level, we will interpret the result as evidence to reject the null and support the hypothesis that $\beta_2 > 0$. If β_2 is negative and is statistically significant at a 5% level, we will interpret the result as evidence to reject the null hypothesis and support the hypothesis that $\beta_2 < 0$.

Hypothesis 2 posits that the rate of increase in accurate answers increases with numeracy when interpreting data more congenial with one's ideological beliefs, i.e., $\beta_3 > 0$. A positive coefficient on the interaction term of congenial and numeracy suggests that higher numeracy subjects will have greater differences in correct answers between congenial and uncongenial conditions compared to lower numeracy subjects. A two-sided z-test will verify whether to reject the null $\beta_3 = 0$. If β_3 is not statistically discernible at 5% level, we interpret the result as the lack of evidence to reject the null. If β_3 is positive and statistically significant at 5% level, this will be evidence to reject the null hypothesis and support the hypothesis that $\beta_3 > 0$. If β_3 is negative and statistically significant at a 5% level, we will interpret the result as evidence to reject the null and support the hypothesis that $\beta_3 < 0$.

With a sample size of 1000 for the non-incentivized condition (split equally between $Covid\_increases_i$ and $Covid\_decreases_i$ conditions), we will be able to reject the null hypothesis $\beta_2 = 0$ at 5% level of significance with 100% power and the null hypothesis $\beta_3 = 0$ at a 5% level of significance with 99.5% power.

Even if the result is statistically discernible, it may not be substantively meaningful. To discuss the substantive significance, we will conduct the inferiority test (Lakens et al., 2018). Since Hypotheses 1 and 2 are conceptual replication of Kahan et al. (2017), we will test whether the observed effect size is smaller than the smallest effect size that the original study could have detected. Using simulated data of that study, we find that the minimum effect size of β_2 in equation (1) that could have been detected with 81.2% power using their sample of 1111 is 0.04 in a linear probability model. We will set the inferiority bound for the test to be $\Delta = 0.04$ for β_2. Thus, we would conclude that a meaningful effect is absent if the observed effect size is reliably lower than the smallest effect size of interest, which is 0.04 (one-sided z-test).

Similarly, we find that the minimum effect size of β_3 in equation (1) that could have been detected with 82.9% power using their sample size of 1111 is 0.021 in a linear probability model. We will set the inferiority bound for the test to be $\Delta = 0.021$ for β_3. Thus, we would conclude that a meaningful effect is absent if the observed effect size is reliably smaller than the smallest effect size of interest, which is 0.021 (one-sided z-test).

Pre-analysis for hypothesis 3

$$\text{Correct}_i = \beta_0 + \beta_1 \text{Incentive}_i + \epsilon_i \tag{2}$$

To test hypothesis 3, we will use a linear probability model to estimate parameters in equation (2). We estimate the effect of incentives on correct answers using all observations.

We hypothesize that incentivized participants are more likely to answer correctly the contingency table question compared to those unincentivized. A two-sided z-test will indicate if we can reject the null $\beta_1 = 0$. If β_1 is not statistically discernible at 5% level, we will interpret the result as the lack of evidence to reject the null hypothesis. If β_1 is positive and statistically significant, we will interpret the result as evidence to reject the null hypothesis and support the hypothesis that $\beta_1 > 0$. If β_1 is negative and statistically discernible, we will interpret the result as evidence to reject the null hypothesis and support the hypothesis that $\beta_1 < 0$. With a sample size of 2000 for incentivized conditions and 1000 for unincentivized conditions, we will be able to reject the null $\beta_1 = 0$ at 5% level with 100% power.[4]

Pre-analysis for hypothesis 4

$$\begin{aligned}\text{Correct}_i = \beta_0 &+ \beta_1\text{Numeracy}_i \\ &+ \beta_2\text{Congenial}_i \\ &+ \beta_3\text{Numeracy}_i \times \text{Congenial}_i \\ &+ \beta_4\text{Numeracy}_i^2 \\ &+ \beta_5\text{Incentive}_i \\ &+ \beta_6\text{Incentive}_i \times \text{Congenial}_i \\ &+ \beta_7\text{Incentive}_i \times \text{Numeracy}_i \\ &+ \beta_8\text{Incentive}_i \times \text{Numeracy}_i \times \text{Congenial}_{ii} + \epsilon_i\end{aligned} \tag{3}$$

To test hypothesis 4, we will use a linear probability model to estimate parameters in equation (3). Hypothesis 4 posits that the congeniality bias among incentivized respondents increases at a *lower* rate with one's numeracy, compared to the rate of congeniality bias increase among unincentivized respondents., i.e., the coefficient $\beta_3 > 0$ and $\beta_8 < 0$ in equation (3). Two-sided z-tests will verify whether to reject the null $\beta_3 = 0$ *and* $\beta_8 = 0$. If β_3 (β_8) is not statistically discernible at 5% level, we will interpret the result as the lack of evidence to reject the null. If β_3 (β_8) is positive and statistically significant at a 5% level, we will interpret the result as evidence to reject the null and support the hypothesis that $\beta_3 > 0$ ($\beta_8 > 0$). If β_3 (β_8) is negative and statistically significant at 5% level, this will be evidence to reject the null hypothesis and support the hypothesis that $\beta_3 < 0$ ($\beta_8 < 0$).

With a sample size of 1000 for the non-incentivized condition and 2000 for the incentivized condition, we can reject the null $\beta_3 = 0$ at 5% level with 99.8% power and $\beta_8 = 0$ at 5% level with 98.2% power.

[4] This experiment is well-powered: we collect 3,000 responses as this survey precedes another study within the same project.

Detailed description of data exclusion criteria based on quality checks

After the treatment, respondents answered the factual manipulation check question, i.e., an objective question about the study's content (Kane and Barabas, 2019); those who failed the check were removed from the sample.

Additionally, Qualtrics removed 'careless or insufficient effort (C/IE)' responses by identifying speeders (who took less than half the median time), those who took too long or took long breaks mid-survey, duplicates (based on demographics and IP/geo data), "straightliners" (respondents answering multiple consecutive questions in the same manner or those who create patterns in grids), likely bots (based on the algorithm developed by Qualtrics, which checks for specific demographic profiles and attitudinal responses), and responses from outside the U.S. (checked via IP/geo data).

Nested quotas

Research finds that many human subject samples tend to overrepresent educated minorities and politically active people (Pew Research Center 2016). For a more accurate representation of U.S. residents, we calculated nested quotas of ethnicity and education based on census data and data from the American Council on Education. Qualtrics was unable to fulfill all planned nested quotas, nevertheless, our sample more accurately reflects the ethnic and educational composition of the population of U.S. residents compared to other studies that do not take these issues into consideration. For example, Kahan et al. (2017) use a “nationally diverse” sample without specification of which, if any, markers of the sample are representative. Similarly, Bullock et al. and Khanna and Sood (2018) use two types of samples: 1) MTurk samples, which are more liberal, young, white, male than the general population, and 2) a Qualtrics sample that is more representative in party and ideology, but still more educated, white, and female than the general population. In short, our sample with nested quotas is more representative of the population than many existing studies similar by topic.

Table A1: Planned vs. Actual Overall Quotas for 2016 political behavior:

Vote in 2016	% of Total	N of Total (out of 3050)	Actual N	Actual % of Total
A. Non-voters	45%	3050*0.45=1372.5 ≈ **1373**	**861**	**28%**
B. Voted for Trump in 2016	45.9% of the vote = 25%	3050*0.25=762.5 ≈ **763**	**938**	**31%**
C. Voted for Clinton in 2016	48% of the vote = 26%	3050*0.26=**793**	**978**	**32%**
D. Voted for other in 2016	4%	3050*0.04=**122**	**149**	**5%**
E. Prefer not to say/ Others			**124**	**4%**

Table A2: Planned vs. Actual Overall Quotas for Race Ethnicity:

Race/Ethnicity	% of Total	Planned N of Total (out of 3050)	Actual N of Total	Actual % of Total
A. Non-Hispanic White	61.9%	1887.95 ≈ **1888**	**2126**	**70%**
B. Non-Hispanic Black	12.3%	375.15 ≈ **375**	**163**	**5%**
C. Hispanic	17.4%	530.7 ≈ **531**	**492**	**16%**
D. Asian	5.3%	161.65 ≈ **162**	**147**	**5%**
E. American Ind./Alaskan Native	0.7%	21.35 ≈ **21**	**14**	**1%**
F. Other Race	2.5%	76.25 ≈ **76**	**73**	**2%**
G. Prefer not to say			**35**	**1%**

Table A3. Non-Hispanic White—Planned vs. Actual Nested Quotas by Education

Education Level	% of White non-Hispanic	Calculation for the planned N of Total	Planned N of Total	Planned % of Total (out of 3050)	Actual N	Actual % of Total
Less than High School	5.9%	0.059*1888=111.392	**111**	**4%**	**70**	**2%**
High School Graduate / GED	28.3%	0.283*1888=534.304	**534**	**17%**	**584**	**19%**
Some College but No Degree	16.6%	0.166*1888=313.408	**313**	**10%**	**539**	**18%**
Associate's Degree	11.2%	0.112*1888=211.456	**211**	**7%**	**-**	**-**
Bachelor's Degree	23.7%	0.237*1888=447.456	**447**	**15%**	**571**	**19%**
Graduate Degree (combines Master's, Professional, and Doctoral Degree)	14.3%	0.143*1888=269.984	**270**	**9%**	**353**	**12%**
Others					**9**	

Table A4: Non-Hispanic Black—Planned vs. Actual Nested Quotas by Education

Education Level	% of Black	Calculation for the planned N of Total	Planned N of Total	Planned % of Total	Actual N	Actual % of Total
Less than High School	11.90%	0.119*375=44.625	**45**	**1%**	**7**	
High School Graduate / GED	33%	0.330*375=123.75	**124**	**4%**	**42**	**1%**
Some College but No Degree	20.40%	0.204*375=76.5	**77**	**3%**	**38**	**1%**
Associate's Degree	10.30%	0.103*375=438.625	**39**	**1%**	**-**	**-**
Bachelor's Degree	15.30%	0.153*375=57.375	**57**	**2%**	**50**	**2%**
Graduate Degree (combines Master's, Professional, and Doctoral Degree)	8.90%	0.089*375=33.375	**33**	**1%**	**26**	**1%**
Others					**0**	

Table A5: Hispanic—Planned vs. Actual Nested Quotas by Education

Education Level	% of Hispanic	Calculation for the N of Total	Planned N of Total	Planned % of Total	Actual N	Actual % of Total
Less than High School	29.5%	0.295*531=156.645	**157**	**5%**	**26**	**1%**
High School Graduate / GED	31%	0.310*531=164.61	**165**	**5%**	**164**	**5%**
Some College but No Degree	14.4%	0.144*531=76.464	**76**	**2%**	**125**	**4%**
Associate's Degree	8%	0.080*531=42.48	**42**	**1%**	**-**	
Bachelor's Degree	12.2%	0.122*531=64.782	**65**	**2%**	**125**	**4%**
Graduate Degree (combines Master's, Professional, and Doctoral Degree)	5.1%	0.051*531=27.081	**27**	**1%**	**46**	**2%**
Others					**6**	

Table A6: Asian—Planned vs. Actual Nested Quotas by Education

Education Level	% of Asian	Calculation for N of Total	Planned N of Total	Planned % of Total	Actual N	Actual % of Total
Less than High School	9.1%	0.091*162=14.742	**15**		**3**	
High School Graduate / GED	19.9%	0.199*162=32.238	**32**	**1%**	**15**	
Some College but No Degree	9.3%	0.093*162=15.066	**15**		**29**	**1%**
Associate's Degree	6.4%	0.064*162=10.368	**10**		**-**	
Bachelor's Degree	30.7%	0.307*162=49.734	**50**	**2%**	**57**	**2%**
Graduate Degree (combines Master's, Professional, and Doctoral Degree)	24.7%	0.247*162=40.014	**40**		**43**	**1%**
Others					**0**	

Descriptive statistics

Table A7: Participant Characteristics

	Mean	Standard Deviation
Outcome		
Correct answer(%)	42.23	49.40
Treatments and Theoretical Variables		
Covid-Increase treatment(%)	50.69	50.00
Covid-Decrease treatment(%)	49.31	50.00
Numeracy (No. of correct answers out of 6)	2.70	1.65
Numeracy (standardized: -2.70 to 3.30, centered at 0)	1.19e-08	1.65
Incentive(%)	66.69	47.14
Conservative (-3 to 3)	-0.13	1.66
Conservative (standardized: -1.73 to 1.88, centered at 0)	-8.94e-09	1
Congenial	0.00	1
Controls		
Age	52.38	17.99
Female(%)	52.54	49.94
Non-hispanic white(%)	69.70	45.96
Hispanic(%)	16.13	36.79
Other races(%)	14.16	34.87
High school or less(%)	31.34	46.40
College grad/Some college(%)	52	49.97
Post grad/Others(%)	16.66	37.26
Vote Donald Trump(%)	30.76	46.16
Vote Hillary Clinton(%)	32.07	46.68
No Vote(%)	28.23	45.02
Observations	3050	

Balance on observable covariates

Table A8: Balance on observable covariates among treatment groups

	Variable Mean and Standard Error				Difference in means **p-values**					
	No incentive COV-dec (1)	**No incentive COV-inc** (2)	**Incentive COV-dec** (3)	**Incentive COV-inc** (4)	**(1)vs(2)**	**(1)vs(3)**	**(1)vs(4)**	**(2)vs(3)**	**(2)vs(4)**	**(3)vs(4)**
Age	52.34 (0.82)	51.62 (0.82)	52.65 (0.56)	52.51 (0.55)	0.531	0.758	0.866	0.299	0.366	0.859
Female	0.53 (0.02)	0.55 (0.02)	0.51 (0.02)	0.53 (0.02)	0.518	0.334	0.842	0.088	0.343	0.346
Non-Hispanic White	0.66 (0.02)	0.68 (0.02)	0.70 (0.01)	0.72 (0.01)	0.623	0.151	0.029	0.386	0.108	0.366
Hispanic	0.18 (0.02)	0.18 (0.02)	0.16 (0.01)	0.15 (0.01)	0.964	0.199	0.095	0.220	0.108	0.640
Other races	0.15 (0.02)	0.14 (0.02)	0.14 (0.01)	0.13 (0.01)	0.547	0.603	0.275	0.864	0.690	0.484
High school or less	0.33 (0.02)	0.32 (0.02)	0.32 (0.01)	0.29 (0.01)	0.670	0.628	0.093	0.996	0.238	0.145
College grad/Some college	0.50 (0.02)	0.53 (0.02)	0.51 (0.02)	0.53 (0.02)	0.313	0.508	0.165	0.617	0.830	0.380
Post grad/Others	0.17 (0.02)	0.15 (0.02)	0.16 (0.01)	0.17 (0.01)	0.406	0.783	0.816	0.495	0.229	0.534
Donald Trump	0.31 (0.02)	0.30 (0.02)	0.32 (0.01)	0.30 (0.01)	0.705	0.844	0.741	0.528	0.914	0.519
Hillary Clinton	0.29 (0.02)	0.34 (0.02)	0.33 (0.01)	0.32 (0.01)	0.073	0.096	0.112	0.679	0.613	0.914
No Vote	0.31 (0.02)	0.28 (0.02)	0.27 (0.01)	0.28 (0.01)	0.225	0.089	0.207	0.755	0.889	0.578
Sample Size	512	504	992	1042						

Standard errors of means are in paratheses in columns 1-4.
These standard errors are calculated using the formula $stddev/\sqrt{(samplesize)}$.

Differences-in-means

Differences in accuracy by congeniality among unincentivized respondents

Table A9 describes mean levels of the congeniality bias among unincentivized respondents by comparing accuracy rates between respondents facing less and more congenial data tasks. While *Congenial*$_i$ is a continuous measure of consistency of the data with one's ideological outlook, we dichotomized this variable at three cutoff points to obtain differences-in-means displayed in Table A9. The three cutoffs are: less than 25th percentile (strongly uncongenial data), more than 50th, and more than 75th (strongly congenial data).

Table A9: Differences in accuracy by levels of congeniality (unincentivized participants)

		Strongly uncongenial data			
	Congeniality > 25pct	Congeniality < 25pct	Difference	t-statistic	p-value
Accuracy rate	0.45	0.34	0.10	2.84	0.00
		Rather congenial data			
	Congeniality < 50pct	Congeniality > 50pct	Difference	t-statistic	p-value
Accuracy rate	0.39	0.46	-0.07	-2.12	0.03
		Strongly congenial data			
	Congeniality < 75pct	Congeniality > 75pct	Difference	t-statistic	p-value
Accuracy rate	0.41	0.47	-0.06	-1.72	0.09

Consider the top row of Table A9: 34% of individuals facing uncongenial data (less than 25th percentile) interpreted the table correctly, which is 10 percentage points (pp) lower (the difference is statistically discernible at 0.05 level in a two-tail test) than respondents given congenial data. The bottom two rows reveal the same trend: accuracy rates were by 7 and 6 pp higher among respondents interpreting congenial data (above the 50th and above the 75th percentile of congeniality); the latter result is not statistically discernible at 0.05 level (p=0.09).

Differences in accuracy by congeniality and numeracy among unincentivized respondents

Table A10 replicates Table A9 while further dividing participants into less numerate (half a standard deviation below the mean) and more numerate (half a standard deviation above the mean). Table A10, therefore, describes mean levels of the congeniality bias among unincentivized respondents accounting for one's numeracy. While there is no difference between less and more numerate respondents facing strongly uncongenial data (the top scenario of Table A10), more numerate individuals were by 13 and 11 pp more likely to answer correctly interpreting congenial data, relative to less numerate individuals respectively (the effect is not statistically significant for the 75th percentile cutoff at 0.05 level, p=0.07). Unincentivized, less numerate respondents exhibited no congeniality bias.

Table A10: Differences in accuracy by levels of congeniality and numeracy (unincentivized participants)

	Strongly uncongenial data				
	Congeniality > 25pct	Congeniality < 25pct	Difference	t-statistic	p-value
Accuracy rate for less numerate	0.46	0.36	0.10	1.37	0.17
Accuracy rate for more numerate	0.43	0.34	0.09	1.50	0.13
	Rather congenial data				
	Congeniality < 50pct	Congeniality > 50pct	Difference	t-statistic	p-value
Accuracy rate for less numerate	0.44	0.43	0.01	0.20	0.84
Accuracy rate for more numerate	0.35	0.48	−0.13	−2.35	0.02
	Strongly congenial data				
	Congeniality < 75pct	Congeniality > 75pct	Difference	t-statistic	p-value
Accuracy rate for less numerate	0.44	0.43	0.01	0.10	0.92
Accuracy rate for high numeracy	0.38	0.49	−0.11	−1.80	0.07

Differences in accuracy by incentive

The third expectation posits that monetary incentives increase accuracy (this is the only hypothesis that can be tested by obtaining the difference-in-means). The data are not consistent with this proposition, as there is no difference in the accuracy rates of incentivized and unincentivized respondents (42% of both groups interpreted the contingency table correctly).

Table A11: Differences in accuracy by incentive

	Incentive=0	Incentive=1	Difference	t-statistic	p-value
Accuracy rate	0.42	0.42	0.00	0.07	0.94

Differences in accuracy by congeniality and numeracy among incentivized respondents

Tables A10 (unincentivized respondents) and A12 (incentivized respondents) replicate Table 1 while further differentiating between less and more numerate participants (half a standard deviation below and above the mean respectively). Unincentivized, more numerate individuals exhibit the congeniality bias of 9 to 13 pp (Table A10), while incentivized, more numerate individuals—of 7 to 17pp (Table A12). The subsequent sections employ statistical tests to infer whether these descriptive differences between unincentivized and incentivized more numerate respondents' levels of congeniality bias are statistically distinct from each other.

Table A12: Differences in accuracy by levels of congeniality and numeracy (incentivized participants)

	Strongly uncongenial data				
	Congeniality > 25pct	Congeniality < 25pct	Difference	t-statistic	p-value
Accuracy rate for less numerate	0.39	0.42	-0.03	-0.57	0.57
Accuracy rate for more numerate	0.51	0.38	0.13	2.82	0.00
	Rather congenial data				
	Congeniality < 50pct	Congeniality > 50pct	Difference	t-statistic	p-value
Accuracy rate for less numerate	0.40	0.39	0.01	0.32	0.75
Accuracy rate for more numerate	0.44	0.51	-0.07	-1.86	0.06
	Strongly congenial data				
	Congeniality < 75pct	Congeniality > 75pct	Difference	t-statistic	p-value
Accuracy rate for less numerate	0.38	0.43	-0.05	-1.02	0.31
Accuracy rate for more numerate	0.43	0.60	-0.17	-3.87	0.00

Inferiority tests (based on Lakens et al. 2018) for the multivariate analysis presented in the main paper

Section “Multivariate analysis” of the main paper refers to inferiority tests conducted to determine whether the obtained effects are substantively meaningful.

The inferiority test for hypothesis 1: Is the effect of congeniality on accuracy among unincentivized respondents substantively important?

The inferiority test (Lakens et al., 2018) helps determine if this effect is substantively meaningful. Since hypotheses 1 and 2 are conceptual replications of Kahan et al. (2017), we test whether the observed effect size is smaller than the smallest effect size that the original study could have detected. Using simulated data of that study, we find that the minimum effect size of β_1 in equation (1) that could have been detected with 82% power using their sample of 1111 is 0.04 in a linear probability model, i.e., the inferiority bound for the test is $\Delta = 0.04$ for β_1 in equation (1). The obtained $\beta_1 - hat$ equals 0.048; we thus cannot reject the null that the estimated effect of $Congenial_i$ on $Correct_i$ is statistically smaller than the smallest effect size of interest (p = 0.62). Thus, we cannot conclude that a meaningful effect is absent.

The inferiority test for hypothesis 2: Is the effect of congeniality and numeracy on accuracy among unincentivized respondents substantively important?

While the coefficients on the interaction term of congenial and numeracy, and congenial and numeracy squared is not statistically discernible at 0.05 level, is it substantively meaningful? We find that the minimum effect size of β_4 and β_5 in equation (2) that could have been detected with 86.7% and 80.3% power using Kahan et al.’s (2017) sample size of 1111 is 0.018 and -0.0056 respectively in a linear probability model. The coefficient point estimates we obtain of 0.017 and -0.0035 are not statistically smaller than the effect size of interest of 0.018 and -0.0056 (p=0.526 and p =0.704 in a one-sided t-test). Thus, we cannot conclude that a meaningful effect is absent.

Logistic regressions

In the main paper, a linear probability model estimates the parameters of equations (1)–(4). This section of the appendix also includes the logistic regression models that estimate these equations. Equations (1) and (2) are estimated on unincentivized observations to test hypotheses 1 and 2; equations (3) and (4) test hypotheses 3 and 4, utilizing the full sample. All results from logistic regressions are consistent with the linear probability models shown in the main paper.

Table A13: The impact of numeracy and congeniality on accuracy
(unincentivized participants)

	(1)	(2)	(3)	(4)
Congenial	0.197***	0.186***	0.234**	0.219**
	(0.065)	(0.066)	(0.091)	(0.093)
Numeracy	-0.017	-0.023	-0.021	-0.027
	(0.040)	(0.043)	(0.040)	(0.043)
Numeracy^2	0.013	0.020	0.015	0.021
	(0.022)	(0.023)	(0.022)	(0.023)
Numeracy × Congenial			0.073*	0.062
			(0.041)	(0.042)
Numeracy^2 × Congenial			-0.015	-0.013
			(0.023)	(0.024)
Constant	-0.351***	-0.418	-0.354***	-0.416
	(0.088)	(0.423)	(0.088)	(0.421)
Observations	1016	1016	1016	1016
Controls	No	Yes	No	Yes

Standard errors in parentheses

Note:Logit regression with heterscedasticity robust standard errors.

Control variables in the regression are age, gender, race, education, and voting behavior in 2016.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A14: The impact of incentives, numeracy, and congeniality on accuracy
(all participants)

	(5)	(6)	(7)	(8)	(9)	(10)
Incentive	-0.006	-0.007	-0.005	-0.004	-0.014	0.007
	(0.078)	(0.078)	(0.078)	(0.079)	(0.108)	(0.108)
Numeracy			-0.022	-0.030	-0.021	-0.032
			(0.039)	(0.040)	(0.040)	(0.041)
Congenial			0.190**	0.184**	0.234**	0.225**
			(0.074)	(0.074)	(0.091)	(0.091)
Numeracy × Congenial			0.064	0.058	0.073*	0.066
			(0.040)	(0.040)	(0.041)	(0.041)
Numeracy^2			0.017	0.017	0.015	0.020
			(0.013)	(0.013)	(0.022)	(0.022)
Numeracy^2 × Congenial			0.001	0.001	-0.015	-0.013
			(0.013)	(0.013)	(0.023)	(0.023)
Incentive × Congenial			-0.111	-0.107	-0.176	-0.167
			(0.079)	(0.080)	(0.111)	(0.111)
Incentive × Numeracy			0.088*	0.085*	0.087*	0.088*
			(0.048)	(0.048)	(0.049)	(0.049)
Incentive × Numeracy × Congenial			0.003	0.011	-0.009	-0.000
			(0.048)	(0.049)	(0.050)	(0.050)
Incentive × Numeracy^2					0.003	-0.004
					(0.027)	(0.027)
Incentive × Numeracy^2 × Congenial					0.023	0.022
					(0.028)	(0.028)
Constant	-0.310***	-0.419	-0.359***	-0.418	-0.354***	-0.425
	(0.064)	(0.259)	(0.073)	(0.263)	(0.088)	(0.267)
Observations	3050	3050	3050	3050	3050	3050
Controls	No	Yes	No	Yes	No	Yes

Standard errors in parentheses

Note:Logit regression with heterscedasticity robust standard errors.

Control variables in the regression are age, gender, race, education, and voting behavior in 2016.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Order effects

Controlling for order effects in the pooled sample

In Tables A15-A18, *Order_numeracy_treatment* is 1 if numeracy is measured before the treatment and is 0 otherwise. We find that the measure of numeracy is not affected by the order of randomization of numeracy and treatment condition. However, the rate of correct answers in the contingency table data interpretation task increases by 6.8pp if numeracy is measured before the treatment condition as observed in the tables A5 and A6.

Table A15: Dependent variable is Numeracy (number of correctly solved questions out of 6)

	Entire sample	Non-incentivized	Incentivized
Order_numeracy_treatment	-0.041	-0.048	-0.038
	(0.060)	(0.104)	(0.073)
Constant	2.719***	2.697***	2.730***
	(0.042)	(0.073)	(0.052)
Observations	3050	1016	2034

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A16: Dependent variable is Correct (whether the participant answered correctly to the contingency table)

	Entire Sample	Non-incentivized	Incentivized
Order_numeracy_treatment	0.068***	0.073**	0.065***
	(0.018)	(0.031)	(0.022)
Constant	0.389***	0.387***	0.390***
	(0.012)	(0.021)	(0.015)
Observations	3050	1016	2034

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

We control for the order effect in our main regressions and find that the coefficients of interest do not change when we control for these effects. Thus, even though we find that the order affected the percentage of correct answers, they did not affect our regression results.

Table A17: Testing for hypothesis 1 and 2 controlling for order effects

	(1)	(2)	(3)	(4)
Treatment order	0.074**	0.070**	0.078**	0.074**
	(0.031)	(0.031)	(0.031)	(0.031)
Congenial	0.048***	0.045***	0.058***	0.053**
	(0.015)	(0.016)	(0.021)	(0.022)
Numeracy	-0.004	-0.006	-0.005	-0.006
	(0.010)	(0.010)	(0.010)	(0.010)
Numeracy^2	0.003	0.005	0.004	0.005
	(0.005)	(0.005)	(0.005)	(0.005)
Numeracy $\times$ Congenial			0.019*	0.016
			(0.010)	(0.010)
Numeracy^2 $\times$ Congenial			-0.004	-0.004
			(0.005)	(0.006)
Constant	0.377***	0.363***	0.375***	0.362***
	(0.026)	(0.104)	(0.026)	(0.104)
Observations	1016	1016	1016	1016
Controls	No	Yes	No	Yes

Standard errors in parentheses

Note:Linear Probability Model with heterscedasticity robust standard errors.

Control variables in the regression are age, gender, race, education, and voting behavior in 2016.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A18: Testing for hypothesis 3 and 4 controlling for order effects

	(5)	(6)	(7)	(8)	(9)	(10)
Treatment order	0.068***	0.068***	0.070***	0.069***	0.070***	0.069***
	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)
Incentive	-0.001	-0.002	-0.002	-0.002	-0.004	0.000
	(0.019)	(0.019)	(0.019)	(0.019)	(0.026)	(0.026)
Numeracy			-0.005	-0.007	-0.005	-0.007
			(0.009)	(0.010)	(0.010)	(0.010)
Congenial			0.047***	0.046***	0.058***	0.055***
			(0.018)	(0.018)	(0.021)	(0.021)
Numeracy × Congenial			0.016*	0.015	0.019*	0.017*
			(0.009)	(0.009)	(0.010)	(0.010)
Numeracy^2			0.004	0.004	0.004	0.005
			(0.003)	(0.003)	(0.005)	(0.005)
Numeracy^2 × Congenial			-0.000	-0.000	-0.004	-0.004
			(0.003)	(0.003)	(0.005)	(0.005)
Incentive × Congenial			-0.027	-0.025	-0.042	-0.040
			(0.019)	(0.019)	(0.026)	(0.026)
Incentive × Numeracy			0.021*	0.020*	0.021*	0.021*
			(0.011)	(0.011)	(0.012)	(0.012)
Incentive × Numeracy × Congenial			0.001	0.003	-0.003	-0.001
			(0.011)	(0.011)	(0.012)	(0.012)
Incentive × Numeracy^2					0.001	-0.001
					(0.007)	(0.007)
Incentive × Numeracy^2 × Congenial					0.006	0.005
					(0.007)	(0.007)
Constant	0.390***	0.365***	0.377***	0.366***	0.379***	0.364***
	(0.018)	(0.063)	(0.019)	(0.064)	(0.023)	(0.065)
Observations	3050	3050	3050	3050	3050	3050
Adjusted R^2	0.004	0.007	0.013	0.014	0.012	0.013
Controls	No	Yes	No	Yes	No	Yes

Standard errors in parentheses

Note:Linear Probability Model with heterscedasticity robust standard errors.

Control variables in the regression are age, gender, race, education, and voting behavior in 2016.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Split sample analyses

Preview of results from split sample analyses

Our study tests the following hypotheses:

> Hypothesis 1: Among unincentivized respondents, the rate of correct data interpretation increases as the data become more congenial to one's ideological beliefs.
>
> Hypothesis 2: Among unincentivized respondents, the congeniality bias increases with one's numeracy.
>
> Hypothesis 3: Relative to unincentivized respondents, those incentivized will exhibit greater accuracy in all conditions.
>
> Hypothesis 4: The congeniality bias among incentivized respondents increases at a *lower* rate with one's numeracy, compared to the rate of bias increase among unincentivized respondents.

The main paper tests these expectations against the pooled sample. The results from the pooled sample indicate that we reject the first null of no effect (i.e., each one-unit increase in $Congenial_i$ increases accuracy by 4–5pp among unincentivized respondents) and fail to reject the null hypotheses 2, 3, and 4.

The subsequent sections of the appendix retest these four hypotheses against the subsample "numeracy after treatment" (i.e., the order of the survey in which numeracy questions appeared followed the contingency table-based data interpretation task) and "numeracy before treatment" (i.e., the order of the survey in which numeracy questions preceded the contingency table). To preview the results of the split sample analyses:

- Sample "numeracy after treatment" – all conclusions from the pooled dataset stand:
    - We reject the null hypothesis 1. The substantive impact of $Congenial_i$ on accuracy increases relative to the pooled sample: each one-unit increase in $Congenial_i$ raises accuracy by 7–8pp among unincentivized respondents in the "numeracy after treatment" subsample.
    - We fail to reject the null hypotheses 2, 3, and 4.

- Sample "numeracy before treatment" – the conclusion regarding hypothesis 1 is not retained in this subsample, but all other conclusions are:

- We fail to reject the null hypothesis 1 in this subsample. The substantive impact of *Congenial*$_i$ on accuracy (although in the expected positive direction) does not rise to the level of statistical significance at 0.05 level. This result implies that our rejection of the first null hypothesis in the pooled sample is driven by the subsample that received numeracy questions after the contingency table.
- We fail to reject the null hypotheses 2 and 3 of no effect.
- Incentives increase the congeniality bias among more numerate individuals in the subsample "numeracy before treatment" which (although implies that there is an effect of congeniality in interaction with numeracy on the congeniality bias), the observed changes are the opposite of what hypothesis 4 expected.

Order: Treatment first, then numeracy (numeracy after treatment)

Multivariate analysis for hypotheses 1 and 2 (numeracy after treatment)

Table A19: The impact of numeracy and congeniality on accuracy (unincentivized participants, subsample "numeracy after treatment")

	(1)	(2)	(3)	(4)
Congenial	0.080***	0.074***	0.066**	0.051*
	(0.021)	(0.021)	(0.029)	(0.030)
Numeracy	-0.019	-0.015	-0.019	-0.015
	(0.013)	(0.014)	(0.013)	(0.014)
Numeracy^2	0.008	0.010	0.008	0.010
	(0.007)	(0.007)	(0.007)	(0.007)
Numeracy × Congenial			0.011	0.009
			(0.013)	(0.013)
Numeracy^2 × Congenial			0.005	0.008
			(0.007)	(0.007)
Constant	0.363***	0.489***	0.362***	0.484***
	(0.029)	(0.145)	(0.029)	(0.143)
Observations	519	519	519	519
Controls	No	Yes	No	Yes

Standard errors in parentheses

Note:Linear Probability Model with heterscedasticity robust standard errors.

Control variables in the regression are age, gender, race, education, and voting behavior in 2016.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Hypothesis 1 states that, among unincentivized respondents, accuracy is higher when interpreting more congenial data. Models 1–2 of Table A19 estimate equation (1) of the main paper. The coefficient estimate on $Congenial_i$ implies that one unit increase in $Congenial_i$ (ranges from -1.88 to 1.88, mean=0, standard deviation=1) generates a 7–8 percentage point (pp) increase in accuracy (statistically discernible from 0). We thus reject the null hypothesis 1 of congeniality having no impact on accuracy in the subsample "numeracy after treatment."

To test hypothesis 2 (among unincentivized respondents, the congeniality bias increases with one's numeracy), we use Figures A1 and A2 and Table A21 to estimate if $Numeracy_i$ has a range of values for which $Congenial_i$ has a statistically meaningful impact among unincentivized respondents.

Multivariate analysis for hypotheses 3 and 4 (numeracy after treatment)

Table A20: The impact of incentives, numeracy, and congeniality on accuracy (all participants, subsample “numeracy after treatment”)

	(5)	(6)	(7)	(8)	(9)	(10)
Incentive	0.003	0.004	0.004	0.006	0.013	0.021
	(0.026)	(0.026)	(0.026)	(0.026)	(0.036)	(0.036)
Numeracy			-0.018	-0.019	-0.019	-0.020
			(0.013)	(0.013)	(0.013)	(0.014)
Congenial			0.071***	0.068***	0.066**	0.061**
			(0.024)	(0.024)	(0.029)	(0.029)
Numeracy × Congenial			0.012	0.011	0.011	0.009
			(0.013)	(0.013)	(0.013)	(0.013)
Numeracy^2			0.006	0.006	0.008	0.010
			(0.004)	(0.004)	(0.007)	(0.007)
Numeracy^2 × Congenial			0.003	0.004	0.005	0.006
			(0.004)	(0.004)	(0.007)	(0.007)
Incentive × Congenial			-0.055**	-0.055**	-0.048	-0.045
			(0.026)	(0.026)	(0.036)	(0.036)
Incentive × Numeracy			0.029*	0.026*	0.030*	0.029*
			(0.016)	(0.016)	(0.016)	(0.016)
Incentive × Numeracy × Congenial			-0.004	-0.002	-0.002	-0.000
			(0.015)	(0.015)	(0.016)	(0.016)
Incentive × Numeracy^2					-0.003	-0.006
					(0.009)	(0.009)
Incentive × Numeracy^2 × Congenial					-0.003	-0.004
					(0.009)	(0.009)
Constant	0.387***	0.340***	0.369***	0.332***	0.362***	0.321***
	(0.021)	(0.086)	(0.024)	(0.087)	(0.029)	(0.088)
Observations	1558	1558	1558	1558	1558	1558
Controls	No	Yes	No	Yes	No	Yes

Standard errors in parentheses

Note:Linear Probability Model with heterscedasticity robust standard errors.

Control variables in the regression are age, gender, race, education, and voting behavior in 2016.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Hypothesis 3 expects that incentivized participants are more likely to correctly interpret the contingency table compared to those unincentivized. Models 5 and 6 of Table A20 test this expectation, estimating parameters in equation (3), using subsample “numeracy after treatment”. The coefficient estimates on $Incentive_i$ in models 3 and 4 are substantively and statistically negligible; we thus fail to reject the null hypothesis of monetary incentives having no impact on accuracy.

Visualization of interaction effects (numeracy after treatment)

Figure A1 demonstrates that the congeniality bias increases with numeracy. The predicted accuracy rates for respondents facing uncongenial, neutral, and congenial data constitute 1) 36%, 42%, and 48%, i.e., the congeniality bias equals 12pp for the less numerate unincentivized respondents (top left), 2) 26%, 35%, and 45%, i.e., the congeniality bias is 19pp for the more numerate unincentivized (top right), 3) 36%, 37%, and 38%, i.e., the bias equals 2pp for the less numerate incentivized (bottom left), and 4) 37%, 41%, and 45%, i.e., the bias constitutes 8pp for the more numerate incentivized (bottom right).

Figure A1: Predicted probabilities of correctly interpreting the data (subsample numeracy after treatment)

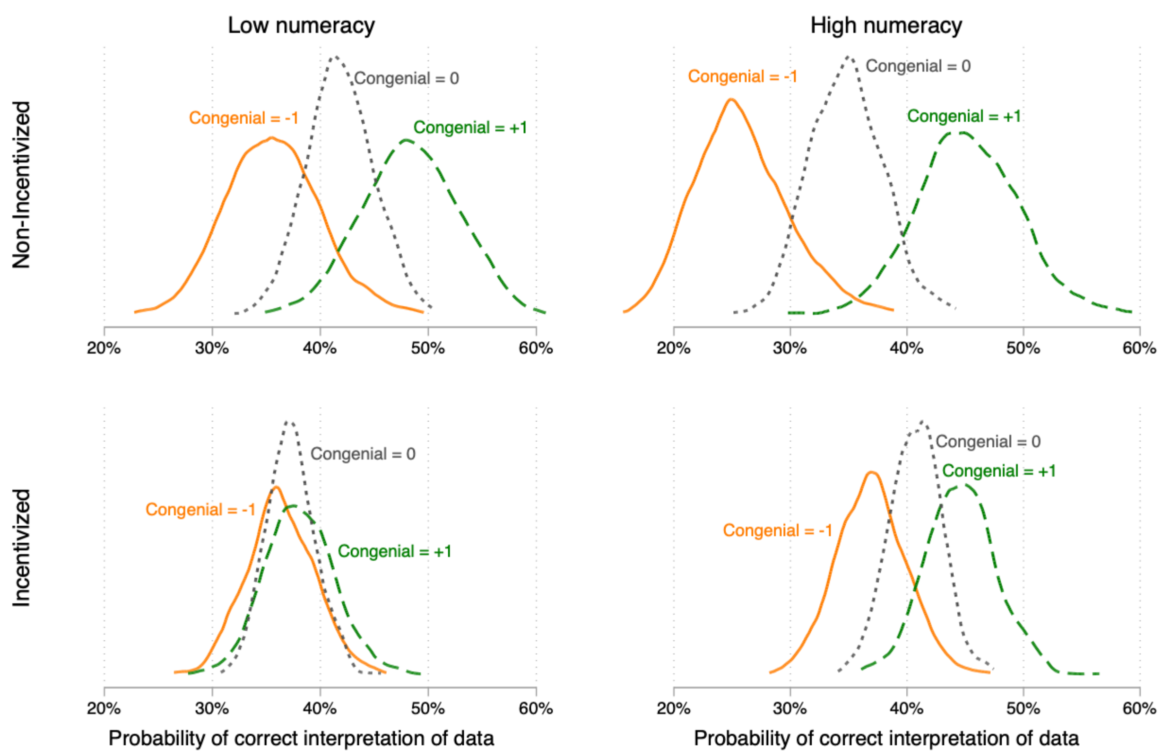


Figure A2 demonstrates that the congeniality bias increases with numeracy. The predicted accuracy rates for respondents facing uncongenial and congenial data at -2SD and +2SD constitute 1) 30%,and 55%, i.e., the congeniality bias equals 25pp for the less numerate unincentivized respondents (top left), 2) 18% and 56%, i.e., the congeniality bias is 38pp for the more numerate unincentivized (top right), 3) 35% and 39%, i.e., the bias equals 4pp for the less numerate incentivized (bottom left), and 4) 34% and 49%, i.e., the bias constitutes 15pp for the more numerate incentivized (bottom right).

Figure A2: Predicted probabilities of correctly interpreting the data (+/-2SD above/below the mean, subsample numeracy after treatment)

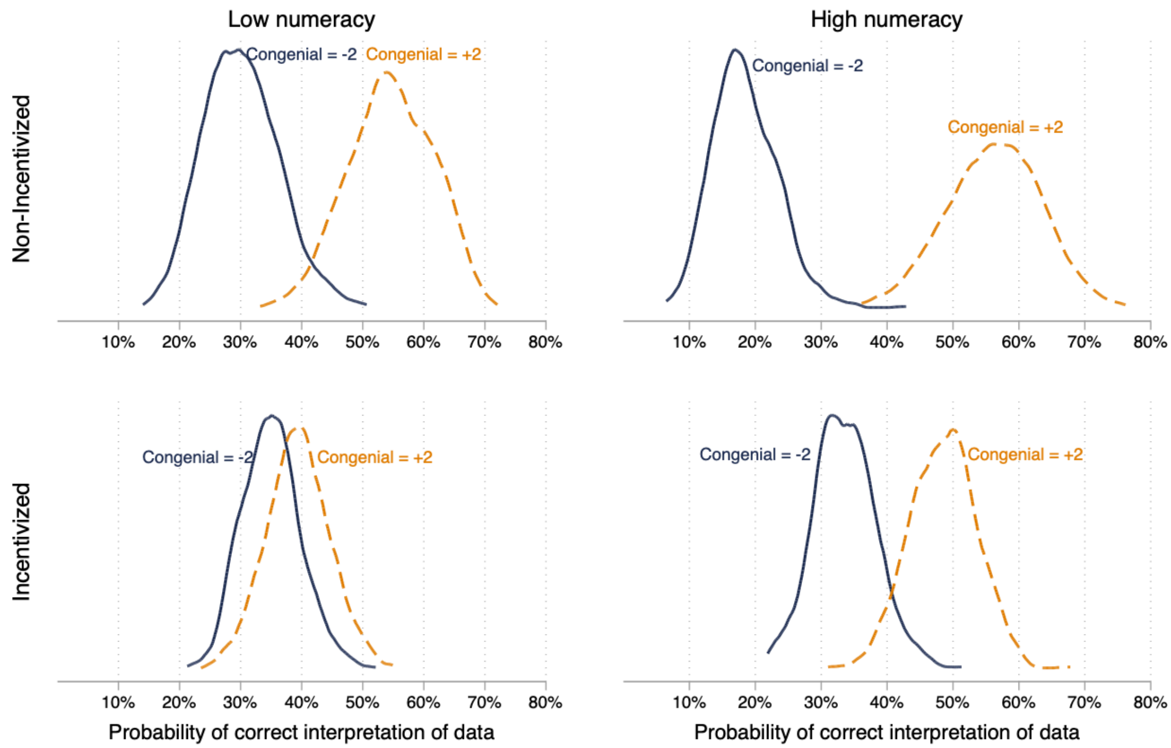


Table A21: Differences in the predicted congeniality bias between less and more numerate individuals at various levels of numeracy (subsample numeracy after treatment)

	Congeniality Bias Low Num	Congeniality Bias High Num	Difference	t-statistic	p-value
Low vs high numeracy (+/-1 SD), no inc	-.12	-.20	.08	0.84	0.403
Low vs high numeracy (+/-1 SD), inc	-.02	-.08	.06	0.91	0.361
Low vs high numeracy (+/-1.5 SD), no inc	-.14	-.25	.11	0.80	0.425
Low vs high numeracy (+/-1.5 SD), inc	-.02	-.11	.09	0.91	0.365
Low vs high numeracy (+/- 2 SD), no inc	-.16	-.31	.15	0.66	0.508
Low vs high numeracy (+/- 2 SD), inc	-.03	-.14	.11	0.73	0.466

Note: Congeniality Bias = Pr(correct=1|congenial=-1SD) - Pr(correct=1|congenial=1SD); Numeracy is set at either +/-1SD or +/-1.5 SD or +/-2 SD

Table A21 uses differences-in-means to test whether the differences in the congeniality bias between less and more numerate individuals are statistically distinct in the subsample "numeracy after treatment." To test hypothesis 2, consider respondents in the no incentive condition first. Highly numerate respondents exhibit greater congeniality bias than less numerate subjects by 8pp, 11pp, and 11 pp for respondents with numeracy of -/+1SD, -/+1.5SD, and +/-2SD above/below the mean, however these difference are not statistically distinct at 0.05 level. We thus fail to reject the second null hypothesis of no effect of numeracy on congeniality bias among unincentivized respondents in the subsample "numeracy after treatment."

Incentives shrink the gap in motivated numeracy from 12pp to 2pp (10pp reduction) among less numerate (+/-1SD) and from 20pp to 8pp (12pp reduction) among more numerate individuals. Similar reductions in the congeniality bias are observed at other levels of numeracy. Hypothesis 4 expected that incentives would generate a larger reductions in the bias among more numerate individuals, however the gains we observe are not statistically distinct between less and more numerate individuals; we fail to reject the fourth null hypothesis of no effect in the subsample "numeracy after treatment."

Order: Numeracy first, then treatment (numeracy before treatment)

Multivariate analysis for hypotheses 1 and 2 (numeracy before treatment)

Table A22: The impact of numeracy and congeniality on accuracy
(unincentivized participants, subsample 'numeracy before treatment')

	(1)	(2)	(3)	(4)
Congenial	0.017	0.020	0.048	0.053*
	(0.023)	(0.023)	(0.031)	(0.031)
Numeracy	0.010	0.004	0.007	0.000
	(0.014)	(0.015)	(0.014)	(0.015)
Numeracy^2	-0.001	-0.001	0.001	0.001
	(0.008)	(0.008)	(0.008)	(0.008)
Numeracy × Congenial			0.025*	0.020
			(0.014)	(0.015)
Numeracy^2 × Congenial			-0.012	-0.012
			(0.008)	(0.008)
Constant	0.463***	0.326**	0.462***	0.324**
	(0.031)	(0.147)	(0.031)	(0.149)
Observations	497	497	497	497
Controls	No	Yes	No	Yes

Standard errors in parentheses

Note:Linear Probability Model with heterscedasticity robust standard errors.

Control variables in the regression are age, gender, race, education, and voting behavior in 2016.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Hypothesis 1 states that, among unincentivized respondents, accuracy is higher when interpreting more congenial data. Models 1–2 of Table A22 estimate equation (1) of the main paper. The coefficient estimate on $Congenial_i$ implies that one unit increase in $Congenial_i$ (ranges from -1.88 to 1.88, mean=0, standard deviation=1) generates a 2 percentage point (pp) increase in accuracy, however this effect is not statistically distinct from 0. We thus fail to reject the null hypothesis 1 of congeniality having no impact on accuracy in the subsample "numeracy before treatment."

To test hypothesis 2 in the subsample "numeracy before treatment" (among unincentivized respondents, the congeniality bias increases with one's numeracy), we use Figures A3 and A4 and Table A24 to estimate if $Numeracy_i$ has a range of values for which $Congenial_i$ has a statistically meaningful impact among unincentivized respondents.

Multivariate analysis for hypotheses 3 and 4 (numeracy before treatment)

Table A23: The impact of incentives, numeracy, and congeniality on accuracy
(all participants)

	(5)	(6)	(7)	(8)	(9)	(10)
Incentive	-0.005	-0.008	-0.008	-0.009	-0.017	-0.017
	(0.027)	(0.028)	(0.027)	(0.028)	(0.038)	(0.038)
Numeracy			0.007	0.006	0.007	0.005
			(0.014)	(0.014)	(0.014)	(0.014)
Congenial			0.025	0.025	0.048	0.048
			(0.026)	(0.026)	(0.031)	(0.031)
Numeracy × Congenial			0.021	0.019	0.025*	0.023
			(0.014)	(0.014)	(0.014)	(0.014)
Numeracy^2			0.003	0.003	0.001	0.002
			(0.005)	(0.005)	(0.008)	(0.008)
Numeracy^2 × Congenial			-0.004	-0.004	-0.012	-0.012
			(0.005)	(0.005)	(0.008)	(0.008)
Incentive × Congenial			0.002	0.003	-0.034	-0.031
			(0.028)	(0.028)	(0.038)	(0.039)
Incentive × Numeracy			0.015	0.014	0.015	0.014
			(0.017)	(0.017)	(0.017)	(0.017)
Incentive × Numeracy × Congenial			0.006	0.008	-0.001	0.001
			(0.017)	(0.017)	(0.017)	(0.018)
Incentive × Numeracy^2					0.003	0.002
					(0.010)	(0.010)
Incentive × Numeracy^2 × Congenial					0.013	0.012
					(0.010)	(0.010)
Constant	0.461***	0.460***	0.456***	0.476***	0.462***	0.479***
	(0.022)	(0.093)	(0.025)	(0.095)	(0.031)	(0.097)
Observations	1492	1492	1492	1492	1492	1492
Controls	No	Yes	No	Yes	No	Yes

Standard errors in parentheses

Note:Linear Probability Model with heterscedasticity robust standard errors.

Control variables in the regression are age, gender, race, education, and voting behavior in 2016.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Hypothesis 3 expects that incentivized participants are more likely to correctly interpret the contingency table compared to those unincentivized. Models 5 and 6 of Table A23 test this expectation, estimating parameters in equation (3) of the main paper, using subsample "numeracy before treatment". The coefficient estimates on *Incentive*$_i$ in models 3 and 4 are substantively and statistically negligible; we thus fail to reject the null hypothesis of monetary incentives having no impact on accuracy in the subsample "numeracy before treatment."

Visualization of interaction effects (numeracy before treatment)

Figure A3 demonstrates that the congeniality bias increases with numeracy. The predicted accuracy rates for respondents facing uncongenial, neutral, and congenial data constitute 1) 48%, 46%, and 43%, i.e., the congeniality bias equals -5pp for the less numerate unincentivized respondents (top left), 2) 42%, 48%, and 53%, i.e., the congeniality bias is 11pp for the more numerate unincentivized (top right), 3) 45%,42%, and 40%, i.e., the bias equals -5pp for the less numerate incentivized (bottom left), and 4) 43%, 49%, and 55%, i.e., the bias constitutes 12pp for the more numerate incentivized (bottom right).

Figure A3: Predicted probabilities of correctly interpreting the data (subsample numeracy before treatment)

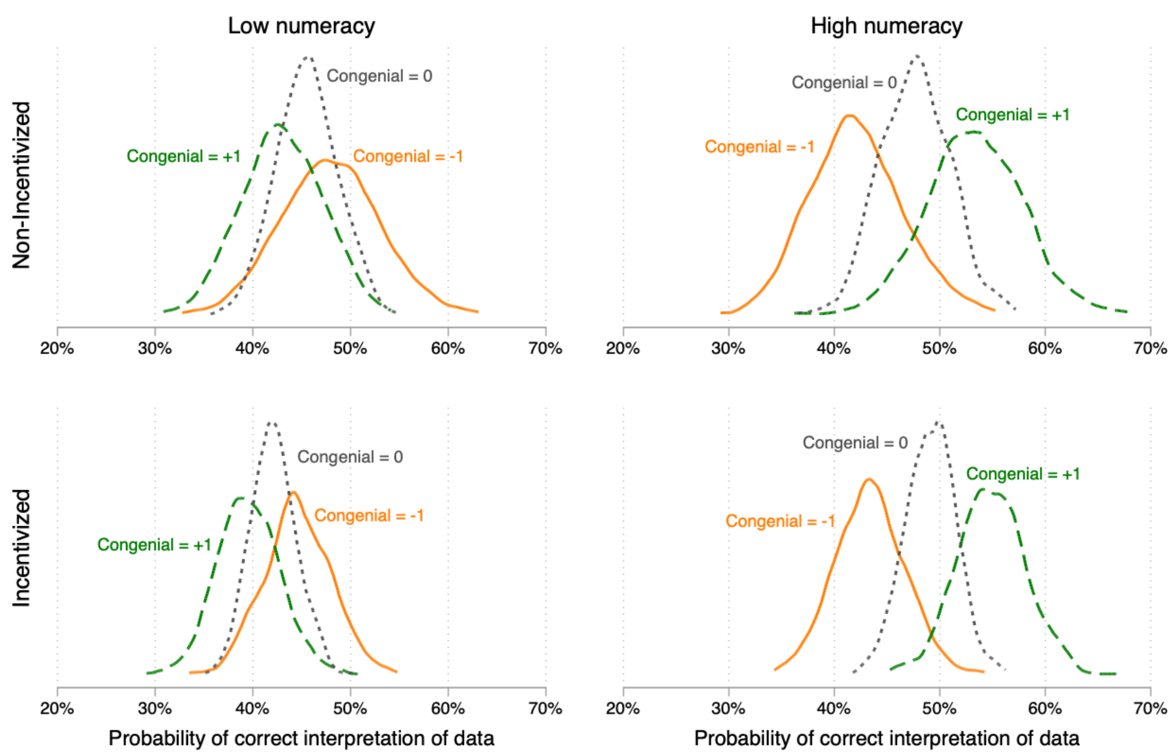


Figure A4 demonstrates that the congeniality bias increases with numeracy. The predicted accuracy rates for respondents facing uncongenial and congenial data at -2SD and +2SD constitute 1) 50%,and 41%, i.e., the congeniality bias equals -9pp for the less numerate unincentivized respondents (top left), 2) 36% and 59%, i.e., the congeniality bias is 23pp for the more numerate unincentivized (top right), 3) 47% and 38%, i.e., the bias equals -9pp for the less numerate incentivized (bottom left), and 4) 38% and 60%, i.e., the bias constitutes 22pp for the more numerate incentivized (bottom right).

Figure A4: Predicted probabilities of correctly interpreting the data (+/-2SD above/below the mean, subsample numeracy before treatment)

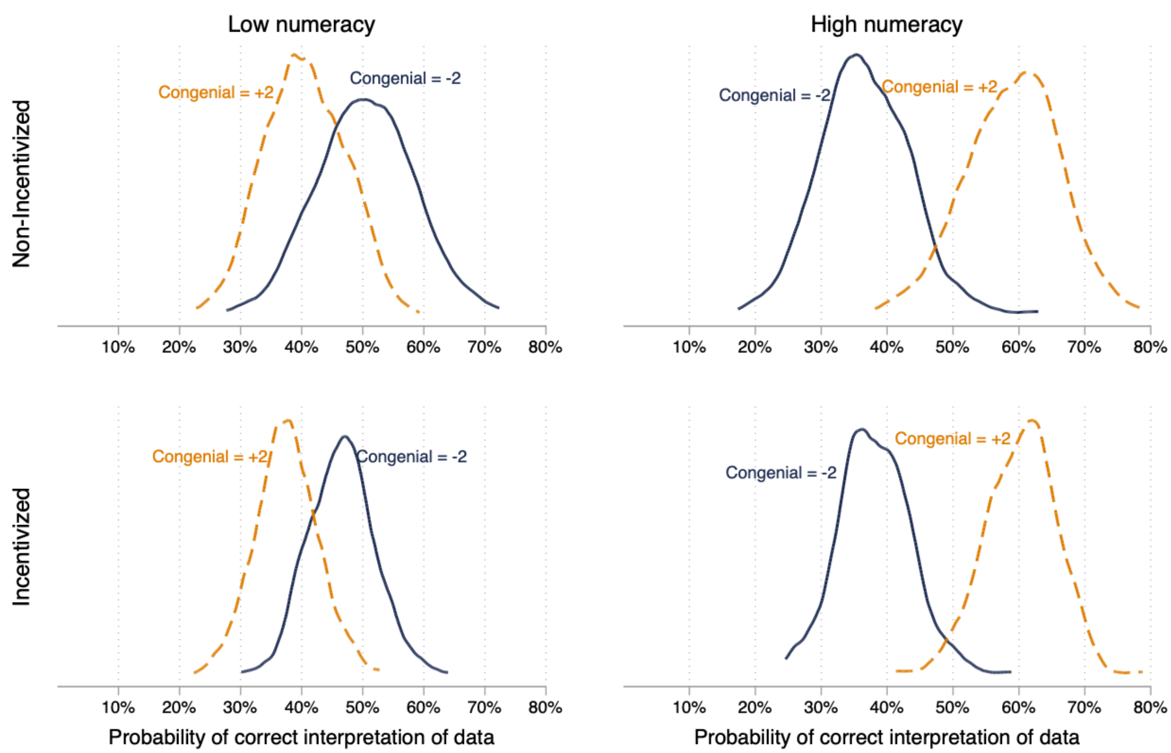


Table A24: Differences in the predicted congeniality bias between less and more numerate individuals at various levels of numeracy (numeracy before treatment)

	Congeniality Bias Low Num	Congeniality Bias High Num	Difference	t-statistic	p-value
Low vs high numeracy (+/-1 SD), no inc	.05	-.12	.17	1.80	0.072
Low vs high numeracy (+/-1 SD), inc	.05	-.11	.16	2.44	0.015
Low vs high numeracy (+/-1.5 SD), no inc	.18	-.07	.25	1.78	0.075
Low vs high numeracy (+/-1.5 SD), inc	.08	-.16	.24	2.35	0.019
Low vs high numeracy (+/- 2 SD), no inc	.31	.00	.31	1.44	0.150
Low vs high numeracy (+/- 2 SD), inc	.11	-.20	.31	1.85	0.064

Note: Congeniality Bias = Pr(correct=1|congenial=-1SD) - Pr(correct=1|congenial=1SD); Numeracy is set at either +/-1SD or +/-1.5 SD or +/-2 SD

Table A24 uses differences-in-means to test whether the differences in the congeniality bias between less and more numerate individuals are statistically distinct in the subsample "numeracy before treatment." To test hypothesis 2, consider respondents in the no incentive condition first. Highly numerate respondents exhibit greater congeniality bias than less numerate subjects by 17pp, 25pp, and 31 pp for respondents with numeracy of -/+1SD, -/+1.5SD, and +/-2SD above/below the mean, however these difference are not statistically distinct at 0.05 level. We thus fail to reject the second null hypothesis of no effect of numeracy on congeniality bias among unincentivized respondents in the subsample "numeracy after treatment."

Incentives shrink the gap in motivated numeracy from by 0pp, 10pp and 20pp for less numerate respondents at various levels of numeracy. Incentives shrink the bias by 1pp and *increase* it by 9pp and 20pp for more numerate respondents. While this implies that numeracy interacts with congeniality to affect more numerate respondents differently than less numerate, this effect is the opposite of what we expected in hypothesis 4, which posited that incentives would generate greater reductions in the congeniality bias among more numerate individuals, however the changes we observe are the opposite – more numerate individuals exhibit greater congeniality bias with numeracy.

Deviations from pre-registration

The main paper includes three deviations from the pre-registered report as described below. Section "Analysis with no deviations from the pre-registration" of the appendix includes the analyses that strictly follow pre-registration with no deviations. All substantive conclusions remain the same as those presented in the main paper.

Updated test of hypothesis 1

The registered report in-principle accepted at *JEPS* in January 2021 (same as the OSF registration) included the following equation to test hypothesis 1:

$$\begin{aligned}\text{Correct}_i = \beta_0 &+ \beta_1 \text{Numeracy}_i \\ &+ \beta_2 \text{Congenial}_i \\ &+ \beta_3 \text{Numeracy}_i \times \text{Congenial}_i \\ &+ \beta_4 \text{Numeracy}_i^2 + \epsilon_i\end{aligned} \tag{1}$$

Hypothesis 1 states that among unincentivized respondents, the rate of correct data interpretation increases as the data become more congenial to one's ideological beliefs. We are grateful to JEPS reviewers who noted during the final review that former equation 1 only estimates the effect of $Congenial_i$ in interaction with $Numeracy_i$, however the hypothesis does not imply a multiplicative interaction. Therefore, the main paper includes two equations that differ from the OSF registration. The updated equations include:

$$\mathbf{Correct}_i = \beta_0 + \beta_1 \mathbf{Congenial}_i + \beta_2 \mathbf{Numeracy}_i + \beta_3 \mathbf{Numeracy}_i^2 + \epsilon_i \tag{1}$$

$$\begin{aligned}\mathbf{Correct}_i = \beta_0 &+ \beta_1 \mathbf{Congenial}_i \\ &+ \beta_2 \mathbf{Numeracy}_i \\ &+ \beta_3 \mathbf{Numeracy}_i \times \mathbf{Congenial}_i \\ &+ \beta_4 \mathbf{Numeracy}_i^2 \times \mathbf{Congenial}_i \\ &+ \beta_5 \mathbf{Numeracy}_i^2 + \epsilon_i\end{aligned} \tag{2}$$

The updated equation 1 allows us to estimate the impact of $Congenial_i$ not interacted with $Numeracy_i$. Substantively, the changes to interpretation are negligible.

Updated test of hypothesis 2

Hypothesis 2 posits that the congeniality bias increases with one's numeracy among unincentivized respondents; the updated equation 2 tests this hypothesis. The original equation 1 (whose updated version is labeled as equation 2 in the main paper) was intended to replicate Kahan et al.'s (2017) results and so our regression model was inspired by the one Kahan et al. (2017) used in which they include a numeracy-sq term in their regression model but do not interact it with other variables. Since we did not conduct any preliminary data analyses we did not know that numeracy will not have a quadratic effect on accuracy in our sample.

To test the second hypothesis, the updated equation 2 now includes the interaction between $Congenial_i$ and $Numeracy_i^2$, which was missing from the pre-registered first equation. The main paper and the additional analyses in the appendix utilize visualizations to interpret the effect of the interaction between $Congenial_i$ and $Numeracy_i$. Substantively, the changes to interpretation are negligible.

Updated test of hypothesis 4

Hypothesis 4 posits that the congeniality bias among incentivized respondents increases at a *lower* rate with one's numeracy, compared to the rate of bias increase among unincentivized respondents. The pre-registered equation designed to test this hypothesis was:

$$\begin{aligned}\mathrm{Correct}_i = \beta_0 &+ \beta_1 \mathrm{Numeracy}_i \\ &+ \beta_2 \mathrm{Congenial}_i \\ &+ \beta_3 \mathrm{Numeracy}_i \times \mathrm{Congenial}_i \\ &+ \beta_4 \mathrm{Numeracy}_i^2 \\ &+ \beta_5 \mathrm{Incentive}_i \\ &+ \beta_6 \mathrm{Incentive}_i \times \mathrm{Congenial}_i \\ &+ \beta_7 \mathrm{Incentive}_i \times \mathrm{Numeracy}_i \\ &+ \beta_8 \mathrm{Incentive}_i \times \mathrm{Numeracy}_i \times \mathrm{Congenial}_{ii} + \epsilon_i\end{aligned} \tag{3}$$

The updated equation included in the main paper is:

$$\begin{aligned}\mathrm{Correct}_i = \beta_0 &+ \beta_1 \mathrm{Numeracy}_i \\ &+ \beta_2 \mathrm{Congenial}_i \\ &+ \beta_3 \mathrm{Numeracy}_i \times \mathrm{Congenial}_i \\ &+ \beta_4 \mathrm{Numeracy}_i^2 \\ &+ \beta_5 \mathrm{Numeracy}_i^2 \times \mathrm{Congenial}_i \\ &+ \beta_6 \mathrm{Incentive}_i \\ &+ \beta_7 \mathrm{Incentive}_i \times \mathrm{Congenial}_i \\ &+ \beta_8 \mathrm{Incentive}_i \times \mathrm{Numeracy}_i \\ &+ \beta_9 \mathrm{Incentive}_i \times \mathrm{Numeracy}_i^2 \\ &+ \beta_{10} \mathrm{Incentive}_i \times \mathrm{Numeracy}_i \times \mathrm{Congenial}_i \\ &+ \beta_{11} \mathrm{Incentive}_i \times \mathrm{Numeracy}_i^2 \times \mathrm{Congenial}_i + \epsilon_i\end{aligned} \tag{4}$$

Again, we are grateful to *JEPS* reviewers who noted that we need to add the interaction between Numeracy-squared and Congenial to test hypothesis 4. The original equation was intended to replicate Kahan et al.'s (2017) results, so our regression model was inspired by the one Kahan et al. (2017) used in which they include a numeracy-sq term in their regression model but do not interact it with other variables. Since we did not conduct any preliminary data analyses we did not know that numeracy will not have a quadratic effect on accuracy in our sample.

To test the fourth hypothesis, the updated equation 4 now includes the interaction between $Congenial_i$ and $Numeracy_i^2$ and $Congenial_i$ and $Numeracy_i^2$ and $Incentive_i$ which were missing from the pre-registered third equation. Substantively, the changes to interpretation are negligible.

Analysis with no deviations from pre-registration

Multivariate analyses

The impact of congeniality on accuracy among unincentivized respondents (hypothesis 1)

To test hypotheses 1 and 2, Models 1 and 2 of Table A25 estimate equation (1), using a linear probability model for data from the unincentivized conditions; Model 2 also adds the controls (age, ethnicity, gender, education, and 2016 voting behavior). Hypothesis 1 states that, among unincentivized respondents, accuracy is higher when interpreting more congenial data. The coefficient estimate on $Congenial_i$ is positive and statistically discernible from 0, which means that for respondents of average numeracy (those whose *Numeracy* equals 0), each one unit increase in $Congenial_i$ (ranges from -1.88 to 1.88, mean=0, standard deviation=1) generates a 4.3–4.7pp increase in accuracy.

The inferiority test (Lakens et al., 2018) helps determine if this effect is substantively meaningful. Since Hypotheses 1 and 2 are conceptual replications of Kahan et al. (2017), we test whether the observed effect size is smaller than the smallest effect size that the original study could have detected. Using simulated data of that study, we find that the minimum effect size of β_2 in equation (1) that could have been detected with 81.2% power using their sample of 1111 is 0.04 in a linear probability model, i.e., the inferiority bound for the test is $\Delta = 0.04$ for β_2 in equation (1). The obtained β_2 equals 0.043 –0.047; we thus conclude that the estimated effect of $Congenial_i$ on $Correct_i$ for respondents of average numeracy is substantively meaningful in addition to being statistically significant.

Table A25: The impact of numeracy and congeniality on accuracy (unincentivized participants)

	(1) Correct	(2) Correct
Numeracy	-0.005	-0.006
	(0.010)	(0.010)
Congenial	0.047***	0.043***
	(0.016)	(0.016)
Numeracy x Congenial	0.015*	0.013
	(0.009)	(0.009)
Numeracy-sq	0.004	0.005
	(0.005)	(0.005)
Constant	0.414***	0.401***
	(0.021)	(0.101)
Observations	1016	1016
Controls	No	Yes

Standard errors in parentheses

Note: Linear probability model with heteroscedastic robust standard errors.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The impact of congeniality and numeracy on accuracy among unincentivized respondents (Hypothesis 2)

Hypothesis 2 posits that, among unincentivized respondents, the congeniality bias increases with one's numeracy, i.e., the coefficient estimate for the multiplicative interaction term *Numeracy x Congenial* is expected to be positive. The obtained positive coefficient on the interaction term of congenial and numeracy suggests that higher numeracy subjects will have greater differences in correct answers between congenial and uncongenial conditions compared to lower numeracy subjects. While this coefficient estimate is not statistically discernible at 5% level (p=0.099), we use visualizations to understand if there are ranges of values of $Numeracy_i$ for which $Congenial_i$ has a statistically meaningful impact.

While the coefficient on the interaction term of congenial and numeracy is not statistically discernible at 0.05 level, is it substantively meaningful? We find that the minimum effect size of β_3 in equation (1) that could have been detected with 82.9% power using Kahan et al.'s (2017) sample size of 1111 is 0.021 in a linear probability model. The coefficient point estimate we obtain of 0.015 is not statistically smaller than the effect size of interest of 0.021 (p=0.726 in a one-sided t-test). Thus, we cannot conclude that a meaningful effect is absent.

The impact of incentives on accuracy (hypothesis 3)

Hypothesis 3 expects that incentivized participants would be more likely to correctly interpret the contingency table question compared to those unincentivized. Models 3 and 4 of Table A26 test this expectation, estimating parameters in equation (2), using all observations. The coefficient estimates on $Incentive_i$ in Models 3 and 4 are substantively and statistically negligible, which we interpret as lack of evidence to reject the null hypothesis of monetary incentives having no effect on accuracy rate.

Table A26: The impact of incentives, numeracy, and congeniality on accuracy
(all participants)

	(3) Correct	(4) Correct	(5) Correct	(6) Correct
Incentive	-0.001	-0.002	-0.002	-0.002
	(0.019)	(0.019)	(0.019)	(0.019)
Numeracy			-0.005	-0.007
			(0.009)	(0.010)
Congenial			0.047***	0.045***
			(0.015)	(0.015)
Numeracy x Congenial			0.015*	0.014
			(0.009)	(0.009)
Numeracy-sq			0.004	0.004
			(0.003)	(0.003)
Incentive x Congenial			-0.026	-0.025
			(0.019)	(0.019)
Incentive x Numeracy			0.021*	0.020*
			(0.011)	(0.011)
Incentive x Numeracy x Congenial			0.001	0.003
			(0.011)	(0.011)
Constant	0.423***	0.397***	0.412***	0.399***
	(0.016)	(0.063)	(0.018)	(0.063)
Observations	3050	3050	3050	3050
Controls	No	Yes	No	Yes

Standard errors in parentheses

Note: Linear probability model with heteroscedastic robust standard errors.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

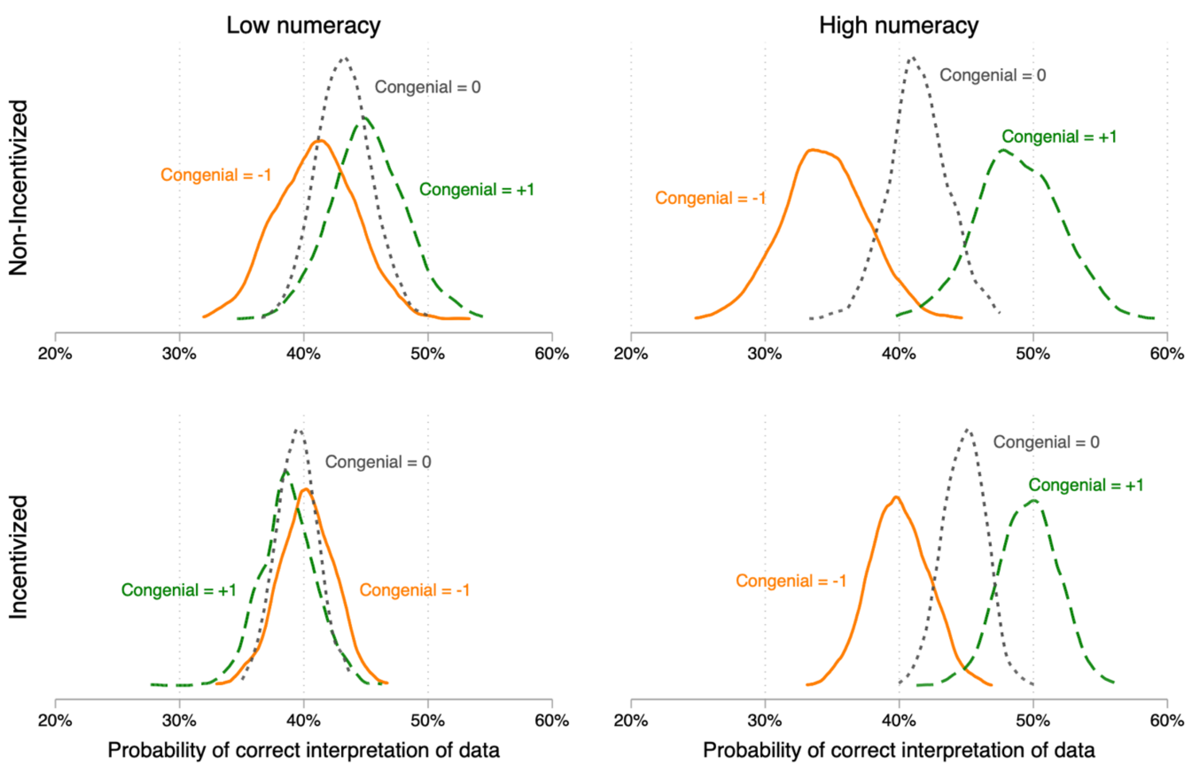
The impact of incentives, numeracy, and congeniality on accuracy (hypothesis 4)

Hypothesis 4 posits that the congeniality bias among incentivized respondents increases at a *lower* rate with one's numeracy, compared to the rate of congeniality bias increase among unincentivized respondents. To test hypothesis 4, linear probability models (Models 5 and 6 of Table A26) estimate parameters in equation (3). This hypothesis expects a positive coefficient estimate on the interaction of numeracy and congeniality (β_3 of equations 3) and a negative coefficient estimate on the interaction of incentive and numeracy and congeniality (β_8 of equations 3). The interaction of numeracy and congeniality is positive; however, it does not reach statistical significance at the 0.05 level. The three-way interaction coefficient has a statistically and substantively negligible effect on correct answers. To summarize, the increase in congeniality bias with numeracy is not different between incentivized and unincentivized respondents. We also use visualizations to understand if there are ranges of values of $Congenial_i$ and $Numeracy_i$ for which $Incentive_i$ has a statistically meaningful impact.

Visualization of the interaction effects

We supplement the linear probability model results with logistic regression results in the appendix, utilized to estimate the out-of-sample predicted probabilities of correct answer (based on MC simulations) to create Figure A5, which visualizes how incentives, numeracy, and congeniality interact to shape motivated numeracy.[5]

Figure A5: Predicted probabilities of correctly interpreting the data



Note: Density distributions derived via MC simulation from logistic regression that estimates equation 3 (output is shown in the appendix), when *Congenial* is set at –1 SD (i.e., respondents facing data contradicting their beliefs), at mean (i.e., ideological moderates on the ideology-party affiliation spectrum), and +1 SD (i.e., data are consistent with beliefs), and numeracy set at -1SD (1 out of six correct questions) for 'low numeracy and +1SD (4.35 out of six correct questions) for 'high numeracy.'

[5] The appendix also contains visualizations of locally weighted regressions of correct interpretation of contingency tables on numeracy scores and the differences in estimated probabilities of correct interpretation between liberal democrats and conservative republicans (following Kahan et al.'s 2017 Figures 6 and 8)

Figure 2 demonstrates that the congeniality bias increases with numeracy. The predicted accuracy rates for respondents facing uncongenial, neutral, and congenial data constitute 1) 41%, 43%, and 45%, i.e., the congeniality bias equals 4pp for the less numerate unincentivized (top left), 2) 34%, 41%, and 49%, i.e., the congeniality bias is 15pp for the more numerate unincentivized (top right), 3) 40%, 39%, and 39%, i.e., the bias equals 0pp for the less numerate incentivized (bottom left), and 4) 40%, 45%, and 50%, i.e., the bias constitutes 10pp for the more numerate incentivized (bottom right).

Table 7 uses differences-in-means to test whether the differences in the congeniality bias between less and more numerate individuals are statistically distinct. The difference of 10pp for respondents with numeracy of -/+1SD below/above the mean is not statistically distinct at 0.05 level (p=0.1025). However, the difference of 16pp for respondents with numeracy of -/+1.5SD below/above the mean is marginally statistically significant (p=0.056) and the difference for respondents with numeracy of -/+2 SD below/above the mean is 21pp (p=0.04). Our data, therefore, indicate that numeracy has a smaller impact on the congeniality bias of unincentivized respondents than Kahan et al. (2017) found; we reject the second null hypothesis of no effect of numeracy on congeniality bias.

Table A27: Differences in the exhibited congeniality bias between less and more numerate individuals at various levels of numeracy

	Congeniality Bias Low Num	Congeniality Bias High Num	Diffe rence	t-statistic	p-value
Low vs high numeracy (+/-1 SD), no inc	-.0415	-.1460	0.1045	1.6336	0.1025
Low vs high numeracy (+/-1 SD), inc	0.0148	-0.0973	0.1121	2.5418	0.0111
Low vs high numeracy (+/-1.5 SD), no inc	-.0148	-.1723	.1576	1.91	0.0560
Low vs high numeracy (+/-1.5 SD), inc	.0420	-.1261	.1681	2.996	0.0028
Low vs high numeracy (+/- 2 SD), no inc	.0122	-.1986	.2108	2.056	0.0399
Low vs high numeracy (+/- 2 SD), inc	.0692	-.1539	.2231	3.222	0.0013

Note: Congeniality Bias = Pr(correct=1|congenial=-1SD) - Pr(correct=1|congenial=1SD);
Numeracy is set at either +/-1SD or +/-1.5 SD or +/-2 SD

Incentives shrink the gap in motivated numeracy from 4pp to 0pp (4pp reduction) among less numerate (-1SD) and from 15pp to 10pp (5pp reduction) among more numerate individuals. Similar reductions in the congeniality bias are observed at other levels of numeracy. Hypothesis 4 expected that incentives would generate a slightly larger gain in accuracy among more numerate individuals, however the gains we observe are not statistically distinct between less and more numerate; we fail to reject the fourth null hypothesis of no effect.

Logistic regression results (no deviations from pre-registration)

The impact of congeniality on accuracy among unincentivized respondents (hypothesis 1) and the impact of congeniality and numeracy on accuracy among unincentivized respondents (hypothesis 2)

Table A28: The impact of numeracy and congeniality on accuracy

	(1) Correct	(2) Correct
Numeracy	-0.021	-0.026
	(0.040)	(0.044)
Congenial	0.194***	0.184***
	(0.065)	(0.067)
Numeracy x Congenial	0.065	0.056
	(0.040)	(0.041)
Numeracy-sq	0.014	0.020
	(0.022)	(0.023)
Constant	-0.352***	-0.417
	(0.088)	(0.421)
Observations	1016	1016
Controls	No	Yes

Standard errors in parentheses

Note:Logit regression with heterscedasticity robust standard errors.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The impact of incentives on accuracy (hypothesis 3) and the impact of incentives, numeracy, and congeniality on accuracy (hypothesis 4)

Table A29: The impact of incentives, numeracy, and congeniality on accuracy

	(3) Correct	(4) Correct	(5) Correct	(6) Correct
Incentive	-0.006	-0.007	-0.005	-0.004
	(0.078)	(0.078)	(0.078)	(0.079)
Numeracy			-0.022	-0.030
			(0.039)	(0.040)
Congenial			0.194***	0.188***
			(0.065)	(0.065)
Numeracy x Congenial			0.065	0.059
			(0.040)	(0.040)
Numeracy-sq			0.017	0.017
			(0.013)	(0.013)
Incentive x Congenial			-0.111	-0.107
			(0.079)	(0.080)
Incentive x Numeracy			0.088*	0.085*
			(0.048)	(0.048)
Incentive x Numeracy x Congenial			0.003	0.011
			(0.048)	(0.049)
Constant	-0.310***	-0.419	-0.360***	-0.418
	(0.064)	(0.259)	(0.073)	(0.263)
Observations	3050	3050	3050	3050
Controls	No	Yes	No	Yes

Standard errors in parentheses

Note:Logit regression with heterscedasticity robust standard errors.

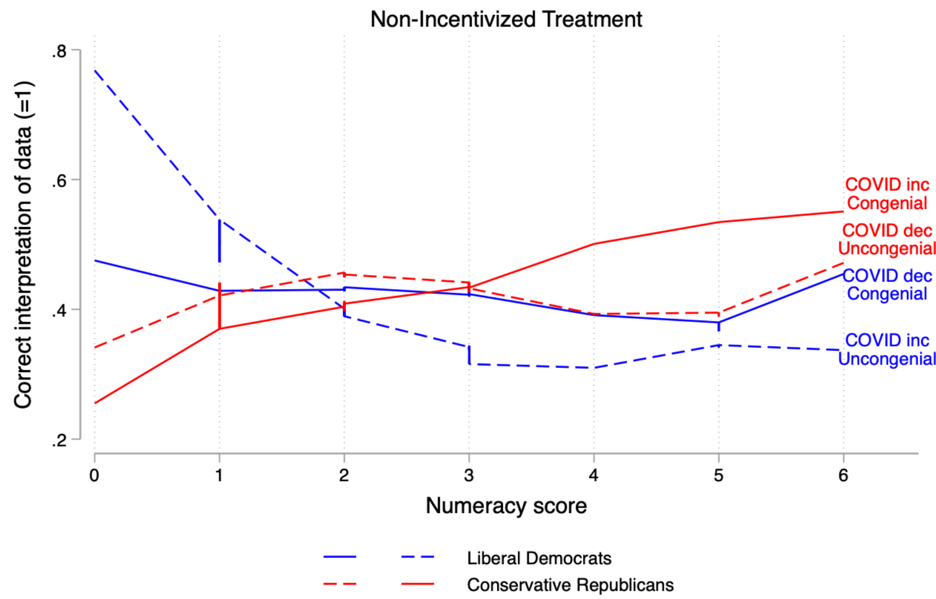
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

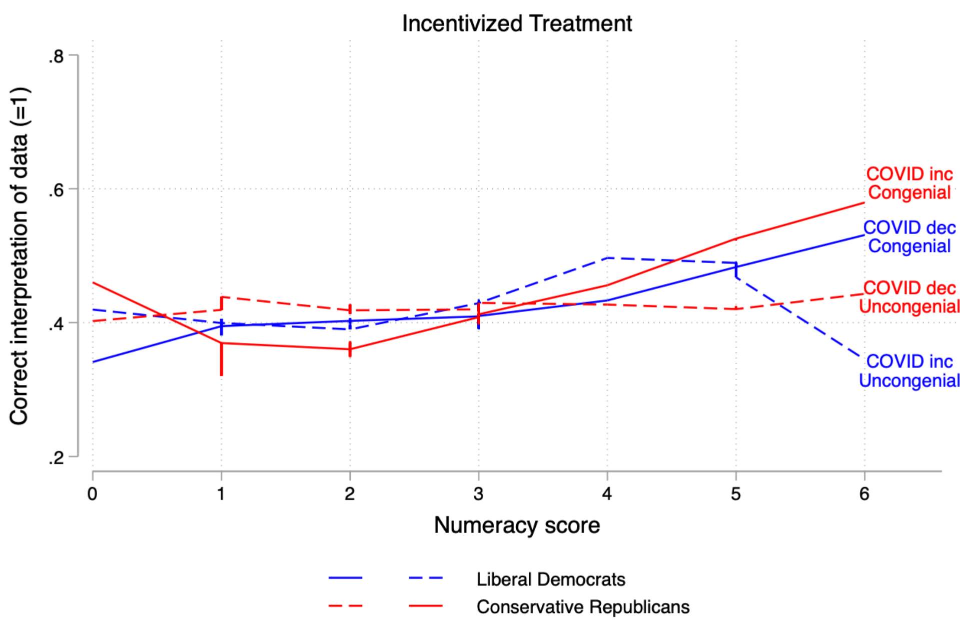
Additional visualizations of the interaction effects

Responses by subjects of opposing ideological outlooks

We use weighted regressions to create Figure A6 that follows Kahan et al.'s (2017) Figure 6 to help further interpret the interaction effects that incentives, numeracy, and congeniality have on one's motivated numeracy.

Figure A6: Responses by subjects of opposing ideological outlooks





Note: Locally weighted regression lines track the proportion of subjects answering the contingency table correctly in relation to numeracy levels in the various conditions. Blue lines plot relationships for subjects who score below the mean and red ones are for subjects who score above the mean on *Conservative*, the composite measure of ideology and identification with one or the other major party. Solid lines are used for subjects in the congenial condition, while dashed lines—in uncongenial.

Figure A6 charts locally weighted regressions of correct interpretation for the full range of numeracy scores for liberal Democrats and conservative Republicans for each of the four treatments. First, consider the top part of Figure A1 that displays the results for the unincentivized respondents. Visual inspection suggests that political outlooks interact with one's numeracy when interpreting ideologically tinged data: liberal Democrats facing the congenial condition *Covid_decreases* exhibit no relationship between numeracy and correct answer, while liberal Democrats placed in the uncongenial condition *Covid_increases* become increasingly unlikely to answer correctly as they become more numerate. The result appears to be the opposite for conservative Republicans: those facing the congenial condition *Covid_increases* become increasingly likely to answer correctly as they become more numerate. By contrast, numeracy appears to have no impact on correct answer for conservative Republicans placed in the uncongenial condition *Covid_decreases.*

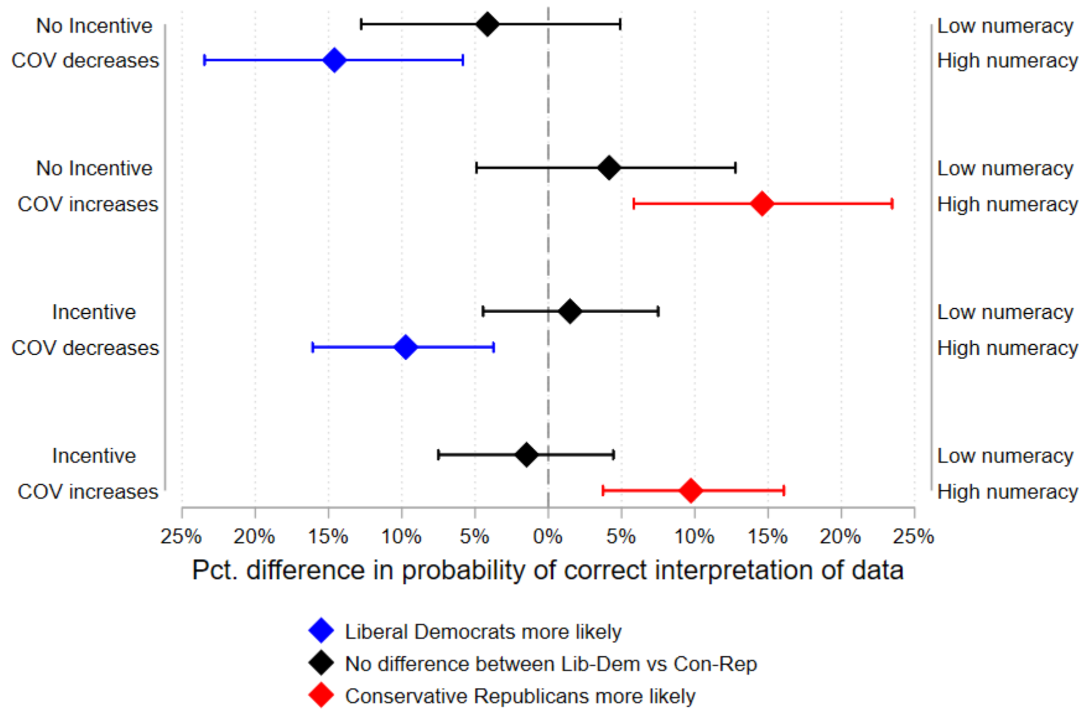
Figure A6's top chart suggests, therefore, two takeaways. First, we confirm the finding established by the descriptive analysis of differences-in-means as well as multivariate analyses: unincentivized respondents exhibit strong congeniality bias with greater numeracy: both Democrats and Republicans of above-average numeracy showed lower rates of accuracy in the uncongenial condition than in the congenial condition. However, Kahan et al.'s (2017) main conclusion from weighted regression visualization was that within each ideological camp, more numerate individuals were less likely to interpret uncongenial data correctly than less numerate respondents. While this pattern appears to be true for Liberal Democrats, Figure 2's top chart shows that this is not the case for Conservative Republicans.

The bottom graph of Figure A6 replicates the same weighted regressions for the incentivized respondents. Visual inspection suggests that both ideological camps exhibit a modest increase in accuracy with numeracy when facing congenial data. When facing uncongenial conditions, incentivized Republicans exhibit no increase in accuracy with numeracy, while Democrats appear to demonstrate a modest increase until the numeracy score of 4 and a sharp decline in accuracy for those who answered all numeracy problems correctly. We caution against overinterpreting this result, because there were very few Democrats who scored 6 on the numeracy test.

Predicted differences in probability that partisans will correctly interpret the data

Figure A7 graphs the differences in predicted probabilities of giving the correct answer in each treatment for partisans with low numeracy and high numeracy, using MC simulations with the same parameter assumptions employed to generate Figure 2 of the paper. Figure A7 allows for additional evaluation of hypotheses 2 and 4.

Figure A7: Predicted differences in probability that partisans will correctly interpret the data



Note: Predicted differences in probabilities derived via Monte Carlo simulation from logistic regression model that estimates equation 3 in the appendix. Predictors for *Conservative* set at –1 SD and +1 SD for 'liberal Democrat' and 'conservative Republican', respectively, and numeracy set at -1SD (1 out of six correct questions) for 'low numeracy and +1SD (4.35 out of six correct questions) for 'high numeracy'. Confidence intervals indicate 0.95 levels of confidence.

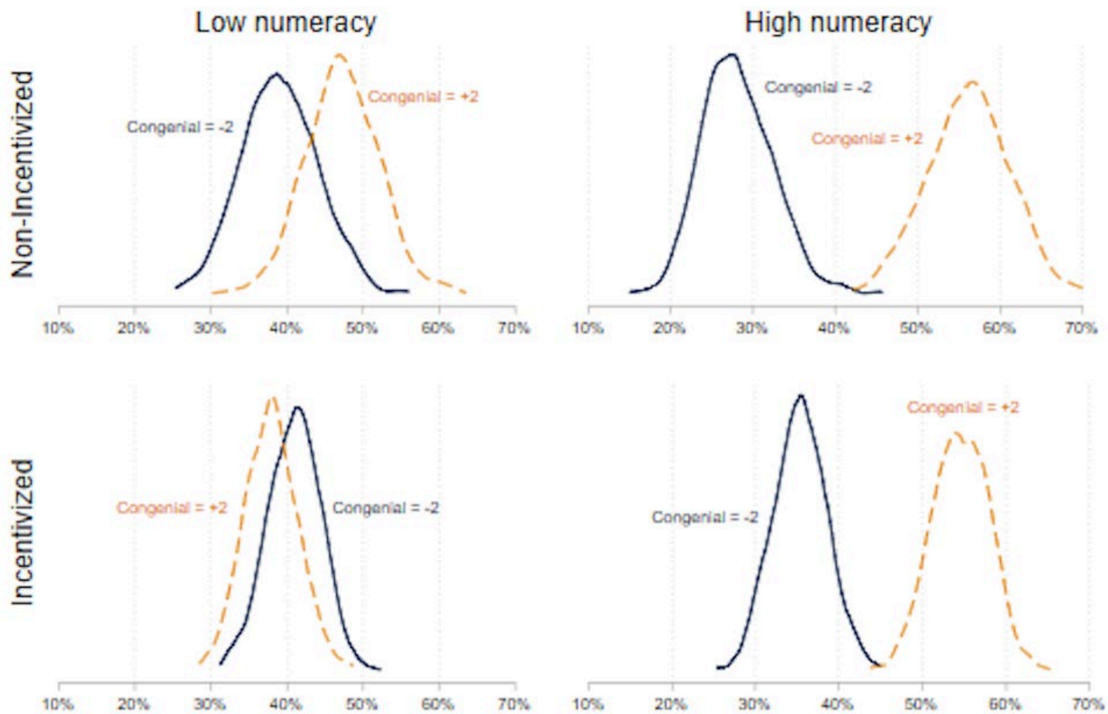
Figure A7 presents four pairs of differences. First, consider the differences in the estimated probability of correct answer between low and high numeracy unincentivized respondents (top two pairs of differences). We observe that when facing congenial data (Democrats in the COVID decreases condition and Republicans in the COVID increases condition), more numerate respondents are by 14.6 percentage points more likely to interpret the contingency table correctly, while less numerate respondents were only 4.1 percentage points more likely to exhibit the congeniality bias. That is, greater numeracy adds 10.5 percentage points to one's accuracy rate when interpreting congenial data without incentives for accuracy. The difference between the low and high numeracy individuals is statistically significant in a one-tail but not in a two-tail test.

Consider the bottom two pairs of estimates for incentivized participants in Figure A2. We observe that incentives for accuracy do not eliminate the tendency of more numerate respondents to exhibit greater congeniality bias. We observe that when facing congenial data (Democrats in the COVID decreases condition and Republicans in the COVID increases condition) incentivized highly numerate respondents are by 9.8 percentage points more likely to interpret the contingency table correctly, while less numerate respondents were 1.5 percentage point less likely to interpret the contingency table correctly (though this number is not statistically discernible at the 5% level). That is, the gap between high and low numeracy respondents rises under incentives (it equals 11.2 percentage points and is statistically discernible in a two-tail test).

The fourth hypothesis expected that incentives would boost accuracy to a greater extent among more numerate than less numerate respondents. Incentives reduce the congeniality bias from 4.1pp to -1.5pp, i.e., a drop of 5.6pp among less numerate and from 14.6pp to 9.8pp among more numerate individuals, i.e., a drop of 4.8pp. These gains in accuracy between low and high numeracy respondents are not statistically distinct, therefore we fail to reject the fourth null hypothesis.

Predicted probabilities of correctly interpreting the data for congeniality set at +/-2SD scenarios

Figure A8: Predicted probabilities of correctly interpreting the data



Note: Density distributions derived via Monte Carlo simulation from logistic regression that estimates equation (3, output is shown in the appendix), when *Congenial* is set at –2 SD and +2 SD, and numeracy set at -1 SD (1 out of six correct questions) for 'low numeracy and +1SD (4.35 out of six correct questions) for 'high numeracy'.

In Figure A8, we compare how higher levels of congeniality and uncongeniality with data (set at +/-2 SD for Congenial) affects the probability of correct interpretation. We find that the probability distributions of correctly interpreting congenial and uncongenial data are similar and overlapping for low numerate participants (for both incentivized and unincentivized treatment conditions). On the other hand, the probability distributions of correctly interpreting the congenial and uncongenial data are very different for high numerate participants (for both incentivized and unincentivized treatment conditions.) Congenial data is interpreted at much higher levels of accuracy than uncongenial data for the high numeracy participants.

Differences in accuracy between supporters and opponents of mask mandates

Table A30: Differences in accuracy between presumed supporters and opponents of mask mandates (high numeracy respondents)

	Uncongenial	Congenial	Difference	t	p-value	DiD	t	p-value
Conservative Republicans								
Conservative, high num, no incentive	0.33	0.72	-0.39	-2.20	0.04	-0.24	-1.09	0.28
Conservative, high num, incentive	0.33	0.48	-0.15	-1.13	0.26			
Liberal Democrats								
Liberal, high num, no incentive	0.22	0.37	-0.15	-0.96	0.34	0.15	0.76	0.45
Liberal, high num, incentive	0.38	0.67	-0.29	-2.71	0.01			

Note: The congenial condition for Conservative Republicans (respondents who exceed 1 on *Conservative)* is assumed to be the condition "Covid increases" and for Liberal Democrats (respondents who are below –1 on *Conservative)*— "Covid decreases." Numeracy is set at –1SD (1 out of six correct questions) for 'low numeracy and +1SD (4.35 out of six correct questions) for 'high numeracy.'

Table A31: Differences in accuracy between presumed supporters and opponents of mask mandates (low numeracy respondents)

	Uncongenial	Congenial	Difference	t	p-value	DiD	t	p-value
Conservative Republicans								
Conservative, low num, no incentive	0.21	0.29	-0.07	-0.46	0.65	-0.30	-1.55	0.12
Conservative, low num, incentive	0.54	0.32	0.23	2.16	0.03			
Liberal Democrats								
Liberal, low num, no incentive	0.46	0.67	-0.21	-1.46	0.15	-0.06	-0.35	0.73
Liberal, low num, incentive	0.38	0.52	-0.15	-1.44	0.15			

Note: The congenial condition for Conservative Republicans (respondents who exceed 1 on *Conservative)* is assumed to be the condition "Covid increases" and for Liberal Democrats (respondents who are below –1 on *Conservative)*— "Covid decreases." Numeracy is set at –1SD (1 out of six correct questions) for 'low numeracy and +1SD (4.35 out of six correct questions) for 'high numeracy.'

References

Kahan, Dan M., Ellen Peters, Erica Cantrell Dawson, and Paul Slovic. 2017. "Motivated numeracy and enlightened self-government." *Behavioural Public Policy* 1(1): 54–86.

Lakens, Daniël. Anne M. Scheel , and Peder M. Isager. 2018. "Equivalence Testing for Psychological Research: A Tutorial." *Advances in Methods and Practices in Psychological Science*, 1(2) 259–269.

Pew Research Center. 2016. "Evaluating online nonprobability surveys – 3. Demographic, political and interest profiles." May 2. URL: https://www.pewresearch.org/methods/2016/05/02/demographic-political-and-interest-profiles/

The effect of incentives on motivated numeracy amidst COVID-19

Eunbin Chung[1] Pavitra Govindan Anna Pechenkina

Recommended Reporting Standards for Experiments (Survey)

A. Hypotheses

- Specific objectives or hypotheses.
  - What question(s) was (were) the experiment designed to address?
    - How do incentives for accuracy and numeracy affect the processing of uncongenial data?
  - What are the specific hypotheses to be tested?
    - Hypothesis 1: Among unincentivized respondents, the rate of correct data interpretation increases as the data become more congenial to one's ideological beliefs. (The congeniality bias exists.)
    - Hypothesis 2: Among unincentivized respondents, the congeniality bias increases with one's numeracy.
    - Hypothesis 3: Relative to unincentivized respondents, those incentivized will exhibit greater accuracy in all conditions.
    - Hypothesis 4: The congeniality bias among incentivized respondents increases at a *lower* rate with one's numeracy, compared to the rate of bias increase among unincentivized respondents.

B. Subjects and Context

- Eligibility and exclusion criteria for participants.
  - Why was this subject pool selected?
    - The research question would be ideally tested in a random sample of the US population. For feasibility reasons, US-based internet users.
  - Who was eligible to participate in the study?
    - Any US-based internet user over 18 years old.
  - What would result in the exclusion of a participant?
    - After the treatment, respondents answered the factual manipulation check question, i.e., an objective question about the study's content (Kane and Barabas, 2019); those who failed the check were removed from the sample.
    - Additionally, Qualtrics removed 'careless or insufficient effort (C/IE)' responses by identifying speeders (who took less than half the median time), those who took too long or took long breaks mid-survey, duplicates (based on demographics and IP/geo data), "straightliners" (respondents answering multiple consecutive questions in the same manner or those who create patterns in grids), likely bots (based on the algorithm developed by Qualtrics, which checks for specific demographic profiles and attitudinal responses), and responses from outside the U.S. (checked via IP/geo data).
  - Were any aspects of recruitment changed (such as the exclusion criteria) after recruitment began?
    - Before recruitment, we expected to pre-screen participants based on demographic quotas once certain quotas were filled. Most online samples overrepresent better

[1] Names appear alphabetically, authorship is equal.

educated and more politically engaged individuals. After a month of data collection, Qualtrics informed us that based on trends, the planned quotas could not be met, i.e., if we wanted screen out better educated and more politically engaged individuals, we would not be able to fill the planned quotas by mid-fall of 2021. Although not all planned quotas were filled, the resulting sample is more balanced than most online samples. Tables A1–A6 of the appendix compare the planned quotas versus the obtained.

- Procedures used to recruit and select participants.
  - <u>If there is a survey</u>: Identify the survey firm used and describe how they recruit respondents.
    - The survey was implemented by Qualtrics.
    - Qualtrics targets potential participants through third party vendors that recruit respondents. For this project, Qualtrics used two panel providers that helped collect these responses. Depending on the method of recruitment, respondents are typically invited or allowed to self-select entry into surveys. This method of recruitment varies from app notifications or mobile apps, or different websites our panels may use for recruitment. These surveys are either sent to respondents based on their demographic characteristics or have screeners at the start in which respondents must state different targetable demographics to access the survey.

- Recruitment dates defining the periods of recruitment and when the experiments were conducted:
    - April 15 to June 30, 2021.

  - Also list dates of any repeated measurements as part of a follow-up:
    - N/A

- Settings and locations where the data were collected.
  - In the field, lab, classroom, or some other specialized setting?
    - In the field via Qualtrics's online platform: participants could take the survey on any device that supports an internet browser at any location of their choice.

  - Other relevant specifics of the population: e.g., large public university vs. small private university; geographic location; etc.
    - No specific characteristics.

- <u>If there is a survey</u>: Provide response rate and how it was calculated.
    - 13864 individuals were invited to take the survey, of whom 2522 did not consider taking the survey and 11342 individuals started taking the survey. Of the 11342, 8292 were excluded from the sample for various reasons and 3050 provided high-quality responses and passed the factual manipulation check (this is our sample). The response rate was 22% (i.e., 3050/13864).
    - Data collection was terminated after 3050 quality responses were reached.

C. Allocation Method

- Details of the procedure used to generate the assignment sequence (e.g., randomization procedures).
    - Randomization procedure was done through the Qualtrics survey taking software that has a built-in randomizer.
    - The study randomly assigned participants to either the no-incentives (1/3 of the sample) or incentives treatment (2/3 of the sample). Within each group, respondents were randomly assigned with probability 0.5 to one of

two data tasks.

- If random assignment used, then details of procedure (e.g., any restrictions, blocking).
  - Note the unit of randomization (individuals, groups, households, etc). Pay careful attention to report clustered random assignment if subjects were assigned at some level other than the individual subject.
    - The unit of randomization is individual subject; no restrictions or clustering was used.

- If random assignment used, provide evidence of random assignment.
  - If demographic or other pretreatment variables were collected, a table (in text or appendix) showing baseline means and standard deviations for demographic characteristics and other pre-treatment measures by experimental group.
    - Table A8 of the online appendix shows that all observable attributes are balanced across treatment groups.

  - table for each of the blocks. If there are too many blocks for this to be practical, combine blocks to present weighted averages of covariates using inverse probability weighting.
    - N/A

- Blinding: Were participants, those administering the interventions, and those assessing the outcomes unaware of condition assignments?
  - If blinding took place, include a statement regarding how it was accomplished and how the success of blinding was evaluated.
    - Subjects did not know that there were multiple treatments and to which treatment they were assigned. That is, each subject saw only one of four possible versions of the survey (incentive vs. no-incentive, data task 1 vs. data task 2).

D. Treatments

- Description of the interventions in each treatment condition, as well as a description of the control group.
  - Descriptions should be sufficient to allow replication: Summary or paraphrasing of experimental instructions in the article text; verbatim instructions and/or other treatment materials provided in an appendix.
    - Incentive treatment:
      [**Incentive Treatment**:] We will give you a bonus of $1.00 for the correct answer.
      [**No-Incentive Treatment**:] Your answer to this question will not affect the total amount you earn in this survey.

    - Data task treatment has two conditions: *Covid_decreases* and *Covid_increases*

Condition 1: COVID-19 cases decrease

	COVID-19-positive cases increased	COVID-19-positive cases decreased
Cities with a mask-wearing mandate	223	75
Cities without a mask-wearing mandate	107	21

Condition 2: COVID-19 cases increase

	COVID-19-positive cases decreased	COVID-19-positive cases increased
Cities with a mask-wearing mandate	223	75
Cities without a mask-wearing mandate	107	21

-

- How and when manipulations or interventions were administered.
  - Method of delivery: Pen-and-paper vs. computer or internet vs. face-to-face communication vs. over the telephone.
    - Internet. Participants could use any device that supports an internet browser.
  - If computerized, the software should be described and cited. (If possible, programs should be included in appendix so as to be available for purposes of replication.)
    - Qualtrics survey tool.
  - For lab experiments (and other experiments, when relevant):
    - Report the number of repetitions of the experimental task and the group rotation protocol. Report the ordering of treatments for within-subject designs. Any piggybacking of other protocols should be reported. Report any use of experienced subjects or subjects used in more than one session or treatment.
      - No repetitions
      - Within-subject: we randomly assigned half the sample to take numeracy questions before the data task treatment and half—after. The order effects are discussed in section "Order effects" of the paper.
      - No other protocols.
      - No experienced subjects—Qualtrics screens out repeating IP addresses.
    - Time span: How long did each experiment last? How many sessions were subjects expected to attend? If there were multiple sessions, how much time passed between them?
      - Duration: between 8 and 30 minutes, average completion time was under 15 min.
      - The survey could be taken only once.
    - Total number of sessions conducted and number of subjects used in each session.
      - NA
    - Was deception used?
      - No.
    - Treatment fidelity: Evidence on whether the treatment was delivered as intended.
      - Report any instructional anomalies or inaccuracies.
        - NA
      - Were subjects given quizzes on the experimental instructions?
        - No.

- Were there practice rounds? If so, how many and what were the results?
  - No.

- Did subjects complete a post-experiment debriefing, interview, or questionnaire? If so, is there evidence that subjects understood the instructions and treatments?
  - No post-experiment debriefing was done.

- Did the experimental team observe aspects of the intervention?
  - N/A

- Provide description of manipulation checks, if any.
  - The manipulation check followed the data task:

    The previous question asked me about whether cities with a mask-wearing mandate saw an increase or a decrease in COVID-19 cases.
    True (1)
    False (0)

- Were incentives given? If so, what were they and how were they administered.

- Two types of incentives were administered:
  1. All participants respondents were motivated to answer the numeracy questions correctly: one of six numeracy questions was randomly selected for each participant and if the participant's answer was correct in that question, they received $1.
  2. 2/3 of the sample assigned to the incentive condition also received $1.00 for correctly solving the contingency table task.
- Qualtrics sent rewards to respondents as gift cards.

E. Results

1. Outcome Measures and Covariates

- Provide precise definition of all primary and secondary measures and covariates.
  - For indices, provide exact description of how they are formed. For survey items provide exact question wording in an appendix. Please provide a copy of the complete survey questionnaire (in an on-line appendix if it is long).

The exact question wording for how the following outcome measures and covariates were measured are provided in the survey questionnaire in the paper's appendix.

Primary Outcome Measure: $Correct_i$ equals 1 when a respondent i correctly answered the contingency table question, 0 otherwise.

Covariates:

$Numeracy_i$ is the number of correct answers the participant i gives to the six questions aimed at measuring numeracy (Peters et al., 2007; Schwartz et al., 1997; Lipkus et al., 2001). Numeracy was standardized to be centered at '0' for ease of interpretation. Items on disease/infection were removed considering their relevance to the treatment.

$Congenial_i$ is a continuous measure of attitude-consistent message that captures the degree of congeniality of the contingency table's data with one's ideology. This variable is constructed using, first, the continuous measure of ideology ($Conservative_i$) and, second, two binary indicators of the condition to which a respondent was assigned ($Covid\_decreases_i$ or $Covid\_increases_i$). Using Cronbach's α, two 7-point Likert scales of party affiliation and ideology form an aggregate scale, $Conservative_i$, where negative values indicate liberal Democrats and positive values indicate conservative Republicans (α=0.795 which is similar to the α value of 0.83 in Kahan et al. 2017). $Conservative_i$ is identical to the measure of ideology in Kahan et al. (2017). Second, conservatives received uncongenial data in $Covid\_decreases_i$ and liberals—in $Covid\_increases_i$. The $Congenial_i$ variable follows the formula:

$$Congenial = \begin{matrix} Conservative * (Covid\_increases - Covid\_decreases) & \text{if } Conservative \geq 0 \\ -Conservative * (Covid\_decreases - Covid\_increases) & \text{if } Conservative < 0 \end{matrix}$$

For a conservative respondent ($Conservative_i$>0), the $Congenial_i$ variable is positive in the $Covid\_increases_i$ condition and negative in the $Covid\_decreases_i$ condition, while for a liberal ($Conservative_i$<0), $Congenial_i$ is positive in $Covid\_decreases_i$ and negative in $Covid\_increases_i$ condition. The absolute magnitude of $Congenial_i$ increases with the strength of one's ideology.
$Incentive_i$ is a binary indicator of whether a participant is assigned to that condition.

Indicators of age, ethnicity, gender, education, and voting behavior are the other covariates that were collected.

- Clearly state which of the outcomes and subgroup analyses were specified prior to the experiment and which were the result of exploratory analysis.
  All the outcomes and subgroup analyses were pre-registered prior to the experiment. Deviations from the pre-registration are described in Table 1 of the paper.

2. Complete CONSORT Participant Flow Diagram

- An example of a CONSORT flow diagram is attached. The flow diagram records the initial number of subjects deemed eligible for the experiment and all losses of subjects during the course of the experiment. The flow chart follows the subjects from initial recruitment to the sample used in the main analyses, providing readers clear information on the amount of attrition and exclusions. The chart also reports the portion of each treatment group that received the allocated intervention and if not, why this was not accomplished. Naturally, in the event that there is zero or very trivial non-compliance with group assignment or zero or very trivial attrition, researchers may decide it is more convenient to report the information that would otherwise be shown in the CONSORT diagram in the text and omit the diagram.
  Flow diagram is filled up and appended at the end of the document.

3. Statistical Analysis

- Researchers will conduct statistical analysis and report their results in the manner they deem appropriate. We recommend that this reporting include the following:
  - Report sample means and standard deviations for the outcome variables using intent-to treat (ITT) analysis (means for the entire collection of subjects assigned to a group, whether the treatment is successfully delivered or not).
    - If the experiment uses block randomization with unequal assignment rates, present ITT analysis by block or present overall means using inverse probability weighting.

    N/A. All the subjects that completed the survey were assigned a treatment group and they received the treatment.

  - Note if level of analysis differs from level of randomization and estimate appropriate standard errors.
    Level of analysis and the level of randomization are both at the individual level.

  - If there is attrition, discuss reasons for attrition and examine if attrition is related to pre-treatment variables.
    N/A. One time survey and there was no returning of participants to the experiment.

- Report for other missing data (not outcome variables):
  - Frequency or percentages of missing data by group.
    No missing data.
  - Methods for addressing missing data (e.g., listwise deletion, imputation methods).
    N/A
  - For each primary and secondary outcome and for each subgroup, provide a summary of the number of cases deleted from each analysis and rationale for dropping the cases.
    N/A
- For survey experiments: Describe in detail any weighting procedures that are used.
  No weighting procedures were used for the data analysis.

F. Other Information

- Was the experiment reviewed and approved by an IRB?
  The study was approved by the University of Utah Institutional Review Board (#IRB_00132903).

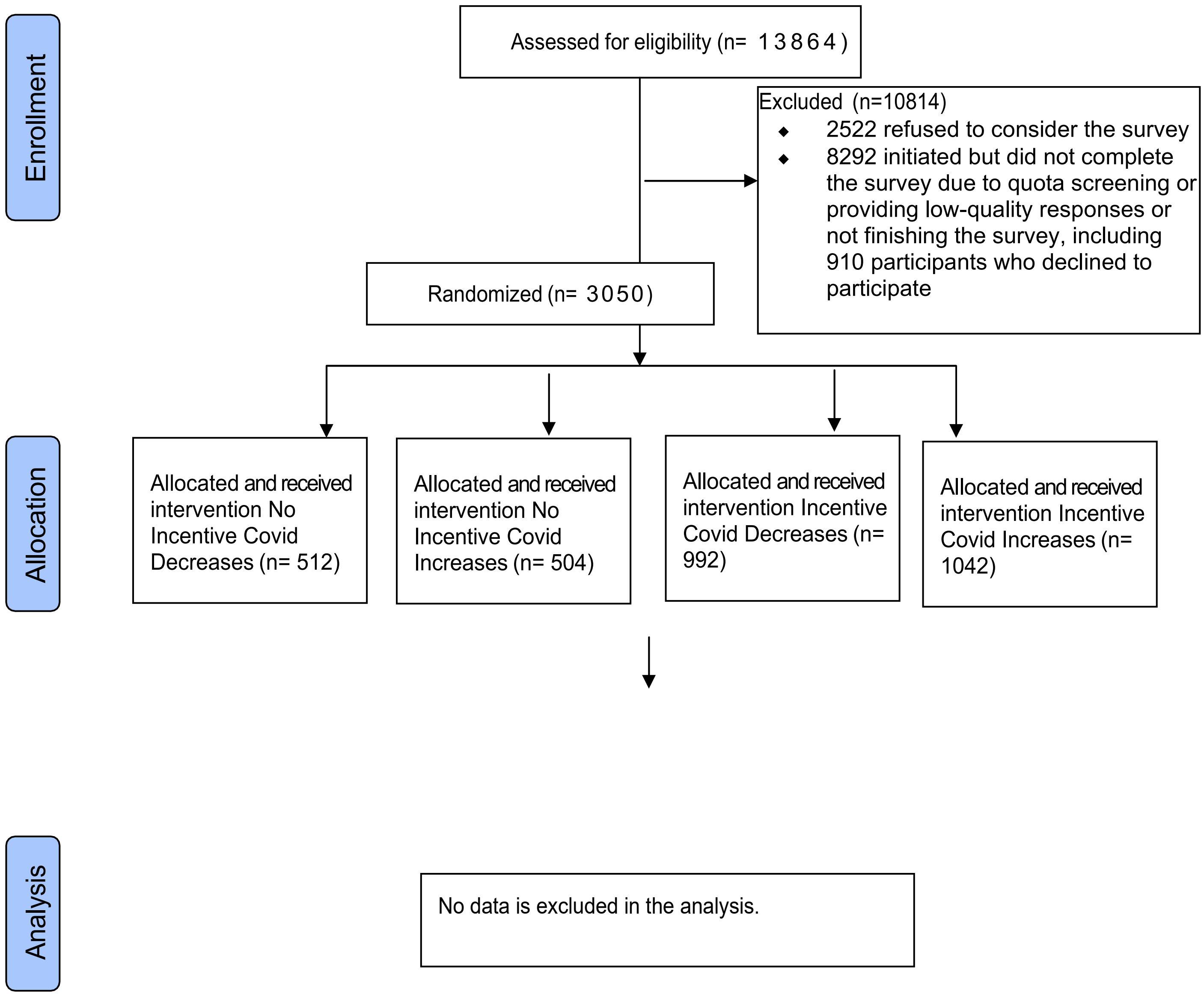
- If the experimental protocol was registered, where and how can the filing be accessed?

The experimental protocol was registered on OSF. It can be assessed at the following:
Chung, Eunbin, Anna Pechenkina, and Pavitra Govindan. 2020. "Psychology of Ideology, Xenophobia, and Motivated Numeracy Amidst COVID-19." OSF. May 21. doi:10.17605/OSF.IO/KVZ9T.

- What was the source of funding? What was the role of the funders in the analysis of the experiment?
  The sources of funding were the: 1) Department of Political Science, College of Social and Behavioral Science, University of Utah, 2) College of Humanities and Social Sciences, Utah State University, and the 3) Department of Economics, College of Social and Behavioral Science, University of Utah. The funders did not have a role in the analysis of the experiment.

  - Were there any restrictions or arrangements regarding what findings could be published?
    None.
  - Any funding sources where conflict of interest might reasonably be an issue? None.
- If a replication data set is available, provide the URL.
  https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/EAL6AE

CONSORT Statement 2010 Flow Diagram



From Schulz KF, Altman DG, Moher D, for the CONSORT Group. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *BMJ* 2010;340:c332.

For more information, visit www.consort-statement.org.