Essays on Belief Formation and Political Polarization

Dissertation Advisor:
Professor Matthew Rabin

Author:
Michael Thaler

Essays on Belief Formation and Political Polarization

Abstract

This dissertation consists of essays in behavioral, experimental, and political economics. It studies why people form polarized and biased beliefs, and how these beliefs lead to politically-polarized behavior in public health. The first chapter constructs a new experimental design to show how people, when faced with new information, engage in *motivated reasoning* by distorting their inference in the direction of beliefs they are more motivated to hold. Its results show that motivated reasoning is widespread and helps explain persistent belief polarization along political dimensions about immigration, income mobility, racial discrimination, crime, gender-based math ability, climate change, and gun laws. The second chapter delves into the limits of motivated reasoning, showing that the bias is *not* a major factor in the absence of self-image concerns. It shows that people do not systematically motivatedly reason to think that the world is a better place for others. The third chapter, coauthored with Hunt Allcott, Levi Boxell, Jacob Conway, Matthew Gentzkow, and David Yang, studies the public-health implications for political polarization in beliefs and behavior. It shows in a survey that there are significant gaps between Republicans and Democrats in beliefs about the severity of the COVID-19 pandemic in the United States, and shows using smartphone data that areas with more Republicans engage in significantly less social distancing.

Contents

0 Introduction 1

1 The "Fake News" Effect: An Experiment on Motivated Reasoning and Trust in News 3
- 1.1 Introduction . . . 3
- 1.2 Model and Predictions . . . 11
- 1.3 Experimental Design . . . 21
- 1.4 Data and Experiment Details . . . 30
- 1.5 Results . . . 31
- 1.6 Demographic Heterogeneity . . . 51
- 1.7 Conclusion . . . 56

2 The Limits of Motivated Reasoning When Self Image Is Not at Stake 58
- 2.1 Introduction . . . 58
- 2.2 Theory and Experimental Design . . . 61
- 2.3 Data . . . 65
- 2.4 Results . . . 66
- 2.5 Discussion . . . 72
- 2.6 Conclusion . . . 74

3 Polarization and Public Health: Partisan Differences in Social Distancing during the Coronavirus Pandemic 77
- 3.1 Introduction . . . 77
- 3.2 Motivating Facts . . . 81
- 3.3 Stylized Model . . . 82
- 3.4 Data . . . 87
- 3.5 SafeGraph Empirical Specification and Results . . . 89
- 3.6 Survey Results . . . 97
- 3.7 Conclusion . . . 102

References 110

Appendix A Appendix for Chapter 1 111
- A.1 Additional Results . . . 111
- A.2 Supplementary Appendix: Demand for News, susceptibility, and Structurally Estimating Motives . . . 122
- A.3 Study Materials: Exact Question Wordings . . . 133
- A.4 Replication . . . 140
- A.5 Online Appendix: Additional Robustness for Table 1.2 . . . 146
- A.6 Online Appendix: Study Materials . . . 158

Appendix B Appendix for Chapter 2 **175**
B.1 Additional Tables 175
B.2 Study Materials: Exact Question Wordings 177
B.3 Study Materials: Pages in Experiment 180

Appendix C Appendix for Chapter 3 **188**
C.1 Additional Figures 188
C.2 Additional Details 200

List of Figures

1.1 Politically-Motivated Reasoning: Perceived Veracity by News Direction and Subject Partisanship . . . 33
1.2 Motivated Reasoning and Trust in Fake News: Perceived Veracity by News Direction and Actual Veracity . . . 34
1.3 Motivated Reasoning Across Topics: Effect of Pro-Party News on Perceived Veracity by Topic . . . 37
1.4 Gender Heterogeneity in Motivated Reasoning Across Topics: Effect of News Direction on Perceived Veracity by Topic and Gender . . . 55

2.1 Histogram of Perceived Veracity of Positive and Negative News . . . 68
2.2 Histogram of Perceived Veracity of True and Fake News About Positivity . . . 68
2.3 Comparing Motivated Reasoning About Positivity to Politics and Performance . 73
2.4 People Systematically Expect Motivated Reasoning About Politics and Performance, But Do Not About Positivity . . . 75
2.5 People Expect Happiness Due to Beliefs About Positivity, Not Just About Politics and Performance . . . 76

3.1 Partisan Differences in Perceived Risk and Social Distancing . . . 82
3.2 Geographic Variation in Social Distancing, Partisanship, COVID-19, and Public Policy . . . 91
3.3 Social Distancing and COVID-19 Prevalence . . . 93
3.4 Partisan Differences in Social Distancing . . . 94
3.5 Partisan Differences in Social Distancing by 2-Digit NAICS Code Industry . . . 97
3.6 Partisan Differences in Social Distancing, Daily . . . 98
3.7 Partisan Differences in Social Distancing, Alternative Measures . . . 99
3.8 Partisan Differences in Beliefs and Actions . . . 100

A.1 Histogram of Perceived Veracity of Pro-Party and Anti-Party News. . . . 114
A.2 Histogram of Perceived Veracity of True News and Fake News on Politicized Topics. . . . 115
A.3 Motivated Reasoning Across Topics: Effect of Pro-Party News on Perceived Veracity by Topic and Partisanship . . . 118
A.4 Bin-Scatter Plot of Expected Performance by Gender and True Performance . . 121
A.5 Round-by-Round Effects of News Direction on Perceived Veracity . . . 157
A.6 Crime Under Obama question page. . . . 164
A.7 Crime Under Obama news assessment page. . . . 167
A.8 Crime Under Obama news assessment page: Second Guess question. . . . 168

C.1 POI Visits in 2019 . . . 189
C.2 Partisan Differences in Social Distancing, 2019 . . . 190
C.3 Partisan Differences in Social Distancing, Robustness . . . 191

C.3 (continued) Partisan Differences in Social Distancing, Robustness. 192
C.3 (continued) Partisan Differences in Social Distancing, Robustness. 193
C.4 Partisan Differences in Social Distancing, Precinct 194
C.5 Partisan Differences in Social Distancing, Precinct 2019 195
C.6 Partisan Differences in Beliefs and Actions: Unweighted 196
C.7 Effect of Incentives on Beliefs 197
C.8 Partisan Differences in Beliefs and Actions: County Fixed Effects 198
C.9 Partisan Differences in Social Distancing with Controls for Beliefs and News . . 199

List of Tables

1.1 Topics and Hypothesized Motives in the Experiment . . . 6
1.2 The Effect of News Direction and Actual Veracity on Perceived Veracity . . . 36
1.3 Changing Guess to Follow Message Given News Direction . . . 39
1.4 Effects of Topic and Partisanship on News Assessment Scores . . . 47
1.5 Performance and Expected Performance by Partisanship . . . 49
1.6 Heterogeneity in the Partisan Direction of Motivated Reasoning: Horse Race Regression . . . 52
1.7 Heterogeneity in the Magnitude of Motivated Reasoning: Horse Race Regression 53

2.1 Topics and Hypothesized Motives in the Experiment . . . 59
2.2 The list of topics and positivity motives; the exact wording of each question is in Section B.2. . . . 59
2.3 The Effect of News Direction and Actual Veracity on Perceived Veracity . . . 69
2.4 Heterogeneity in Positivity-Motivated Reasoning: Horse Race Regression . . . 71

A.1 The Effect of News Direction, Actual Veracity, and Skewed Prior Distributions on Perceived Veracity . . . 119
A.2 The Effect of News Direction, Actual Veracity, and Previous News Directions on Perceived Veracity . . . 120
A.3 Determinants of Willingness-To-Pay . . . 131
A.4 Estimated Motives: By Direction, By Party, and By Prior . . . 132
A.5 The Effect of News Direction and Actual Veracity on Perceived Veracity: Replication . . . 143
A.6 Changing Guess to Follow Message Given News Direction: Replication . . . 144
A.7 Overprecision and Partisanship: Replication . . . 145
A.8 The Effect of News Direction and Actual Veracity on Perceived Veracity: Second-Guess Treatment . . . 147
A.9 The Effect of News Direction and Actual Veracity on Perceived Veracity: Willingness-to-Pay Treatment . . . 148
A.10 The Effect of News Direction and Actual Veracity on Perceived Veracity: Given 50-50 Prior Treatment . . . 149
A.11 The Effect of News Direction and Actual Veracity on Perceived Veracity: Not Given 50-50 Prior Treatment . . . 150
A.12 The Effect of News Direction and Actual Veracity on Perceived Veracity: Including Subjects Who Fail Comprehension Checks . . . 152
A.13 The Effect of News Direction and Actual Veracity on Perceived Veracity: Logit Veracity Assessments . . . 154
A.14 The Effect of News Direction and Actual Veracity on Perceived Veracity: Rounds 1-6 . . . 155
A.15 The Effect of News Direction and Actual Veracity on Perceived Veracity: Rounds 7-12 . . . 156

Acknowledgments

This dissertation would not have been possible without the incredible support of my advisors, family, friends, and colleagues.

My dissertation committee — Matthew Rabin, Alberto Alesina, David Laibson, and Christine Exley — provided invaluable advice and support at every stage of this process. I have been very fortunate to have their guidance over the course of my PhD. I am also grateful to many informal mentors who have given me feedback along the way, including Benjamin Enke, Edward Glaeser, Eric Maskin, Gautam Rao, and many other participants of the behavioral economics and political economy lunch seminars.

I would like to thank Hunt Allcott, Levi Boxell, Jacob Conway, Matthew Gentzkow, and David Yang, my excellent coauthors, for fascinating discussions and work on Chapter 3.

My family has been with me every step of the way, and it is impossible to thank them enough. My parents, Fred and Linda Thaler, have kept me grounded and always reminded me of the important things in life. They have also listened to every bad research idea I've ever had, and for that I apologize. My sister, Emily Thaler, has always been someone I could talk to about anything at any time.

Every experience over the past several years has been a little brighter because of the constant support of my partner, Mattie Toma; I feel very lucky to have had her to share the highs and lows of graduate school with. (Even when she beats me at board games.)

Funding for papers in this dissertation was generously provided by the Harvard Business School Research Fellowship, the Eric M. Mindich Research Fund for the Foundations of Human Behavior, the Bradley Graduate Fellowship, the Stanford Institute for Economic Policy Research (SIEPR), the John S. and James L. Knight Foundation, the Sloan Foundation, the Institute for Humane Studies, and the National Science Foundation.

Last but not least, I would like to thank Pippi, who has always helped with a motivating meow or a reassuring purr. She also assisted with the writing of this section by walking on my keyboard, but this ended up being less helpfuklilsjgklfmjmasfdjsdgfljl;sfdl;lll.

0

Introduction

This dissertation consists of essays on behavioral, experimental, and political economics, with an emphasis on understanding how people make inferences from information. It analyzes the causes and consequences of biased and polarized beliefs. To better understand why people maintain erroneous beliefs, I study how they make inferences from new information. I construct a new experimental design to show how people engage in *motivated reasoning* by distorting their inference in the direction of beliefs they are more motivated to hold. My results show that motivated reasoning is widespread and helps explain persistent belief polarization along political dimensions. I then delve into the limits of motivated reasoning, showing that the bias is *not* a major factor in the absence of self-image concerns. Lastly, I study the consequences of political belief polarization, showing that this polarization manifests itself in public health behavior.

The first chapter, "The 'Fake News' Effect: An Experiment on Motivated Reasoning and Trust in News," provides a model of motivated reasoning and a new experimental design to identify the bias from Bayesian updating. I reject Bayesian updating in favor of politically-driven motivated reasoning on eight topics: immigration, income mobility, racial discrimination, crime, gender-based math ability, climate change, gun laws, and the performance of other subjects. My measure of motivated reasoning can explain why people hold biased and politically-polarized beliefs about these issues. The bias additionally leads to overconfidence and overprecision, particularly among partisans who hold stronger politically-motivated beliefs.

The second chapter, "The Limits of Motivated Reasoning When Self Image Is Not at Stake," follows the framework of the first chapter but instead delves into *what* beliefs people are motivated to hold. I experimentally study whether people engage in motivated reasoning towards "positive" beliefs in the absence of self-image concerns. In particular, I ask whether people motivatedly reason to think that the world is a better place for others on topics like infant mortality, cancer survival rates, and others' happiness. Using the same measure of motivated reasoning as in my first chapter, I starkly find that there is no systematic evidence for positivity- (or negativity-) motivated reasoning, and can rule out modest effects.

The third chapter, "Polarization and Public Health: Partisan Differences in Social Distancing during the Coronavirus Pandemic" considers the public-health implications for political polarization in beliefs and behavior. This paper, coauthored with Hunt Allcott, Levi Boxell, Jacob Conway, Matthew Gentzkow, and David Yang, studies partisan differences in Americans' responses to the COVID-19 pandemic. We develop a simple model of a pandemic response with heterogeneous agents that clarifies the causes and consequences of divergent responses. We use location data from a large sample of smartphones to show that areas with more Republicans engage in less social distancing, controlling for other factors including state policies, population density, and local COVID-19 cases and deaths. We then present new survey evidence of significant gaps between Republicans and Democrats in beliefs about personal risk and the future path of the pandemic.

So far as I can see, all political thinking for years past has been vitiated in the same way. People can foresee the future only when it coincides with their own wishes, and the most grossly obvious facts can be ignored when they are unwelcome.

George Orwell (Partisan Review, 1945)

1

The "Fake News" Effect: An Experiment on Motivated Reasoning and Trust in News

1.1 Introduction

On many topics, people extensively disagree about the answers to factual questions, and their beliefs are often inaccurate in predictable directions. People have differing beliefs about questions related to income mobility, crime rates, and racial discrimination in labor markets; tend to be biased in the direction that is more representative of their political party's stances; and often overestimate their own political knowledge (e.g. Alesina, Stantcheva, and Teso 2018; Flynn, Nyhan, and Reifler 2017; Ortoleva and Snowberg 2015). As shown by Meeuwis et al. (2019) and Gerber and Huber (2009), these beliefs can affect consumer, financial, and political behavior. Given the importance of these issues, why does such bias and belief polarization persist? This paper helps answer this question by analyzing how beliefs change when people receive new information.

After receiving a piece of news, people form their posterior beliefs by incorporating their prior beliefs and their perceived informativeness of the information. If we only observe beliefs at a snapshot in time, two people's disagreement can be consistent with several explanations: for instance, they may have had different priors, they may differently perceive the informativeness of the news, or they may have different inference processes. The first two channels are often relevant in politicized settings. First, Democrats and Republicans often have different preconceived notions, leading to differences in posteriors; this can be consistent with Bayes' rule and with prior-confirming behavioral biases. Second, Democrats and Republicans often consume different news sources, and may find arguments akin to those from MSNBC and from Fox News differentially informative.[1]

This paper studies the third channel: people form different posteriors, even if they have the same priors and receive the same information, because they distort their updating process. When people receive information, they are often reminded of what beliefs they currently hold, and of particular beliefs they find more attractive to hold. In the model of *motivated reasoning* developed in this paper, people distort their updating process in the direction of these particular beliefs (*motives*). The model defines motives to be a function that maps beliefs to the real numbers. Agents make inferences using a modified Bayes' rule, weighting priors and likelihoods as a Bayesian would, but act as if they receive an extra signal that puts more weight on higher-motive states. Motives are often heterogeneous, such as on politicized issues. In such a setting, the model shows how motivated reasoning can lead agents to over-trust news that reinforces their biases, can cause belief polarization, and can lead to miscalibrated and overconfident beliefs.

While there is an intuition in the literature that motivated reasoning plays a role in inference, particularly in political domains, designing an experiment that identifies the bias has been a challenge in domains where people enter the experiment with different beliefs (as discussed in Kahan 2016a and Sunstein et al. 2017). This paper's main experimental contribu-

[1] There is ample evidence consistent with these channels (e.g. Taber and Lodge 2006; Kahan, Hoffman, et al. 2012; Nyhan and Reifler 2010; Nyhan and Reifler 2013; Nyhan, Reifler, and Ubel 2013).

tion is to construct a new design to disentangle motivated reasoning from Bayesian inference in such settings. In the experiment, subjects make inferences about the veracity of messages that are equally likely to tell them that their current median beliefs are biased upward or biased downward. Because subjects report their median beliefs, they believe that both a source that sends a truthful message and a source that sends a false message are equally likely to send a "greater than" or a "less than" message. Therefore, there is nothing for a Bayesian to infer. However, motivated reasoners will trust messages more if the messages better align with their motivated beliefs.[2]

The context of this experiment concerns Americans' assessments of the veracity of news about economic, political, and social issues. The news veracity setting is useful for identifying motivated reasoning, and is particularly salient in the United States today. According to Gallup (2018a) and the Knight Foundation (2018), fewer than one in four Americans has confidence in the news media, a sizable majority believe that "fake news is a threat to democracy," and less than half can even "think of a news source that reports the news objectively."[3]

I run a within-subject experiment on Amazon Mechanical Turk with approximately 1,000 Americans. Subjects are given factual questions about nine *politicized* topics (on economic, political, and social issues), three neutral topics, and one question about on own performance in the experiment. The list of topics and pre-hypothesized motives is in Table 1.1.

As previewed above, the experimental design has two main steps. First, subjects are given a variety of factual questions with numerical answers. On each question, the medians of subjects' belief distributions are elicited, so that subjects think the true answer is equally likely to be above or below their medians. Second, subjects are given one binary message that is chosen randomly from either a True News source or a Fake News source; the message tells them whether the answer was above or below their median. If the message is from True News, it is always accurate. If the message is from Fake News, it is always inaccurate. Subjects are not

[2]Because there is nothing to infer, the result cannot be due to general over- or under-weighting of priors or likelihoods. As discussed later, this design also shows that motivated reasoning is hard to reconcile with utility-maximizing beliefs like those in Brunnermeier and Parker (2005); Benabou and Tirole (2011); or Mobius et al. (2014).

[3]Among those who *can* name an objective news source, there is not a single outlet that both 5 percent of Democrats and 5 percent of Republicans think of as objective.

Table 1.1: Topics and Hypothesized Motives in the Experiment

Topic	**Pro-Democrat Motives**	**Pro-Republican Motives**
US crime	Got better under Obama	Got worse under Obama
Upward mobility	Low in US after tax cuts	High in US after tax cuts
Racial discrimination	Severe in labor market	Not severe in labor market
Gender	Girls better at math	Boys better at math
Refugees	Decreased violent crime	Increased violent crime
Climate change	Scientific consensus	No scientific consensus
Gun reform	Decreased homicides	Didn't decrease homicides
Media bias	Media not dominated by Dems	Media is dominated by Dems
Party performance	Higher for Dems over Reps	Higher for Reps over Dems
Own performance	Higher for self over others	Higher for self over others
Random number	Neutral	Neutral
Latitude of US	Neutral	Neutral
Longitude of US	Neutral	Neutral

Notes: The first nine topics are called politicized topics. On the computer, each topic is a hyperlink that links to the exact question wording in Section A.3.

told which source the message came from; instead, they make inferences about the source's veracity from the message.

Since messages relate the true answer to subjects' subjective median, Bayesian subjects would believe that it is equally likely for either source to report either message. That is, the subjective likelihood that a True News source would report that the answer is greater than the median is 1/2, and the subjective likelihood that a Fake News source would report that the answer is greater than the median is also equal to 1/2. Therefore, a "greater than" message is uninformative about the veracity of the news source to a Bayesian. Likewise, a "less than" message is also uninformative to a Bayesian.

On the other hand, a subject who engages in motivated reasoning will trust the news more if it sends a message that supports what he is more motivated to believe. The main hypothesis in this paper is that the direction of motivated beliefs is driven by political preferences on these topics. In other words, it predicts that people will assess messages that align with beliefs of their political party (*Pro-Party news*) to be more truthful, while assessing messages that misalign (*Anti-Party news*) to be less truthful.

The main result of the experiment is that Bayesian updating is rejected in favor of politically-motivated reasoning on these topics. While a Bayesian would believe that Pro-Party and Anti-Party news are equally likely to be True News on the politicized topics, subjects in the experiment believe that Pro-Party messages are 9 percentage points (standard error (s.e.) 0.7 percentage points) more likely than Anti-Party messages to come from the True News source. This gap increases in the partisanship of the subject, and assessments on neutral topics lie in between Pro-Party news and Anti-Party news veracity assessments. This design allows for enough statistical power to test motivated reasoning for each topic individually; for eight of the nine politized topics, the main effect is significant at the $p = 0.001$ level. On each of these topics, this experiment provides novel evidence for motivated reasoning; unlike in prior studies, these results are not confounded by alternative explanations involving Bayesian updating or prior-confirming biases.[4] In addition, there is evidence for performance-driven motivated rea-

[4] Papers that find asymmetric responses to information on these topics include: Taber and Lodge (2006) [gun laws]; Alesina, Stantcheva, and Teso (2018) [upward mobility]; Cappelen, Haaland, and

soning on a question asking subjects to rate their performance on the experiment relative to others. These main results are robust to a host of alternative explanations.[5]

Secondly, results support the hypothesis that the error in subjects' current beliefs is due in part to motivated reasoning. The theory predicts that, since people who motivatedly reason about an issue will form directionally-biased beliefs on average, we can partly infer what people's motivated beliefs are by looking at their current beliefs. That is, under this hypothesis, *error* predicts *motive*. In the experiment, this hypothesis means that people will give higher veracity assessments to news that (falsely) reinforces their error compared to news that (truthfully) brings them closer to the correct answer. Indeed, in the experiment, people trust the error-reinforcing Fake News more than the error-correcting True News, and only on topics where motivated reasoning is expected to play a role. This gap persists when controlling for whether the news is Pro-Party or Anti-Party.

Thirdly, the theory explains how motivated reasoning can lead to other behavioral biases. Motivated reasoning may provide a link between overprecision and partisanship, a relationship documented in Ortoleva and Snowberg (2015). In an environment with normally-distributed priors and likelihoods, people's belief distributions are more likely to be miscalibrated when they have stronger motives, and their 50-percent confidence intervals will contain the true answer less than 50 percent of the time. This result is borne out in the experiment: subjects' intervals are significantly overprecise on politicized and performance questions, while they are not overprecise on neutral topics. The model also discusses how motivated reasoning can lead

Tungodden (2018) [responses to taxation]; Haaland and Roth (2019) [racial labor market discrimination]; Sarsons (2017), Kunda and Sinclair (1999), and Iyengar and Westwood (2015) [gender and performance]; Alesina, Miano, and Stantcheva (2018), Haaland and Roth (2018), and Druckman, Peterson, and Slothuus (2013) [impact of immigrants]; Nyhan and Reifler (2013) and Nyhan, Reifler, and Ubel (2013) [perceptions of elected officials]; and Sunstein et al. (2017) [climate change]. Many results from these papers can be explained by motivated reasoning.

[5]The main predictions are identical if subjects mistakenly believe Fake News sends random messages instead of always-false messages, and results are not driven by subjects who have skewed distributions and may misreport their median. Importantly, there is also evidence for asymmetric updating regarding their beliefs about the initial question. Subjects are significantly more likely to change their beliefs in the direction of the message if the news is Pro-Party than if the news is Anti-Party, and this asymmetry is entirely captured by the differences in news veracity assessments of those sources. This suggests that the results cannot be explained by expressive preferences or mistakenly treating Fake News as Anti-Party news.

to both underperformance and overconfidence, since the bias leads to erroneously extreme beliefs that news sources are either very likely to be True News or are very likely to be Fake News. Indeed, in the experiment subjects perform significantly worse than if they had always said there was a 50 percent chance of the source being True News or Fake News.

Motivated reasoning not only affects how people trust or distrust news, but also impacts how people change their beliefs about the politicized topics themselves, and leads to belief polarization. Subjects are significantly more likely to revise their beliefs away from the population mean than towards it. This form of polarization is entirely accounted for by the news veracity assessments, suggesting that subjects are polarizing because of their misinference of the veracity of Pro-Party and Anti-Party news; it also shows that informational content is *not* a necessary condition for polarization.[6] Politically-motivated reasoning helps reconcile the notions that the ideological polarization of beliefs may be high, even if the ideological polarization of information acquisition is modest (the latter shown by Gentzkow and Shapiro 2011).[7]

There is no other sizable demographic heterogeneity in motivated reasoning on the politicized topics, neither in direction or magnitude, once party preference is controlled for.[8] Differences in treatment effects across subjects of different demographic groups are statistically indistinguishable from zero, and estimates are precise enough to rule out even modest effect sizes. This result suggests that motivated reasoning is homogeneous across demographic groups — and that even on many issues that are explicitly about particular groups, such as gender and math ability, racial discrimination, and income mobility — motivated beliefs are principally driven by politics.

However, when subjects are asked about their own performance, there is substantial heterogeneity in motivated reasoning by gender. Men motivatedly reason to believe they outper-

[6]There is a related literature that discusses the relationship between trust in news and political partisanship (Nisbet, Cooper, and Garrett 2015; Levendusky 2013; Druckman, Levendusky, and McLain 2018).

[7]Gentzkow and Shapiro (2006) and Gentzkow, Wong, and Zhang (2018) provide alternative theoretical explanations with Bayesian agents who have different priors, but this experiment's results are not predicted by their models.

[8]Demographics include race, gender, income, age, education, religion, and whether one's home state voted for Trump or Clinton in the 2016 presidential election.

formed others, and women do not motivatedly reason in either direction on average.[9] Motivated reasoning can help explain the gender gap in overconfidence, and more broadly suggests that politically-motivated reasoning may be a more universal phenomenon than performance-motivated reasoning in this context.

Finally, this paper contributes methodologically to the growing experimental literature on the identification of motivated reasoning. As summarized by Daniel Benjamin (2019), the current experimental evidence for motivated reasoning has been mixed: Mobius et al. (2014); Eil and Rao (2011); and Charness and Dave (2017) find that people update more from ego-positive news than ego-negative news, while Ertac (2011); Kuhnen (2014); and Coutts (2018) find the opposite.[10] The design for these papers typically involves giving subjects informative signals and testing for asymmetric updating from "Good" and "Bad" news, and thus requires noise-inducing strategies to disentangle motivated biases from non-motivated biases such as under-inference from information and prior-confirming biases. My design aims to better isolate the motivated reasoning channel by constructing an environment in which misweighting priors and likelihoods plays no role, as messages are uninformative about source veracity. As such, statistical power is large, results are precise, and the design can be portably used to test motivated reasoning on a wide variety of topics.

The rest of the paper proceeds as follows: Section 1.2 develops the model of motivated reasoning, generating testable predictions. Section 1.3 introduces the experimental design and hypotheses corresponding to these predictions. Section 1.4 discusses further details of the experiment and data.

Section 1.5 analyzes experimental results. Section 1.5.2 provides results about news veracity assessments in support of the main motivated reasoning hypotheses. Section 1.5.3 shows that

[9] This relates to results found in Coffman, Collis, and Kulkarni (2019), which was run contemporaneously and also finds differences in updating by gender.

[10] It is worth noting that there is more consistent evidence for choice-based implications of motivated beliefs. This includes information avoidance and moral wiggle room (Oster, Shoulson, and Dorsey 2013; Dana, Weber, and Kuang 2007; Gino, Norton, and Weber 2016), and both risk- and ambiguity-driven distortions (Exley 2015; Haisley and Weber 2010). Yet in the setting in this paper, I do not see evidence for information avoidance: In Section A.2.2 I show that subjects are willing to pay positive amounts for information, and pay similar amounts for information about both motivated and neutral states.

motivated reasoning and belief polarization occur about the original question. Section 1.5.4 goes through several categories of robustness checks. Section 1.5.5 relates motivated reasoning to beliefs and overprecision. Section 1.5.6 relates motivated reasoning to underperformance and overconfidence. Section 1.6.1 discusses treatment effect heterogeneity. Section 1.6.2 discusses gender, overconfidence, and performance-driven motivated reasoning.

Section 1.7 concludes and proposes directions for future work. Section A.1 provides a proof as well as several tables and figures that are omitted from the main text. Section A.2 considers a version of the motivated reasoning model in which subjects form posteriors with noise, and the strength of the motivated-reasoning signal is equal to the standard deviation of this noise term. It then discusses results from a willingness-to-pay treatment consistent with the extended model's predictions, and structurally estimates this model. Section A.3 lists the exact questions, answers, and sources that subjects see. Section A.4 provides results from a preregistered replication of all the main results and many of the secondary results. The online appendices include additional robustness checks, as well as the entire experiment flow with screenshots of each page.

1.2 Model and Predictions

This section introduces and develops a model of motivated reasoning in which agents distort their updating process in the direction of their motivated beliefs when they receive information. The model predicts that people will over-trust news that supports their motivated beliefs compared to a Bayesian, and that we can infer what people are motivated to believe from the directional error in their *current* beliefs. In the political context, this implies that both current beliefs and strength of party preference affect the bias in information processing. It also generates secondary predictions under additional functional form assumptions, showing how motivated reasoning can lead to belief polarization, overprecision, underperformance, and overconfidence.

1.2.1 A Model of Motivated Reasoning

Motivated reasoning posits that agents distort their updating process to put higher likelihood on events that they are more motivated to believe. In this paper, we will often study agents who are motivated to hold beliefs that better support their preferred political party's stances. For example, we will posit that Republicans are politically motivated to believe that murder and manslaughter rates increased during the presidency of Barack Obama, and that Democrats are politically motivated to believe that rates decreased. I will define motivated reasoning by formalizing and extending the framework of Kahan (2016a) in which agents update from information using a modified Bayes' rule. They act as if they put appropriate weights on their prior and the signal likelihood, but receive an additional signal that puts more weight on beliefs that they are more motivated to hold.

To formalize, suppose that agents are inferring about the probability that an event is true (T) or false ($\neg T$), and have prior $\mathbb{P}(T)$. We compare inference from a Bayesian agent (she) to a motivated-reasoning agent (he) when they receive the same signal $x \in X$ about the probability that the event is T.[11] The Bayesian sets her posterior to be proportional to her prior times the likelihood of the signal:

$$\underbrace{\mathbb{P}(T|x)}_{\text{posterior}} \propto \underbrace{\mathbb{P}(T)}_{\text{prior}} \cdot \underbrace{\mathbb{P}(x|T)}_{\text{likelihood}}$$

Taking log odds ratios of both sides gives the Bayesian logit updating process:

$$\text{logit } \mathbb{P}(T|x) = \text{logit } \mathbb{P}(T) + \log\left(\frac{\mathbb{P}(x|T)}{\mathbb{P}(x|\neg T)}\right), \tag{1.1}$$

The motivated reasoner updates similarly, but he incorporates his prior, likelihood, and a motivated reasoning term:

$$\underbrace{\mathbb{P}(T|x)}_{\text{posterior}} \propto \underbrace{\mathbb{P}(T)}_{\text{prior}} \cdot \underbrace{\mathbb{P}(x|T)}_{\text{likelihood}} \cdot \underbrace{M(T)^{\varphi(x)}}_{\text{mot. reasoning}},$$

where $M(T) : \{T, \neg T\} \to \mathbb{R}_+$.[12] Define $m(T) \equiv \log M(T)$ and take log odds ratios to get the

[11] This can be straightforwardly generalized to any discrete state space of events $\{E_1, E_2, \ldots\}$, where agents infer about the probability of events E_1 versus $\neg E_1$, E_2 versus $\neg E_2$, ...

[12] Note that there is also a change in the proportionality constant between Bayes and motivated rea-

motivated-reasoning logit updating process, which will be central to the rest of this section:

$$\text{logit } \mathbb{P}(T|x) = \text{logit } \mathbb{P}(T) + \log\left(\frac{\mathbb{P}(x|T)}{\mathbb{P}(x|\neg T)}\right) + \varphi(x)(m(T) - m(\neg T)). \tag{1.2}$$

The motivated reasoner acts as if he receives both the actual signal (x) and a signal whose relative likelihood corresponds to how much he is motivated to believe the state is T. $m(T) : \{T, \neg T\} \to \mathbb{R}$ is denoted the **motive** function.

We assume that the motive function does not depend on the signal structure. Motives may also be indirect; for instance, an agent may be motivated to believe that a news source is truthful because it reports something in support of her political party. It will also be useful to treat the motive function cardinally in order to study distributions of beliefs. That is, m can be thought of as an *expected* motive function to mirror the standard expected utility function u.

The agent weights the motive signal by parameter $\varphi(x) \geq 0$, called **susceptibility**. When $\varphi(x) = 0$, the agent is Bayesian; when $\varphi(x) > 0$, the agent motivatedly reasons. φ may be a function of signal x and the *perceived informativeness* of x, but does not depend on m.[13]

Closing the model requires additional assumptions on $\varphi(x)$. In the main text of this paper, we will not probe further and instead focus on one particular type of signal structure. For further discussion of a specific definition of φ that depends on the noisiness of the updating process, see Section A.2. Experimentally, this paper studies one type of signal structure, but a future experiment could expand the space of signals to identify perceptions of informativeness by exogenously varying φ.

soning, but this is not a function of T. A similar definition arises for a continuous state ω. Bayes rule sets $f(\omega|z) \propto p(z|\omega) \cdot f(\omega)$, and motivated reasoning sets $f(\omega|z) \propto p(z|\omega) \cdot f(\omega) \cdot m(\omega)^{\varphi(z)}$.

[13]This can be different from the *actual* informativeness of x in important ways. The experiment shows an environment in which signals are uninformative, but are perceived as informative, and still lead to motivated reasoning. Within the class of uninformative signals, there is heterogeneity in perceptions, and this can drive susceptibility.

1.2.2 Identifying Motivated Reasoning

We now use the above framework to identify φ when we assume something about people's motives. We consider an environment in which priors are fixed and Bayesians do not infer anything, but motivated reasoning can play a role.

Consider an agent with prior $F(\theta)$ about a state in Θ. Denote by $\mu \equiv F^{-1}(1/2)$ the median of $F(\theta)$. For simplicity, we assume that F has no atom at μ and that $\mathbb{P}(\mu = \theta) = 0$. That is, the agent believes that the answer has probability zero of being exactly equal to μ, and the true probability is indeed zero.

To preview the experimental design that is developed in Section 1.3, suppose the agent now receives a message from one of two news sources, True News (TN) or Fake News (FN), and does not know which. Both news sources send a binary message $x^{TN}, x^{FN} \in \{G, L\}$ that compares θ to μ. G says that θ is *greater than* μ and L says that θ is *less than* μ. TN always sends the "true" message and FN always sends the "fake" message:

	$\theta > \mu$	$\theta < \mu$
True News sends	G	L
Fake News sends	L	G

The agent has a prior about the news source $p \equiv \mathbb{P}(TN)$ that does not depend on θ, and infers about $\mathbb{P}(TN)$ given the message received. The agent receives quadratic utility from stating probability a:

$$u(a|TN) = 1 - (1-a)^2 \text{ and}$$
$$u(a|FN) = 1 - a^2,$$

such that she maximizes utility by stating her subjective belief a.

We can now look at how a Bayesian and a motivated reasoner update their beliefs about the news source. Given message G, the Bayesian updates according to Equation (1.1):

$$\begin{aligned}\text{logit } a|G = \text{logit } \mathbb{P}(TN|G) &= \text{logit } \mathbb{P}(TN) + \log\left(\frac{\mathbb{P}(G|TN)}{\mathbb{P}(G|FN)}\right) \\ &= \text{logit } p + \log\left(\frac{\mathbb{P}(\theta > \mu)}{\mathbb{P}(\theta < \mu)}\right) \\ &= \text{logit } p.\end{aligned}$$

$$\text{Therefore: } a|G = p = a|L.$$

Since the Bayesian thinks that both messages are equally likely ex ante, she doesn't update in any direction. In the experiment, this will be the main null hypothesis, and the hypothesis for unmotivated topics: $a|G = a|L$.

However, the motivated reasoner updates according to Equation (1.2):

$$\begin{aligned}\text{logit } a|G &= \text{logit } \mathbb{P}(TN) + \log\left(\frac{\mathbb{P}(G|TN)}{\mathbb{P}(G|FN)}\right) + \varphi\left(m(\theta|\theta > \mu) - m(\theta|\theta < \mu)\right) \\ &= \text{logit } p + \varphi\left(m(\theta|\theta > \mu) - m(\theta|\theta < \mu)\right).\end{aligned}$$

This implies the following:

Fact 1 (Identifying motivated reasoning using news veracity assessments)

The procedure above identifies motivated reasoning from Bayesian updating:

- *For a Bayesian ($\varphi = 0$), $a|G = a|L$.*
- *For a motivated reasoner ($\varphi > 0$), $a|G > a|L \iff m(\theta|\theta > \mu) > m(\theta|\theta < \mu)$.*

More specifically, this design identifies whether agents have greater expected motive for believing that the true state is above their median belief μ or for believing that the true state is below μ.

In this paper, states are real numbers and motives are typically assumed to be monotonic in the state, so that sign$\left(\frac{\partial m}{\partial \theta}\right)$ does not depend on θ. For simplicity, we will sometimes make the further restriction that motives are *linear*. In the linear case, $m(\theta) = m \cdot \theta$, so that the prediction does not rely on the distribution $F(\theta)$: that is, $a|G > a|L$ if and only if $m \cdot \varphi(x) > 0$.[14]

Predictions involve jointly hypothesizing that agents motivatedly reason and hypothesizing something about their motive function. In the context of the experiment, the main hypothesis will be that observables (such as political preferences) predict $m(\theta|\theta > \mu) - m(\theta|\theta < \mu)$, and therefore predict $\text{logit}(a|G) - \text{logit}(a|L)$.

It is worth noting that the null hypothesis is the same for many *non-Bayesian* models of inference. Consider the following class of updating rules defined by general misweighting of priors or likelihoods:

$$\begin{aligned}
\text{logit } a|G &= \zeta \text{ logit } \mathbb{P}(TN) + \kappa \log\left(\frac{\mathbb{P}(G|TN)}{\mathbb{P}(G|FN)}\right) \\
&= \zeta \text{ logit } p + \kappa \cdot 0, \text{ and} \\
\text{logit } a|L &= \zeta \text{ logit } p \text{ as well.}
\end{aligned}$$

This class of updating rules includes a form of prior-confirming bias ($\zeta > 1$), conservatism ($\kappa < 1$), base-rate neglect ($\zeta < 1$), and over-inference ($\kappa > 1$). In all these cases, there is no differential updating from G and L. These biases may also affect inference in many settings in which motivated reasoning plays a role. In such cases, the motivated reasoning term can simply be separately added to other models.

1.2.3 Inferring Motives from Beliefs

When motives are unobservable, an experimenter can learn about agents' motives by looking at their initial beliefs μ. Conceptually, an agent's error in beliefs can be partly explained by motivated reasoning, and therefore the direction of the error predicts the direction of the mo-

[14] Strictly monotonic motives posit that people are more motivated to hold extreme beliefs. An example of a more "moderate" motive function is *quadratic loss*: $m(\theta) = -m_{\text{quad}}(\theta^* - \theta)^2$, where $m_{\text{quad}} > 0$ so that θ^* is the highest-motive belief. One parametrization sets θ^* equal to μ; this motive suggests a similar psychology to prior-confirming bias. Experimentally, the quadratic term could be identified by giving people binary messages that say that the answer is within or outside their 50-percent confidence interval.

tive function. A motivated reasoner with an increasing motive function will be more likely to hold a belief that $\mu > \theta$, and a motivated reasoner with a decreasing motive function will be more likely to hold a belief that $\mu < \theta$, if they receive a signal drawn from the same distribution. This implies that an agent who believes $\mu > \theta$ is more likely to have an increasing motive function than is an agent who believes $\mu < \theta$.

When the two agents then make news assessments using the structure above, agents will trust news that *reinforces* the error in their beliefs more than news that *mitigates* the error. This occurs even though signals are designed exactly so that their interpretation is distinct from μ.

More formally, there is a state $\theta \in \mathbb{R}$. Consider a Bayesian (she) and a motivated reasoner (he) with the same prior $\theta \sim F_\theta$ and who receive a public signal $z \sim F_z$. We assume that the motivated reasoner has motive $m(\theta')$ that is strictly monotonic in θ', and $\varphi(z, F_z) > 0$. We also assume that the signal leads the Bayesian's posterior median μ_B to take values close to θ with positive probability, but $\mathbb{P}(\mu_B = \theta) = 0$. That is, for all $\delta > 0$, there exists some $\delta' > 0$ such that $\mathbb{P}(|\mu_B - \theta| < \delta) > \delta'$.

Without loss of generality, consider a motivated reasoner who has $m(\theta')$ strictly increasing in θ'. Since the log-likelihood of the motive is strictly increasing, his posterior distribution first-order stochastically dominates the Bayesian's posterior distribution. In addition, for every such motive function, there exists a δ such that for all signals leading to the Bayesian having a posterior median $\mu_B \in (\theta - \delta, \theta)$, the motivated reasoner has posterior median $\mu_M > \theta$. Since there is a probability of at least $\delta' > 0$ of such a signal, this high-θ-motivated reasoner is *strictly* more likely than the Bayesian to state $\mu > \theta$. By the same argument, a low-θ motivated reasoner is strictly less likely than the Bayesian to state $\mu > \theta$.

Now suppose that μ is observable and the true θ is known, but z and $m(\theta)$ are unobservable. If some people have monotonically-increasing motives and others have monotonically-decreasing motives, then:

$$\mathbb{P}(m(\theta) \text{ increasing} \mid \mu > \theta) > \mathbb{P}(m(\theta) \text{ increasing} \mid \mu < \theta).$$

If we look at how agents respond to the procedure above to a new message G or L, this implies that $\mathbb{E}[a|G, \mu > \theta] > \mathbb{E}[a|G, \mu < \theta]$ and $\mathbb{E}[a|L, \mu > \theta] < \mathbb{E}[a|L, \mu < \theta]$ when motives are heterogeneous.

Now, recall that message G says that $\theta > \mu$ and L says $\theta < \mu$. Since G and L are equally likely, the prediction is that subjects trust error-reinforcing messages more than error-mitigating messages when motivated reasoning plays a role.

In this design, error-mitigating messages are exactly True News and error-reinforcing messages are exactly Fake News. Therefore, agents give higher assessments to Fake News than True News, with and without controlling for observable party preference:

Fact 2 (Motivated reasoning leads to over-trusting Fake News, under-trusting True News) *Suppose that agents motivatedly reason with a strictly monotonic motive. Then:*

- $a|$*Fake News* $> a|$*True News.*
- $a|$*Fake News; Pro-Party news* $\geq a|$*True News; Pro-Party news.*
- $a|$*Fake News; Anti-Party news* $\geq a|$*True News; Anti-Party news.*

Suppose also that the sign of the slope of the motive function is heterogeneous within party. That is, the probability of an agent having $\frac{\partial m(\theta)}{\partial \theta} > 0$ *is strictly between 0 and 1, conditional on the agent's party. Then:*

- $a|$*Fake News; Pro-Party news* $> a|$*True News; Pro-Party news.*
- $a|$*Fake News; Anti-Party news* $> a|$*True News; Anti-Party news.*

The stark result that motivated reasoners will trust Fake News more than True News is particular to the uninformativeness of the messages. However, the prediction that agents will trust Fake News more than a *Bayesian* will is quite general, only relying on unobservable inputs into current beliefs. It is also worth noting that this prediction only holds for motivated states, psychologically differentiating this theory from unmotivated explanations of over-trusting Fake News (such as a general prior-confirming bias). Practically, it suggests that excessive trust in disinformation will be more prominent when people hold stronger motivated beliefs.

1.2.4 Motivated Reasoning, Overprecision, and Overconfidence

There are two ways in which motivated reasoners may have excess confidence in their beliefs compared to Bayesians in this setting. First, motivated reasoners may form miscalibrated confidence intervals about the initial questions: *overprecision.* Second, motivated reasoners may form more extreme beliefs about the veracity of the news sources, leading them to overestimate their news veracity assessment accuracy: *overconfidence.*

These consequences of motivated reasoning require functional form assumptions. Unlike the previous subsection, we now suppose that agents have a normally-distributed prior $\theta \sim \mathcal{N}(\mu_0, 1/\tau_0^2)$, and that agents receive a noisy signal $z = \theta + \epsilon_z$, where $\epsilon_z \sim \mathcal{N}(0, 1/\tau_z^2)$.

Suppose also that motivated reasoners have $m(\theta) = m \cdot \theta$ and $\varphi(z) = \varphi(\tau_z)$. That is, agents have a linear motive and signals only affect susceptibility through their level of precision.[15] In the political context, $|m|$ can be thought of as increasing in political partisanship.

A Bayesian forms the posterior:

$$f(\theta|z) = \mathcal{N}\left(\frac{\tau_0\mu_0 + \tau_z z}{\tau_0 + \tau_z}, \tau_0 + \tau_z\right),$$

and a motivated reasoner forms the posterior:

$$f(\theta|z) = \mathcal{N}\left(\frac{\tau_0\mu_0 + \tau_z z + \varphi(\tau_z) \cdot m}{\tau_0 + \tau_z}, \tau_0 + \tau_z\right).$$

Notably, the two agents have the same posterior variance, but the motivated reasoner's distribution is miscalibrated. Consider their $(1-Q)/2$- and $(1+Q)/2$-quantile beliefs, and call this the *Q-confidence interval.* Then:

Fact 3 (Motivated reasoning and overprecision)

Suppose that a motivated reasoner has normally-distributed priors and receives a signal normally distributed with mean equal to θ, as above. When $\varphi > 0$, the probability that his Q-confidence interval contains θ is equal to Q for $m = 0$ and strictly decreases in $|m|$.

[15] One appealing functional form is $\varphi(z) = \min\{\varphi \cdot \tau_z, \bar{\varphi}\}$. Using this form, the susceptibility of two weak signals is equal to the sum of their precisions, but there is a maximum level of susceptibility.

Since Bayesian updating is equivalent to motivated reasoning with $m = 0$, this says that Bayesians are appropriately precise and motivated reasoners are overprecise. Note that the direction of overprecision relies not just on linear motives, but also on the normal-normal functional form.[16]

Next, we consider underperformance and overconfidence. Recall that this theory posits that agents who motivatedly reason may do so at a cost to their utility.[17] Specifically, motivated-reasoning agents underperform by having lower decision utility from their assessments than Bayesians do. This expected utility decreases as the motive function becomes steeper. However, *anticipated* expected utility often will *increase* in motive steepness, since agents become (erroneously) more confident about their assessments. This discrepancy leads to overconfidence.

Using the quadratic utility from above, agents' assessments lead them to attain utility that is decreasing in $|m|$. This implies that motivated reasoners underperform compared to Bayesians, who update the same way as motivated reasoners who have $m = 0$.

Fact 4 (Motivated reasoning and underperformance)

For all $\varphi > 0$ and priors $p \in (0, 1)$, $\mathbb{E}[u(a; m)]$ decreases in $|m|$.

Though agents with steeper motives will receive lower utility on average, they will *expect* to receive *higher* utility, denoted by $\tilde{\mathbb{E}}$, as long as their priors on news veracity are not too extreme.

Fact 5 (Motivated reasoning and confidence)

For all $\varphi > 0$, $\tilde{\mathbb{E}}[u(a; m)]$ increases in $|m|$ if $p \in [\frac{1}{2} - \frac{\sqrt{3}}{6}, \frac{1}{2} + \frac{\sqrt{3}}{6}] \approx [0.211, 0.789]$.

The proof involves more algebra than insight, so it is relegated to Section A.0.1.

To intuitively understand why this is true, consider a partisan (with a steeper motive) and a moderate (with a less steep motive). The partisan will move her assessments substantially

[16] For instance, suppose that the state space contains two values and a Bayesian infers from a signal that either one value has a $(1 + Q)/2$ likelihood of occurring or the other value has a $(1 + Q)/2$ likelihood of occurring. Then, her confidence interval would contain one point, and a motivated reasoner may have a confidence interval that contains both points.

[17] This is in contrast to models in which people deviate from Bayes' rule because they choose *utility-maximizing beliefs* and strategically self-deceive, as in Brunnermeier and Parker 2005; Benabou and Tirole 2002; and Mobius et al. 2014.

upwards when she receives Pro-Party news and expect to score highly, and she will move her assessments substantially downward when she receives Anti-Party news — and still expect to score highly. The moderate will have more tempered expectations given that his assessments are less extreme. Exceptions can occur when p is close to 0 or 1 and φ is not too large, because when partisans update more from Pro-Party (Anti-Party) news, their posteriors may end up below (above) 1/2.

We can conceptually combine these two results by defining overconfidence as the agent's anticipated expected utility minus her actual expected utility. The implication is that political partisans become more overconfident in the accuracy of their news veracity assessments because of motivated reasoning.

1.3 Experimental Design

1.3.1 Summary, Timeline, and Topics

This section details the experiment — introduced in Section 1.2.2 — that is designed to test predictions of the motivated-reasoning model. This design focuses on how subjects infer the veracity of a message that says that their current median belief is erroneously high or erroneously low.

To fix ideas, consider the following question, taken verbatim from the experiment:

Some people believe that the Obama administration was too soft on crime and that violent crime increased during his presidency, while others believe that President Obama's pushes towards criminal justice reform and reducing incarceration did not increase violent crime.

This question asks how murder and manslaughter rates changed during the Obama administration. In 2008 (before Obama became president), the murder and manslaughter rate was 54 per million Americans.

In 2016 (at the end of Obama's presidency), what was the per-million murder and manslaughter rate?

The main test of motivated reasoning involves three steps:

1. **Beliefs:** Subjects are asked to guess the answers to politicized questions like the one above. Importantly, they are asked and incentivized to guess their median belief (i.e. such that find it equally likely for the answer to be above or below their guess). They are also asked and incentivized for their interquartile range. Screenshots of instruction pages are in the Online Appendix.

2. **News:** Subjects receive a binary message from one of two randomly-chosen news sources: True News and Fake News. The message from True News is always correct, and the message from Fake News is always incorrect. This is the main (within-subject) treatment variation.

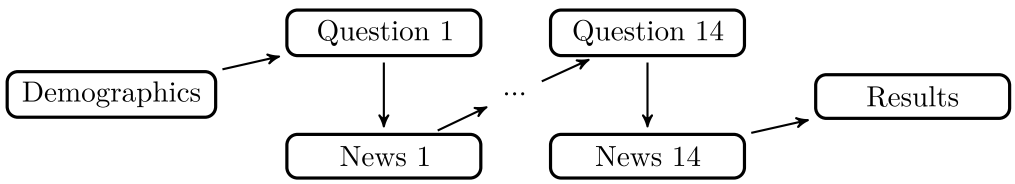
   The message says either "The answer is **greater than** your previous guess of [previous guess]." or "The answer is **less than** your previous guess of [previous guess]." Note that the exact messages are *different* for each subject since subjects have different guesses. These customized messages are designed so that they have the same *subjective* likelihood of occurring.

   For the Crime Under Obama question above, "greater than" corresponds to Pro-Republican News and "less than" to Pro-Democratic News. For subjects who support the Republican Party than the Democratic Party, "greater than" is Pro-Party news and "less than" is Anti-Party news, and vice versa for subjects who support the Democratic Party more.

3. **Assessment:** After receiving the message, subjects assess the probability that the message came from True News using a scale from 0/10 to 10/10, and are incentivized to state their true belief. This news veracity assessment is the main outcome measure. The effect of variation in news direction on veracity assessments is the primary focus for much of this paper. An example of the News / Assessment page is in the Online Appendix.

Since subjects receive messages that compare the answer to their median, a Bayesian would not change her assessment based on the message. Likewise, general over- and under-weighting of priors and likelihoods (such as forms of prior-confirming biases and conservatism) do not predict a treatment effect of message direction on assessment.

Subjects see 14 questions in the experiment. 13 are in Table 1.1 and one is a comprehension check. The experiment has the following general structure:



The Demographics page includes questions about party ratings (which will be used to determine subjects' relative party preference), party affiliation, ideology, gender, age, race and ethnicity, annual income, highest education level, state or territory of residence, religion, nine opinion questions (one each about eight topics in the study and one about Donald Trump's performance), and a 4-item multiple-choice quiz about current events.

The Results page tells subjects what their overall performance was, what their score on each question and assessment was, and the correct answer to each question and assessment. Importantly, subjects are told that they will see this page at the beginning of the experiment, and they are forced to go through it before exiting the study and receiving payment.[18] Being forced to learn the true answers at the end of the experiment substantially limits the scope for strategic self-deception, differentiating motivated reasoning from theories of utility-maximizing inference.

The order of Questions 1-12 is randomized between subjects, but Questions 13 and 14 are the same for each subject. These last two questions are "meta-questions" that rely on previous questions: Question 13 asks subjects about their performance on the first 12 questions relative to 100 other (pilot) subjects, and Question 14 asks about other Democratic (pilot) subjects'

[18]Subjects spend an average of 71 seconds on the Results page, suggesting that they are indeed looking at it. They spend about as long on the Results page as on one Question page and one Info page combined.

performance compared to other Republican (pilot) subjects' performance on Questions 1-12.[19]

Each of the politicized and neutral topics are equally likely to be selected in each round, but the comprehension check is restricted to be between Question 2-11. This restriction is to make sure subjects are still paying attention after the first question, and to make sure the willingness-to-pay treatment (discussed in Section A.2.2), which occurs for Question 12, does not overlap with the comprehension check.

All of the specific question wordings are in Section A.3. Screenshots for every page are in the Online Appendix.

1.3.2 Pages and Scoring Rules

Overall Scoring Rule

At the end of the experiment, subjects earn a show-up fee of $3 and either receive a bonus of an additional $10 or nothing. As will be elaborated below, in each round of the experiment subjects earn between 0-100 "points" based on their performance. These points correspond to the probability that the subject wins the bonus: a score of x points corresponds to an $x/10$ percent chance of winning the bonus.[20]

Questions Page

On question pages, subjects are given the round number (Question x of 14), the topic, the text of the question, and are asked to input three numbers about their initial beliefs:

- *My Guess*: This elicits the median of the subjects' prior distribution.
- *My Lower Bound*: This elicits the 25th percentile of the subjects' prior distribution.
- *My Upper Bound*: This elicits the 75th percentile of the subjects' prior distribution.

The scoring rule for guesses is piecewise linear. Subjects are given $\max\{100 - |c - g|, 0\}$ points for a guess of g when the correct answer is c. Subjects maximize expected points by

[19] Half of subjects are given the Democrats' score and asked to predict the Republicans'; half are given the Republicans' score and asked to predict the Democrats'.

[20] This lottery system is designed to account for risk aversion; directly mapping points to earnings could lead to subjects hedging their guesses. The probability distribution is identical to randomly choosing a round for payment and subsequently playing the lottery based on the points in that round.

stating the median of their belief distribution. They are told the scoring rule in the instructions and given the following message:

It is in your best interest to guess an answer that is in the 'middle' of what you believe is likely. For example, if you think the answer is equally likely to be 10, 40, and 60, you should guess 40.[21]

The scoring rule for bounds is piecewise linear with different slopes. For upper bound ub, subjects are given $\max\{100 - 3(c - ub), 0\}$ points if $c \geq ub$ and $\max\{100 - (ub - c), 0\}$points if $c \leq ub$. For lower bound lb, subjects are given $\max\{100 - (c - lb), 0\}$ points if $c \geq lb$ and $\max\{100 - 3(lb - c), 0\}$ points if $c \leq lb$. Subjects maximize expected points by setting ub to be the 75^{th} percentile and lb to be the 25^{th} percentile of their belief distribution. They are told the scoring rule in the instructions and given the following message:

It is in your best interest to choose a lower bound such that you think it's 3 times more likely to be above the bound than below it, and an upper bound such that it's 3 times more likely to be below the bound than above it. For example, if you think the answer is equally likely to be any number from 100 to 200, you should set a lower bound of 125 and an upper bound of 175.

In addition, subjects are restricted to only give answers such that My Lower Bound $\leq$ My Guess $\leq$ My Upper Bound.

News Assessments Page

After submitting their initial beliefs, subjects are given a second page about the same question. At the top of the page is the exact text of the original question. Below the question is a message relating the correct answer to the number they submit for *My Guess*. This message says either:

"The answer is **greater than** your previous guess of [**My Guess**]." or

"The answer is **less than** your previous guess of [**My Guess**]."

[21] This example is chosen intentionally because the mean and median are different.

Subjects are told that True News *always* tells the truth and Fake News *never* tells the truth, and that sources are iid. The message saying "greater than" or "less than" is the main treatment variation. Below the message, subjects are asked: "Do you think this information is from True News or Fake News?" and can choose one of eleven radio buttons that say "x/10 chance it's True News, (10-x)/10 chance it's Fake News" from each x=0, 1, ..., 10 in increasing order.

The scoring rule for assessments is quadratic. For assessment a, subjects are given $100(1-(1-a)^2)$ points if the source is True News and $100(1-a^2)$ points if it is Fake News. The optimal strategy is to answer with the closest multiple of 0.1 to the true belief. In the instructions, subjects are given a table with the points earned as a function of each assessment and news source.

Occasionally, a subject will correctly guess the answer. If this happens, she skips the news assessment page and moves on to the next question.[22]

Second-Guess Treatment

Half of subjects are in the "Second Guess" treatment. For these subjects, immediately below the news assessment question they are asked an additional question: "After seeing this message and assessing its truthfulness, what is your guess of the answer to the original question?"

Subjects are given the same linear scoring rule as on the initial guess. They are given $\max\{100-|c-g|, 0\}$ points for a guess of g when the correct answer is c. See the Online Appendix for a screenshot of the Crime Under Obama news assessment page that subjects in the Second Guess treatment see, with the second guess part highlighted.

Willingness-to-Pay (WTP) Treatment

The other half of subjects are in the WTP treatment. These subjects see an additional page between Question 12 and News 12, on which they are given instructions and asked to submit a WTP for a message. Results suggest that subjects (erroneously) value the message for the

[22] This is true except for the comprehension check question, where the message says "The answer is **equal / not equal** to your previous guess of [**My Guess**]."

purpose of assessing veracity and that they do not differentially value messages on politicized and neutral topics (indicating naivete about their motivated reasoning). For more detailed instructions and results, see Section A.2.

1.3.3 Hypotheses

This subsection summarizes the predictions from Section 1.2 in the context of the experiment to generate testable hypotheses.

The main hypothesis is that a news veracity assessment will be larger when it leads to a higher motive. This is therefore a joint test that (1) people motivatedly reason, giving higher assessments to news in the direction of higher motives than to news in the direction of lower motives, and (2) the predicted direction of motives is as in Table 1.1. Since we will be mostly considering politicized topics, the degree of partisanship will affect the steepness of the motive function.

Hypothesis 1 (Motivated reasoning with political motives)

- $a|$*Pro-Party news* $>$ $a|$*Anti-Party news.*
- $a|$*Neutral topic news* $\in$ *(*$a|$*Anti-Party news,* $a|$*Pro-Party news).*
- *(*$a|$*Pro-Party news* $-$ $a|$*Anti-Party news) increases in partisanship.*

A similar prediction is that changes in beliefs are differentially affected by the news category, i.e. that people are more likely to follow the message if it is Pro-Party than Anti-Party. This is tested using the Second Guess subsample. The hypothesis is the same for Pro-Own Performance news and Anti-Own Performance news.

Second, we can test whether the direction of the error in subjects' beliefs can be explained in part by their motives. This implies that they will give higher assessments to error-reinforcing news compared to error-mitigating news, as in Fact 2. Recalling that error-reinforcing news is Fake News and error-mitigating news is True News, this leads to the following prediction:

Hypothesis 2 (Motivated reasoning and trust in Fake News)

- $a|$*Fake News* $>$ $a|$*True News on politicized topics, but not on neutral topics.*

- *$a|$\{Fake News, Pro-Party news\} $>$ $a|$\{True News, Pro-Party news\}.*

- *$a|$\{Fake News, Anti-Party news\} $>$ $a|$\{True News, Anti-Party news\}.*

These are the main hypotheses of the experiment.

There are two secondary hypotheses that focus on the Second-Guess treatment. By comparing subjects' first and second guesses, we can replicate the main politically-motivated reasoning prediction and study a form of belief polarization.

First, there is an equivalent hypothesis to Hypothesis 1: controlling for news direction, subjects more frequently adjust their guesses in the direction of their political party preference. Second, by a similar logic to Hypothesis 2, motivated reasoning will lead subjects to be more likely to adjust their guesses towards the population mean than away from it on politicized topics.

Hypothesis 3 (Motivated reasoning and second guesses)
Define Follow Message as the ternary variable that takes value:

- *1 if the message says G and $\mu|G > \mu$ or if it says L and $\mu|L < \mu$;*

- *0 if $\mu|message = \mu$; and*

- *-1 if the message says G and $\mu|G < \mu$ or if it says L and $\mu|L > \mu$.*

Define Polarizing news as news that says G if μ is greater than the population mean guess or L if μ is less than the population mean guess. Define Anti-Polarizing news as the opposite.

- *Follow Message $|$ Pro-Party news $>$ Follow Message $|$ Anti-Party news.*

- *Follow Message $|$ Polarizing news $>$ Follow Message $|$ Anti-Polarizing news on politicized topics.*

Next, we consider the predictions from Section 1.2.4 about the consequences of motivated reasoning. Namely, the model predicts overprecision, underperformance, and overconfidence.

The overprecision hypothesis uses subjects' 50-percent confidence intervals, claiming that motivated reasoning can lead the belief distribution to become miscalibrated.

Hypothesis 4 (Overprecision and partisanship)

- *On politicized questions, subjects' 50-percent confidence intervals contain the correct answer less than 50 percent of the time.*

- *The likelihood that confidence intervals contain the correct answer decreases in partisanship.*

Next, as a direct implication of Hypothesis 2, motivated reasoners will earn fewer points in the experiment. In the political realm, the severity of underperformance is hypothesized to be increasing in partisanship. Additionally, it predicts that partisans will give more certain answers on assessments of politicized news, leading to greater confidence in answers. This is measured by subjects' predictions of their performance relative to others'.

Hypothesis 5 (Underperformance and overconfidence)

- *Average points scored on news assessments is less than the points earned by assessing that the probability of True News is 50 percent.*

- *Average points scored on news assessments is decreasing in partisanship on politicized topics.*

- *Predicted overconfidence — expected performance relative to other subjects minus actual performance — will be increasing in partisanship.*

In the theory, the last part of this hypothesis holds when subject priors on P(True) are between 0.21 and 0.79. In the experiment, nearly all subjects have average assessments within this range, including the ones who are not told explicitly that the probably of True News is 1/2.

The main experiment tests each of these hypotheses. It is worth noting that while each of the primary results were hypothesized ex ante, some of the consequences results involved ex-post analysis. As such, Section A.4 discusses results from a replication exercise conducted one year later on a smaller sample. In the replication, I pre-registered each of the above hypotheses with the exception of overconfidence, due to insufficient statistical power. The replication results are very similar to the main results.

1.4 Data and Experiment Details

The experiment was conducted on June 25, 2018 on Amazon's Mechanical Turk (MTurk) platform. MTurk is an online labor marketplace in which participants choose "Human Intelligence Tasks" to complete. MTurk has become a very popular way to run economic experiments (e.g. Horton, Rand, and Zeckhauser 2011; Kuziemko et al. 2015), and Levay, Freese, and Druckman (2016) find that participants generally tend to have more diverse demographics than students in university laboratories with respect to politics. The experiment was coded using oTree, an open-source software based on the Django web application framework developed by D. Chen, Schonger, and Wickens (2016).

The study was offered to MTurk workers currently living in the United States. 1,387 subjects were recruited and answered at least one question, and 1,300 subjects completed the study. Of these subjects, 987 (76 percent) passed simple attention and comprehension checks, and the rest are dropped from the analyses.[23]

All subjects are asked to rate the Democratic Party and the Republican Party using a scale from 0-100; this scale is modeled after the feeling thermometer used in the American National Election Studies. 627 subjects (64 percent) give a higher rating to the Democratic Party; 270 (27 percent) give a higher rating to the Republican Party; and 90 (9 percent) give identical ratings to both parties.[24] These subjects are labeled as "Pro-Dem," "Pro-Rep," and "Neutral," respectively, and for most analyses the Neutral subjects will be dropped. Results are similar if liberal-conservative ideology or party affiliation is used instead, though many more subjects are then classified as neutral. Results are also similar when weighting by party preference, ideology, or party registration, or when using demographic weights for gender, age categories, race, religion, and location to make the sample representative of the U.S. population.

Treatments were cross-randomized so that 2/3 of subjects would not receive a prior about

[23] In order to pass these checks, subjects needed to perfectly answer the comprehension check question in A.3 (by giving a correct answer, correct bounds, and answering the news assessment with certainty). In addition, many questions had clear maximum and minimum possible answers (such as percentages, between 0 and 100). Subjects were dropped if any of their answers did not lie within these bounds. The Online Appendix shows that main results are robust to inclusion of these subjects.

[24] Levay, Freese, and Druckman (2016) also find that the MTurk subject pool is mostly Democratic.

the veracity of the news source, and 1/3 of subjects would be told that True News and Fake News were equally likely. Independently, 1/2 of subjects would receive the willingness-to-pay treatment and 1/2 would be in the second-guess treatment. Indeed, of the non-Neutral subjects, 66 percent do not receive a prior and 34 percent do; 49 percent are in the WTP treatment and 51 percent are in the second-guess treatment.

Each subject answers 13 questions; there are a total of 11,661 guesses to questions for the 897 non-neutral subjects. There are 11,443 news assessments. The discrepancy between these numbers is due to 143 subjects in the WTP treatment who did not receive a message in one round, and due to there being 75 (0.7 percent) correct guesses.[25] I drop these 218 observations for news assessment analyses. There are 7,902 news assessments on politicized topics, 891 on the question about own performance, and 2,650 on neutral topics.

The balance table for the Pro-Party / Anti-Party treatment is in Section A.0.5. Since this randomization is within subject, treatments are expected to be balanced across demographics. Importantly, the overall shares of Pro-Party and Anti-Party news are not noticeably different. This suggests that there was no differential attrition in the experiment by treatment.

1.5 Results

1.5.1 Raw Data

This subsection shows that the raw data supports the main predictions of the model, and the following subsections show the relevant regressions. To validate that these questions are politicized, Section A.0.4 compares initial guesses by party and finds that there are systematic differences in beliefs between Pro-Rep and Pro-Dem subjects in the direction predicted in Table 1.1. Recall Hypothesis 1 and Hypothesis 2: subjects will trust Pro-Party news more than Anti-Party news; this gap will be larger for partisans than moderates; Neutral news will lie in between; and error-accentuating Fake News will be trusted more than error-mitigating True

[25] The low frequency of correct guesses is an indicator that the vast majority of subjects were not looking up the answers. It is also a sign that the model's assumption of an atomless belief distribution is reasonable.

News on politicized topics.

The mean assessment of Pro-Party news is 62.0 percent (s.e. 0.5 percent), the mean assessment of Neutral news is 57.9 percent (s.e. 0.6 percent), and the mean assessment of Anti-Party news is 52.9 percent (s.e. 0.5 percent).[26] The difference between every pair of these is highly significant ($p < 0.001$ each).[27]

In support of Hypothesis 1, Figure 1.1 shows the subject-demeaned assessments by news direction (Pro-Party; Anti-Party; Neutral) and subject type (Partisan and Moderate, as defined by the absolute difference in party ratings). Subjects indeed give higher average assessments to Pro-Party than to Neutral news, and higher to Neutral than to Anti-Party news, and these differences are larger for partisans. Appendix Figure A.1 shows the empirical distribution of assessments for Pro-Party and Anti-Party news on politicized topics. The empirical distribution of Pro-Party assessments first-order stochastically dominates the distribution of Anti-Party assessments.

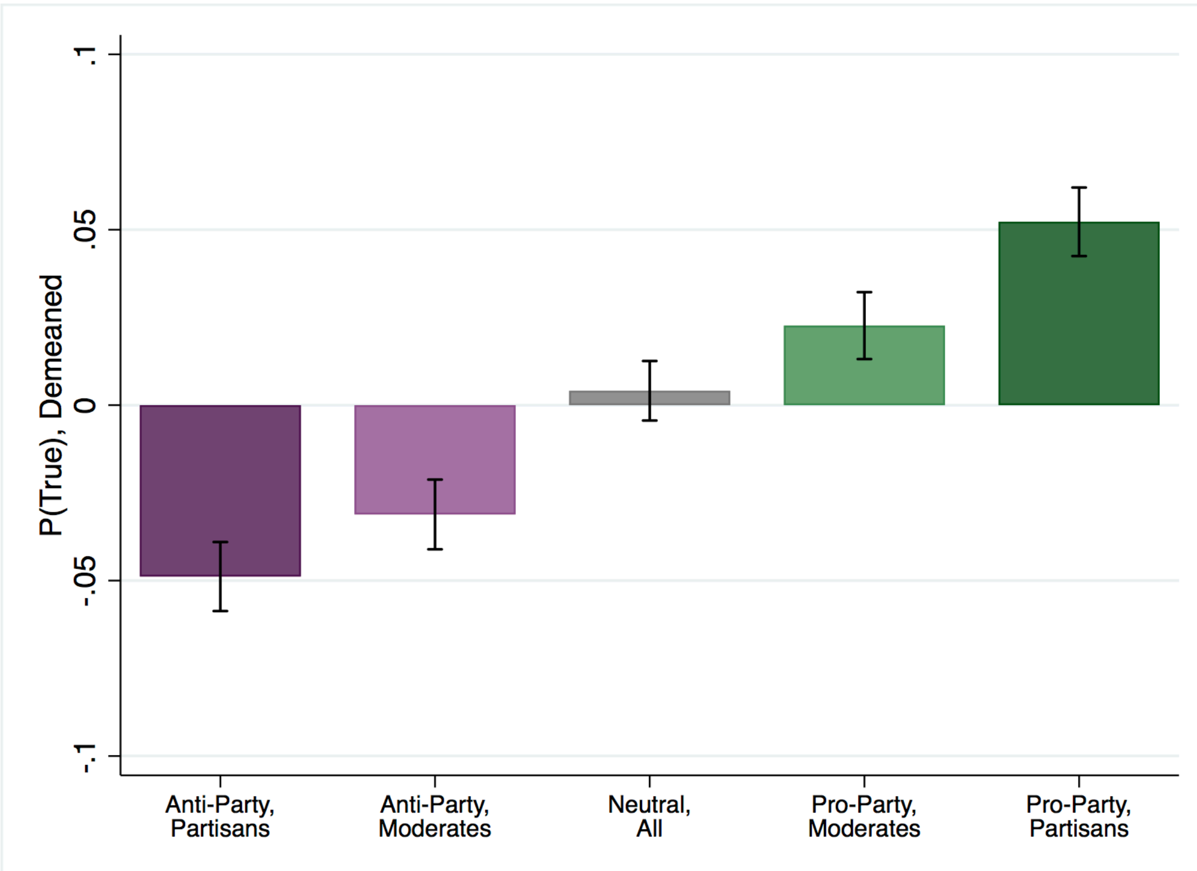
In support of Hypothesis 2, Figure 1.2 shows the subject-demeaned assessments by news direction (Pro-Party; Anti-Party; Neutral) and news veracity (True News; Fake News). Subjects indeed give higher assessments to Fake News than to True News on politicized topics, but they do not on neutral topics.[28] Similar results hold if we look at where subjects' initial guesses lie compared to the median subject instead of compared to the truth. Appendix Figure A.2 shows the empirical distribution of assessments for True News and Fake News on politicized questions. The empirical distribution of Fake News assessments first-order stochastically dominates the distribution of True News assessments.

[26] All standard errors are clustered at the individual level.

[27] All these percentages are significantly greater than 50, even for subjects who are given a prior that True News and Fake News are equally likely. There are two potential explanations that are beyond the scope of this paper. First, perhaps subjects ignore the stated prior and set their own prior around 58 percent. Second, and more suggestively, subjects may motivatedly reason to trust what they are told. Further work can explore this latter channel.

[28] If anything, assessments are higher for neutral True News than neutral Fake News. This suggests that reflecting on the question again may lead to adjusting estimates *towards* the truth in the absence of motivated beliefs.

Figure 1.1: Politically-Motivated Reasoning: Perceived Veracity by News Direction and Subject Partisanship



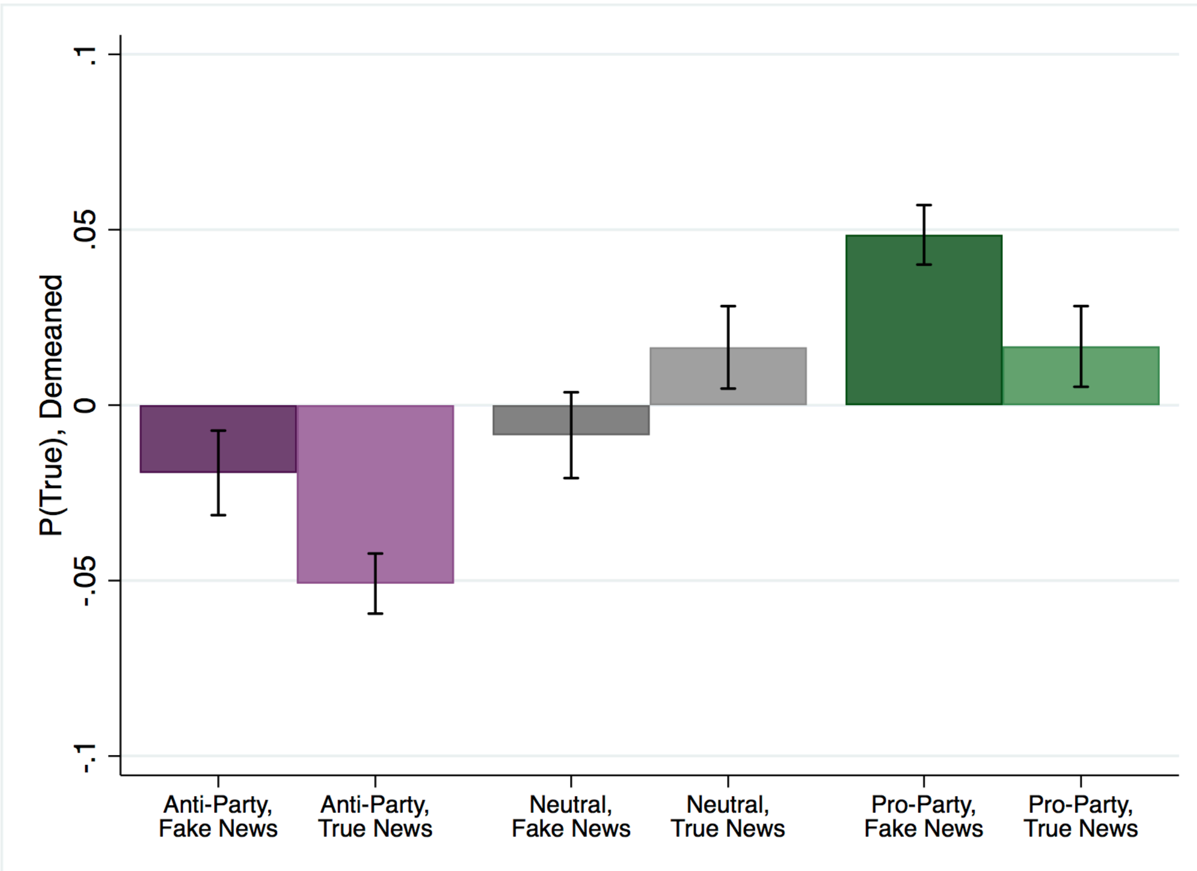
Notes: The y-axis is stated P(True), demeaned at the subject level. News on partisan topics is classified as Pro-Party (Anti-Party) if it is more (less) representative of the subject's preferred political party, as defined in Table 1.1. A subject who is above the median value for abs(Republican Party rating - Democratic Party rating) is classified as Partisan; a subject who is not is classified as Moderate. Error bars correspond to 95 percent confidence intervals.

1.5.2 Regression Specifications for News Assessments

The primary regression specifications are within subject; 892 of the 897 non-neutral subjects receive at least one piece of Pro-Party news, Anti-Party news, and Neutral news.[29]

[29] Three subjects randomly receive no Pro-Party news; two subjects randomly receive no Anti-Party news; and all subjects receive some Neutral news.

Figure 1.2: Motivated Reasoning and Trust in Fake News: Perceived Veracity by News Direction and Actual Veracity



Notes: The y-axis is stated P(True), demeaned at the subject level. News on partisan topics is classified as Pro-Party (Anti-Party) if it is more (less) representative of the subject's preferred political party, as defined in Table 1.1. Fake News sends messages that reinforce the direction of subjects' error; True News sends messages that mitigate subjects' error. Error bars correspond to 95 percent confidence intervals.

In particular, the main specification for politically-motivated reasoning is in Table 1.2, column 2. The regression looks at assessments a for subject i, question topic q, and round r with fixed effects for i, q, and r when all news is Pro-Party or Anti-Party:[30]

$$a_{iqr} = \alpha + \beta \cdot 1(\text{Pro-Party})_{iqr} + \gamma FE_i + \delta FE_q + \zeta FE_r + \epsilon_{iqr}$$

[30]As seen in the Online Appendix, all results are qualitatively identical if we use logit(a_{iqr}) instead of a_{iqr}. The linear specification is used for ease of interpretation.

Hypothesis 1 claims that the Pro-Party / Anti-Party gap is increasing in partisanship, so column 3 interacts partisanship (the absolute difference in party ratings) with Pro-Party news. It also claims that motivated reasoning leads to both higher assessments for Pro-Party news and lower assessments for Anti-Party news; as such, column 4 includes indicators for both Pro-Party (vs. Neutral) news and Anti-Party (vs. Neutral) news.

Hypothesis 2 claims that for politicized news, subjects will trust Fake News more than True News, so the final two specifications in Table 1.2 regress assessments on a dummy for True News, controlling for and not controlling for Pro-Party news.

Table 1.2 shows that Hypotheses 1 and 2 are strongly supported. Assessments for Pro-Party news are substantially higher than for Anti-Party news, and this effect increases in partisanship. There is evidence for motivated reasoning on both Pro-Party and Anti-Party news, and controlling for news type, Fake News assessments are significantly higher than True News assessments.

Table 1.2: The Effect of News Direction and Actual Veracity on Perceived Veracity

	(1)	(2)	(3)	(4)	(5)	(6)
Pro-Party News	0.092***	0.088***	0.040***	0.037***		0.077***
	(0.006)	(0.007)	(0.013)	(0.007)		(0.007)
Partisanship x			0.050***			
Pro-Party			(0.012)			
Anti-Party News				-0.048***		
				(0.007)		
True News					-0.059***	-0.034***
					(0.006)	(0.007)
Neutral News	No	No	No	Yes	No	No
Question FE	Yes	Yes	Yes	No	Yes	Yes
Round FE	Yes	Yes	Yes	Yes	Yes	Yes
Subject controls	Yes	No	No	No	No	No
Subject FE	No	Yes	Yes	Yes	Yes	Yes
Observations	7902	7902	7902	10552	7902	7902
R^2	0.05	0.25	0.25	0.21	0.23	0.25
Mean	0.573	0.573	0.573	0.574	0.573	0.573

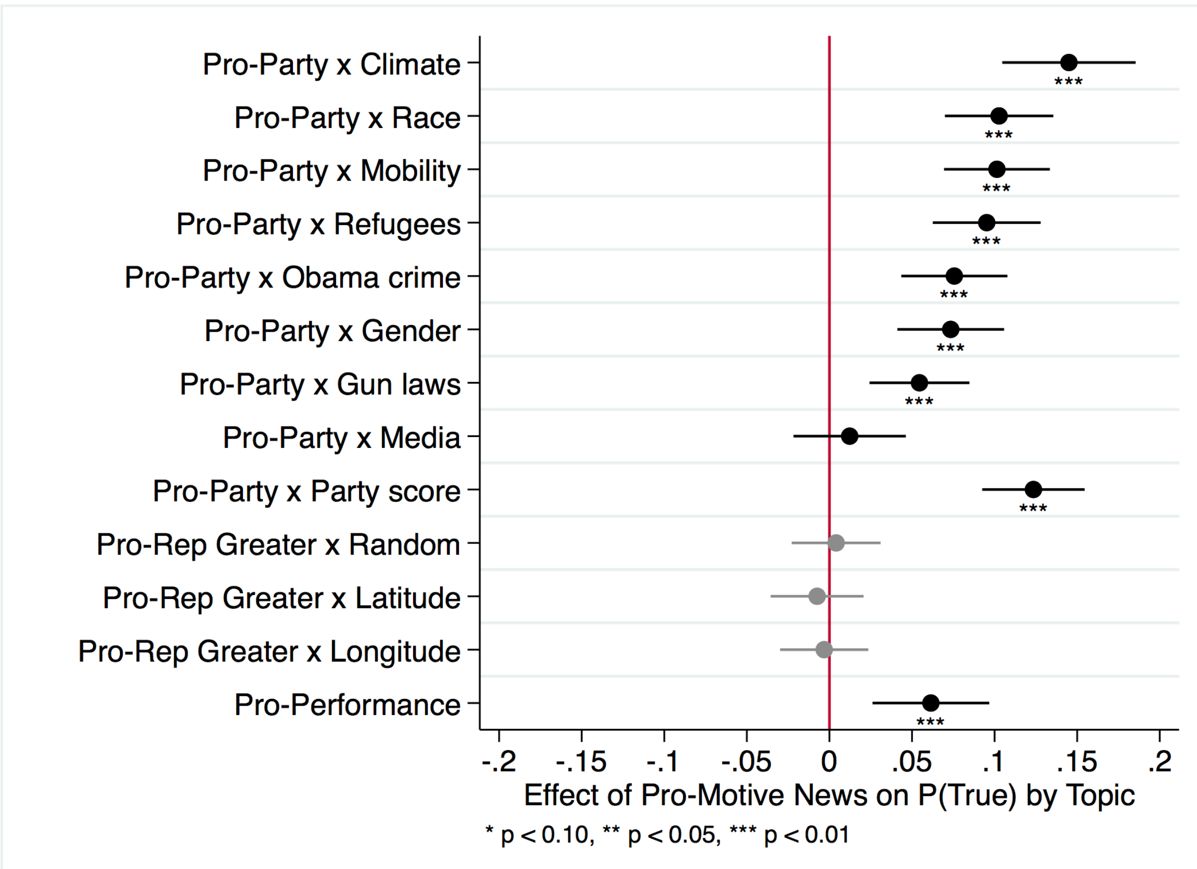
Standard errors in parentheses

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

Notes: OLS, errors clustered at subject level. Neutral News indicates that Pro-Party / Anti-Party news assessments are compared to assessments on Neutral topics. These classifications are defined in Table 1.1. Controls: race, gender, log(income), years of education, religion, and whether state voted for Trump or Clinton in 2016. Partisanship is the absolute difference between ratings of the Republican and Democratic parties.

Next, to show that motivated reasoning is a general phenomenon across domains, we look at each topic separately by regressing on the interaction of topic dummies and news type. Figure 1.3 shows that there is strong evidence to support politically-motivated reasoning on eight of the nine hypothesized topics. Each of these eight coefficients are statistically significant from zero at the $p = 0.001$ level. I also analyze placebo tests that Pro-Rep subjects support high answers for the neutral questions compared to Pro-Dem subjects, and find no evidence supporting a difference in assessments by party. On the performance topic, the average effect supports the hypothesis that people motivatedly reason towards believing they outperformed others.[31]

Figure 1.3: Motivated Reasoning Across Topics: Effect of Pro-Party News on Perceived Veracity by Topic



Notes: OLS regression coefficients, errors clustered at subject level. FE included for round number and topic. Pro-Party (vs. Anti-Party) news is defined in Table 1.1. Pro-Rep Greater is a placebo check to test whether Pro-Rep and Pro-Dem subjects give different assessments on neutral topics. Error bars correspond to 95 percent confidence intervals.

[31] Unlike on the politicized topics, this effect is entirely driven by men. See Section 1.6.2.

Consistent with hypotheses, partisans engage in at least as much politically-motivated reasoning than moderates on each topic where motivated reasoning plays a role. Appendix Figure A.3 interacts the topic-by-topic treatment effects from Figure 1.3 with a dummy for being partisan and for being moderate. There is a consistent difference on politicized topics, but not on neutral topics.

1.5.3 Changing Guesses and Polarization

Recall that half of subjects are randomly assigned to give a second guess to the initial question after receiving news. While the predictions here are not as well-identified, motivated reasoning should play the same role. In particular, the related hypothesis is that subjects are more likely to update in the Pro-Party direction than in the Anti-Party direction. This test is useful as a robustness check, but also helps us better understand how these messages affect subjects' beliefs about these issues.

Table 1.3 shows that subjects are more likely to update in the direction of Pro-Party messages than they are from Anti-Party messages. As hypothesized, on politicized topics subjects are also more likely to change their guesses in the direction of a polarizing message (one that tells them that their guess is further away from the mean) than from an anti-polarizing message.

Columns 4-6 of Table 1.3 show that discrepancies in both motivated reasoning and belief polarization are largely explained by differences in news assessments. After controlling for assessments, guess changes are not significantly affected by Pro-Party / Anti-Party messages, nor are they significantly affected by polarizing messages. This indicates that belief changes and news assessments are consistent with each other, validating the news assessment measure of motivated reasoning. That is, how someone assesses the veracity of a news source is the main determining predictor of how she directionally changes her beliefs.[32]

[32]It is worth noting that the converse is not true. For instance, after controlling for the direction of guess changes, subjects give statistically significantly higher assessments to Pro-Party news than to Anti-Party news.

Table 1.3: Changing Guess to Follow Message Given News Direction

	(1)	(2)	(3)	(4)	(5)	(6)
Pro-Party News	0.122***		0.114***	0.018		0.024
	(0.021)		(0.021)	(0.018)		(0.018)
Polarizing News		0.061***	0.032*		-0.017	-0.022
		(0.019)	(0.019)		(0.016)	(0.016)
P(True)				1.126***	1.139***	1.131***
				(0.062)	(0.061)	(0.063)
Question FE	Yes	Yes	Yes	Yes	Yes	Yes
Round FE	Yes	Yes	Yes	Yes	Yes	Yes
Subject FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4085	4085	4085	4085	4085	4085
R^2	0.28	0.28	0.28	0.45	0.45	0.45
Mean	0.659	0.659	0.659	0.659	0.659	0.659

Standard errors in parentheses

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

Notes: OLS, errors clustered at subject level. Only subjects from the Second-Guess treatment. Only Pro-Party / Anti-Party news observations, as defined in Table 1.1. Polarizing News is defined as news that tells subjects that, compared to their initial guess, the answer is in the opposite direction from the population mean. Dependent variable is 1 if subjects change their guess upwards when the message says "Greater Than" or downwards when the message says "Less Than," -1 if they change their guess in the opposite direction, and 0 if they do not change their guess.

More broadly, this gives a stark prediction about how people change their beliefs. Motivated reasoning leads people to engage in belief polarization from *uninformative* messages. This suggests that, in environments where signals serve to remind people of their motivated beliefs, not only do people not need different news sources to polarize their beliefs, informational content is not a necessary condition either.

1.5.4 Alternative Explanations and Robustness Checks

Misunderstanding "median" and skewed priors

It is reasonable to expect that subjects do not fully understand the concept of a median. For instance, they may answer with their mean belief instead. This would not directionally impact the news assessment results in a systematic direction, unless the prior distribution were notably skewed. We can use where the initial guess μ_q lies in subjects' confidence intervals as a proxy for skewness, and see that the main results hold for subjects who have zero skewness.

When looking at the politicized questions, 32 percent of subjects' guesses are exactly halfway between their upper and lower bounds. In the appendix, Table A.1 uses the same within-subject specification as the main regression but interacts Pro-Party news, Anti-Party news, and True News with a dummy that equals 1 for such "unskewed" priors. The treatment effects are essentially both qualitatively and quantitatively identical, indicating that skewness does not directionally effect results.

The independence of news sources

The interpretation of P(True News) in the model and analysis assumes that subjects treat news sources as being drawn from independent distributions. While subjects are explicitly told to do this in the instructions, it is useful to show that they are not using previous pieces of news to update about current pieces of news.

In Appendix Table A.2, I modify the main regression table to account for the relative number of Pro- and Anti-Party news in previous rounds. The effect of previous rounds' Pro- and Anti-Party news have precisely zero effect on current beliefs, and the main coefficients of interest remain unchanged, suggesting that subjects indeed treat news sources as independent.

Misunderstanding "Fake News"

First, suppose that subjects believe that messages from Fake News are actually from "Random News" and are equally likely to send correct and incorrect messages, instead of *always* sending incorrect messages. In this experiment, that would not affect any predictions about assessments. A Bayesian would still have an ex-ante prior that Pro-Party and Anti-Party messages are equally likely, and would not infer anything about P(True News) given either message. A motivated reasoner who is motivated to believe that the answer is large would still infer that P(True | Pro-Party) > P(True | Anti-Party).

A more complicated situation involves subjects believing that messages from Fake News are actually from a news source that is biased against their party. That is, suppose that subjects believe that Fake News was politically asymmetric, and is more likely to report Anti-Party news given Pro-Party truth than Pro-Party news given Anti-Party truth.

To test this, we can again look at how subjects change their guesses in Table 1.3. In particular, suppose that subjects were Bayesian but had this asymmetrically wrong definition of Fake News. Then, they would find Pro-Party "Fake News" messages to be more informative than Anti-Party "Fake News" messages, since "Fake News" is expected to usually send the Anti-Party message. (The quotes here indicate that these subjects are using the wrong definition of Fake News.) So, such subjects would update more from Pro-Party than Anti-Party news, *conditional on their assessment of P(True News).*

In Table 1.3, we see that subjects are similarly likely to update from Pro-Party and Anti-Party news after controlling for their assessments. While the data are too imprecise to rule out that there may exist subjects who treat Fake News as biased, this explanation is insufficient to drive results.

Incorrect initial guesses

While it can sometimes be in subjects' best interests to strategically misreport their median in order to earn more points on news assessment questions, I find no evidence of this.

In Round 1 of the experiment, subjects do not yet know that they will be seeing a news assessment page. If subjects were strategically mis-guessing to earn more assessment points, they would perform worse in Round 1 than subsequent rounds on assessments and better in Round 1 than subsequent rounds on guesses.

There are no significant differences in assessment scores in Round 1. Subjects score 67.2 points (s.e. 0.9) in Round 1 and 66.4 points (s.e. 0.3) in Rounds 2-12;[33] the difference is 0.8 points (s.e. 1.0) and insignificant ($p = 0.383$). The null result remains when using within-subject tests, controlling for topic, and controlling for linear round trends.

There are also no significant differences in guess scores in Round 1. Subjects score 76.2 points (s.e. 1.0) in Round 1 and 75.9 points (s.e. 0.2) in Rounds 2-12; the difference is 0.3 points (s.e. 1.0) and insignificant ($p = 0.758$). Within-subject tests, controlling for topic, and controlling for linear round trends do not change the null result.

Non-strategic forms of incorrect initial guesses are more complicated to rule out. If there is symmetric noise such that the probability that a subject is equally likely to state her Q quantile and her $1 - Q$ quantile for $Q \neq 1/2$, then the main results do not change. Results are also not consistent with subjects biasing their initial guesses towards the population mean. While this behavior can explain why subjects trust error-reinforcing news more than error-mitigating news on politicized and performance topics (and why they trust Pro-Party news more than Anti-Party news), it incorrectly predicts the same pattern on neutral topics. The one form of misreporting that can be consistent with both Bayesian updating and results from the experiment involves subjects systematically misreporting medians in a way that is biased in the opposite direction from their party.

In theory, one potential reason for an Anti-Party-biased first guess is that subjects do not sufficiently think about the question; and, given more time, they update towards their true (more Pro-Party) belief. A version of this explanation in which purely time spent affects the extremity of beliefs seems unlikely to explain these results, as the main treatment effect does not noticeably affect time spent on the question page.[34] An alternative version in which seeing the second screen causes subjects to think harder about the original question, and thinking harder leads to more Pro-Party beliefs, is more plausible. The psychology behind this expla-

[33] I exclude scoring on Rounds 13-14 since the questions are not randomly assigned in those rounds; the result is identical if they are included. I also exclude scoring on comprehension check questions.

[34] The mean time spent on the assessment page with Pro-Party news is 14.6 seconds (s.e. 0.3 seconds), and the mean time spent on the assessment page with Anti-Party news is 14.8 seconds (s.e. 0.3 seconds).

nation is very similar to this theory of motivated reasoning, as the second page evokes the motive, and further work could better elucidate the contours of what qualifies as a signal for motivated reasoning.

Expressive preferences

Bursztyn et al. (2019) provides recent evidence showing that people in an experiment may forgo payment in order to make political statements. In this experiment, if subjects have a preference for stating Pro-Party signals, then both their initial guesses and their news assessments will be biased in the Pro-Party direction, consistent with the data. However, if they are Bayesian, how they *change* their guesses will not be directional, since they have already stated their preferred belief.

Recall that in Table 1.3, subjects are more likely to update their guesses in the Pro-Party direction than in the Anti-Party direction, even though they are equally likely to receive Pro-Party and Anti-Party messages. This is consistent with subjects genuinely trusting the Pro-Party messages more; it is not consistent with Bayesian updating and expressive preferences.

Motivated reasoning by treatment and round

It is possible to construct alternative hypotheses in which some treatments lead to more biased updating processes than others. For instance, perhaps the subjects who were not told in the instructions that P(True News) = 1/2 behave differently than those who are told this, and the latter group does not motivatedly reason because of this prior. Or perhaps the subjects who are told to give a second guess to the initial question are reminded of their initial median more and this leads to a correction of motivated reasoning.

In the Online Appendix, I restrict the regressions from Table 1.2 to subjects in the willingness-to-pay treatment, the second-guess treatment, the received-prior treatment, and the did-not-receive-prior treatment. Estimates naturally become noisier, but the direction of every estimate is identical. There is no evidence that any treatment significantly affected the observed magnitude of motivated reasoning.

It is also possible that subjects learn over the course of the experiment that they motivatedly reason and debias themselves. I find no evidence for this. In the Online Appendix, I interact the main effect with dummies for each round number; in every single round, subjects give larger assessments to Pro-Party news than Anti-Party news. I also restrict Table 1.2 to Rounds 1-6 and Rounds 7-12, and effects are in the same direction.

1.5.5 Motivated Reasoning and Initial Beliefs

The previous results have shown that subjects differentially update their beliefs about the veracity of news sources based on the direction of messages received, and that an observer can infer something more about people's motives from their erroneous beliefs.

This subsection discusses two consequences of motivated reasoning that are manifested in beliefs about the questions themselves: polarization and overprecision. First, I show that variation in this experiment's measure of motivated beliefs can explain a sizable fraction of variation in actual beliefs about these questions.

I look at the relationship between motives and beliefs by correlating the normalized answers to politicized questions with the normalized differences in assessments between Pro-Rep and Pro-Dem news. For each politicized question, subjects' initial guesses are winsorized (at the 5-percent level), normalized, and signed; positive numbers correspond to more Pro-Rep. Next, for each subject, these normalized guesses are averaged (and re-normalized) to give a measure of how Pro-Rep her beliefs are. I correlate this value with the normalized average difference between Pro-Rep news assessments and Pro-Dem news assessments.[35]

Variation in news assessments explain 13 percent of the variation in beliefs. By comparison, non-political demographics collected in this experiment — age, gender, race, education, logged income, whether one is religious, whether one is from a state that voted for Trump or Clinton in 2016 — explain 7 percent of the variance in beliefs.[36]

[35] There are five subjects who, by chance, do not receive one of the two news types. I drop these subjects, but the estimate is exactly the same if they are instead set to zero.

[36] These are unadjusted R^2 values. Adjusted R^2 is 13 percent for news assessments and 6 percent for demographics.

In support of Hypothesis 4, there is evidence that subjects are overprecise in their initial beliefs about questions that evoke motivated beliefs, but no evidence for overprecision on neutral questions. On politicized topics, subjects' confidence intervals contain the correct answer 46.6 percent of the time (s.e. 0.6 percent); this is statistically significantly different from 50 percent ($p < 0.001$). Overprecision on these topics is primarily driven by partisans, whose intervals contain the correct answer 44.2 percent of the time (s.e. 0.9 percent). Moderates' intervals contain the correct answer 48.8 percent of the time (s.e. 0.8 percent). Partisans' level of overprecision is statistically significantly larger than moderates' ($p < 0.001$). On the performance question, subjects' confidence intervals contain the correct answer 42.0 percent of the time (s.e. 1.6 percent), which is also statistically significantly different from 50 percent ($p < 0.001$).

This evidence for overprecision cannot be explained by a more universal bias towards overly narrow confidence intervals. On the neutral topics, subjects are actually somewhat underprecise. A natural test looks at subjects' confidence intervals for the "Random Number" question, which asks them to guess what a random number drawn from 0 to 100 will equal. On this question, subjects' intervals contain the correct answer 54.6 percent of the time, which is statistically significantly *larger* than 50 ($p = 0.004$).[37]

1.5.6 Motivated Reasoning, Underperformance, and Overconfidence

Next, we consider implications discussed in Hypothesis 5: underperformance and overconfidence. On news assessment questions, subjects typically score worse than if they had ignored the message entirely. This is primarily explained by two factors:

1. **Noisy updating lowers performance.** Subjects score worse on neutral topic news assessments than if they had always guessed their prior P(True).

2. **Motivated beliefs lower performance.** Subjects score worse on news assessments about politicized topics than about neutral topics. This is a logical consequence of Hy-

[37] Similarly, the average interval subjects have is 56.4, while a correctly-calibrated subject would have an interval of 50. On the two other neutral questions, subjects exhibit moderate underprecision as well.

pothesis 2, since subjects are more likely to believe Fake News than True News on politicized topics compared to neutral topics.

If subjects had always answered P(True) = 1/2 on news assessment questions, they would score 75 points. Yet, on average, subjects score lower than 75 points on every question. Table 1.4 shows that scores are especially lower for politicized topics compared to neutral topics.

The lower-than-75 scores on neutral topics can be explained by subjects updating with noise.[38] The further gap between neutral and politicized topics can be explained by motivated reasoning. The difference in subjects' news assessment scores between politicized and neutral topics increases in subjects' partisanship.

In fact, partisanship can explain nearly the entirety of subjects' scoring gap between politicized and neutral questions. In Table 1.4, column 1 shows that scores are lower for politicized topics than neutral topics, column 2 shows that the gap between neutral and political assessment scores increases in partisanship and column 3 shows that this is due to decreasing political scores more than increasing neutral scores.

[38]An alternative explanation is that prior beliefs about P(True) may be substantially different from 1/2 for some subjects. However, even subjects whose average assessment is exactly 1/2 score significantly lower than 75 points on both partisan and neutral questions.

Table 1.4: Effects of Topic and Partisanship on News Assessment Scores

	(1)	(2)	(3)
Politicized Topic	-4.14***	-1.92*	-1.92*
	(0.62)	(1.08)	(1.03)
Partisanship x Politicized Topic		-5.09**	-3.58***
		(1.98)	(1.14)
Partisanship x Neutral Topic			1.53
			(1.62)
Round FE	Yes	Yes	Yes
Subject FE	Yes	Yes	No
Observations	11612	11612	11612
R^2	0.12	0.12	0.01
Mean	69.47	69.47	69.47

Standard errors in parentheses

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

Notes: OLS, errors clustered at subject level. Party-indifferent subjects included. News assessment scores range from 0 to 100; subjects can guarantee a score of 75 by saying that the source is equally likely to be True News or Fake News. Partisanship is the absolute difference between subjects' ratings of the Republican and Democratic parties. Politicized topics and neutral topics as defined in Table 1.1. Subject controls are party preference, age, race, gender, log(income), education, religion, and whether state voted for Trump or Clinton in 2016.

Similarly, subjects' average scores across all pages are negatively correlated with their partisanship. These scores are hard to interpret on their own, but they are compared to 100 pilot participants to establish a Relative Performance percentile.[39] On the relative performance

[39]These overall scores are an average of scores on assessments, guesses, bounds, and either second

question, subjects are asked to predict how many of these 100 they outscored. Hypothesis 5 posits that more partisan subjects will have lower Performance scores and be more overconfident in how they scored relative to others.

Table 1.5 gives evidence for both parts of this hypothesis. Expected performance significantly increases in partisanship. Points scored significantly decrease in partisanship, though the Relative Performance percentile is a noisy estimate of this, so this measure is only significant at the 10% level. There does not appear to be substantial overconfidence overall; subjects on average expect to perform at the median. But, partisans on score worse than the median and expect to score better than the median on average.

---

guesses or the willingness-to-pay round. They are calculated after round 12.

Table 1.5: Performance and Expected Performance by Partisanship

	News Pts		Performance		Expectation	
	(1)	(2)	(3)	(4)	(5)	(6)
Partisanship	-3.58***	-3.53***	-4.74*	-4.89*	6.13***	8.19***
	(1.13)	(1.15)	(2.60)	(2.68)	(2.18)	(2.16)
Male		-0.81		3.65**		11.98***
		(0.71)		(1.70)		(1.31)
Subject controls	No	Yes	No	Yes	No	Yes
Observations	8696	8696	987	987	987	987
R^2	0.00	0.00	0.00	0.04	0.01	0.13
Mean	69.01	69.01	47.64	47.64	50.36	50.36

Standard errors in parentheses

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

Notes: OLS. Party-indifferent subjects included. A subject's News Pts is her points scored on news questions on politicized topics. A subject's Performance is equal to how many pilot subjects (out of 100) she outscored. Calculations are made after round 12 of the experiment. A subject's Expectation is equal to her median belief of how many pilot subjects (out of 100) she outscored. Partisanship is the absolute difference between subjects' ratings of the Republican and Democratic parties. Subject controls are party preference, age, race, gender, log(income), education, religion, and whether state voted for Trump or Clinton in 2016.

1.5.7 Discussion

These results strongly support the hypothesis of motivated reasoning with politically-motivated beliefs compared to Bayesian updating. Subjects significantly over-trust Pro-Party news and Fake News in an environment with uninformative signals, real monetary stakes, and little room for self-deception. This bias leads to other errors and biases such as underperformance, overconfidence, and overprecision.

Motivated reasoning may explain one form of prior-confirming bias, one in which people update further in the direction of their prior than a Bayesian would. Evidence supporting this prior-confirming bias would show a positive correlation between over-updating and prior beliefs. Motivated reasoning suggests that prior beliefs are often reflective of motivated beliefs, and that detection of prior-confirming biases may in fact be detecting motivated reasoning.

The results in Section 1.5.3 also relate to the effect of motivated reasoning on political polarization. Not only do subjects polarize in beliefs about the veracity of news, they polarize in their beliefs about the questions themselves, despite receiving uninformative news. Gentzkow and Shapiro (2011) find only modest differences in the media that liberals and conservatives consume, and motivated reasoning can help explain why people polarize even if they consume similar media outlets.

Results also suggest that motivated beliefs are *even further apart* than current beliefs, and that people have not yet reached their highest-motive beliefs. The reason for this is that the amount of distortion in updating is constrained by the actual informational content of signals. Motivated reasoners who receive precise signals would in fact become less polarized.

Methodologically, news assessments seem to be a more precise measure of motivated reasoning than changing guesses. With a continuous state, there is much heterogeneity in how *Bayesian* subjects would update their beliefs from information, so the null hypothesis is harder to reject and the magnitude of bias is hard to compare across domains. By using this experimental paradigm, subjects' priors are standardized, heterogeneity across issues and subjects is testable, and the Bayesian null is more easily falsifiable.

1.6 Demographic Heterogeneity

1.6.1 Heterogeneity in Motivated Reasoning

There are two types of heterogeneity to consider: heterogeneity in the direction of motivated reasoning, and heterogeneity in its magnitude. The main finding of this section is that motivated reasoning on the politicized topics does not noticeably depend on any non-political demographics, and that we can rule out even moderately large effects.

First, we consider the direction of heterogeneity. To do this, Table 1.6 runs a horse race regression that regresses news assessments on the interaction of the political direction of the news (Pro-Rep vs. Pro-Dem) and observable demographics. Non-political demographics in this study are race, gender, income, age, education, whether the subject's state voted for Trump or Clinton in 2016, and religious affiliation. Controlling for party preference, *none* of the other demographics have any significant effect on the direction of motivated reasoning.

Table 1.6: Heterogeneity in the Partisan Direction of Motivated Reasoning: Horse Race Regression

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Rep News x Pro-Rep	0.13***								0.12***
	(0.01)								(0.01)
Rep News x (Age>32)		0.00							-0.01
		(0.01)							(0.01)
Rep News x Male			0.00						0.00
			(0.01)						(0.01)
Rep News x White				0.03*					0.02
				(0.02)					(0.02)
Rep News x College					0.01				0.02
					(0.01)				(0.01)
Rep News x (Inc>50K)						-0.01			-0.02
						(0.01)			(0.01)
Rep News x Red State							0.02*		0.01
							(0.01)		(0.01)
Rep News x Religious								0.05***	0.02
								(0.01)	(0.01)
Rep News	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Question FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Round FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	7902	7902	7902	7902	7902	7902	7902	7902	7902
R^2	0.26	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.26
Mean	0.57	0.57	0.57	0.57	0.57	0.57	0.57	0.57	0.57

Standard errors in parentheses

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

Notes: OLS regression coefficients, errors clustered at subject level. FE included for round number and topic. Only Pro-Party / Anti-Party news observations, as defined in Table 1.1. Pro-Rep: higher rating for Republican than Democratic Party. Red State: coted for Trump in 2016. Religious: subject affiliates with any religion.

Not only are other demographics not statistically significantly different from zero, they are all statistically significantly different from +/- 0.05. This does not seem to be an artifact of aggregating across questions; even on questions about particular demographics (e.g. gender and math ability; racial discrimination), there are not statistically significant demographic effects.

Next, we consider the *magnitude* of motivated reasoning, acknowledging that this design does not enable us to disentangle magnitude of bias and strength of motive. Table 1.7 runs another horse race regression, regressing the news assessments on the interaction of the motivated direction of the news (Pro-Party / Pro-Performance vs. Anti-Party / Anti-Performance) and observable demographics.

Table 1.7: Heterogeneity in the Magnitude of Motivated Reasoning: Horse Race Regression

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Pro-Motive x Pro-R	0.06***								0.08***
x Partisan	(0.02)								(0.03)
Pro-Motive x Pro-R	-0.02								0.00
x Moderate	(0.01)								(0.03)
Pro-Motive x Pro-D	0.11***								0.12***
x Moderate	(0.01)								(0.02)
Pro-Motive x Pro-D	0.13***								0.14***
x Partisan	(0.01)								(0.02)
Pro-Motive x (Age>32)		0.00							0.01
		(0.01)							(0.01)
Pro-Motive x Male			0.00						0.01
			(0.01)						(0.01)
Pro-Motive x White				-0.02					-0.01
				(0.01)					(0.01)
Pro-Motive x College					0.01				0.01
					(0.01)				(0.01)
Pro-Motive x (Inc>50K)						-0.02**			-0.02*
						(0.01)			(0.01)
Pro-Motive x Red State							-0.02*		-0.01
							(0.01)		(0.01)
Pro-Motive x Religious								-0.04***	-0.01
								(0.01)	(0.01)
Pro-Motive News	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
Question FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Round FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	8793	8793	8793	8793	8793	8793	8793	8793	8793
R^2	0.24	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.24
Mean	0.58	0.58	0.58	0.58	0.58	0.58	0.58	0.58	0.58

Standard errors in parentheses

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

Notes: OLS regression coefficients, errors clustered at subject level. FE for round number and topic. Only observations relevant to motives as defined in Table 1.1. Pro-Motive indicates Pro-Party or Pro-Performance. Pro-R: higher rating for Rep than Dem Party. Partisan: above median for abs(Rep rating - Dem rating). Red State: state voted for Trump in 2016. Religious: subject affiliates with any religion.

There is strong evidence that partisans of both parties motivatedly reason; the discrepancy between partisans and moderates seems to be a difference in motive strength, not in the level of bias. Interestingly, there is a notable difference between Pro-Rep and Pro-Dem moderates, the former of which do not motivatedly reason in the predicted direction on average. This party difference may be better explained by direction instead of magnitude of bias, as the sample is non-representative *conditional on party.* For instance, only 76 percent of Republicans in this sample approved of President Trump's performance; in a Gallup poll conducted contemporaneously (from June 25-July 1), 87 percent of Republicans approved of his performance (Gallup 2018b).

As with the direction of motivated reasoning, non-political demographics do not notably affect the magnitude of the bias; all effects are again between +/- 0.05 once party preference is controlled for.

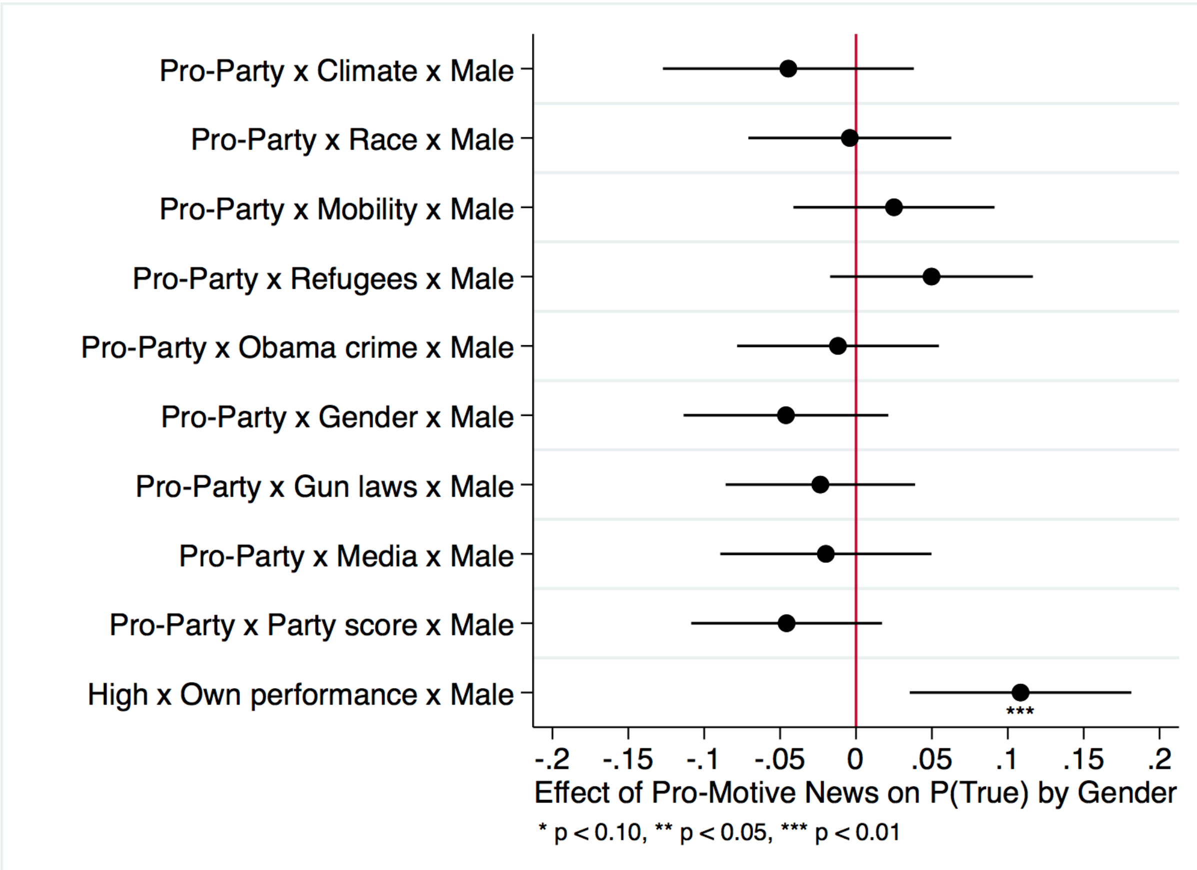
These results suggest that the degree of bias of motivated reasoning is somewhat consistent across demographics, but that the direction of motivated beliefs are heterogeneous. In particular, the sign of the slope of the motive function about many of these issues is exactly the opposite for Democrats and Republicans.

1.6.2 Gender, Performance-Motivated Reasoning, and Confidence

Section 1.6.1 showed that the direction of motivated reasoning was similar across non-party demographics on politicized topics, but this is not the case for performance-motivated reasoning. Overall, men significantly motivatedly reason in the direction of believing they outperformed others, while women do not systematically motivatedly reason in either direction.

Figure 1.4 shows gender differences in the magnitude of the treatment effect by question. On the performance topic, men give Pro-Performance news 11 percentage points higher assessments than Anti-Party news (s.e. 2 percent) and women give Pro-Performance news 0.2 percentage points lower assessments than Anti-Performance news (s.e. 2 percent). But on the politicized topics, men give Pro-Party news 9 percentage points higher assessments than Anti-Party news (s.e. 0.9 percentage points), and women give Pro-Party news 10 percentage points higher assessments than Anti-Party news (s.e. 0.9 percentage points).

Figure 1.4: Gender Heterogeneity in Motivated Reasoning Across Topics: Effect of News Direction on Perceived Veracity by Topic and Gender



Notes: OLS regression coefficients, errors clustered at subject level. FE included for round number and topic, interacted with gender. Only subjects who identify as male or female included. Only news observations that are relevant to motives as defined in Table 1.1. Figure shows interaction between Pro-Party / Pro-Performance news and male, controlling for Pro-Party / Pro-Performance news. Error bars correspond to 95 percent confidence intervals.

Similar patterns emerge for overconfidence. Table 1.5 shows that men are more overconfident than women. In fact, only men are overconfident. The magnitude of the overconfidence discrepancy is especially stark when comparing expected performance by gender when controlling for actual performance, as seen in Appendix Figure A.4. Except for the highestperforming women, women of all performance levels expect to score below the median, and men of all performance levels expect to score above the median.

These results suggest that gender differences in confidence are related to gender differences in motivated reasoning. In particular, there is a clear gender asymmetry in the magnitude of both biases: women are not systematically biased about either performance-related motivated reasoning nor confidence on average, while men are systematically biased about both in the direction of believing they outperformed others.

1.7 Conclusion

Motivated reasoning plays a substantial role in people's assessment of the veracity of news and helps explain why people form inaccurate and polarized beliefs. This paper demonstrates its importance across numerous varied topics with a novel experimental design, showing that motivated reasoning is a unique phenomenon from Bayesian updating, prior- and likelihood-misweighting biases, and utility-maximizing beliefs. Furthermore, these results have shown how motivated reasoning leads to further belief polarization, overprecision, and an excess trust in Fake News.

One interpretation of this paper is unambiguously bleak: people form polarized and biased beliefs because of motivated reasoning, motivatedly reason even farther in the polarized direction in the experiment, and make particularly biased inferences on issues they find important. However, there is another, complementary, interpretation of this paper: this experimental design takes a step towards better identifying and understanding motivated reasoning, and makes it easier for future work to attenuate the bias. Using this design, we can identify and estimate the magnitude of the bias; future projects that use interventions to debias people can use this estimated magnitude as a dependent variable. Since the bias often decreases utility, people may have demand for such interventions.

A potential path to attenuate the bias involves understanding the determinants of susceptibility. There is no evidence that susceptibility depends much on an individual's characteristics; however, it could depend on the signal structure. Intuitively, an agent who receives an arbitrarily precise signal or a clearly irrelevant signal will likely have very low susceptibility, while an agent who receives a hard-to-interpret signal will likely have higher susceptibility and deviate more from Bayes' rule. Future work can experimentally and empirically estimate this parameter in contexts with fixed motives but varying signal structures. And, if we can decrease susceptibility, we can limit the bias in people's updating process.

Many of these results also suggest further exploration of what motives actually represent. This paper identifies a few specific parts of the motive function distribution, but extending this design can identify the shape of the distribution and generate utility-like properties such as concavity and risk motives. It also can provide insight on how motives and choices interact.

Finally, while one definition of motivated beliefs posits that they are beliefs that increase utility, this paper provides no evidence that people are motivated to believe "good things" about the world, and such an interpretation gives perverse implications about peoples' preferences. What does it mean if Republicans are motivated to believe that more Americans were murdered during Obama's presidency? What does it mean if Democrats are motivated to believe that there is rampant racial discrimination in labor markets? These are controversial questions, but they are ones that are crucial for understanding how people form beliefs in a highly politicized society.

2

The Limits of Motivated Reasoning When Self Image Is Not at Stake

2.1 Introduction

There is a common intuition in economics that people find it more attractive to believe that they are in a "good" state of the world than in a "bad" state of the world, and that this can lead them to form beliefs that are directionally distorted in favor of good states. However, tests of such over-optimism focus on states where self-image is at stake, such as about one's future prospects, one's ability, one's altruism, or one's politics (e.g. Weinstein 1980; Mobius et al. 2014; Eil and Rao 2011; Chapter 1).

This paper argues that motivated reasoning — the distortion of new information in the direction of more attractive beliefs — is not solely about "good" and "bad" states. I run an experiment to explore how people make inferences about states of the world that are good or

Table 2.1: Topics and Hypothesized Motives in the Experiment

Topic	**Positive Motives**	**Negative Motives**
Infant mortality	Low / Decreasing	High / Increasing
Others' reported happiness	High / Increasing	Low / Decreasing
Leukemia survival rate for children	High / Increasing	Low / Decreasing
Global poverty rate	Low / Decreasing	High /Increasing
Deaths in armed conflicts	Low / Decreasing	High / Increasing
Latitude of US	Neutral	Neutral

Table 2.2: The list of topics and positivity motives; the exact wording of each question is in Section B.2.

bad for others, and in which self-image does not play a role: *positivity-motivated reasoning.* I find evidence that this form of positivity-motivated reasoning does *not* play much of a role in inference.

In a large online experiment, I test whether people engage in positivity-motivated reasoning on five topics: the survival rate of children with leukemia, global poverty rates, annual deaths in armed conflict, others' happiness levels, and infant mortality rates. The topics are in Table 2.2 below.

To identify motivated reasoning, I conduct a large online experiment that builds off of the design of Chapter 1. That paper found evidence of motivated reasoning in political and performance domains. Forms of motivated reasoning have also been found in other self-image domains such as about one's altruism (Exley 2015; Di Tella et al. (2015)), intelligence (Mobius et al. 2014; Eil and Rao 2011), financial earnings (Mayraz 2013), and attractiveness (Eil and Rao 2011). Related theories often focus on image-relevant settings or treat motivated beliefs as a form of utility (e.g. Kunda 1990; Benabou and Tirole 2002; Brunnermeier and Parker 2005).

The main result in this paper is that – across several settings in which self-image is not relevant – there is no evidence for positivity- or negativity-motivated reasoning. I show that, aggregating across these questions, even modest effects can be ruled out. In fact, by comparing the magnitude of positivity-motivated reasoning to results in Chapter 1, we can rule out an effect of positivity or negativity that is half as large as politically-driven or performance-driven motivated reasoning. This evidence shows that positivity, by itself, is insufficient for motivated reasoning.

The second result is that there is no evidence that subjects' current beliefs are reflective of past positivity-motivated reasoning. That is, subjects whose beliefs are overly positive are no more likely to engage in positivity-motivated reasoning in this experiment. This suggests that there is limited heterogeneity in subjects' positivity-motivated reasoning. Relatedly, there are not substantial differences in motivated reasoning by demographic factors like gender, education, or income.

The third result is that people do not *expect* to see evidence for positivity-motivated reasoning, but do expect positivity to affect happiness. In a separate survey, I ask participants what they expect the direction of motivated reasoning to be about positivity, politics, and own performance. While the majority of participants expect others to engage in pro-party and pro-performance motivated reasoning, they are similarly likely to expect to see positivity-motivated reasoning, negativity-motivated reasoning, or no notable difference. Yet a clear majority of participants expect positive news to make people happier.

Taken together, these results are consistent with the notion that motivated reasoning is not only driven by belief-based utility. Subjects may attain higher utility by learning that the world is good for other people, and yet not systematically distort their inference process in favor of these beliefs. That is, the beliefs that people find more attractive do not necessarily make them happier. Rather, it may be limited to belief-based utility that relates to one's self-image.

The rest of the paper proceeds as follows: Section 2.2 discusses the main theory and experimental design that identifies motivated reasoning, adapted from Chapter 1. Section 2.3 discusses the data. Section 2.4 presents the main experimental results. Section 2.5 discusses interpretations of the main experiment and presents survey evidence about what people expect about others' behavior and utility. Section 2.6 concludes and proposes directions for future work. The appendices provide a table that is omitted from the main text, and list the exact questions and pages that subjects see.

2.2 Theory and Experimental Design

2.2.1 Theory and Predictions

The theory of motivated reasoning follows Chapter 1. Further details are in that paper. When a motivated-reasoning agent infers about the probability that an event is true (T) or false ($\neg T$), with prior $\mathbb{P}(T)$, the agent forms his posterior by incorporating prior, likelihood, and a motivated beliefs term:

$$\underbrace{\mathbb{P}(T|x)}_{\text{posterior}} \propto \underbrace{\mathbb{P}(T)}_{\text{prior}} \cdot \underbrace{\mathbb{P}(x|T)}_{\text{likelihood}} \cdot \underbrace{M(T)^{\varphi(x)}}_{\text{mot. reasoning}},$$

We take log odds ratios to attain the additive form:

$$\text{logit } \mathbb{P}(T|x) = \text{logit } \mathbb{P}(T) + \log\left(\frac{\mathbb{P}(x|T)}{\mathbb{P}(x|\neg T)}\right) + \varphi(x)(m(T) - m(\neg T)). \tag{2.1}$$

The motivated reasoner acts as if he receives both the actual signal (x) and a signal whose relative likelihood corresponds to how much he is motivated to believe the state is T. $m(T) : \{T, \neg T\} \to \mathbb{R}$ is denoted the **motive** function. The weight put on this signal is $\varphi(x) \geq 0$, called **susceptibility**. When $\varphi(x) = 0$, the agent is Bayesian; when $\varphi(x) > 0$, the agent motivatedly reasons.

This paper will assume $\varphi(x) > 0$ for the particular information structure in the experiment below, and back out $m(T) - m(\neg T)$ from the inference process. We will be interested in the psychology of the motive function. In this paper, either T will correspond to positivity (the world being a better place) and $\neg T$ to negativity (the world being a worse place), or vice versa.

The experiment provides people with not-obviously-uninformative signals about the veracity of news sources. To fix ideas, consider the following question, taken verbatim from the experiment:

Acute Myeloid Leukemia (AML) is a devastating illness in which cancerous cells emerge in the bone marrow, invade the blood stream, and may spread to the rest of the body. Tragically, hundreds to thousands of children under the age of 15 are diagnosed with AML each year; it is one of the most common cancers among children.

Of children under the age of 15 who are diagnosed with AML, what percent survive for at least 5 years?

This is a question for which higher-valued states are more positive. The main test of motivated reasoning then involves three steps:

1. **Beliefs:** Subjects are asked to guess the answers to questions like the one above. Importantly, they are asked and incentivized to guess their median belief (i.e. such that they find it equally likely for the answer to be above or below their guess).

2. **News:** Subjects receive a binary message from one of two randomly-chosen news sources: True News and Fake News. The message from True News is always correct, and the message from Fake News is always incorrect. This is the main (within-subject) treatment variation.

   The message says either "The answer is **greater than** your previous guess of [previous guess]." or "The answer is **less than** your previous guess of [previous guess]." Note that the exact messages are *different* for each subject since subjects have different guesses. These customized messages are designed so that they have the same *subjective* likelihood of occurring.

   For the cancer question above, "greater than" corresponds to Positive news and "less than" to Negative news.

3. **Assessment:** After receiving the message, subjects assess the probability that the message came from True News using a scale from 0/10 to 10/10, and are incentivized to

state their true belief. This news veracity assessment is the main outcome measure. The effect of variation in news direction on veracity assessments is the primary focus for much of this paper.

More formally, consider an agent with prior $F(\theta)$ about a state in Θ. Denote by $\mu \equiv F^{-1}(1/2)$ the median of $F(\theta)$. For simplicity, we assume that F has no atom at μ and that $\mathbb{P}(\mu = \theta) = 0$. That is, the agent believes that the answer has probability zero of being exactly equal to μ, and the true probability is indeed zero.[1]

The agent receives a source that is either from True News (T) or Fake News ($\neg T$). Both report one of two binary messages G or L: "The answer θ is greater than your median μ" or "The answer θ is less than your median μ." Prior beliefs $P(T)$ are fixed, and $\log\left(\frac{\mathbb{P}(G|T)}{\mathbb{P}(G|\neg T)}\right) = \log\left(\frac{\mathbb{P}(L|T)}{\mathbb{P}(L|\neg T)}\right) = 0$ by definition of a median.

	$\theta > \mu$	$\theta < \mu$
True News sends	G	L
Fake News sends	L	G

The agent has a prior about the news source $p \equiv \mathbb{P}(T)$ that does not depend on θ, and infers about $\mathbb{P}(T)$ given the message received.

We can now look at how a motivated reasoner updates his beliefs about the news source after receiving G:

$$\begin{aligned}\text{logit } P(T|G) &= \text{logit } \mathbb{P}(T) + \log\left(\frac{\mathbb{P}(G|T)}{\mathbb{P}(G|\neg T)}\right) + \varphi\left(m(\theta|\theta > \mu) - m(\theta|\theta < \mu)\right) \\ &= \text{logit } p + \varphi\left(m(\theta|\theta > \mu) - m(\theta|\theta < \mu)\right).\end{aligned}$$

Therefore, if m is strictly monotonically increasing in θ, then $P(T|G) > P(T|L)$, and if m is strictly monotonically decreasing in θ, then $P(T|G) < P(T|L)$. By contraposition, if $P(T|G) = P(T|L)$, then m is neither strictly monotonically increasing nor decreasing in θ.

[1] In this experiment, zero answers are correct, so the assumption appears reasonable.

Additionally, if m is monotonic in θ for all agents but there is heterogeneity in its slope, then the average slope may be zero because some agents have upward-sloping motives ("positivity motives") and some agents have downward-sloping motives ("negativity motives"). In this case, if agents have received information drawn from the same distribution in the past, then their current beliefs will reflect their motives. A positivity-motivated reasoner will be more likely to hold a belief that $\mu > \theta$, and a negativity-motivated reasoner with a decreasing motive function will be more likely to hold a belief that $\mu < \theta$. This implies that an agent who believes $\mu > \theta$ is more likely to believe that $P(T|G) > P(T|L)$ in the experiment, and an agent who believes $\mu < \theta$ is more likely to believe that $P(T|G) < P(T|L)$ in the experiment.

That is, if motive direction is heterogeneous, subjects will trust Fake News more than True News. For further details, see Chapter 1. By contraposition, if subjects trust Fake News and True News equally, then there is no evidence for heterogeneity in positivity- versus negativity-motivated reasoning.

2.2.2 Experimental Details

The experiment follows the structure and incentive scheme of Chapter 1, which contains further details. Screenshots of a version of each page in the experiment, including instructions and scoring rules, can be found in Section B.3.

Subjects first see an Introduction page for consent, then a Demographics page, and then the instructions and point system for Question pages. On each Question page, subjects are asked and incentivized to give a median guess, a lower bound (equal to their 25th-percentile belief), and an upper bound (equal to their 75th-percentile belief). The median is incentivized using a linear scoring rule, and the bounds using piecewise-linear scoring rules. For details, see Section B.3.

Next, subjects see the instructions and point system for News pages. On each News page, subjects see the message that says whether the answer is greater than or less than their previous median guess, and are asked and incentivized to assess the probability that the message comes from True News versus Fake News using a quadratic loss scoring rule. Subjects are told that the ex ante probability of True News is 1/2. They are also asked to give an updated median guess after seeing the message, and are again incentivized with a linear scoring rule. For details, see Section B.3.

Subjects see News pages after their corresponding Question page, in the order: Question 1, News 1, Question 2, News 2, ... At the end of the experiment, they see a Results page with details on all the correct answers, points scored, and money earned.

At the end of the experiment, subjects' points earned on each part of the experiment are averaged. Subjects are paid a $3 show-up fee and have a probability of winning a $10 bonus equal to their average score divided by 1000. This probabilistic bonus is designed to eliminate potential hedging and risk-aversion confounds.

2.3 Data

The experiment was conducted on Amazon's Mechanical Turk (MTurk) platform. MTurk is an online labor marketplace in which participants choose "Human Intelligence Tasks" to complete. MTurk has become a very popular way to run economic experiments (e.g. Horton, Rand, and Zeckhauser 2011; Kuziemko et al. 2015), and Levay, Freese, and Druckman (2016) find that participants generally tend to have more diverse demographics than students in university laboratories on dimensions like age and politics. The experiment was coded using oTree, an open-source software based on the Django web application framework developed by D. Chen, Schonger, and Wickens (2016).

Wave 1 was conducted on July 8-9, 2019, and asked about the leukemia survival rates question. Wave 1 additionally included political and performance questions that were part of a separate experiment. Wave 2 was conducted on October 1-2, 2019, and asked about the other four questions. Both waves were offered to MTurk workers currently living in the United States who had not previously taken one of my motivated reasoning experiments. 522 participants from Wave 1 and 508 participants from Wave 2 passed simple attention and comprehension checks.[2] Wave 1 also included politicized questions and tested a debiasing treatment

[2]In order to pass these checks, subjects needed to perfectly answer the comprehension check ques-

that is unrelated to this paper; only participants in the control group are included here, and only observations on the positivity questions are kept.

Subjects in Wave 1 answer one question about positivity; subjects in Wave 2 answer four questions about positivity and one neutral question. There are a total of 3,062 guesses to these questions. Zero guesses were exactly correct.[3] There are therefore a total of 3,062 news assessments. 2,554 assessments are Positive or Negative, and 508 assessments are on the neutral topic.

The balance table for the Positive / Negative treatment is in Section B.1.1. Since this randomization is within subject, treatments are expected to be balanced across demographics. The overall shares of Positive and Negative are not statistically significantly different, indicating that there is not substantially different attrition.

2.4 Results

2.4.1 Raw Data

This subsection shows that the raw data does not support positivity- or negativity-motivated reasoning, and the following subsection shows the relevant regressions.

The mean assessment of Positive news is 57.7 percent (s.e. 0.7 percent) and the mean assessment of Negative news is 58.5 percent (s.e. 0.8 percent).[4] The difference between these is -0.7 percentage points; this point estimate is statistically insignificantly different from zero ($p = 0.457$).

Figure 2.1 shows the empirical distributions of assessments for Positive and Negative news, which substantially overlap.

Likewise, there is no evidence that current beliefs are reflective of motivated beliefs on these questions. The mean assessment of Error-Accentuating news is 58.0 percent (s.e. 0.7 percent)

tion in Section B.2 (by giving a correct answer, correct bounds, and answering the news assessment with certainty). In addition, many questions had clear maximum and minimum possible answers (such as percentages, between 0 and 100). Subjects were dropped if any of their answers did not lie within these bounds.

[3] This suggests, reassuringly, that subjects did not look up the correct answers.

[4] All standard errors are clustered at the individual level.

and the mean assessment of Error-Mitigating news is 58.2 percent (s.e. 0.8 percent). The difference between these is -0.3 percentage points; this point estimate is statistically insignificantly different from zero ($p = 0.800$).

Figure 2.2 shows the empirical distributions of assessments for True News and Fake News, which substantially overlap.

2.4.2 Main Specifications

Due to subjects in the two waves seeing different questions, the main specification is between subjects. In particular, subjects in Wave 1 only see one positivity-related question, so the within-subject test essentially ignores this sample.

In particular, the main specification for positivity-motivated reasoning is in Table 2.3, column 1. The regression looks at assessments a for subject i, question topic q, and round r with fixed effects for q and r when all news is Positive or Negative:

$$a_{iqr} = \alpha + \beta \cdot 1(\text{Pro-Party})_{iqr} + \gamma z_i + \delta FE_q + \zeta FE_r + \epsilon_{iqr}$$

z_i is a vector of controls. The controls used are age, an indicator for political party, an indicator for race, an indicator for gender, log(income), years of education, and an indicator for whether the subject is part of a religious group.

Column 2 uses the within-subject design; the standard error is larger, but the coefficient does not substantially change. In order to test whether there are differences compared to Neutral news, column 3 includes indicators for both Positive (versus Neutral) news and Negative (versus Neutral) news.

Columns 4 and 5 regress assessments on an indicator for True News (as opposed to Fake News), with and without controls for positivity. Recall from Section 2.2 that this measures whether directional errors in current beliefs are partly explained by past motivated reasoning on these topics, and therefore whether we should expect much heterogeneity in motive direction.

Figure 2.1: Histogram of Perceived Veracity of Positive and Negative News

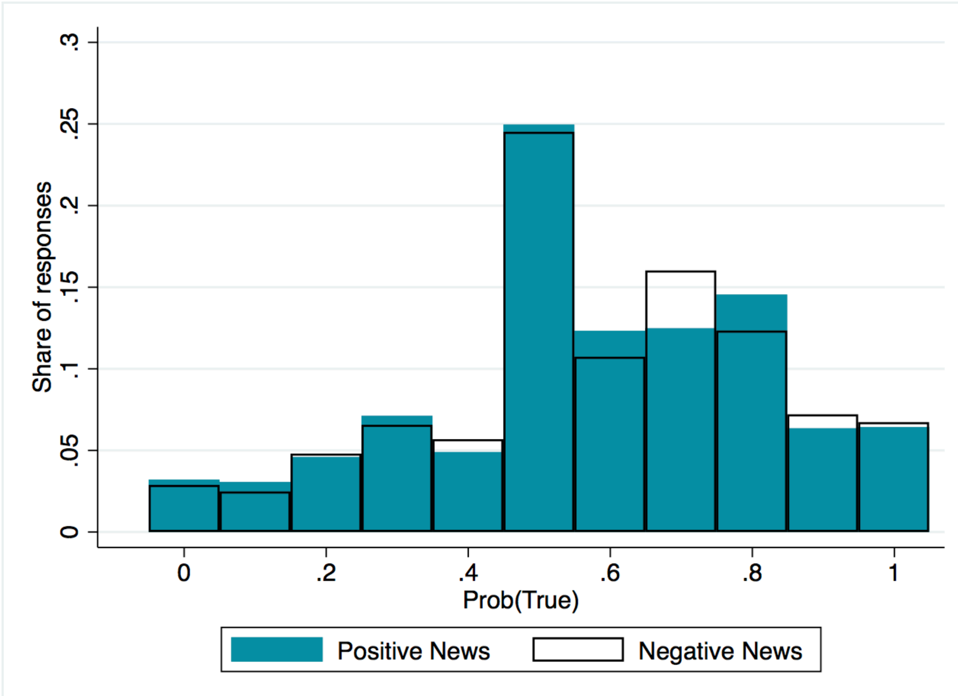
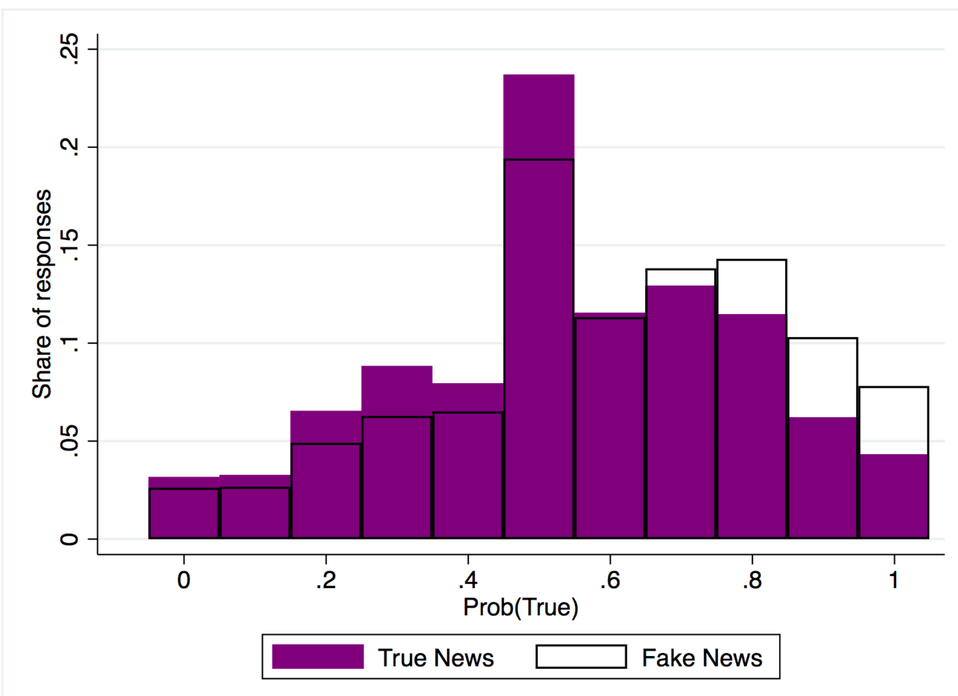


Figure 2.2: Histogram of Perceived Veracity of True and Fake News About Positivity



Notes: Only Positive / Negative news observations, as defined in Table 2.1. Messages are customized so that Bayesians give the same assessment for Positive and Negative news.

Table 2.3: The Effect of News Direction and Actual Veracity on Perceived Veracity

	(1)	(2)	(3)	(4)	(5)
Positive News	-0.005	-0.017	0.007		-0.005
	(0.010)	(0.016)	(0.016)		(0.010)
Negative News			0.012		
			(0.016)		
True News				-0.002	0.000
				(0.010)	(0.011)
Neutral News	No	No	Yes	No	No
Question FE	Yes	Yes	Yes	Yes	Yes
Round FE	Yes	Yes	Yes	Yes	Yes
Subject controls	Yes	No	Yes	Yes	Yes
Subject FE	No	Yes	No	No	No
Observations	2554	2554	3062	2554	2554
R^2	0.03	0.49	0.03	0.03	0.03
Mean	0.581	0.581	0.577	0.581	0.581

Standard errors in parentheses

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

Notes: OLS, errors clustered at subject level. Neutral News indicates that Positive / Negative news assessments are compared to assessments on Neutral topics. These classifications are defined in Table 2.1. Controls: age, political party, race, gender, log(income), years of education, and member of religious group.

Every single coefficient in Table 2.3 is insignificant and each point estimate is within 2 pp of zero. Modest effect sizes can be ruled out at the 95-percent significance level. There is no evidence for aggregate-level positivity-motivated reasoning or negativity-motivated reasoning. There is no evidence that subjects infer differently on positivity-related topics compared to the neutral topic. There is no evidence that subjects have formed erroneous beliefs in the direction of their motivated reasoning.

An alternative measure of motivated reasoning, which looks at subjects update their beliefs about the original question, generates the same prediction. After seeing the message, subjects' updated median belief is elicited. 38 percent of the time, subjects update in the positive direction (s.e. 1 percent). 38 percent of the time, subjects update in the negative direction (s.e. 1 percent). 24 percent of the time, subjects stay with their original guess (s.e. 1 percent). Clearly, there is no systematic updating in the positive or negative direction.

2.4.3 Heterogeneity

The results above show that there is no aggregate evidence for positivity- or negativity-motivated reasoning. This may be because nobody engages in motivated reasoning or because there is some heterogeneity with mean zero. This section discusses two forms of heterogeneity: heterogeneity across people, and heterogeneity across questions.

As shown in columns 4 and 5 of Table 2.3, there is no evidence that subjects have formed their current beliefs on these topics because of past motivated reasoning. This suggests that the degree of heterogeneity across people is not likely to be large. In support of this, Table 2.4 shows that treatment effects are not especially heterogeneous across demographic groups in systematic ways.

Table 2.4: Heterogeneity in Positivity-Motivated Reasoning: Horse Race Regression

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Pos News x Male	0.03							0.02
	(0.02)							(0.02)
Pos News x (Age>32)		-0.04*						-0.03*
		(0.02)						(0.02)
Pos News x White			-0.01					-0.01
			(0.02)					(0.02)
Pos News x College				-0.02				-0.01
				(0.02)				(0.02)
Pos News x (Inc>50K)					-0.01			-0.01
					(0.02)			(0.02)
Pos News x Democrat						-0.03		-0.03
						(0.02)		(0.02)
Pos News x Republican						0.04		0.05*
						(0.03)		(0.03)
Pos News x Religious							-0.00	-0.01
							(0.02)	(0.02)
Pos News	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Question FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Round FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Subject controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2554	2554	2554	2554	2554	2554	2554	2554
R^2	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03
Mean	0.58	0.58	0.58	0.58	0.58	0.58	0.58	0.58

Standard errors in parentheses

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

Notes: OLS regression coefficients, errors clustered at subject level. FE included for round number and topic. Only Positive / Negative news observations, as defined in Table 2.1. Religious: subject affiliates with any religion. Controls: age, political party, race, gender, log(income), years of education, and religious.

The evidence is not precise enough to clearly argue in favor of or against between-topic heterogeneity. On one question (regarding global poverty rates), there is statistically significant evidence for *negativity*-motivated reasoning: the difference between Negative and Positive news assessments is 11 pp (s.e. 3 pp).[5] The treatment effect on news assessments for the other questions is neither large nor significant at 95 percent confidence levels.[6] One direction for future work is to extend the domain of topics analyzed to determine whether the global poverty question is an outlier or in a different motivated category such as social comparisons.

2.5 Discussion

Results from the experiment indicate that there is no evidence for positivity- or negativity-motivated reasoning. To better understand the precision of these results, it is helpful to compare the results to those in Chapter 1. Since the experimental design is the same, the comparison has the same units. Treatment effects for the three categories are plotted in Figure 2.3.

It is easily apparent that the effect of Positive news is significantly different than the effect of seeing Pro-Party news or Pro-Performance news.

There are two sets of explanations for the experimental results. First, motives may be systematically different from belief-based utility, leading people to distort how they process information differently in the self-image-relevant domains from the positivity domains. Second, people may not actually receive utility from holding beliefs from positivity.

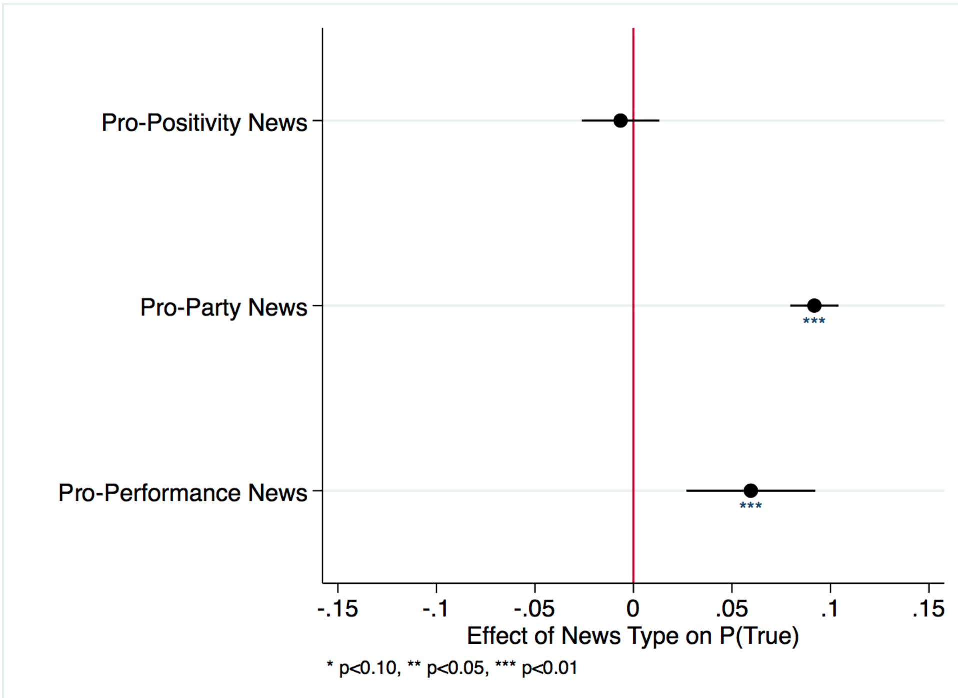
As a suggestive test to separate these different hypotheses, I run two follow-up surveys among a new group of subjects — drawn from different Mechanical Turk samples — on January 8-9 and 13, 2019. There are 303 participants in Survey 1 and 167 in Survey 2.[7] The ev-

[5]There are several potential explanations for this. It may be noise, and it may also be a question on which self-image does play a role. In particular, people may engage in social comparison and be motivated to believe that many others are less well-off.

[6]The overall effect is still close to zero if we remove the global poverty question. In the between-subject test, the coefficient on Positive news is 0.019 (s.e. 0.11; $p = 0.068$); in the within-subject test, the coefficient is 0.009 (s.e. 0.19; $p = .647$). Effects of larger than 4 pp lie outside the 95-percent confidence interval.

[7]This number does not include 16 subjects in Survey 1 and 5 subjects in Survey 2 who failed a simple attention check question. Results do not qualitatively change with the inclusion of these subjects.

Figure 2.3: Comparing Motivated Reasoning About Positivity to Politics and Performance



Notes: Treatment effects for the effect of Positive versus Negative, Pro-Party versus Anti-Party, and Pro-Performance versus Anti-Performance on news veracity assessments. Pro-Party and Pro-Performance coefficients from Chapter 1. Error bars correspond to 95 percent confidence intervals.

idence is consistent with the hypothesis that motives are systematically different from belief-based utility, and that survey participants are aware of this.

In Survey 1, participants were given a definition of motivated reasoning and asked to predict the direction of motivated reasoning about positivity, politics, and performance, and given sample topics on all three. On positivity, they were asked whether they thought that most people motivatedly reasoned in the direction of believing that the world was a better place for others, most people motivatedly reasoned in the direction of believing the world was a worse place for others, or about the same. The example topics were infant mortality, happiness, and cancer survival rates.

The results from this survey are shown in Figure 2.4. 65 percent of subjects expect motivated-reasoning distortions in the Pro-Party direction, versus only 16 percent who expect distortions in the Anti-Party direction. Similarly, 56 percent expect Pro-Performance distortions, and only 18 percent expect Anti-Performance distortions. Subjects, however, were similarly likely to predict distortions in the Positive (36 percent) and Negative (30 percent) directions; this difference is not statistically significant ($p = 0.231$). They were also more likely to predict no directional distortions about positivity (34 percent) than about party (18 percent) or performance (26 percent). Each answer about positivity is statistically significantly different from each answer about party and performance at the 5-percent level.

This suggests that the experimental results — that motivated reasoning occurs about politics and performance, but not directionally about positivity — are anticipated by the sample as a whole.

In Survey 2, participants were given the same categories and examples, but now asked to predict whether such beliefs would lead people to be *happier*. The results are shown in Figure 2.5. Unlike with motivated reasoning, a clear majority of 69 percent believes that positivity makes people happier, and only 10 percent believes that negativity increases happiness. Politics and performance have similar and statistically indistinguishable point estimates.

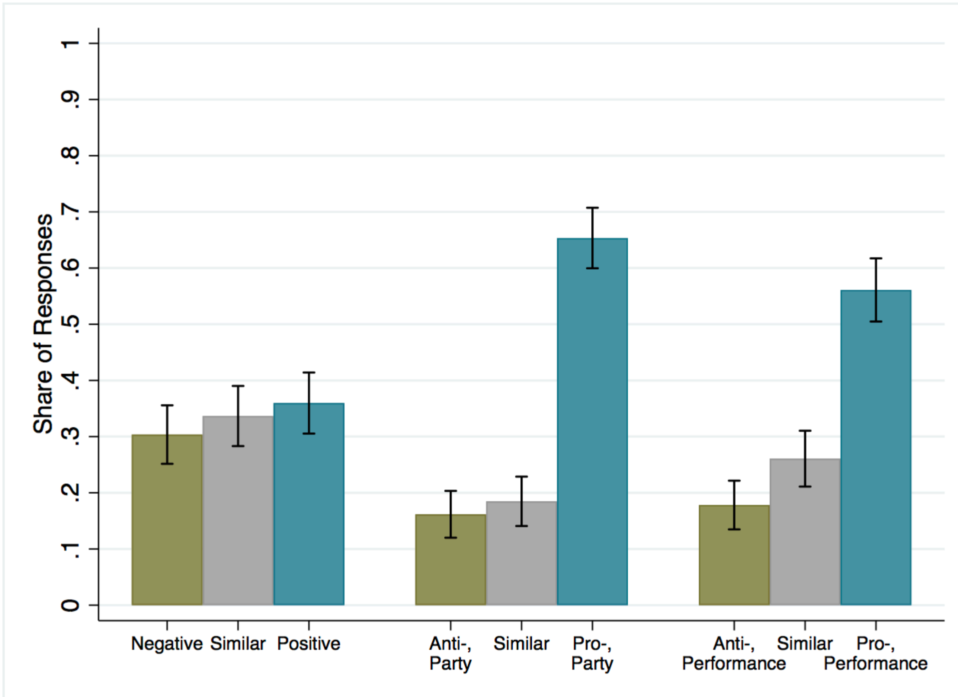
Taken as a whole, the evidence is in support of the explanation that people do receive utility (in the form of happiness) from believing that the world is good for others, but not in support of this form of positivity influencing motives. That is, motives do not include other-regarding belief-based utility. This explanation may also help explain action-induced motivated beliefs, where people distort their beliefs as an excuse to not be generous to others (Exley 2015; Di Tella et al. (2015)).

2.6 Conclusion

This paper has shown that people do not necessarily motivatedly reason in the direction of "good" states when their self-image is not at stake. When asked about positive or negative news about others, this experiment finds no evidence of systematic directional distortions of how people process the information. It also does not indicate any evidence that people's current beliefs are distorted due to such positivity- or negativity-motivated reasoning.

Survey results additionally show that people think that believing in positive states of the

Figure 2.4: People Systematically Expect Motivated Reasoning About Politics and Performance, But Do Not About Positivity

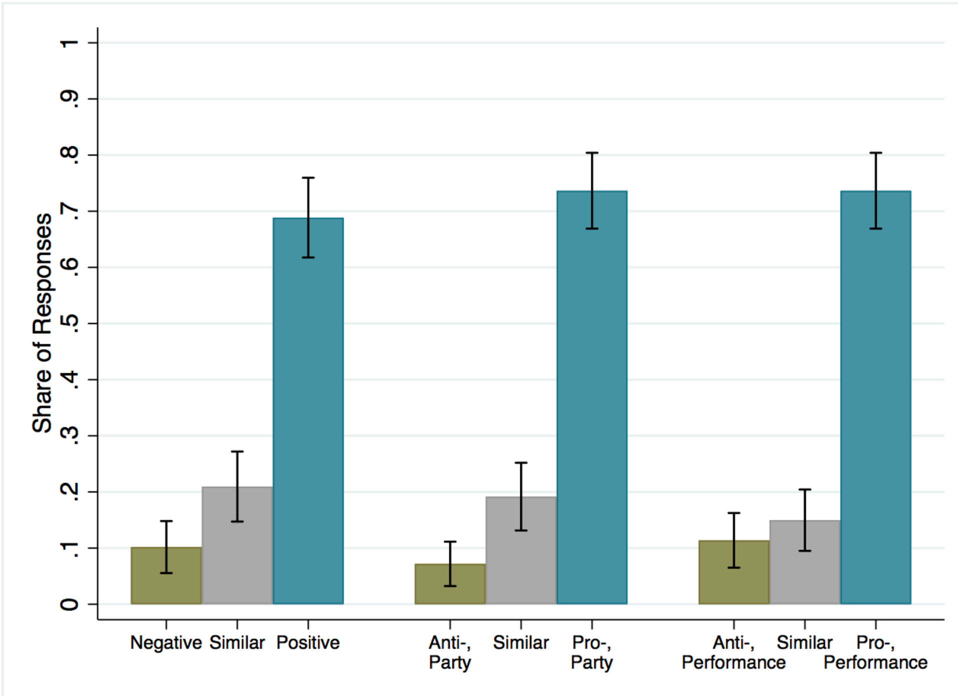


Notes: The y-axis is the share of respondents who stated that they expect most people to have motivatedly reason in one direction, the other direction, or a similar amount in both directions. 0.4 percent of questions are left unanswered, and are coded as "Similar." Error bars correspond to 95 percent confidence intervals. The differences between each of the Positivity bars and their corresponding bars in the Party and Performance columns are statistically significant at the 95 percent level.

world does not induce motivated reasoning. However, people do think that believing in these positive states leads to increased happiness. The results from this paper also suggest that utility-maximizing beliefs do not necessarily explain why people form persistently inaccurate beliefs.

One direction for future work is to better understand the relationship between the belief-based utility function and the motive function. For instance, do particular emotions affect utility and motives differently? These results suggest that happiness is insufficient for motives; however, in many self-image domains, pride and identity confirmation play larger roles than happiness. An alternative hypothesis in psychology, proposed by von Hippel and Trivers (2011), is that self-deception is a mechanism by which people can deceive others. Convincing others that the world is a good place can be less impactful than convincing others that one is

Figure 2.5: People Expect Happiness Due to Beliefs About Positivity, Not Just About Politics and Performance



Notes: The y-axis is the share of respondents who stated that they expect most people to be happier when receiving news in one direction, the other direction, or a similar amount in both directions. Error bars correspond to 95 percent confidence intervals. The differences between each of the Positivity bars and their corresponding bars in the Party and Performance columns are not statistically significant at the 95 percent level.

smarter, more altruistic, or more correct.

Once the tether between the motive function and the utility function is removed, future work can treat the motive function as a separate object for study. The objective is for this experimental design to then be used to elicit motives and analyze their role in other economic contexts.

3

Polarization and Public Health: Partisan Differences in Social Distancing during the Coronavirus Pandemic

3.1 Introduction

Mobilizing an effective public response to an emerging pandemic requires clear communication and trust (Holmes 2008; Taylor et al. 2009; van der Weerd et al. 2011; Vaughn and Tinker 2011). Risk reduction measures such as social distancing and self-quarantine can rarely be enforced entirely by coercion, particularly in democratic societies. The public must understand what is required of them and be persuaded of the importance of complying.

Partisian differences could play a key role in determining how Americans respond to the COVID-19 pandemic. Prominent officials have sent conflicting messages about the crisis, with

President Trump and other Republican officials sometimes saying it was less severe, and Democrats giving more emphasis to its dangers (Beauchamp 2020; Stanley-Becker and Janes 2020; Coppins 2020; McCarthy 2020). Partisan media have tended to echo this division (Aleem 2020; Kantrowitz 2020). This could cause differences between people on the right and left in the extent of risk reduction measures such as social distancing, with potentially important effects on human health and the economy.

In this paper, we combine GPS location data from a large sample of smartphones with a new survey to study partisan differences in the early response to COVID-19. The GPS data are collected by the company SafeGraph, and record daily and weekly visits to points of interest (POIs), including restaurants, hotels, hospitals, and many other public and private businesses. Our primary analysis focuses on the period from January 26, 2020 to April 4, 2020.

We begin with two motivating facts. First, recent nationwide surveys have shown that Democrats are more concerned about the spread of COVID-19 than Republicans. Second, Democrats report taking more steps to avoid infection than Republicans. We note, however, that Democratic areas also have had more coronavirus cases and implemented stay-at-home policies earlier. The raw differences observed on surveys could simply be the expected result of local differences in risk or regulation, rather than an effect of partisanship per se.

We then present a simple model that clarifies the potential causes and consequences of divergent social-distancing behavior. It combines a standard epidemiological model of a pandemic with an economic model of optimizing behavior by heterogeneous agents. The model clarifies that divergent responses between groups need not be inefficient. One group might engage in less social distancing because their costs of distancing are greater (e.g., they would lose more income as a result) or because their benefits of distancing are smaller (e.g., they are at lower risk of infection). However, differences in behavior resulting from divergent *beliefs* of otherwise similar agents do suggest systematic inefficiency, as optimizing based on different beliefs means that the marginal costs of social distancing are not equated across people. In that case, society gets less social distancing at higher cost than if agents had the same beliefs.

Our main GPS results show that the strong partisan differences in social distancing behavior that emerged with the rise of COVID-19 are not merely an artifact of differences in state policies or observed risks. Controlling for state-time fixed effects to account for heterogenous policy responses by state governments only attenuates the partisan gap slightly. Including controls to proxy for health and economic variables interacted flexibly with time attenuates the gap more substantially, but it remains statistically and economically significant. After including our full set of controls, we estimate that moving from the 10th to the 90th percentile of Republican county vote share is associated with an 18.6 percent increase in the number of POI visits during the week of March 29.

Our findings are robust to the inclusion or exclusion of control variables, excluding states with early COVID-19 outbreaks, or dropping highly populated counties. Replacing the continuous measure of partisanship with discrete indicators for portions of the Republican vote share distribution or restricting the sample to counties from certain portions of the distribution does not change our qualitative conclusions. Furthermore, there is no evidence of a similar partisan gap during the same period in 2019.

To complement the data showing county-level differences in behavior, we use a nationally-representative survey to show that beliefs about social distancing are partisan as well. We collect participants' demographics (including party affiliation), beliefs regarding the efficacy of social distancing, self-reported distancing due to COVID-19, and predictions about future COVID-19 cases. Compared to Republicans, we find that Democrats believe the pandemic is more severe and report a greater reduction in contact with others. In our survey, we also randomly vary whether predictions about future COVID-19 cases are incentivized, and do not find evidence that incentives reduce the partisan gap, suggesting that these predictions are less likely to be due to partisan cheerleading (as in Bullock et al. 2015 and Prior et al. 2015), and more likely to reflect true differences in beliefs. A back-of-the-envelope calculation of our model estimates that the deadweight loss cost due to these partisan differences is $2.7 billion per year.

Several contemporaneous studies also measure partisan differences in responses to COVID-

19.[1] Gadarian et al. (2020) present survey evidence showing partisan gaps in self-reported responses to the pandemic. Barrios and Hochberg (2020) show differences between Republican and Democratic areas in the frequency of COVID-related queries on Google and in movement patterns as measured in GPS data from a different source than the one we use here. Painter and Qiu (2020) examine partisan heterogeneity in response to state-level, stay-at-home orders.

Our work contributes to a broader literature on what drives responses to pandemics (e.g., Blendon et al. 2008; Vaughan and Tinker 2009; Fineberg 2014). Risk perception, behavior changes, and trust in government information sources change as pandemics progress (Ibuka et al. 2010; Bults et al. 2011). Demographic characteristics, such as gender, income, geography, or social interactions, are important determinants of the adoption of recommended public health behaviors (Bish and Michie 2010; Ibuka et al. 2010; Bults et al. 2011; Chuang et al. 2015; Shultz et al. 2016; Gamma et al. 2017).

A related literature focuses on the consequences of political polarization for health behaviors (e.g., Iyengar et al. 2019 and Montoya-Williams and Fuentes-Afflick 2019). Party affiliation is correlated with physician recommendations on politicized health procedures, enrollment in government exchanges created under the Affordable Care Act, and beliefs in the safety of vaccines (Hersh and Goldenberg 2016; Lerman et al. 2017; Sances and Clinton 2019; Trachtman 2019; Krupenkin 2018; Suryadevara et al. 2019). We show how partisan differences can lead to the inefficient allocation of public health goods, such as social distancing, during pandemics.

Our work also relates to a broader literature on partisan differences in trust and beliefs. For instance, a large body of empirical literature documents partisan differences in beliefs about factual events such as unemployment (Bartels 2002; Gaines et al. 2007; Bullock et al. 2015). There exists a growing literature on building theoretical models of opinion polarization to explain observed partisanship (Dixit and Weibull 2007; Benoit and Dubra 2015; Ortoleva and Snowber 2015; Fryer et al. 2019). Furthermore, a substantial empirical literature studies the

[1]Coverage in the media and some studies examine partisan heterogeneity in response to COVID-19 with no or few controls for differential risk exposure or costs of social distancing (e.g., Economist 2020; Andersen 2020). Baker et al. (2020) use transaction-level data and examine heterogeneity in consumption responses to COVID-19.

link between media markets and political polarization (Glaeser and Ward 2006; McCarty et al. 2006; Campante and Hojman 2013; Prior 2013).

Finally, our work adds to the increasing number of papers using GPS or related data to study social interactions. For example, Dubé et al. (2017) test the effectiveness of mobile targeting with coupons to competing movie theaters based on consumers' real-time location. Hanna et al. (2017) use data from Google Maps to estimate the effects of lifting high-occupancy vehicle restrictions in Jakarta, Indonesia.[2] Chen and Rohla (2018) and Athey et al. (2019) use SafeGraph data to measure the effects of political polarization on the length of Thanksgiving dinners and to estimate a novel measure of racial segregation, respectively.

Sections 3.2, 3.3, 3.4, 3.5, and 3.6, respectively, present our motivating facts, theoretical framework, data, GPS analysis, and survey results.

3.2 Motivating Facts

In this section, we present two basic facts on partisan diferences in social distancing. First, existing surveys document large differences in beliefs and social distancing by political party. Figure 3.1 presents results from previous national polls. Panel A shows that Democrats were consistently more concerned than Republicans about the spread of coronavirus in the United States from January 26 through the most recent polls in early April.

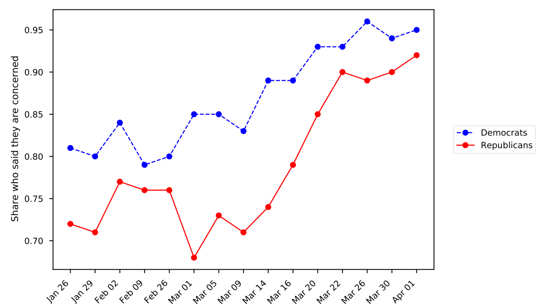
Second, consistent with beliefs, there exist partisan differences in self-reported social distancing behaviors. Panel B presents results from a March 13th poll, showing that Democrats were more likely to say they were eating at home more often, had stocked up on food and supplies, changed travel plans, and cancelled plans to avoid crowds. Panels C and D show that throughout the month of March, Democrats were more likely than Republicans to say that they were avoiding public places and small gatherings.

To interpret these differences, we need a framework to understand why people from the two political parties might behave differently, and why that might matter. For example, Democratic areas also have had more coronavirus cases and implemented stay-at-home policies ear-

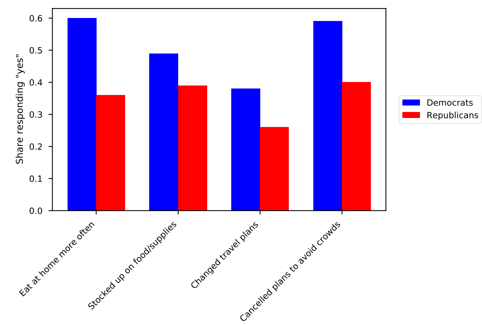
[2]See also Blattman et al. (2018) and Davis et al. (2019).

Figure 3.1: Partisan Differences in Perceived Risk and Social Distancing

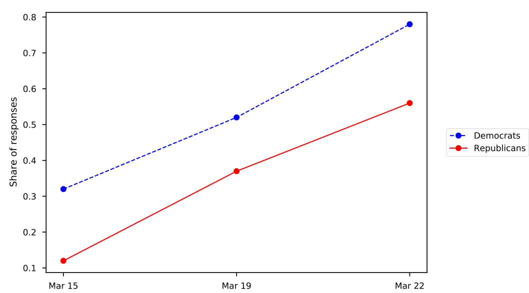
(a) Panel A: Concern over Spread of Coronavirus



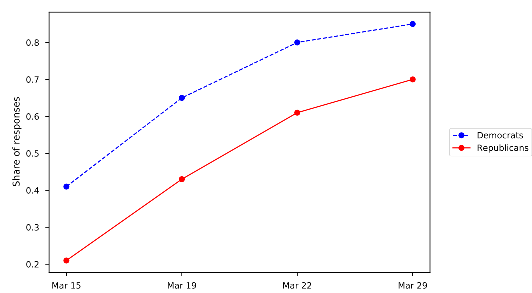
(b) Panel B: Behavior Change from Coronavirus



(c) Panel C: Share Avoiding Public Places



(d) Panel D: Share Avoiding Small Gatherings



Note: This figure shows responses to nationally representative polls by political affiliation. Panel A shows the share of people concerned about coronavirus spreading to the United States (Piacenza 2020). Panel B shows self-reported behavior change as of March 13-14 (Marist 2020). Panel C shows the share of people avoiding public places, such as stores and restaurants (Saad 2020). Panel D shows that share of people avoiding small gatherings, such as with friends and family (Saad 2020).

lier.

3.3 Stylized Model

In this section, we present a stylized model to clarify why it might matter if different types of people choose different amounts of social distancing. We embed an epidemiological model of disease transmission into an economic model with agents who maximize utility considering the expected private cost of disease.

3.3.1 Epidemiological Model

We use a discrete time version of the standard SIR epidemiological model (Kermack and McKendrick 1927). In each period t, each person is in one of four states $\sigma \in \{S, I, R, D\}$, representing Susceptible, Infected, Recovered, and Deceased. The share of the population in each state at time t is s_t, i_t, r_t, and d_t. Let β represent disease infectiousness, and let c_t denote an individual's amount of risky behavior at time t—for example, the amount of travel, dining out, failing to wash hands, and other activities that increase risk of becoming Infected.

All people begin in the Susceptible state. A Susceptible person becomes Infected at time $t+1$ with probability $c_t\beta i_t$ and stays Susceptible with probability $(1 - c_t\beta i_t)$. Infected people stay Infected for one period, after which they become Deceased with probability ψ or Recovered with probability $(1 - \psi)$. Both D and R are absorbing states.

Let θ index different types of people—for example, liberals and conservatives. Let $\omega_{\theta\sigma t}$ be a state variable representing the share of type θ that is in state σ at time t. The population is of measure 1, so $\sum_\theta \sum_\sigma \omega_{\theta\sigma t} = 1$.

3.3.2 Individual Decisions

People of type θ earn flow utility $u_\theta(c_t; \sigma_t)$, which depends on their risky behavior c_t and their state σ_t. People discount the future at rate δ and maximize expected lifetime utility $\sum_{\tau=t}^{\infty} \delta^\tau u_\theta(c_\tau; \sigma_\tau)$. Define $V_\theta(\sigma)$ as the expected lifetime utility of a person currently in state σ; note that this also implicitly depends on current and future population states $\omega_{\theta\sigma t}$. Being infected reduces utility, so we assume $V_\theta(S) > V_\theta(I)$.

We focus on Susceptible people, as they comprise most of the population during the period we study. We can write their maximization problem as a Bellman equation, in which people maximize the sum of utility from risky behavior today and expected future utility:

$$V_\theta(S_t) = \max_{c_t} \left\{ \underbrace{u_\theta(c_t; S_t)}_{\text{current utility from risky behavior}} + \underbrace{\delta \left[c_t\beta i_t V_\theta(I) + (1 - c_t\beta i_t) V_\theta(S) \right]}_{\text{expected future utility}} \right\}. \tag{3.1}$$

The first-order condition for privately optimal risky behavior is

$$\underbrace{u'_\theta}_{\text{marginal utility of risk}} = \underbrace{\beta i_t}_{\text{marginal infection probability}} \underbrace{\delta \left(V_\theta(S) - V_\theta(I)\right)}_{\text{private cost of infection}} . \tag{3.2}$$

The first-order condition shows that people choose their risky behavior to equate marginal benefit (more utility today) with private marginal cost (higher risk of infection, which reduces future utility). The equation illustrates that there are three reasons why risky behavior might vary across types. First is the marginal utility of risk (or equivalently, the marginal cost of social distancing): for example, people vary in how much they like travel and dining out, as well as in how easy it is to work from home. Second is the marginal infection probability: for example, local infection rate i_t differs across geographic areas. Third is the private cost of infection: for example, infection is more harmful for people who are older or have underlying health conditions.

3.3.3 Social Optimum

It is difficult to know for sure whether people take too many or too few steps to reduce disease transmission during our study period. Thus, we do not consider the optimal consumption of c. Instead, we hold constant the total amount of risky behavior and ask whether the allocation across types is optimal. Tangibly, this means that we are not asking, "how much social distancing should people be doing?" Instead, we are asking, "holding constant the amount of social distancing people are doing, would some people ideally be doing less, and others ideally be doing more?"

Social welfare is the sum of utility across all people in all states:

$$W_t = \sum_\theta \sum_\sigma \omega_{\theta\sigma t} V_\theta(\sigma_t). \tag{3.3}$$

Let C_t denote the total risky behavior at time t across all people. The (constrained) socially optimal outcome results from maximizing W_t subject to the constraint that $C_t = \bar{C}_t$. Let λ be the shadow price on that constraint; this reflects the loss from having too much or too little social distancing overall.

Consuming c imposes two types of externalities. First, it imposes a positive pecuniary externality, as travel, dining out, and other risky activities help keep firms in business and workers employed. Second, it imposes a negative externality by increasing the person's infection probability, which increases the expected stock of infected people in the next period (i_{t+1}), which increases other Susceptible people's infection risk. Let ϕ_t denote the net externality per unit of consumption, which may be positive or negative; this becomes more negative as the contagion externality grows. We assume that these externalities are constant across people, and that people do not account for them when setting their c_t^*.

In the constrained social optimum, Susceptible people's consumption of c_t would satisfy the following first-order condition:

$$0 = \underbrace{u'_\theta - \beta i_t \delta \left(V_\theta(S) - V_\theta(I)\right)}_{\text{private marginal utility}} + \underbrace{\phi_t}_{\text{externality}} + \underbrace{\lambda}_{\text{shadow price}} . \tag{3.4}$$

3.3.4 Heterogeneous Risk Misperceptions

We now allow people to misperceive risks. These misperceptions cause people to choose too much or too little risky behavior relative to their private optimum, and heterogeneous misperceptions cause transfers across types and efficiency losses.

We now add θ subscripts to explicitly denote different parameters by type. Let $\mu_{t\theta} := \beta i_t \delta \left(V_\theta(S) - V_\theta(I)\right)$ denote type θ's expected utility cost due to infection from an additional unit of risky consumption. Let $\tilde{\mu}_{t\theta}$ denote type θ's *perception* of that cost. Susceptible type θ consumers then set $c_{t\theta}$ according to the following modified first-order condition:

$$u'_\theta = \tilde{\mu}_{t\theta}, \tag{3.5}$$

giving consumption denoted $c^*_{t\theta}$.

For illustrative purposes, imagine that there are two types $\theta \in \{a, b\}$ in equal proportion, and that period t marginal utility is linear and the same for both types, so $u'_\theta(c) = u'(c)$ for both types and u'' is a constant. Finally, without loss of generality, assume that type a perceives greater risk, so $\tilde{\mu}_{a\theta} > \tilde{\mu}_{b\theta}$. Our survey data show that Democrats perceive greater risk, so one can think of Democrats as type a. We do not take a stand on which type perceives risk more correctly or which type's behavior is closer to the unconstrained social optimum.

Define $\bar{\tilde{\mu}}_t := \frac{1}{2}(\tilde{\mu}_{ta} + \tilde{\mu}_{tb})$ as the average risk perception. With homogeneous risk perceptions, both types would set c_t such that $u' = \bar{\tilde{\mu}}_t$, giving homogeneous consumption denoted $\bar{c}_t$. With heterogeneous misperceptions, type a consumes more and type b consumes less; the consumption difference is $c^*_{tb} - c^*_{ta} = \frac{\tilde{\mu}_{ta} - \tilde{\mu}_{tb}}{-u''}$. These consumption differences cause both transfers across types and efficiency losses.

Risk perceptions affect risky consumption, and risky consumption causes externalities, so the heterogeneous misperceptions cause transfers across groups. The net transfer from type a to type b from heterogeneous instead of homogeneous misperceptions is

$$\underbrace{\frac{\tilde{\mu}_{ta} - \tilde{\mu}_{tb}}{-u''}}_{\text{consumption difference}} \cdot \underbrace{\phi_t}_{\text{externality}} . \tag{3.6}$$

If $\phi_t > 0$, i.e. the positive pecuniary externality from risky consumption outweighs the negative contagion externality, then heterogeneous misperceptions cause a net transfer from type b to type a. Intuitively, we would say that Republicans are doing more to keep the economy going. On the other hand, if $\phi_t < 0$, i.e. the negative contagion externality outweighs the positive pecuniary externality, then heterogeneous misperceptions cause a net transfer from type a to type b. Intuitively, we would say that Democrats are doing more to reduce the spread of disease.

The efficiency cost in period t from heterogeneous instead of homogeneous misperceptions are the two deadweight loss triangles around $\bar{c}_t$, with total area:

$$\Delta W_t = \frac{s_t}{2} \cdot \frac{\left(\overbrace{\tilde{\mu}_{ta} - \bar{\tilde{\mu}}_t}^{\text{misperception}} \right)^2}{\underbrace{-u''}_{\text{slope of private marginal utility}}} . \tag{3.7}$$

Intuitively, type a people (Democrats) are doing too much social distancing, and type b (Republicans) too little, relative to the (constrained) social optimum with homogeneous risk perceptions. The marginal cost of social distancing is increasing: it's easy to start by avoiding going to a bar once a week, but eventually one's only contact with people is going to the grocery store for food, and it is quite costly to stop buying food. Thus, society could achieve the same amount of social distancing at lower cost if type a did less and type b did more.

This model informs the empirical tests in the rest of the paper. In Section 3.6, we ask if Democrats and Republicans have different risk perceptions, which would generate the transfers and efficiency costs described above. In doing so, we control for factors such as population density that could generate difference in actual risks across types. In Sections 3.5 and 3.6, we ask if Democrats and Republicans are reducing risk by different amounts. In doing so, we use proxies to control for differences in actual risks and marginal costs of risk reduction that could cause differential risk reduction to be socially optimal.

3.4 Data

3.4.1 SafeGraph Mobile GPS Location Data

Our analysis uses GPS data from SafeGraph, aggregating GPS pings from numerous mobile applications to measure foot traffic patterns to a collection of points-of-interest (POIs). POIs include retail shops, restaurants, movie theaters, hospitals, and many other public locations individuals may choose to go when leaving their house. For each POI, SafeGraph reports its geographic location, industry, and the total number of visitors in their mobile device panel that have visited each day.[3]

Our primary analysis uses data from a period of ten weeks, from January 26 to April 4, 2020. We aggregate visits across all POIs in a given county for a given week. We also separately aggregate visits by 2-digit NAICS code for each county and week. In a placebo analysis,

[3] SafeGraph removes POIs with fewer than five visitors in a given month for data through February 2020. For the March 2020 data, SafeGraph has released data on a weekly basis, rather than a monthly basis, and include all POIs with at least 1 visitor for these weekly releases.

we analyze data over earlier time periods (starting in January 2019).

We also use data from the SafeGraph Social Distancing data released as a part of their COVID-19 response. This data is available since January 1, 2020 and updated regularly. We use data over the same ten week period. This data contains alternative measures of social distancing beyond POI visits, such as the number of devices leaving their assigned geohash-7 home or the median time spent away from home across devices.

See the Appendix for additional information on the SafeGraph data construction.

We supplement the SafeGraph data with various other sources of county and census block group data. For demographic information on age, race, education, income, and poverty status at the county-level, we aggregate census block group data from SafeGraph Open Census to the county level.[4] For each county, we define county partisanship to be the proportion of total votes received by President Donald Trump in the 2016 election (MIT Election Data and Science Lab 2018). We use county-level data on COVID-19 cases and deaths from The New York Times (2020).

3.4.2 Survey

To supplement these data, we ran an online survey with a sample of American adults to study partisan gaps in beliefs about and responses to COVID-19 at the individual level. The survey was conducted from April 4-7 with Prime Panels from CloudResearch, a market research firm with access to 50 million participants. We recruited 2,000 participants to complete the study; participants are broadly representative of U.S. adults in terms of party affiliation, age, gender, and race.[5] Subjects who completed the survey were paid a show-up fee from CloudResearch and had the chance to earn additional bonus incentives of up to $100.

Participants were asked for their party affiliation on a seven-point scale, ranging from "Strongly Democrat" to "Strongly Republican." We interpret party continuously, where 0 represents "Strongly Democrat" and 6 represents "Strongly Republican." We also classify participants

[4]The SafeGraph Open Census data is derived from the 2016 5-year ACS at the census block group.

[5]In addition, we weighted observations so that age, gender, and race distributions match the 2010 Census data, and party affiliation matches the Gallup survey from March 13-22, 2020 (Gallup 2020).

into Republican (including independents who lean Republican) and Democrats (including independents who lean Democrat) for descriptive analyses.

The survey asked for demographic information (zipcode, age, race, gender, income, education, number of children, and health). It then asked about news consumption habits and trust before and during COVID-19. Then, there were several questions about social distancing: self-reported social distancing in response to COVID-19, beliefs about the risk of not distancing, and the appropriate trade-off between going out more to help the economy versus going out less to avoid spreading COVID-19.

We next elicited beliefs about the number of new COVID-19 cases that would be confirmed in the U.S. in April, 2020, as well as the approval rating of Donald Trump's response to the pandemic on April 30, and randomly vary whether these were incentivized or not. 1,013 (51 percent) subjects made incentivized predictions in which they earn more money if they are closer to the correct answer. They were told that we will randomly select 10 participants who will receive a payment of ($\$100 - \Delta$) where Δ is the percentage point difference between their answer and the true value. The remaining 987 (49 percent) of subjects were not incentivized. The primary four outcome variables are participants' answers to the three social-distancing questions and the one prediction question.

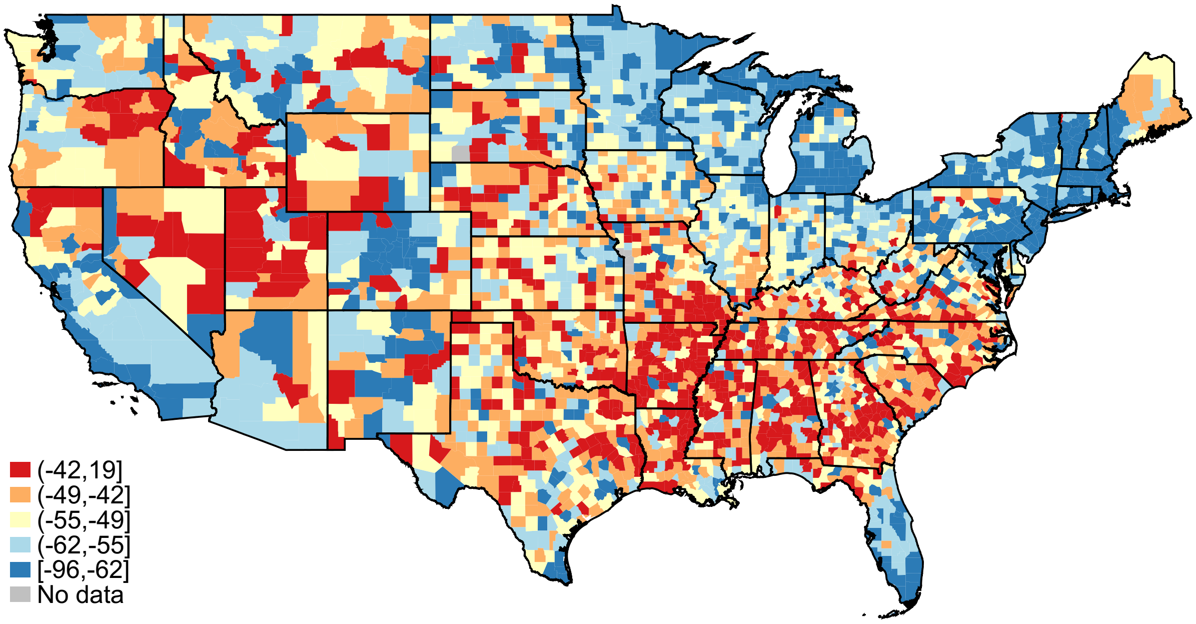
All survey questions are listed in Appendix C.2.3.

3.5 SafeGraph Empirical Specification and Results

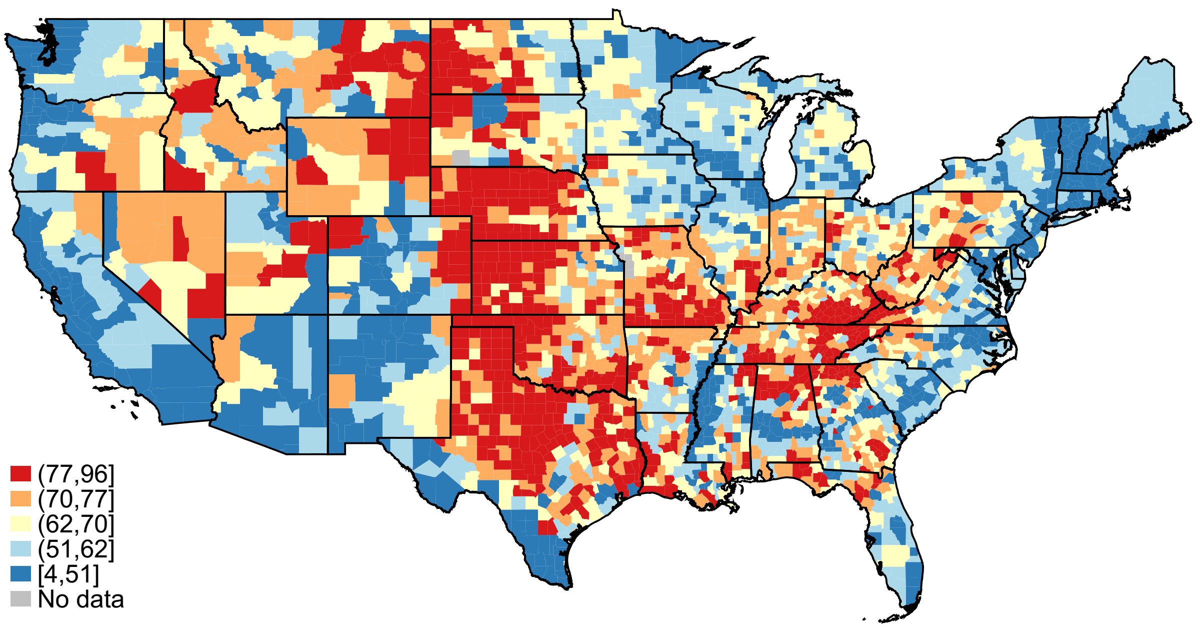
Figure 3.2 visualizes geographic variation in this social distancing response, and compares observed variation to analogous distributions of partisanship, COVID-19 confirmed cases, and public policy responses. Panel A maps the social distancing response observed in each county, as measured by the percent decrease in SafeGraph visits between the week beginning January 26th and the week beginning March 29th, using data described below. Panel B shades counties by their party affiliation, captured by the Republican vote share in the 2016 presidential election. Panel C maps the number of COVID-19 cases confirmed in a given county by April 4th. Panel D shades states by the effective start date for the earliest statewide "stay-at-home" order issued. Panels A and B exhibit a strong geographic correlation between the counties with weaker social distancing responses and those with higher Republican vote shares. Panel D shows that areas with stronger distancing responses also generally instituted earlier statewide, stay-at-home orders. We also observe stronger social distancing responses in counties with more COVID-19 confirmed cases (Panel C). However, these counties also had more coronavirus cases and were in states that initiated stay-at-home policies earlier. Thus, the partisan differences in social distancing could simply be the expected result of local differences in infection risk or regulation. In order to establish partisan difference in social distancing, we next exploit over time variations.

Figure 3.2: Geographic Variation in Social Distancing, Partisanship, COVID-19, and Public Policy

(a) Panel A: % Change in SafeGraph Visits



(b) Panel B: 2016 Republican Vote Share



Note: This figure shows the U.S. geographic distribution of social distancing, political affiliation, COVID-19, and public policy responses. Panel A shows for each county the percent change in aggregate visits between the week beginning January 26, 2020 and the week beginning April 12, 2020. Blue shading denotes a more negative percent change in visits during the latter week relative to the former. Red shading indicates an increase or a smaller decrease in visits. These visits are sourced from SafeGraph's mobile device location data. Panel B maps counties by the percentage of votes Donald Trump received in the 2016 presidential election. Red shading in this panel indicates more Republican counties (higher Trump vote share), and blue shading indicates more Democratic counties (lower Trump vote share).

Figure 3.3 reports trends in social distancing and COVID-19 prevalence separately for Republican and Democratic counties, defined to be counties above or below the median 2016 Republican vote share respectively. Panel A shows that the overall number of POI visits is relatively constant until COVID-19 cases begin emerging in the United States in March. During this same period, Democratic counties exhibited a sharper drop in weekly POI visits than their Republican counterparts. As Panel B demonstrates, Democratic counties also exhibited a much sharper rise in COVID-19 cases and deaths—accounting for nearly all verified COVID-19 cases and deaths through March 29. Appendix Figure C.1 shows that these declining and differential POI trends are not present over the same time period in 2019.

Our main empirical specification takes the following form

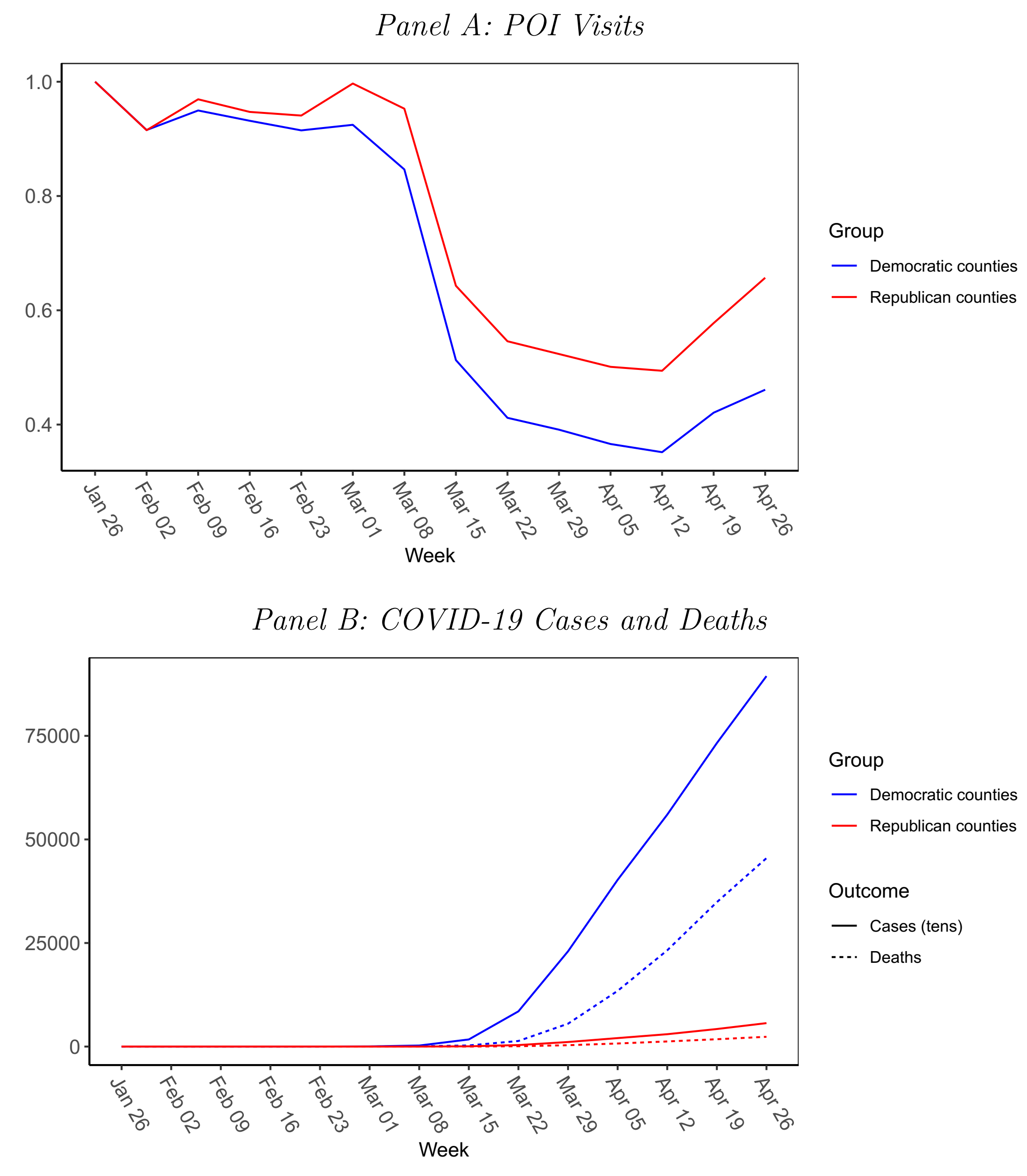
$$\log(c_{it}) = \alpha_t \rho_i + X_{it} \cdot \gamma_t + \epsilon_{it},$$

where c_{it} is the number of POI visits in county i during week t, α_t are the time-varying coefficients on county partisanship ρ_i, X_{it} are potentially time-varying controls, and ϵ_{it} is the county-specific error term.[6] In choosing our control variables X_{it}, we chose variables to flexibly control for the four channels of divergent behavior highlighted in equation (3.2). Standard errors are clustered at the county-level throughout unless specified otherwise.

Figure 3.4 reports our estimates of α_t under various sets of covariates chosen to incrementally control for the mechanisms highlighted by our model.

[6] In implementing, we normalize α_t relative to the first week.

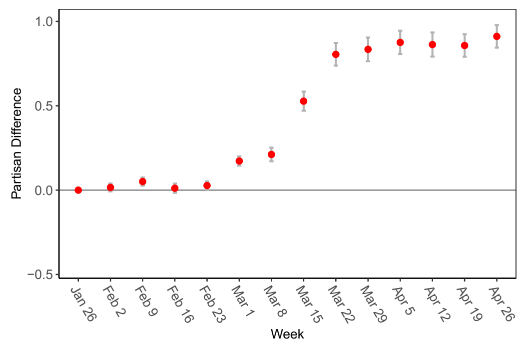
Figure 3.3: Social Distancing and COVID-19 Prevalence



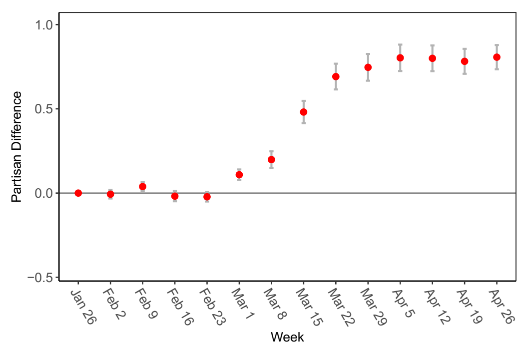
Note: Panel A shows the number of visits (normalized to one) to SafeGraph POIs for each week since January 26, 2020 for Republican counties and Democratic counties separately. Panel B is analagous but plots COVID-19 cases (in tens) and COVID-19 deaths. Republican counties are defined to be those whose 2016 Republican vote share is greater than the median vote share across the counties in our sample.

Figure 3.4: Partisan Differences in Social Distancing

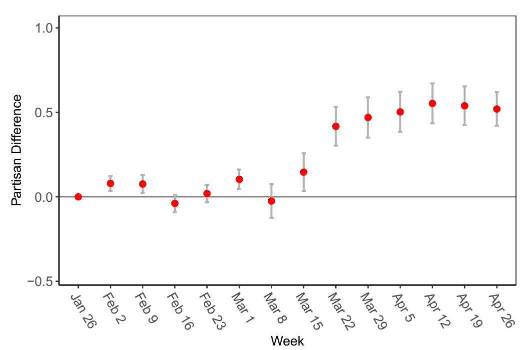
(a) Panel A: Only County & Time FE



(b) Panel B: Adds State-Time FE



(c) Panel C: Adds Health + Econ Controls



Note: Figure shows the estimated coefficients for county partisanship ρ_i on the log number of POI visits in the county using the specification outlined in the main text. For Panel A, only county and time fixed effects are included as controls. Panel B is the same as Panel A except state-time fixed effects replace the time fixed effects. Panel C is the same as Panel B except the health and economic covariates are included. The grey error bars indicate 95 percent confidence intervals constructed using standard errors clustered at the county-level.

In Panel A, we only include county and time fixed effects. This measures the extent to which these two groups' behavior diverges with the rise of COVID-19 via any of the aforementioned channels. Throughout February, there are no significant partisan differences in POI visits relative to the January 26 week baseline. However, as COVID-19 begins to emerge in the United States, partisan differences arise and grow throughout the weeks of March.

These results do not control for differences in state policies, which themselves may be a function of the partisan leanings of government officials. In Panel B, adding state-time fixed effects to control for state-level policies in response to COVID-19 along with other state-level temporal shocks causes the partisan differences to attenuate slightly.

In Panel C, we flexibly control for various health[7] and economic[8] characteristics of the county. We view the health controls as proxies for the marginal infection probability and the private cost of infection, and we view the economic controls as proxies for the marginal cost of social distancing, though each group of controls could proxy for other factors as well. We allow the coefficient on these variables γ_t to vary flexibly across time.

Although these controls attenuates the partisan differences to some degree, they remain economically and statistically significant. By the week of March 29, our estimate of α_t is 0.470. This implies that going from a county with the 10th to the 90th percentile in Republican vote share is associated with an 18.6 percent increase in the number of POI visits during the week of March 29.[9] Appendix Figure C.2 shows that these strong partisan differences do not appear over the same time period in 2019. We view these results as evidence of behavioral differences driven by partisan misperceptions of risks at the group-level, consistent with the survey evidence.

In Appendix Figure C.3, we report sensitivity to various alternative specifications. Panels A and B use alternative sets of controls. Panel C replaces the measure of partisanship with

[7]Health controls include log of one plus the number of confirmed COVID-19 cases in the county, the log of one plus the number of COVID-19 deaths in the county, the log of one plus the county population density (individuals per square kilometer) and the share of the population 65 years or older.

[8]Economic controls include the share of the population with at least a bachelor's degree, the share in poverty, and the shares of white, black, and asians.

[9]The difference between the 90th and 10th percentile of Republican vote share is 0.807 - 0.411 = 0.396.

a discrete indicator for certain quantiles of the Republican vote share distribution. Panel D drops counties with populations above half a million or states with early COVID-19 outbreaks (California, Washington, and New York). Panel E restricts the sample to counties from certain portions of the Republican vote share distribution. And, Panel F weights observations by the county's population, uses standard errors clustered at the state-level, and examines sensitivity to the start date. None of the alternative specifications change the central conclusion regarding partisan differences in social distancing in March.

Appendix Figure C.4 aggregates the number of POI visits at the electoral precinct level and shows that the qualitative conclusion of less social distancing by Republicans holds at the precinct level, even when including county-time fixed effects. Again, these patterns are not present in 2019 (Appendix Figure C.5).

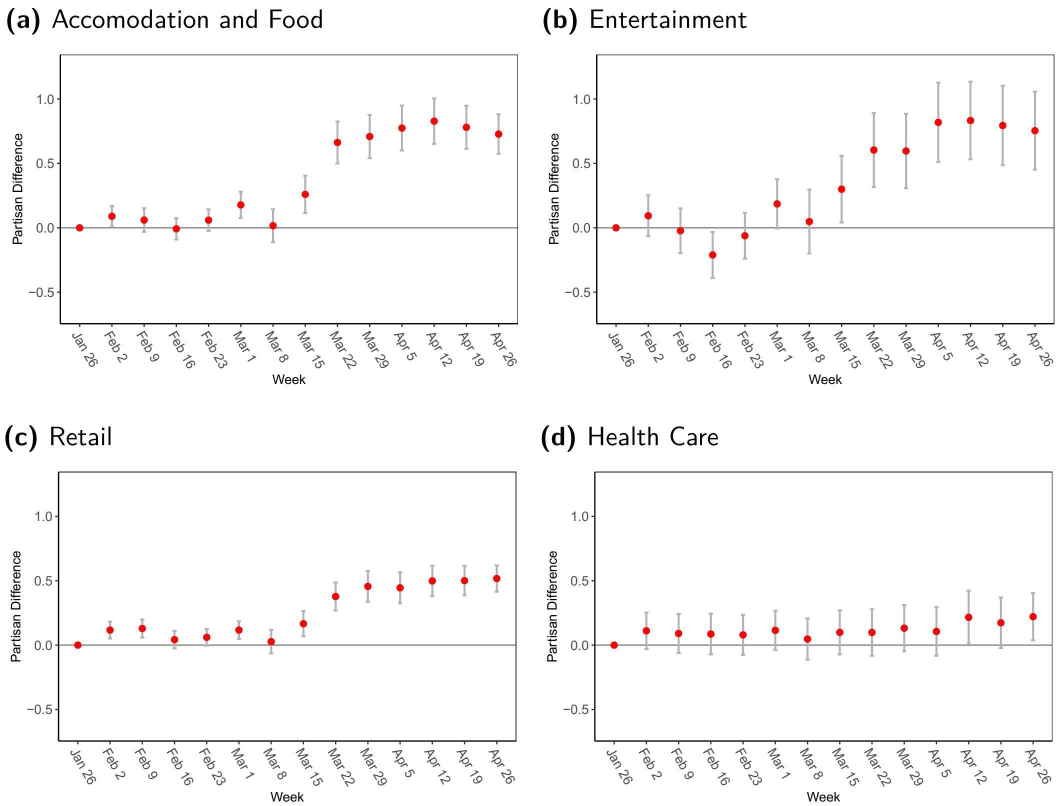
Figure 3.5 examines heterogeneity across 2-digit NAICS codes by re-aggregating POI visits to the county level after restricting to certain NAICS codes. Consistent with the narrative around COVID-19, we see the strongest partisan differences emerge with POIs in the accommodations and food, entertainment, and retail industries. There are no significant partisan differences in visits to health care POIs.

Figure 3.6 repeats Panel C of 3.4, but using POI visits aggregated at the day level. The partisan differences emerge in March for both weekdays and weekends, suggesting these differences are not driven solely by differences in work-from-home policies.

Figure 3.7 considers various alternative measures of social distancing derived from SafeGraph's Social Distancing data release as described in Section 3.4. Statistically significant partisan differences emerge in March for the log number of devices leaving home, the share of devices leaving home, and the total number of active devices.[10] For the log of the median time

[10] A key issue with the SafeGraph social distancing data is sample attrition. SafeGraph restricts the panel to devices with observed location pings in a given time period. For some applications, the frequency of location pings depends on device mobility. If devices are immobile at home or turned off, they may not generate location pings and would then be dropped from the sample. The total number of active devices changes over our sample period in a manner consistent with sample attrition. Given these issues, we prefer measures of social distancing derived solely from *external* activity (e.g., POI visits) that do not contain the same measurement error problems. We attempt to correct for the differential attrition in our measure of the share of devices leaving home (see Figure 3.7 footnotes for correction; see Panel G of Appendix Figure C.3 for estimates using the uncorrected measure).

Figure 3.5: Partisan Differences in Social Distancing by 2-Digit NAICS Code Industry



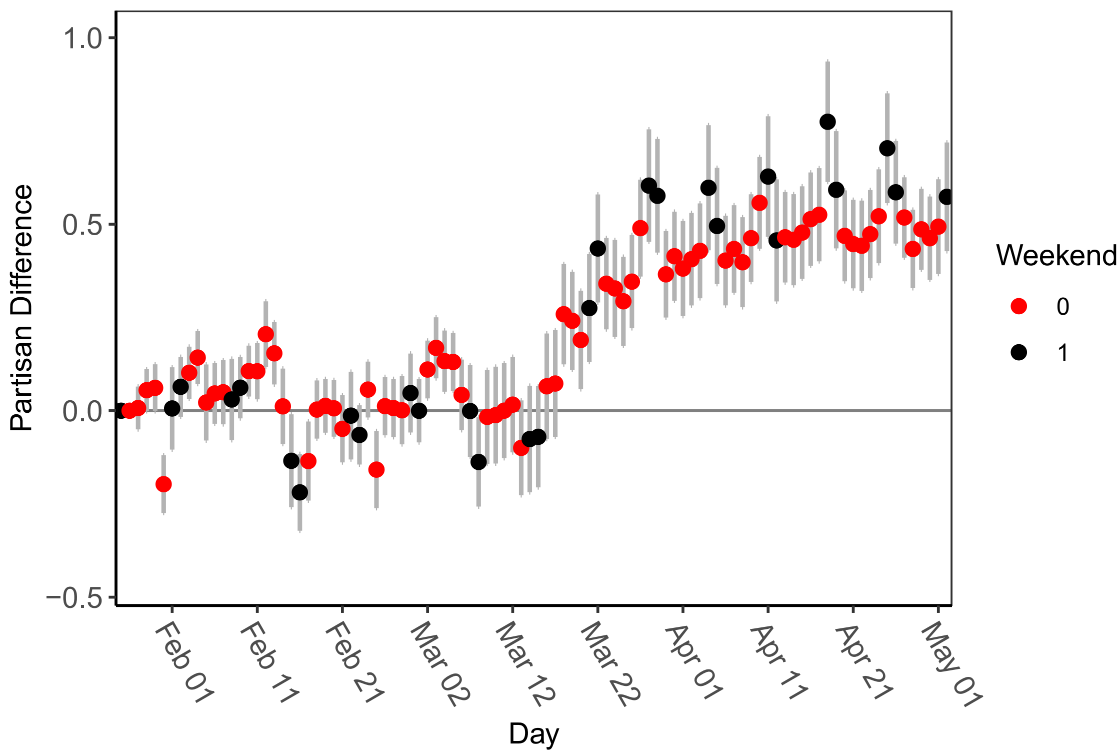
Note: Figure shows the estimated coefficients for county partisanship ρ_i on the log number of POI visits in the county after restricting POI visits to various 2-digit NAICS codes. The NAICS code groups are: Accomodation and Food (NAICS 72), Entertainment (NAICS 71), Retail Trade (NAICS 44 and 45), and Health Care (NAICS 62). The same controls are used as in Panel C of Figure 3.4. The grey error bars indicate 95 percent confidence intervals constructed using standard errors clustered at the county-level.

away from home, we see positive, but insignificant, point estimates.

3.6 Survey Results

Turning to the results of our survey, we first confirm that there indeed exists individual-level partisan differences in (self-reported) social distancing behaviors and attitudes, consistent with the POI visits results presented above, and then show that beliefs about the effectiveness of social distancing and predictions of the spread of COVID-19 follow the same partisan patterns.

Figure 3.6: Partisan Differences in Social Distancing, Daily



Note: Figure shows the estimated coefficients for county partisanship ρ_i on the log number of POI visits in the county. The same controls as in Panel C of Figure 3.4 are used except that state-time fixed effects occur at the day level and the weekday and weekend series are normalized separately. The grey error bars indicate 95 percent confidence intervals constructed using standard errors clustered at the county-level.

Our main empirical specification regresses normalized responses on each question on party:

$$y_i = \kappa + \alpha\rho_i + \gamma X_i + \epsilon_i,$$

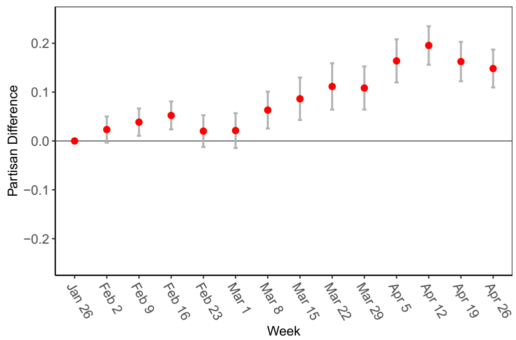
where y_i is the number of standard deviations above the mean for response i, ρ_i is the continuous measure of party lean from 0 to 1, X_i are demographic and location controls, and ϵ_i is an error term.

Figure 3.8 shows consistent evidence for partisan differences in social distancing, both with and without control variables.[11] On average, participants report reducing contact by 70.0 percent, with a standard deviation (SD) of 24.5 percent. After including controls, strong

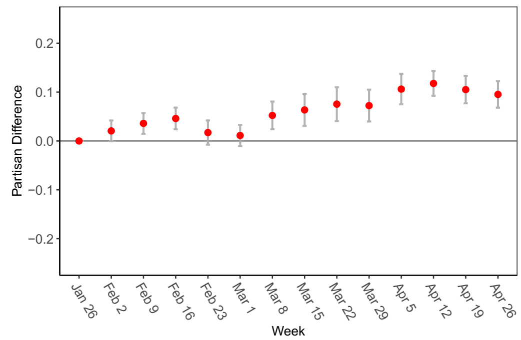
[11]These effects are not due to observation weighting, as shown in Appendix Figure C.6.

Figure 3.7: Partisan Differences in Social Distancing, Alternative Measures

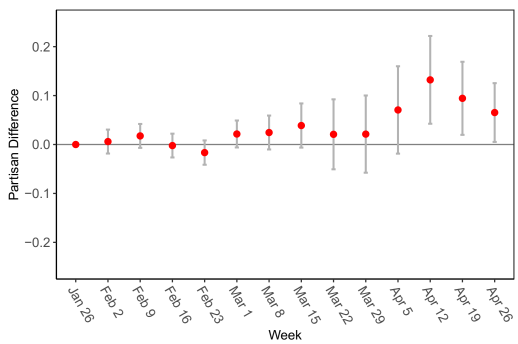
(a) Log Devices Leaving Home



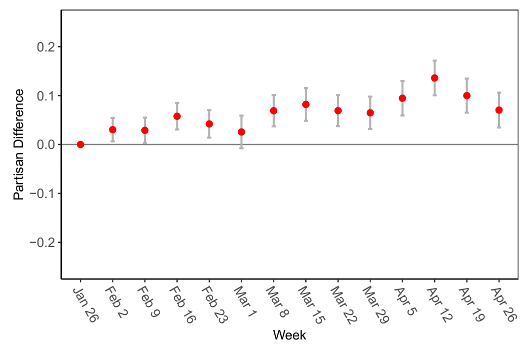
(b) Share Devices Leaving Home



(c) Log Median Time Away



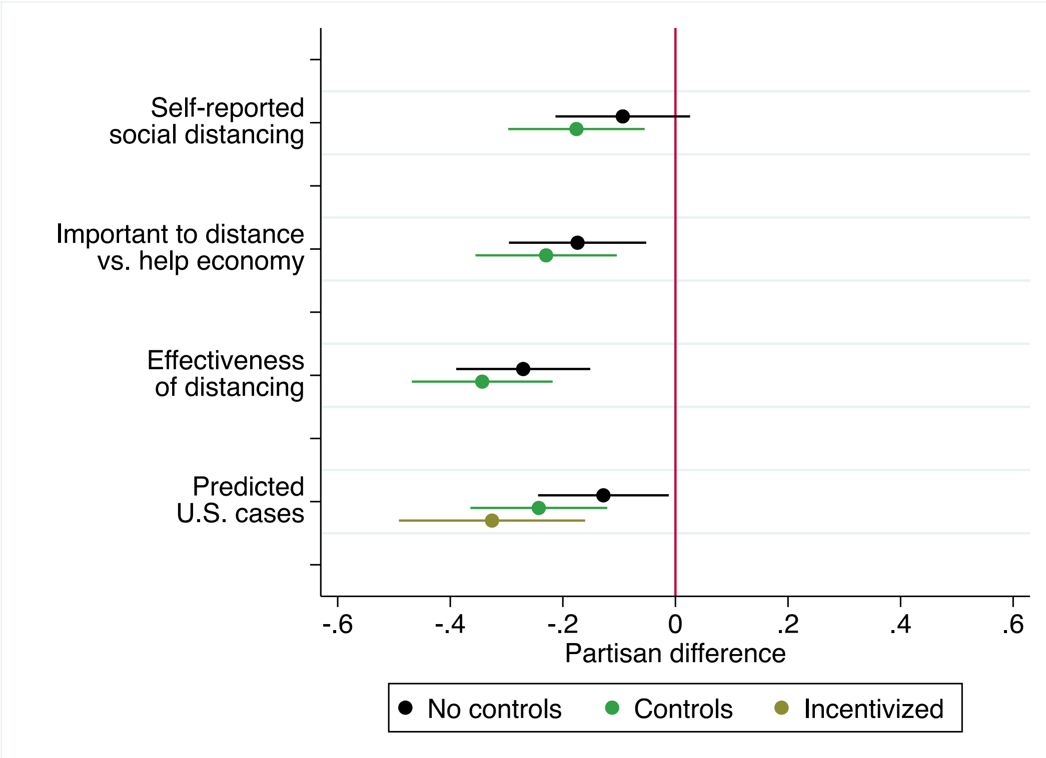
(d) Log Active Devices



Note: Figure shows the estimated coefficients for county partisanship ρ_i on various alternative outcomes constructed from the Daily Social Distancing dataset from SafeGraph. 'Log Devices Leaving Home' is the log of one plus the number of active devices in the panel minus the active devices never observed leaving their geohash-7 home. 'Share Devices Leaving Home' is defined to be $1 - \frac{\max\{0,\text{home devices}+(\text{initital device count}-\text{current device count})\}}{\text{initial device count}}$, where 'home devices' are active devices never observed leaving their geohash-7 home, 'initial device count' is the number of active devices for the week of February 1, and 'current device count' is the number of active devices for the current week. 'Log Median Time Away' is $\log(1 + 1440 - \text{time home})$ where 'time home' is the median observed time at home across devices. 'Log Active Devices' is the log of one plus the number of active devices in the panel. The same controls are used as in Panel C of Figure 3.4. The grey error bars indicate 95 percent confidence intervals constructed using standard errors clustered at the county-level.

Democrats report engaging in 0.18 SD more of a reduction in contact with others as compared to strong Republicans. This corresponds to a gap in reducing contact with others of 72.1 percent for strong Democrats versus 67.8 percent for strong Republicans. Similarly, Democrats find it significantly more important to stay inside to prevent the spread of the virus versus go outside to help the economy, and the difference between strong partisans is 0.23 SD.

Figure 3.8: Partisan Differences in Beliefs and Actions



Note: This figure shows coefficient plots of regressing normalized measures of beliefs and actions on party. Positive values indicate less concern about COVID-19 or social distancing. Demographic controls are age, race, income, education, number of children, ZIP code logged population density, state, county-level deaths and cases. 2 percent of observations are set to the mean due to an invalid ZIP code. Incentivized includes controls and restricts sample to subjects given accuracy incentives. Predicted U.S. cases are predictions about the number of new COVID-19 cases in the U.S. in April; self-reported social distancing is the percent reduction in contact with others over one month; effectiveness of distancing is the estimated likelihood of catching COVID-19 in one month without social distancing; importance of distancing vs. economy is subjects' perception of whether it is more important to go out and stimulate the economy versus staying in and preventing the spread of COVID-19. Observations weighted to mimic a representative sample as described in the text. Error bars represent 95 percent confidence intervals.

We then examine the extent to which partisan differences in social distancing attitudes could be attributed to underlying beliefs regarding COVID-19 severity and efficacy of social distancing. First, we find that Democrats believe that the probability of catching COVID-19 in one month without any social distancing is higher than Republicans do. On average, participants assess this probability to be 55.0 percent (SD 31.9 percent). Strong Democrats hold beliefs that the risk of not socially distancing is 0.34 SD larger as compared to strong Republicans. This corresponds to a gap in beliefs about the probability of catching COVID-19 without social distancing of 60.5 percent for strong Democrats versus 49.6 percent for strong Republicans.

We next consider beliefs about future COVID-19 cases in the entire U.S. We tell participants the number of cases by March 31 and ask them to predict the number of cases in April. We find that Democrats anticipate more future COVID-19 cases. On average, participants predict 202,810 new cases in April 2020 (SD 233,343 cases, due to a long right tail).[12] Strong Democrats predict 0.24 SD more cases as compared to strong Republicans. This corresponds to a gap in beliefs about future cases of 231,283 for strong Democrats versus 174,495 for strong Republicans. Bullock et al. (2015) and Prior et al. (2015) show that partisan differences on factual questions often shrink under incentives due to "partisan cheerleading" rather than differences in true beliefs. As such, we randomize whether subjects' predictions are incentivized for accuracy; we do not finde evidence that the partisan gap decreases.[13] This supports the view that Democrats and Republicans genuinely differ in their beliefs about the severity of COVID-19.

As shown in Appendix Figure C.7, these results are qualitatively similar with county-level controls, although statistical precision is weaker.

Differences in beliefs and news may help explain differences in behavior. In Appendix Figure C.9, we find that the partisan gap in social distancing behaviors attenuates by 60 percent when controlling for respondents' beliefs about the efficacy of social distancing, and that there is no gap when controlling for beliefs and respondents' news sources. While suggestive, we note that controlling for beliefs and news sources do not cleanly separate the causal role of these versus other factors.

Finally, we do a back-of-the-envelope estimation of the deadweight loss from Equation 3.7. We assume that agents have the same flow utility functions $u(c) = \frac{\nu}{2}c^2 + \eta c + k$ and normalize: if $\beta = 0$, all agents choose to consume $c^*(0) = 1$, so $-\nu = \eta \geq 0$. We then consider what

[12] These averages are calculated after winsorizing at the 5-percent level to account for outliers.

[13] Appendix Figure C.7 shows that on an explicitly political question, incentives do significantly reduce the partisan gap, consistent with the previous findings.

happens when partisan perceptions differ about β. From our survey, we find that the median participant's willingness-to-accept for a month of consuming 1 instead of 0 is \$1500, so $\eta = 3000$. From the data above, we approximate that Democrats reduce consumption by 72.1% and Republicans reduce by 67.8%. Therefore, Democrats and Republicans differ in perceived risks by $\tilde{\mu}_{tR} - \tilde{\mu}_{tD} = \$129$ per month.

Plugging this into Equation 3.7, we compare the deadweight loss if partisans have different perceived risks (μ^*_{tD}, μ^*_{tR}) compared to if they have the same percieved risk $(\mu^*_{tD} + \mu^*_{tR})/2$. Using an estimate of 330 million people in the U.S. and 99% of the country being susceptible, we estimate that the partisan inefficiency costs approximately $\Delta W = \$8.24$ per person per year, or \$2.7 billion for the U.S. per year.

3.7 Conclusion

Messages from political leaders and media outlets about the severity of COVID-19 could substantially affect how Americans respond to the pandemic. If Republicans and Democrats disagree about the potential risks, they may also differ in how much they reduce the risk of disease transmission through social distancing and other actions. In this case, our model shows how society ends up with more disease transmission at higher economic cost than if people had the same beliefs.

Our empirical results show that partisan gaps in beliefs and behavior are real. GPS evidence reveals large partisan gaps in actual social distancing behaviors. Survey evidence shows substantial gaps between Republicans and Democrats in beliefs about the severity of COVID-19 and the importance of social distancing. The raw partisan differences partly reflect the fact that Democrats are more likely to live in the dense, urban areas hardest hit by the crisis, and to be subject to policy restrictions—in other words, to face stronger individual incentives for social distancing. Even after controlling carefully for such factors, however, the partisan gaps remain statistically and economically significant. While our evidence does not permit us to conclusively pin down the ultimate causes of partisan divergence, the patterns are consistent with the messaging from politicians and media having played an important role.

Bibliography

Aleem, Zeeshan (2020). “A new poll shows a startling partisan divide on the dangers of the coronavirus”. In: *Vox*. URL: https://www.vox.com/2020/3/15/21180506/coronavirus-poll-democrats-republicans-trump.

Alesina, Alberto, Armando Miano, and Stefanie Stantcheva (2018). “Immigration and Redistribution”. In: *Working Paper*.

Alesina, Alberto, Stefanie Stantcheva, and Edoardo Teso (2018). “Intergenerational mobility and preferences for redistribution”. In: *American Economic Review*.

Allcott, Hunt and Matthew Gentzkow (2017). “Social Media and Fake News in the 2016 Election”. In: *Journal of Economic Perspectives*.

Anderson, Martin (2020). “Early evidence on social distancing in response to COVID-19 in the United States”. In: *Working Paper*.

Athey, Susan et al. (2019). “Experienced Segregation”. In: *Working Paper*.

Augenblick, Ned and Matthew Rabin (2015). “An Experiment on Time Preference and Misprediction in Unpleasant Tasks”. In: *The Review of Economic Studies*.

Baker, Scott R. et al. (2020). “How does household spending respond to an epidemic? Consumption during the COVID-19 pandemic”. In: *Working Paper*.

Barrios, John H. and Yael V. Hochberg (2020). “Risk perception through the lens of politics in the time of the COVID-19 pandemic”. In: *Working Paper*.

Bartels, Larry M. (2002). “Beyond the running tally: Partisan bias in political perceptions”. In: *Political Behavior*.

Beauchamp, Zack (2020). “The stunning contrast between Biden and Trump on coronavirus”. In: *Vox*. URL: https://www.vox.com/policy-and-politics/2020/3/12/21177135/coronavirus-covid-19-pandemic-trump-biden-speeches.

Benabou, Roland (2013). “Groupthink: Collective Delusions in Organizations and Markets”. In: *Review of Economic Studies*.

Benabou, Roland and Jean Tirole (2002). “Self-Confidence and Personal Motivation”. In: *Quarterly Journal of Economics*.

Benabou, Roland and Jean Tirole (2011). “Identity, Morals, and Taboos: Beliefs as Assets”. In: *Quarterly Journal of Economics*.

Benjamin, Dan, Aaron Bodoh-Creed, and Matthew Rabin (2019). “Base-Rate Neglect: Foundations and Applications”. In: *Working Paper*.

Benjamin, Dan, Matthew Rabin, and Colin Raymond (2016). “A Model of Non-Belief in the Law of Large Numbers”. In: *Journal of the European Economic Association*.

Benjamin, Daniel (2019). “Errors in Probabilistic Reasoning and Judgment Biases”. In: *Chapter for the Handbook of Behavioral Economics*.

Benoit, Jean-Pierre and Juan Dubra (2014). “A theory of rational attitude polarization”. In: *Working Paper*.

Bish, Alison and Susan Michie (2010). "Demographic and attitudinal determinants of protective behaviours during a pandemic: A review". In: *British Journal of Health Psychology*.

Blattman, Christopher et al. (2017). "Place-based interventions at scale: The direct and spillover effects of policing and city services on crime". In: *Working Paper*.

Blendon, Robert J. et al. (2008). "Public response to community mitigation measures for pandemic influenza". In: *Emerging Infectious Diseases*.

Bolsen, Toby, James Druckman, and Fay Lomax Cook (2014). "The Influence of Partisan Motivated Reasoning on Public Opinion". In: *Political Behavior*.

Bordalo, Pedro et al. (2019). "Beliefs about Gender". In: *American Economic Review, Forthcoming*.

Brunnermeier, Markus and Jonathan Parker (2005). "Optimal expectations". In: *The American Economic Review*.

Bullock, John G. et al. (2015). "Partisan bias in factual beliefs about politics". In: *Quarterly Journal of Political Science*.

Bults, Marloes et al. (2011). "Perceived risk, anxiety, and behavioural responses of the general public during the early phase of the Influenza A (H1N1) pandemic in the Netherlands: results of three consecutive online surveys". In: *BMC Public Health*.

Bursztyn, Leonardo et al. (2019). "Political Identity: Experimental Evidence on Anti-Americanism in Pakistan". In: *Journal of the European Economic Association, Forthcoming*.

Campante, Filipe R. and Daniel A. Hojman (2013). "Media and polarization: Evidence from the introduction of broadcast TV in the United States". In: *Journal of Public Economics*.

Cappelen, Alexander, Ingar Haaland, and Bertil Tungodden (2018). "Beliefs about Behavioral Responses to Taxation". In: *Working Paper*.

Charness, Gary and Chetan Dave (2017). "Confirmation bias with motivated beliefs". In: *Games and Economic Behavior*.

Chen, Daniel, Martin Schonger, and Chris Wickens (2016). "oTree – An open-source platform for laboratory, online, and field experiments". In: *Journal of Behavioral and Experimental Finance*.

Chen, M. Keith and Ryne Rohla (2018). "Politics Gets Personal: Effects of Political Partisanship and Advertising on Family Ties". In: *Science*.

Chuang, Ying-Chih et al. (2015). "Social capital and health-protective behavior intentions in an influenza pandemic". In: *PloS One*.

Coffman, Katherine, Manuela Collis, and Leena Kulkarni (2019). "Stereotypes and Belief Updating". In: *Working Paper*.

Coppins, McKay (2020). "Trump's Dangerously Effective Coronavirus Propaganda". In: *The Atlantic*. URL: https://www.theatlantic.com/politics/archive/2020/03/trump-coronavirus-threat/607825/.

Coutts, Alexander (2018). "Good news and bad news are still news: Experimental evidence on belief updating". In: *Experimental Economics*.

Dana, Jason, Roberto Weber, and Jason Kuang (2007). "Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness". In: *Economic Theory*.

Data, MIT Election and Science Lab (2018). "County Presidential Election Returns 2000-2016". In: *Harvard Dataverse*. URL: https://doi.org/10.7910/DVN/VOQCHQ.
Davis, Donald R. et al. (2019). "How segregated is urban consumption?" In: *Journal of Political Economy*.
Di Tella, Rafael et al. (2015). "Conveniently Upset: Avoiding Altruism by Distorting Beliefs about Others' Altruism". In: *American Economic Review*.
Dixit, Avinash K. and Jörgen W. Weibull (2007). "Political polarization". In: *Proceedings of the National Academy of Sciences*.
Druckman, James, Matthew Levendusky, and Audrey McLain (2018). "No Need to Watch: How the Effects of Partisan Media Can Spread via Interpersonal Discussions". In: *American Journal of Political Science*.
Druckman, James, Erik Peterson, and Rune Slothuus (2013). "How elite partisan polarization affects public opinion formation". In: *American Political Science Review*.
Duarte, Jose L. et al. (2015). "Political diversity will improve social psychological science". In: *Behavioral and Brain Sciences*.
Dubé, Jean-Pierre et al. (2017). "Competitive price targeting with smartphone coupons". In: *Marketing Science*.
Economist (2020). "Democrats seem to take social distancing more seriously than Republicans". In: *The Economist*. URL: https://www.economist.com/united-states/2020/04/04/democrats-seem-to-take-social-distancing-more-seriously-than-republicans.
Eil, David and Justin Rao (2011). "The good news-bad news effect: asymmetric processing of objective information about yourself". In: *American Economic Journal: Microeconomics*.
Epley, Nicholas and Thomas Gilovich (2016). "The Mechanics of Motivated Reasoning". In: *Journal of Economic Perspectives*.
Ertac, Seda (2011). "Does self-relevance affect information processing? Experimental evidence on the response to performance and non-performance feedback". In: *Journal of Economic Behavior and Organization*.
Exley, Christine (2015). "Excusing Selfishness in Charitable Giving: The Role of Risk". In: *Review of Economic Studies*.
Exley, Christine (2018). "Using Charity Performance Metrics as an Excuse Not to Give". In: *Working Paper*.
Exley, Christine and Judd Kessler (2018). "Motivated Errors". In: *Working Paper*.
Festinger, Leon (1957). "A theory of cognitive dissonance". In: *Stanford University Press*.
Fineberg, Harvey (2014). "Pandemic preparedness and response—lessons from the H1N1 influenza of 2009". In: *New England Journal of Medicine*.
Flynn, D.J., Brendan Nyhan, and Jason Reifler (2017). "The Nature and Origins of Misperceptions: Understanding False and Unsupporting Beliefs About Politics". In: *Advances in Political Psychology*.
Fryer, Roland G., Philipp Harms, and Matthew O. Jackson (2019). "Updating beliefs when evidence is open to interpretation: Implications for bias and polarization". In: *Journal of the European Economic Association*.

Gadarian, Shana Kushner, Sara Wallace Goodman, and Thomas B. Pepinsky (2020). "Partisanship, health behavior, and policy attitudes in the early stages of the COVID-19 pandemic". In: *Working Paper*.
Gagnon-Bartsch, Tristan, Matthew Rabin, and Joshua Schwartzstein (2018). "Channeled Attention and Stable Errors". In: *Working Paper*.
Gaines, Brian J. et al. (2007). "Same facts, different interpretations: Partisan motivation and opinion on Iraq". In: *The Journal of Politics*.
Gallup (2018a). *Military, Small Business, Police Still Stir Most Confidence*. URL: https://news.gallup.com/poll/236243/military-small-business-police-stir-confidence.aspx.
Gallup (2018b). *Presidential Approval Ratings – Donald Trump*. URL: https://news.gallup.com/poll/203198/presidential-approval-ratings-donald-trump.aspx.
Gallup (2019). "Party Affiliation". In: *Gallup*. URL: https://news.gallup.com/poll/15370/party-affiliation.aspx.
Gamma, Anna E. et al. (2017). "Contextual and psychosocial factors predicting Ebola prevention behaviours using the RANAS approach to behaviour change in Guinea-Bissau". In: *BMC Public Health*.
Gentzkow, Matthew and Jesse Shapiro (2006). "Media bias and reputation". In: *Journal of Political Economy*.
Gentzkow, Matthew and Jesse Shapiro (2011). "Ideological Segregation Online and Offline". In: *The Quarterly Journal of Economics*.
Gentzkow, Matthew, Michael Wong, and Allen Zhang (2018). "Ideological Bias and Trust in Information Sources". In: *Working Paper*.
Gerber, Alan and Gregory Huber (2009). "Partisanship and Economic Behavior: Do Partisan Differences in Economic Forecasts Predict Real Economic Behavior?" In: *American Political Science Review*.
Gervais, Simon and Terrance Odean (2001). "Learning to Be Overconfident". In: *The Review of Financial Studies*.
Gino, Francesca, Michael Norton, and Roberto Weber (2016). "Motivated Bayesians: Feeling Moral While Acting Egoistically". In: *Journal of Economic Perspectives*.
Glaeser, Edward L. and Bryce A. Ward (2006). "Myths and realities of American political geography". In: *Journal of Economic Perspectives*.
Grigorieff, Alexis, Christopher Roth, and Diego Ubfal (2018). "Does Information Change Attitudes Towards Immigrants? Representative Evidence from Survey Experiments". In: *Working Paper*.
Haaland, Ingar and Christopher Roth (2018). "Labor Market Concerns and Support for Immigration". In: *Working Paper*.
Haaland, Ingar and Christopher Roth (2019). "Beliefs About Racial Discrimination and Support for Pro-Black Policies". In: *Working Paper*.
Hagmann, David and George Loewenstein (2017). "Persuasion With Motivated Beliefs". In: *Working Paper*.
Haisley, Emily and Roberto Weber (2010). "Self-serving interpretations of ambiguity in other-regarding behavior". In: *Games and Economic Behavior*.

Hanna, Rema, Gabriel Kreindler, and Benjamin A. Olken (2017). “Citywide effects of high-occupancy vehicle restrictions: Evidence from ‘three-in-one’ in Jakarta”. In: *Science*.

Hersh, Eitan D. and Matthew N. Goldenberg (2016). “Democratic and Republican physicians provide different care on politicized health issues”. In: *Proceedings of the National Academy of Sciences*.

Holmes, Bev J. (2008). “Communicating about emerging infectious disease: The importance of research”. In: *Health, Risk & Society*.

Horton, John, David Rand, and Richard Zeckhauser (2011). “The online laboratory: conducting experiments in a real labor market”. In: *Experimental Economics*.

Ibuka, Yoko et al. (2010). “The dynamics of risk perceptions and precautionary behavior in response to 2009 (H1N1) pandemic influenza”. In: *BMC Infectious Diseases*.

Iyengar, Shanto, Yphtach Lelkes, et al. (2019). “The origins and consequences of affective polarization in the United States”. In: *Annual Review of Political Science*.

Iyengar, Shanto and Sean Westwood (2015). “Fear and Loathing across Party Lines: New Evidence on Group Polarization”. In: *American Journal of Political Science*.

Kahan, Dan (2016a). “The Politically Motivated Reasoning Paradigm, Part 1: What Politically Motivated Reasoning Is and How to Measure It”. In: *Emerging Trends in Social and Behavioral Sciences*.

Kahan, Dan (2016b). “The Politically Motivated Reasoning Paradigm, Part 2: Unanswered Questions”. In: *Emerging Trends in Social and Behavioral Sciences*.

Kahan, Dan, David Hoffman, et al. (2012). “They Saw a Protest: Cognitive Illiberalism and the Speech-Conduct Distinction”. In: *Stanford Law Review*.

Kahan, Dan, Ellen Peters, et al. (2017). “Motivated numeracy and enlightened self-government”. In: *Behavioural Public Policy*.

Kantrowitz, Alex (2020). “Conservative media still isn’t sure what to think about the coronavirus”. In: *Buzzfeed News*. URL: https://www.buzzfeednews.com/article/alexkantrowitz/conservative-media-still-isnt-sure-coronavirus.

Kermack, William Ogilvy and Anderson G. McKendrick (1927). “A contribution to the mathematical theory of epidemics”. In: *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*.

Knight Foundation (2018). *American Views: Trust, Media, and Democracy*. URL: https://knightfoundation.org/reports/american-views-trust-media-and-democracy.

Koszegi, Botond (2006a). “Ego utility, overconfidence, and task choice”. In: *Journal of the European Economic Association*.

Koszegi, Botond (2006b). “Emotional agency”. In: *The Quarterly Journal of Economics*.

Krupenkin, Masha (2018). “Does partisanship affect compliance with government recommendations?” In: *Working Paper*.

Kuhnen, Camelia (2014). “Asymmetric Learning from Financial Information”. In: *The Journal of Finance*.

Kunda, Ziva (1990). “The case for motivated reasoning”. In: *Psychological Bulletin*.

Kunda, Ziva and Lisa Sinclair (1999). “Motivated Reasoning With Stereotypes: Activation, Application, and Inhibition”. In: *Psychological Inquiry*.

Kuziemko, Ilyana et al. (2015). “How Elastic Are Preferences for Redistribution? Evidence from Randomized Survey Experiments”. In: *American Economic Review*.

Lee, Alicia (2020). “These states have implemented stay-at-home orders. Here’s what that means for you”. In: *CNN*. URL: https://www.cnn.com/2020/03/23/us/coronavirus-which-states-stay-at-home-order-trnd/index.html.

Lerman, Amy E., Meredith L. Sadin, and Samuel Trachtman (2017). “Policy uptake as political behavior: Evidence from the Affordable Care Act”. In: *American Political Science Review*.

Levay, Kevin, Jeremy Freese, and James Druckman (2016). “The Demographic and Political Composition of Mechanical Turk Samples”. In: *SAGE Open*.

Levendusky, Matthew (2013). “Why Do Partisan Media Polarize Viewers?” In: *American Journal of Political Science*.

Lord, Charles G., Lee Ross, and Mark R. Lepper (1979). “Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence”. In: *Journal of Personality and Social Psychology*.

Marist (2020). “March 13th and 14th survey of American adults”. In: *Marist*. URL: http://maristpoll.marist.edu/wp-content/uploads/2020/03/NPR_PBS-NewsHour_Marist-Poll_USA-NOS-and-Tables_2003151338.pdf.

Mayraz, Guy (2013). “Wishful Thinking in Predictions of Asset Prices”. In: *Working Paper*.

Mayraz, Guy (2018). “Priors and Desires: A Bayesian Model of Wishful Thinking and Cognitive Dissonance”. In: *Working Paper*.

McCarthy, Tom (2020). “Disunited states of America: Responses to coronavirus shaped by hyper-partisan politics”. In: *The Guardian*. URL: https://www.theguardian.com/us-news/2020/mar/29/america-states-coronavirus-red-blue-different-approaches.

McCarty, Nolan, Keith T. Poole, and Howard Rosenthal (2006). “Polarized America: The Dance of Political Ideology and Unequal Riches”. In: *MIT Press: Cambridge, MA*.

Meeuwis, Maarten et al. (2019). “Belief Disagreement and Portfolio Choice”. In: *Working Paper*.

Mobius, Markus et al. (2014). “Managing self-confidence: Theory and experimental evidence”. In: *Working Paper*.

Montoya-Williams, Diana and Elena Fuentes-Afflick (2019). “Political determinants of population health”. In: *JAMA Network Open*.

Moore, Don and Paul Healy (2008). “The Trouble with Overconfidence”. In: *Psychological Review*.

Moore, Don, Elizabeth Tenney, and Uriel Haran (2015). “Overprecision in Judgment”. In: *The Wiley Blackwell Handbook of Judgment and Decision Making*.

Nisbet, Erik, Kathryn Cooper, and R. Kelly Garrett (2015). “The Partisan Brain: How Dissonant Science Messages Lead Conservatives and Liberals to (Dis)Trust Science”. In: *The ANNALS of the American Academy of Political and Social Science*.

Nyhan, Brendan and Jason Reifler (2010). “When Corrections Fail: The Persistence of Political Misperceptions”. In: *Political Behavior*.

Nyhan, Brendan and Jason Reifler (2013). "Which Corrections Work? Research results and practice recommendations". In: *New America Foundation*.

Nyhan, Brendan, Jason Reifler, and Peter Ubel (2013). "The Hazards of Correcting Myths About Health Care Reform". In: *Medical Care*.

O'Donoghue, Ted and Matthew Rabin (2001). "Choice and Procrastination". In: *The Quarterly Journal of Economics*.

Ortoleva, Pietro and Erik Snowberg (2015). "Overconfidence in Political Behavior". In: *The American Economic Review*.

Orwell, George (1945). "London Letter". In: *Partisan Review*.

Oster, Emily, Ira Shoulson, and E. Ray Dorsey (2013). "Optimal Expectations and Limited Medical Testing: Evidence from Huntington Disease". In: *American Economic Review*.

Pennycook, Gordon and David Rand (2019). "Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning". In: *Cognition*.

Piacenza, Joanna (2020). "Tracking public opinion on the coronavirus". In: *Morning Consult*. URL: https://morningconsult.com/form/tracking-public-opinion-on-the-coronavirus/.

Prior, Markus (2013). "Media and political polarization". In: *Annual Review of Political Science*.

Prior, Markus, Gaurav Sood, and Kabir Khanna (2015). "You cannot be serious. The impact of accuracy incentives on partisan bias in reports of economic perceptions". In: *Quarterly Journal of Political Science*.

Rabin, Matthew (2019). "Moral preferences, moral constraints, and self-serving biases". In: *Working Paper*.

Rabin, Matthew and Joel Schrag (1999). "First impressions matter: A model of confirmatory bias". In: *The Quarterly Journal of Economics*.

Saad, Lydia (2020). "Americans step up their social distancing even further". In: *Gallup*. URL: https://news.gallup.com/opinion/gallup/298310/americans-step-social-distancing-even-further.aspx.

Sances, Michael W. and Joshua D. Clinton (2019). "Who participated in the ACA? Gains in insurance coverage by political partisanship". In: *Journal of Health Politics, Policy and Law*.

Sarsons, Heather (2017). "Interpreting Signals in the Labor Market: Evidence from Medical Referrals". In: *Working Paper*.

Schwartzstein, Joshua (2014). "Selective Attention and Learning". In: *Journal of the European Economic Association*.

Shultz, James M. et al. (2016). "The role of fear-related behaviors in the 2013–2016 West Africa Ebola virus disease outbreak". In: *Current Psychiatry Reports*.

Stanley-Becker, Isaac and Chelsea Janes (2020). "As virus takes hold, resistance to stay-at-home orders remains widespread – exposing political and social rifts". In: *Washington Post*. URL: https://www.washingtonpost.com/politics/as-virus-takes-hold-resistance-to-stay-at-home-orders-remains-widespread--exposing-

political-and-social-rifts/2020/04/02/d87314e0-7436-11ea-85cb-8670579b863d%5C_story.html.
Stone, Daniel, Matthew Gentzkow, and Jesse Shapiro (2015). "Media Bias in the Marketplace: Theory". In: *Handbook of Media Economics*.
Sunstein, Cass et al. (2017). "How People Update Beliefs About Climate Change: Good News and Bad News". In: *Cornell Law Review*.
Suryadevara, Manika et al. (2019). "Associations between population based voting trends during the 2016 US presidential election and adolescent vaccination rates". In: *Vaccine*.
Taber, Charles and Milton Lodge (2006). "Motivated Skepticism in the Evaluation of Political Beliefs". In: *American Journal of Political Science*.
Taylor, Melanie et al. (2009). "Public health measures during an anticipated influenza pandemic: factors influencing willingness to comply". In: *Risk Management and Healthcare Policy*.
Tetlock, Philip (1983). "Accountability and the Perserverance of First Impressions". In: *Social Psychology Quarterly*.
Thaler, Michael (2019). "Debiasing Motivated Reasoning Through Learning: Evidence from an Online Experiment". In: *AEA RCT Registry*. URL: https://doi.org/10.1257/rct.4401-1.0.
Times, The New York (2020). "Coronavirus (Covid-19) Data in the United States". In: *The New York Times*. URL: https://github.com/nytimes/covid-19-data.
Trachtman, Samuel (2019). "Polarization, participation, and premiums: How political behavior helps explain where the ACA works, and where it doesn't". In: *Journal of Health Politics, Policy and Law*.
van der Weerd, Willemien et al. (2011). "Monitoring the level of government trust, risk perception and intention of the general public to adopt protective measures during the influenza A (H1N1) pandemic in the Netherlands". In: *BMC Public Health*.
Vaughan, Elaine and Timothy Tinker (2009). "Effective health risk communication about pandemic influenza for vulnerable populations". In: *American Journal of Public Health*.
von Hippel, William and Robert Trivers (2011). "The Evolution and Psychology of Self-Deception". In: *Behavioral and Brain Sciences*.
Voting and Election Science Team (2018). "2016 Precinct-Level Election Results". In: *Harvard Dataverse*. URL: https://doi.org/10.7910/DVN/NH5S2I.
Weinstein, Neil (1980). "Unrealistic Optimism About Future Life Events". In: *Journal of Personality and Social Psychology*.
Zimmermann, Florian (2019). "The Dynamics of Motivated Beliefs". In: *American Economic Review, Forthcoming*.

A

Appendix for Chapter 1

A.1 Additional Results

A.1.1 Proof of Hypothesis 5

First, we calculate the expected utility that an agent anticipates receiving (*agent-expected utility*), given her assessment a:

$$\begin{aligned}\tilde{\mathbb{E}}[u(a)|a] &= a \cdot \left(1 - (1-a)^2\right) + (1-a) \cdot \left(1 - a^2\right) \\ &= 1 - a(1-a) \\ &= 3/4 + (a - 1/2)^2.\end{aligned}$$

Agent-expected utility increases in $(a - 1/2)^2$.

Next, we calculate the average agent-expected utility $\tilde{\mathbb{E}}[u(a)]$, given her motivated reasoning.

The agent motivatedly reasons as follows:

$$\text{logit } a|G = \text{logit } p + \varphi m \left(\mathbb{E}[\theta|\theta > \mu(m)] - \mathbb{E}[\theta|\theta < \mu(m)]\right).$$

$$\text{logit } a|L = \text{logit } p - \varphi m \left(\mathbb{E}[\theta|\theta > \mu(m)] - \mathbb{E}[\theta|\theta < \mu(m)]\right).$$

The average agent-expected utility is $\left(\tilde{\mathbb{E}}[u(a|G)] + \tilde{\mathbb{E}}[u(a|L)]\right)/2$.

Define $\Delta_m \equiv \varphi m \left(\mathbb{E}[\theta|\theta > \mu(m)] - \mathbb{E}[\theta|\theta < \mu(m)]\right)$. Note that since the standard deviation of agents' beliefs does not depend on m, Δ_m is linear in m. Now, take the inverse logit function of both sides, $\text{logit}^{-1}(x) = \frac{e^x}{1+e^x}$, and average:

$$\begin{aligned} 1 - \tilde{\mathbb{E}}[u(a)] &= \frac{1}{4} - P(G)\mathbb{E}\left[\left(\text{logit}^{-1}(\text{logit } p + \Delta_m) - 1/2\right)^2\right] \\ &\quad - P(L)\mathbb{E}\left[\left(\text{logit}^{-1}(\text{logit } p - \Delta_m) - 1/2\right)^2\right] \\ &= \frac{1}{4} - \frac{1}{2}\mathbb{E}\left[\left(\text{logit}^{-1}(\text{logit } p + \Delta_m) - 1/2\right)^2\right] \\ &\quad - \frac{1}{2}\mathbb{E}\left[\left(\text{logit}^{-1}(\text{logit } p - \Delta_m) - 1/2\right)^2\right] \end{aligned}$$

Therefore, average agent-expected utility equals:

$$3/4 + 1/2\left(\text{logit}^{-1}(\text{logit } p + \Delta_m) - 1/2\right)^2 + 1/2\left(\text{logit}^{-1}(\text{logit } p - \Delta_m) - 1/2\right)^2.$$

We can rewrite this using the hyperbolic tan function, $tanh(x) \equiv \frac{\exp(x)-\exp(-x)}{\exp(x)+\exp(-x)}$:

$$3/4 + 1/8\left[tanh(\text{logit } p + \eta_i\Delta_m)/2)^2\right] + 1/8\left[tanh(\text{logit } p - \eta_i\Delta_m)/2)^2\right]$$

Taking the derivative with respect to Δ_m — which is linear in m — and setting equal to zero gives the following:

$$|\varphi\Delta_m| = 2cosh^{-1}\left[1/2\sqrt{sech(\text{logit } p)(2cosh(\text{logit } p) + cosh(2\text{ logit } p) - 3)}\right],$$

where $cosh(x) \equiv 1/2(e^x + e^{-x})$ and $sech(x) \equiv (cosh(x))^{-1}$.

The right-hand side is real only if the term in the square brackets is at least 1, in which case there is such a solution p; that is, if $sech(\text{logit } p)(2cosh(\text{logit } p) + cosh(2 \text{ logit } p) - 3 < 4$, then the first-order condition is never satisfied and anticipated expected utility is always monotonic in $|m|$. In this case, the second-order condition shows that average agent-expected utility is increasing in $|m|$. This condition is met iff $p \in \left(\frac{1}{2} - \frac{\sqrt{3}}{6}, \frac{1}{2} + \frac{\sqrt{3}}{6}\right)$.

A.1.2 Raw Data: Pro-Party and Anti-Party News Assessments

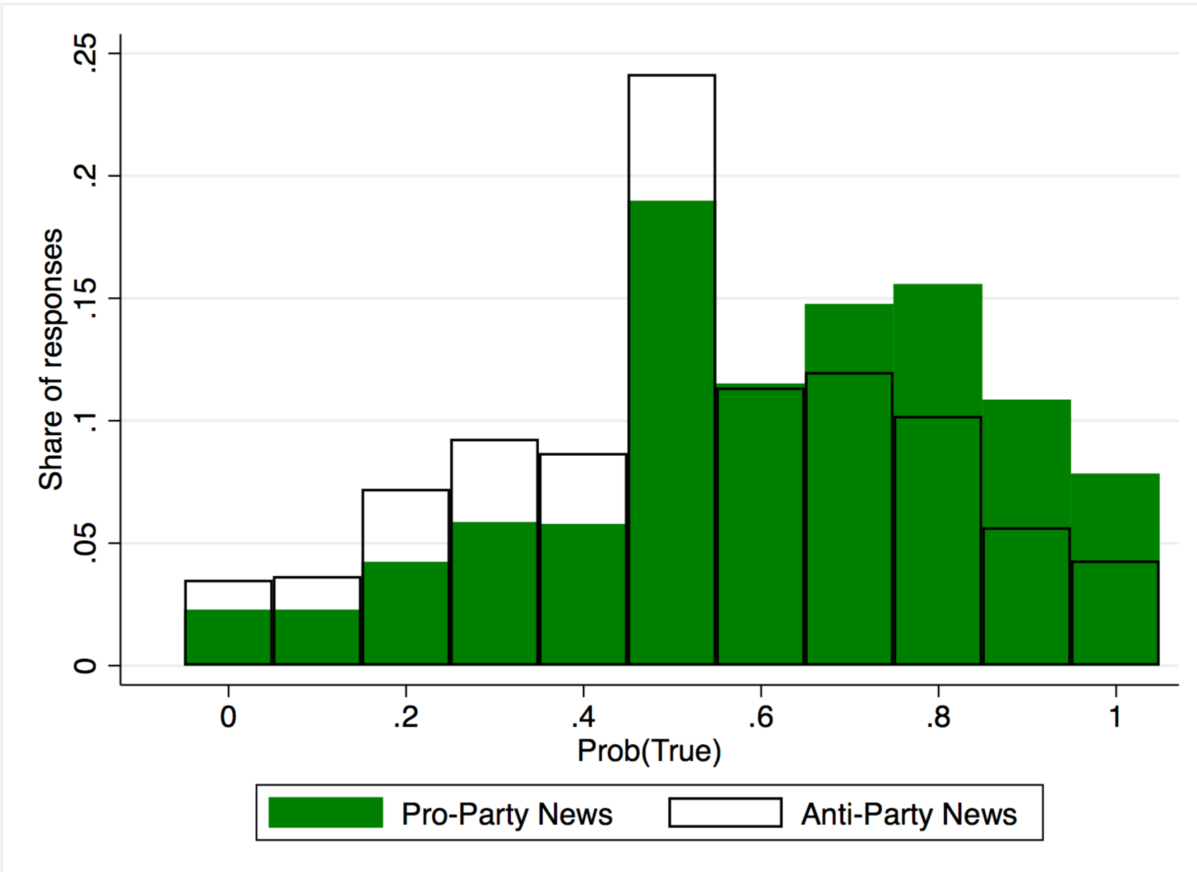


Figure A.1: Histogram of Perceived Veracity of Pro-Party and Anti-Party News.

Notes: Only Pro-Party / Anti-Party news observations, as defined in Table 1. Messages are customized so that Bayesians give the same assessment for Pro-Party and Anti-Party news.

A.1.3 Raw Data: True News and Fake News Assessments

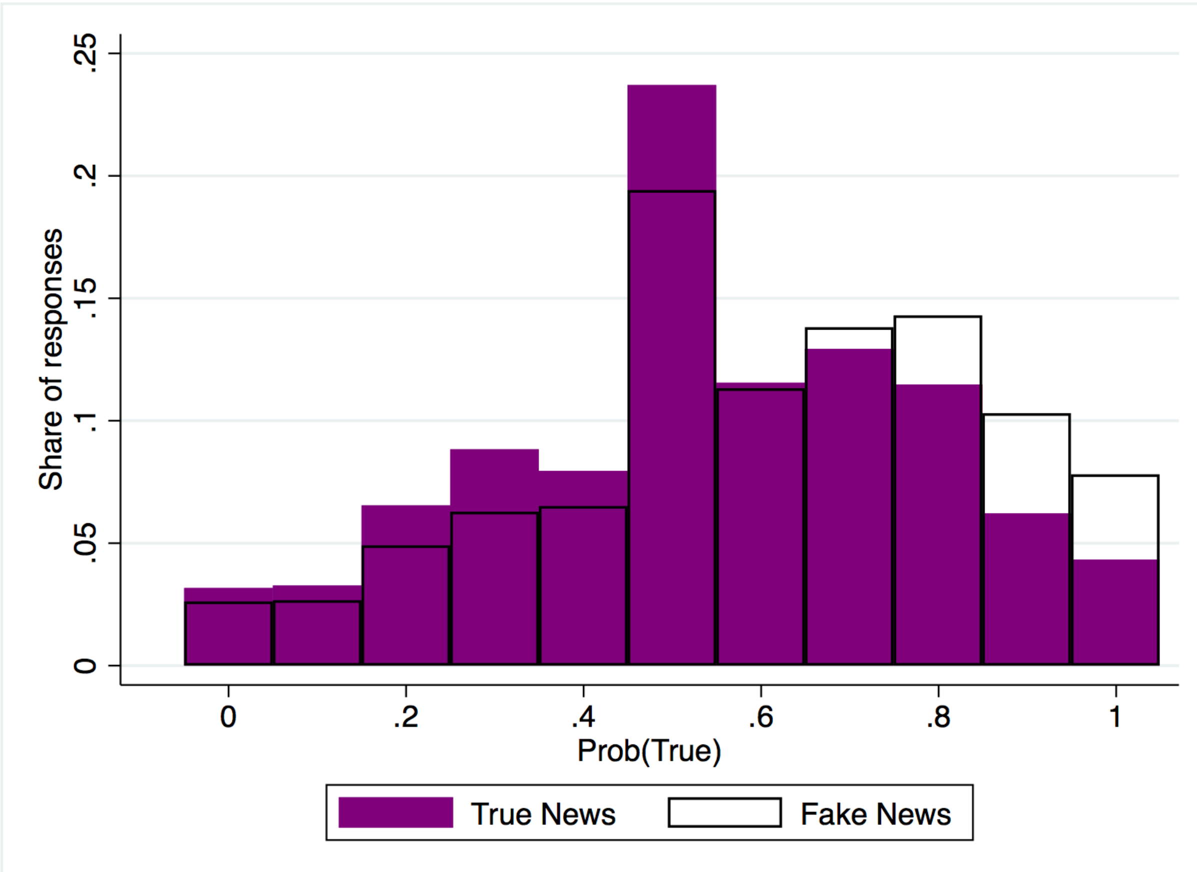


Figure A.2: Histogram of Perceived Veracity of True News and Fake News on Politicized Topics.

Notes: Only Pro-Party / Anti-Party news observations, as defined in Table 1. Messages are customized so that Bayesians give the same assessment for Pro-Party and Anti-Party news.

A.1.4 Relative Prior Beliefs by Party

	Pro-Rep	Pro-Dem	Difference	Answer
Obama Crime Guess	55.907***	49.560***	6.348***	53
	(0.765)	(0.391)	(0.858)	
Mobility Guess	30.185***	22.152***	8.034***	64.9
	(1.048)	(0.611)	(1.211)	
Race Guess	12.349***	8.051***	4.298***	6.45
	(0.874)	(0.436)	(0.975)	
Gender Guess	3.059***	3.086***	-0.027	3.15
	(0.015)	(0.008)	(0.017)	
Refugees Guess	287.640***	239.004***	48.637***	228.2
	(5.894)	(2.353)	(6.335)	
Climate Guess	75.226***	85.366***	-10.140***	87
	(1.056)	(0.572)	(1.200)	
Gun Laws Guess	230.013***	184.478***	45.535***	318.6
	(5.950)	(3.914)	(7.113)	
Media Guess	36.656***	41.850***	-5.195***	19.8
	(1.211)	(0.599)	(1.349)	
Rep Score Guess	71.563***	61.933***	9.630***	70.83
	(0.787)	(0.614)	(0.997)	
Dem Score Guess	64.671***	73.277***	-8.606***	72.44
	(0.771)	(0.415)	(0.875)	
Observations	2430	5643	8073	

Standard errors in parentheses

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

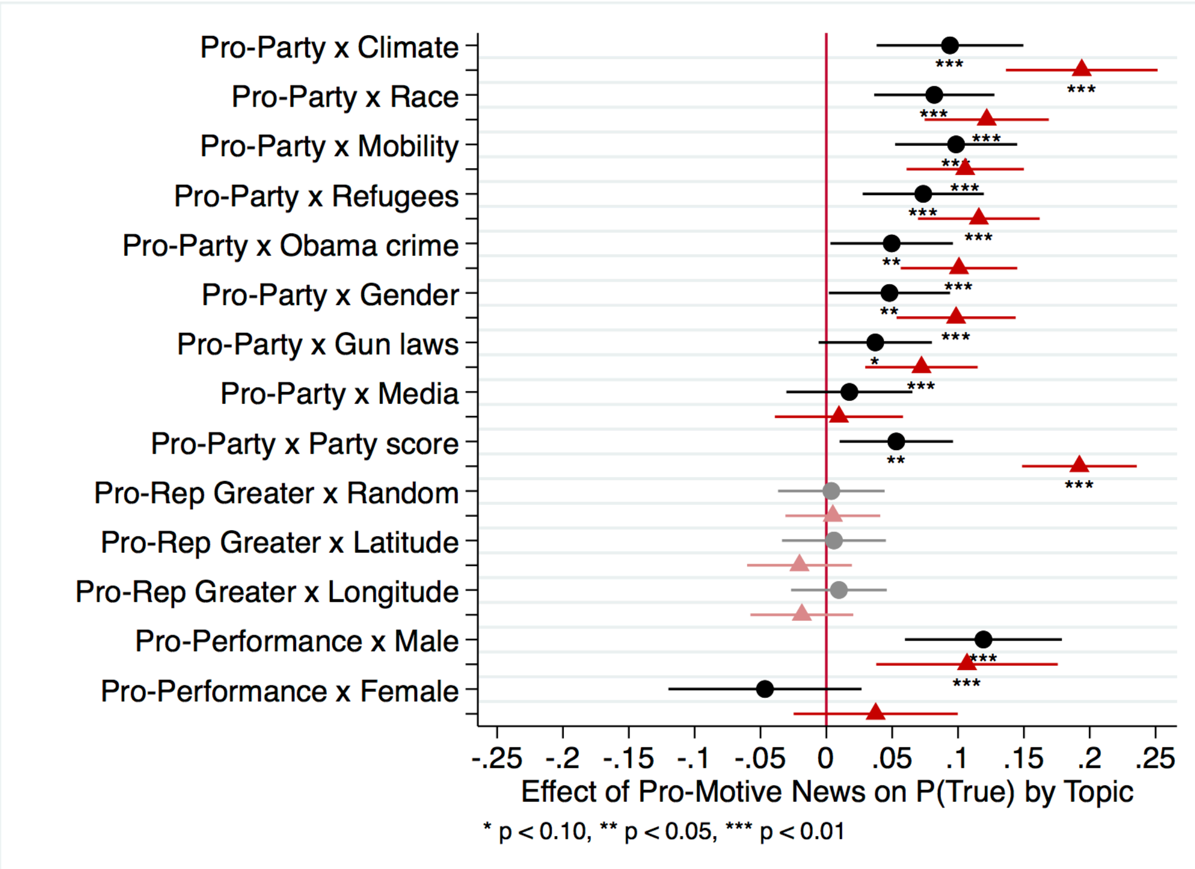
Notes: OLS, robust standard errors. Guesses are winsorized at the 5-percent level. Third column represents mean Pro-Rep guess minus mean Pro-Dem guess. The sign of every coefficient points in the predicted motive direction from Table 1.

A.1.5 Balance Table

	Anti-Party News	Pro-Party News	Anti vs. Pro	p-value
Partisanship	0.484	0.478	0.007	0.312
	(0.005)	(0.005)	(0.007)	
Rep vs. Dem	-0.237	-0.236	-0.001	0.937
	(0.008)	(0.008)	(0.011)	
Male	0.532	0.534	-0.002	0.881
	(0.008)	(0.008)	(0.011)	
Age	35.261	35.400	-0.139	0.573
	(0.175)	(0.173)	(0.246)	
Education	14.716	14.765	-0.049	0.242
	(0.029)	(0.030)	(0.042)	
Log(income)	10.725	10.748	-0.024	0.182
	(0.012)	(0.013)	(0.018)	
White	0.752	0.760	-0.008	0.430
	(0.007)	(0.007)	(0.010)	
Black	0.075	0.081	-0.006	0.303
	(0.004)	(0.004)	(0.006)	
Hispanic	0.066	0.062	0.004	0.499
	(0.004)	(0.004)	(0.006)	
Religious	0.443	0.457	-0.014	0.214
	(0.008)	(0.008)	(0.011)	
Red State	0.567	0.558	0.009	0.431
	(0.008)	(0.008)	(0.0011)	
WTP elicited	0.490	0.476	0.014	0.213
	(0.008)	(0.008)	(0.011)	
Told 1/2 True	0.333	0.344	-0.011	0.309
	(0.007)	(0.008)	(0.011)	
N	3961	3941	7902	

Notes: Standard errors in parentheses. Rep vs. Dem is the rating of the Rep Party minus the rating of the Dem Party and is between -1 and 1. Partisanship is the absolute difference in these ratings. WTP elicited is 1 if subject in the WTP treatment and 0 if in the second-guess treatment. Told 1/2 True is 1 if subject is told that P(True News) is 1/2.

Figure A.3: Motivated Reasoning Across Topics: Effect of Pro-Party News on Perceived Veracity by Topic and Partisanship



Notes: OLS regression coefficients, errors clustered at subject level. Black circles are coefficients for moderates, red triangles are coefficients for partisans. FE included for round number and topic. Pro-Party (vs. Anti-Party) news is defined in Table 1. Pro-Rep Greater is a placebo check to test whether Pro-Rep and Pro-Dem subjects give different assessments on neutral topics. Error bars correspond to 95 percent confidence intervals.

Table A.1: The Effect of News Direction, Actual Veracity, and Skewed Prior Distributions on Perceived Veracity

	(1)	(2)	(3)	(4)
Unskewed	-0.011	-0.003	-0.019*	0.002
	(0.010)	(0.014)	(0.011)	(0.013)
Pro-Party News	0.084***	0.041***	0.028***	0.075***
	(0.008)	(0.015)	(0.008)	(0.008)
Unskewed x Pro-Party	0.014	-0.002	0.021	0.006
	(0.013)	(0.024)	(0.014)	(0.013)
Partisanship x Pro-Party		0.044***		
		(0.014)		
Unskewed x Partisanship x Pro-Party		0.017		
		(0.022)		
Anti-Party News			-0.052***	
			(0.008)	
Unskewed x Anti-Party			0.006	
			(0.014)	
True News				-0.027***
				(0.008)
Unskewed x True News				-0.021
				(0.014)
Neutral News	No	No	Yes	No
Question FE	Yes	Yes	No	Yes
Round FE	Yes	Yes	Yes	Yes
Subject FE	Yes	Yes	Yes	Yes
Observations	7902	7902	10552	7902
R^2	0.25	0.25	0.21	0.25
Mean	0.573	0.573	0.574	0.573

Standard errors in parentheses

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

Notes: OLS, errors clustered at subject level. Neutral News indicates that Pro-Party / Anti-Party compared to Neutral News, as defined in Table 1. Controls: race, gender, log(income), education (in years), religion, whether state voted for Trump or Clinton in 2016. Partisanship is absolute difference between Republican and Democratic ratings. Unskewed is 1 if initial guess is exactly halfway between lower / upper bounds and 0 otherwise.

Table A.2: The Effect of News Direction, Actual Veracity, and Previous News Directions on Perceived Veracity

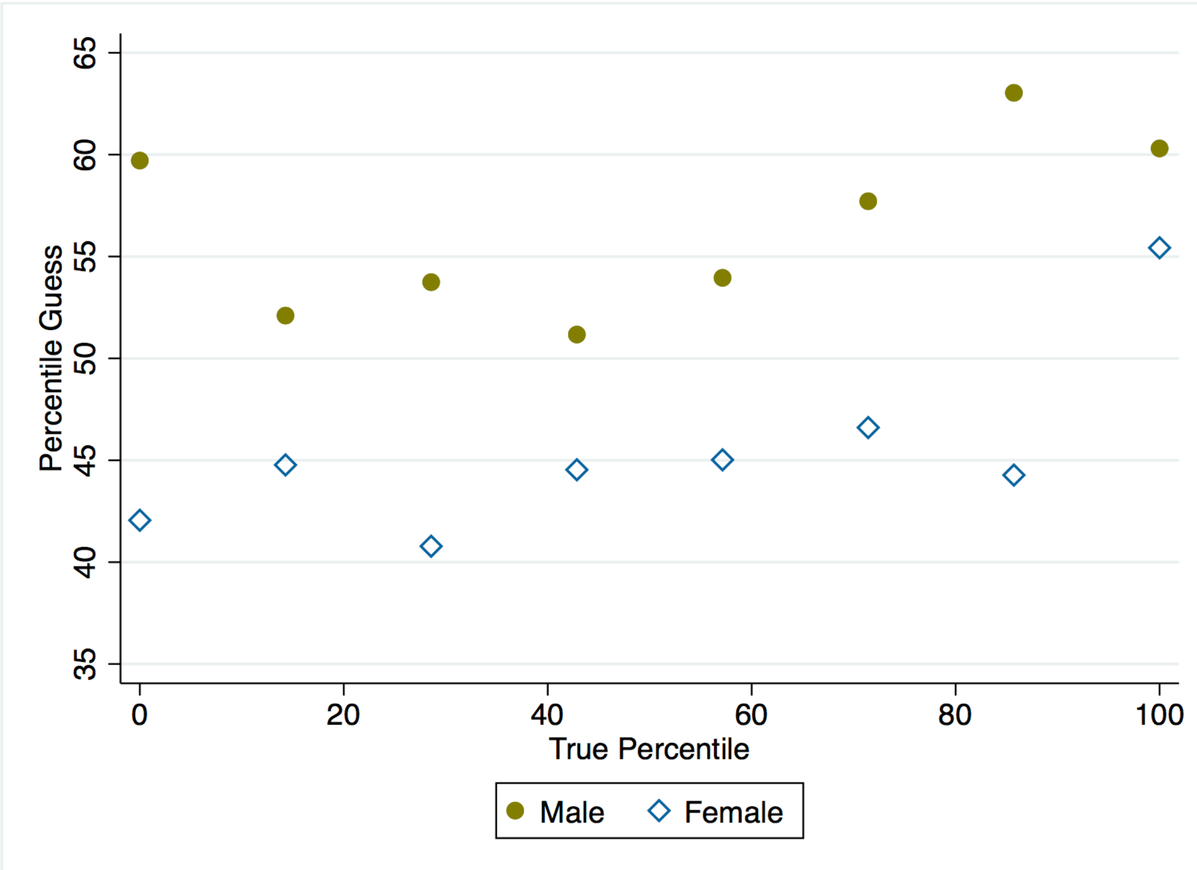
	(1)	(2)	(3)	(4)
Previous Pro-Party	-0.001	-0.001	-0.002	-0.002
	(0.002)	(0.002)	(0.002)	(0.002)
Pro-Party News	0.087***	0.039***	0.036***	0.076***
	(0.007)	(0.013)	(0.007)	(0.007)
Partisanship x Pro-Party		0.050***		
		(0.012)		
Anti-Party News			-0.048***	
			(0.007)	
True News				-0.034***
				(0.007)
Neutral News	No	No	Yes	No
Question FE	Yes	Yes	No	Yes
Round FE	Yes	Yes	Yes	Yes
Subject FE	Yes	Yes	Yes	Yes
Observations	7902	7902	10552	7902
R^2	0.25	0.25	0.21	0.25
Mean	0.573	0.573	0.574	0.573

Standard errors in parentheses

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

Notes: OLS, errors clustered at subject level. Neutral News indicates that Pro-Party / Anti-Party compared to Neutral News, as defined in Table 1. Controls: race, gender, log(income), education (in years), religion, whether state voted for Trump or Clinton in 2016. Partisanship is the absolute difference between Republican and Democratic ratings. Previous Pro-Party is the number of all previous pieces of news that are Pro-Party minus the number that are Anti-Party.

Figure A.4: Bin-Scatter Plot of Expected Performance by Gender and True Performance



Notes: Party-indifferent subjects included. True Percentile compares subjects' score on rounds 1-12 to the scores of 100 pilot subjects. Percentile Guess is subjects' prediction of their True Percentile. Subjects binned by gender into eight True Percentile groups.

A.2 Supplementary Appendix: Demand for News, susceptibility, and Structurally Estimating Motives

This appendix section discusses awareness of motivated reasoning and susceptibility. First, we consider subjects' demand for a message by eliciting willingness to pay (WTP); correlations are consistent with the notion that subjects are aware that they will update from information, but not aware that they motivatedly reason in a way that decreases earnings.

Much of this section relies on an extension of the main model, making the additional assumption that susceptibility is related to the noisiness of the updating process. In particular, we modify Equation (1.2) as follows:

$$\text{logit } \mathbb{P}(\theta|x) = \text{logit } \mathbb{P}(\theta) + \log\left(\frac{\mathbb{P}(x|\theta)}{\mathbb{P}(x|\neg\theta)}\right) + \varphi(m(\theta) - m(\neg\theta)) + \epsilon, \tag{A.1}$$

where $\epsilon \sim \mathcal{N}(0, \varphi^2)$.

Agents update with noise that depends on the signal structure but is independent of the motive. The noise term is normally distributed and its standard deviation is the updated definition of susceptibility.[1]

A.2.1 WTP Treatment Details

In round 12, half of subjects are told that they will either receive the usual message or the message with a black bar over the words "Greater Than" / "Less Than," and are given an example of the black bar message.

They are then asked for their WTP to remove the black bar from the message. WTP is elicited by a standard Becker-DeGroot-Marschak mechanism. The units of payment are points; average points across all rounds in the experiment determine the probability of winning a

[1] If susceptibility is instead assumed linear in φ, it is hard to identify this linear multiple from a linear multiple of the motive function, which is why the extra parameter is not introduced here. Normal noise is used for simplicity, and the choice is fairly arbitrary. Results are qualitatively the same when noise is assumed to be uniform across $[-\varphi, \varphi]$, for instance.

$10 bonus in the experiment.[2] Subjects can choose any valuation between -25 points and 25 points. A noninteger is then chosen uniformly randomly from -25 to 25. If this number is greater than the valuation, it is added to the points on the next page and subjects see a black bar; otherwise, no points are added and the standard message is revealed.

Subjects are told the above, and told that positive numbers indicate that they prefer to see the message, while negative numbers indicate that they prefer not to. Since subjects see the true answers soon after this question, WTP seems to be a reasonable metric for signal valuation. Importantly, these subjects are *not* asked to give a second guess, so the only value of the message is in inferring the veracity of the news source.

A.2.2 Susceptibility and Demand for Messages

This subsection aims to use variance in assessments and demand for messages (WTP) to show that susceptibility, φ, is positive, and to argue that subjects are unaware of their directionally-motivated reasoning. This uses the parametrization from Equation (A.1); in this case, susceptibility can be empirically defined using the standard deviation of the noise in updating about topics absent motivated reasoning.

Importantly, none of the subjects in the WTP treatment is ever asked to give a second guess to any question, as this treatment was intended to capture how subjects valued messages insofar as they provide signals for news assessments. Subjects also know that they soon learn the correct answer, so the only value in seeing the message is for improving their news assessments.

This test helps show that susceptibility is positive and *expected* susceptibility is positive. If $\varphi = 0$, subjects will have WTP = 0 and not vary their answers. If subjects expect to have $\varphi = 0$ but actually have $\varphi > 0$, they will have WTP = 0 but vary their answers. If subjects expect to have $\varphi > 0$ and do have $\varphi > 0$, but don't realize that this is an error, then they will have positive WTP since they expect to perform better with the message.

Meanwhile, there is no evidence that subjects are aware of the motive part of their politically-

[2]More technically, points are added to or subtracted from the news assessment score of that round.

motivated reasoning. This would come through in differences in WTP from politicized and neutral news: if subjects expected to motivatedly reason about politicized news and that this would lead to underperformance, they would have a lower WTP for these signals.

Subjects' WTP are positive and do not seem smaller for politicized topics. 71% of subjects have a strictly positive WTP. Partisanship does not lead to a significantly larger WTP for politicized topic messages. However, a larger standard deviation of previous assessments is highly correlated with WTP, suggesting that subjects genuinely expect to find these messages useful.

There are three main observations from the WTP question, all suggesting that subjects pay for messages based on their perceived expected usefulness but are not aware of the effect of politically-motivated reasoning:

1. WTP is significantly greater than zero for politicized and neutral topics, indicating that subjects do expect messages to be informative. The mean is 9 points (s.e. 1 point); this magnitude is similar to the WTP if subjects expected to move from a prior of P(True) = 1/2 to the empirical P(True | message) distribution (7 points, s.e. 0.2 points).

2. WTP is similar for politicized and neutral topics; that is, in this environment there is no evidence of moral wiggling or awareness about motivated reasoning.

3. WTP significantly increases in variance of P(True | message); that is, subjects are aware of their belief susceptibility.[3]

This adds to the broader literature on meta-awareness of biases, as categorized by Gagnon-Bartsch, Rabin, and Schwartzstein 2018 and Schwartzstein 2014. The literature analyzing sophistication and naivete of other biases include base-rate neglect and present bias (for examples, see Dan Benjamin, Bodoh-Creed, and Rabin 2019, Augenblick and Rabin 2015, and O'Donoghue and Rabin 2001). This result indicates a mixed view of sophistication, in that subjects seem aware that their $\varphi > 0$ but not aware of what their m is.

[3]Similarly, it significantly increases in the measure of subject-expected points in point 1 above.

A.2.3 Structural Estimation

The more precise measure of susceptibility allows for an analytical structural estimation of Equation (A.1). In particular, we restrict to *linear* motive functions $m(\theta) = m \cdot \theta$ and define susceptibility φ as the standard deviation of noise in subjects' updating process as above.

Then, we can estimate m up to a linear multiple under the following additional assumptions:

1. $m(\theta) = 0$ for neutral topics. This allows for identification of φ through variance in assessments on neutral topics.

2. $\varphi(x)$ is fixed across questions and individuals. The former assumes that noisiness is a function of priors and signal likelihood, but not the topic or direction of the message; this assumption is necessary to separately identify $m(\theta)$ and $\varphi(x)$.[4] This assumption posits that subjects first set their φ as a function of the *true* likelihood before considering their motive, and only then bias their updating. If $\varphi(x)$ is allowed to vary across individuals, the model is exactly identified and estimates are unstable.[5]

Assuming that subjects have normally-distributed priors, Equation (A.1) can be rewritten as

$$\begin{aligned}&\epsilon_{iq} = \text{logit } a_{iq} - \text{logit } \hat{p}_i - \hat{\varphi}\hat{m}_{iq}R_{iq},\\ &\text{where } \epsilon \sim \mathcal{N}(0, \hat{\varphi}^2),\end{aligned}$$

where hatted variables are the ones to be estimated, and where $R_{iq} \equiv \mathbb{E}_i[\theta_q | \theta_q > \mu_q] - \mathbb{E}_i[\theta_q | \theta_q < \mu_q]$ is proportional to the difference between the subject's upper and lower bound guesses.[6]

[4] That is, $\varphi(\text{Greater Than message}) = \varphi(\text{Less Than message})$ for each question, but only because the likelihood of receiving each signal is 1/2.

[5] For instance, the maximum likelihood estimate does not exist for agents who happen to give the same assessments for the three neutral questions, as the supremum of the likelihood is achieved when φ_i is arbitrarily small and $|m|$ is arbitrarily large.

[6] $R_{iq} \equiv (\text{Upper Bound}_{iq} - \text{Lower Bound}_{iq}) \cdot \frac{\sqrt{\pi}}{\text{Erfc}^{-1}(1/2)} \approx (\text{Upper Bound}_{iq} - \text{Lower Bound}_{iq}) \cdot 1.183$, where Erfc^{-1} is the inverse complementary error function.

That is, we maximize the following log-likelihood function:

$$\sum_{i,q} \log f_{iq} = \frac{IQ \log(2\pi)}{2} \log \hat{\varphi}$$
$$+ \frac{1}{2\hat{\varphi}^2} \sum_i \left[\sum_n (\text{logit } a_{in} - \text{logit } \hat{p}_i)^2 + \sum_y (\text{logit } a_{iy} - \text{logit } \hat{p}_i - \hat{\varphi} \hat{m}_{iy} R_{iy})^2 \right], \quad \text{(A.2)}$$

where $i = 1, \ldots, I$ indexes subjects, $q = 1, \ldots, Q$ indexes all questions, $y = 1, \ldots, Y$ indexes motivated questions, and $n = 1, \ldots, N$ indexes neutral questions.[7]

To maximize this, we take partial derivatives with respect to the parameters $\hat{m}_{iq}$, logit $\hat{p}_i$, and $\hat{\varphi}$. The following are the equations for each parameter; details are in Section A.2.5.

We end up with the following estimates:

$$\hat{m}_{iy} = \frac{\text{logit } a_{iy} - \text{logit } \hat{p}_i}{\hat{\varphi} R_{iy}}.$$
$$\text{logit } \hat{p}_i = \frac{1}{N} \sum_n \text{logit } a_{in}$$
$$\hat{\varphi}^2 = \frac{1}{IQ} \sum_{i,n} (\text{logit } a_{in} - \text{logit } \hat{p}_i)^2 . \quad \text{(A.3)}$$

Estimated motives are proportional to the change from logit assessment and logit prior, and decrease in susceptibility. Estimated priors are equal to the average logit assessments on neutral questions. Estimated susceptibility is the sum of second moments of a_{iq} about the priors $\hat{p}_i$, divided by the total number of individuals and questions, IQ.[8]

Now, we can solve the set of equations in Equation (A.3) for each i and q. $\hat{m}_{iq}$ are discussed in the next section below. $\hat{\varphi}$ is estimated at 0.47. The mean estimated prior $\hat{p}_i$ is 0.58 (s.e. 0.006), and 80 percent of subjects have estimated priors between $\frac{1}{2} - \frac{\sqrt{3}}{6} \equiv 0.211$ and $\frac{1}{2} + \frac{\sqrt{3}}{6} \equiv 0.789$, the bounds necessary for the hypothesis that confidence increases in partisanship from Hypothesis 5.

[7] Technically, these are Q_i, Y_i, and N_i, since some subjects happen to see slightly different numbers of questions. I don't index to make the structural estimate equations easier to understand.

[8] We divide by IQ instead of IN because, in maximizing the likelihood, each politicized question explains variance in posteriors entirely by motives instead of susceptibility. This feature depends on the motive function chosen.

A.2.4 Comparing Estimated Motives Across Questions

As expected, topic-by-topic results are similar to the more reduced-form measure. We see this in Table A.4 using three variants of the main predictions. First, the sign of the estimated motives are in the hypothesized direction from Table 1.1 on almost every question. Secondly, estimated motives are different for Pro-Rep and Pro-Dem subjects in the hypothesized direction on almost every question. Thirdly, estimated motives are positively correlated with initial guesses on almost every question.

The heterogeneity of estimated motives for one's performance compared to others are stark. The Own Performance motive is only greater than zero for male Pro-Rep subjects (0.040, s.e. 0.012, $p = 0.001$) while almost exactly zero for all other subjects (-0.004, s.e. 0.008, $p = 0.592$).[9]

In general, there is no interpretation of the slope of linear motives, just as there is no interpretation of the slope of a linear utility function. However, we can compare motive slopes to each other. For instance, the average $|m_{i,\text{Refugees and crime}}|$ is 0.045, the average $|m_{i,\text{Obama and crime}}|$ is 0.126, and the average $|m_{i,\text{Guns and crime}}|$ is 0.026.[10] This indicates that a 1-unit increase in crime under Barack Obama is given approximately three times the weight as a 1-unit increase in crime due to Germany's refugee laws, and five times the weight as a 1-unit increase in crime after Australia's gun laws.

Note that these are different scales, however. The refugee question asked about the impact on the per-100,000 violent crime rate in Germany, the Obama question asked about the per-million murder and manslaughter rate in the United States, and the gun laws question asked about the average number of victims in a 5-year period. This indicates that the *signal* of the change in crime is more important than the *number* of victims. While (after adjusting for population) the motives regarding the absolute Germany and United States crime amounts are similarly in magnitude (after correcting for population size), the number of gun deaths in Australia is so comparably small that "motives over number of deaths" would be orders of magni-

[9] In fact, the median estimated motive for subjects in all other gender-party subgroups are exactly zero.

[10] Motives here winsorized at the 5% level due to a few extreme outliers.

tude larger.

In some sense, this is reassuring, since it indicates that Republicans are not motivated to believe people are being violently attacked (due to refugees or Obama's policies) but instead that partisans are motivated to believe in signals that their party is correct. On the other hand, it is telling that partisans have stronger motives over party signals compared to motives over loss of human lives.

A.2.5 Structural Estimation Calculation Details

Recall the log likelihood:

$$\sum_{i,q} \log f_{iq} = \frac{IQ \log(2\pi)}{2} \log \hat{k} + \frac{1}{2\hat{k}^2} \sum_i \left[\sum_n (\text{logit } a_{in} - \text{logit } \hat{p}_i)^2 + \sum_y (\text{logit } a_{iy} - \text{logit } \hat{p}_i - \hat{k}\hat{m}_{iy} R_{iy})^2 \right], \quad \text{(A.4)}$$

where $i = 1, \ldots, I$ indexes subjects, $q = 1, \ldots, Q$ indexes all questions, $y = 1, \ldots, Y$ indexes motivated questions, and $n = 1, \ldots, N$ indexes neutral questions

Solving with respect to $\hat{m}_{iq}$:

$$\frac{\partial \left(\sum \log f_{iq}\right)}{\partial \hat{m}_{iq}} = 0 = \frac{1}{2\hat{\varphi}^2}(-2\hat{\varphi}R_{iy})(\text{logit } a_{iy} - \text{logit } \hat{p}_i - \hat{k}\hat{m}_{iy}R_{iy}) = 0$$

$$\implies \hat{m}_{iy} = \frac{\text{logit } a_{iy} - \text{logit } \hat{p}_i}{\hat{\varphi} R_{iy}}. \quad \text{(A.5)}$$

Solving with respect to logit $\hat{p}_i$:

$$\frac{\partial \left(\sum \log f_{iq}\right)}{\partial (\text{logit } \hat{p}_i)} = 0$$

$$= \frac{1}{2\hat{\varphi}^2} \left[-\sum_n 2(\text{logit } a_{in} - \text{logit } \hat{p}_i) - \sum_y 2(\text{logit } a_{iy} - \text{logit } \hat{p}_i - \hat{\varphi}\hat{m}_{iy}R_{iy}) \right]$$

$$\implies \text{logit } \hat{p}_i = \frac{1}{Q} \left[\sum_q \text{logit } a_{iq} - \hat{\varphi} \sum_y \hat{m}_{iy} R_{iy} \right].$$

Plugging in the estimate for $\hat{m}_{iy}$ shows that priors are entirely identified by neutral assessments:

$$\text{logit } \hat{p}_i = \frac{1}{N} \sum_n \text{logit } a_{in}. \quad \text{(A.6)}$$

Solving with respect to $\hat{\varphi}$:

$$\frac{\partial\left(\sum \log f_{iq}\right)}{\partial \hat{\varphi}} = 0 = \frac{IQ}{\hat{\varphi}}$$

$$-\sum_i \left[\frac{1}{\hat{\varphi}^3}\sum_n (\text{logit } a_{in} - \text{logit } \hat{p}_i)^2 + \frac{1}{\hat{\varphi}^3}\sum_y \left[(\text{logit } a_{iy} - \text{logit } \hat{p}_i)(\text{logit } a_{iy} - \text{logit } \hat{p}_i - \hat{\varphi}\hat{m}_{iy}R_{iy})\right]\right]$$

$$\implies IQ\hat{\varphi}^2 + \left[\sum_{i,y} \hat{m}_{iy}R_{iy}(\text{logit } a_{iy} - \text{logit } \hat{p}_i)\right]\hat{\varphi}$$

$$-\sum_i \left[\sum_n (\text{logit } a_{in} - \text{logit } \hat{p}_i)^2 - \sum_y (\text{logit } a_{iy} - \text{logit } \hat{p}_i)^2\right] = 0$$

$$\implies \hat{\varphi} = -\frac{1}{2IQ}\sum_{i,y} \hat{m}_{iy}R_{iy}(\text{logit } a_{iy} - \text{logit } \hat{p}_i)$$

$$+\sqrt{\left(\frac{1}{2IQ}\sum_{i,y} \hat{m}_{iy}R_{iy}(\text{logit } a_{iy} - \text{logit } \hat{p}_i)\right)^2 + \frac{1}{IQ}\sum_{i,q}(\text{logit } a_{iq} - \text{logit } \hat{p}_i)^2}.$$

Plugging in the estimate for $\hat{m}_{iy}$ and $\hat{p}_i$ simplifies this greatly and shows that φ is also entirely identified by neutral assessments:

$$\hat{\varphi}^2 = \frac{1}{IQ}\sum_{i,n}\left(\text{logit } a_{in} - \frac{1}{N}\sum_{i,n'} \text{logit } a_{in'}\right)^2 = \frac{1}{IQ}\sum_{i,n}\left(\text{logit } a_{in} - \text{logit } \hat{p}_i\right)^2. \qquad \text{(A.7)}$$

Tables for Appendix A.2

Table A.3: Determinants of Willingness-To-Pay

	(1)	(2)	(3)	(4)
Assessment SD		22.655***	22.605***	19.809**
		(8.421)	(8.470)	(8.688)
Politicized topics	1.093		1.052	
	(1.697)		(1.701)	
Constant	8.448***	4.234**	3.466	10.504*
	(1.498)	(2.078)	(2.436)	(5.968)
Question FE	No	No	No	Yes
Subject controls	No	No	No	Yes
Observations	482	482	482	482
R^2	0.00	0.02	0.02	0.06

Standard errors in parentheses

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

Notes: OLS, robust standard errors. Dependent variable is Willingness-To-Pay; this occurs in round 12. Party-indifferent subjects included. Assessment SD is the standard deviation of the subject's news veracity assessments in all other rounds. Politicized topics defined in Table 1.

Table A.4: Estimated Motives: By Direction, By Party, and By Prior

	Hyp. direction	Pro-R vs. Pro-D	Diff. by prior
	(1)	(2)	(3)
Climate topic	0.083***	0.068***	0.085***
	(0.010)	(0.019)	(0.009)
Race topic	0.075***	0.029	0.059***
	(0.019)	(0.041)	(0.009)
Mobility topic	0.032***	0.039***	0.021***
	(0.005)	(0.011)	(0.005)
Refugees topic	0.010***	0.017***	0.004**
	(0.002)	(0.005)	(0.002)
Obama crime topic	0.026***	0.043***	0.009*
	(0.006)	(0.013)	(0.005)
Gender topic	0.605***	0.534	0.279*
	(0.192)	(0.413)	(0.149)
Gun laws topic	0.003**	0.001	0.002
	(0.001)	(0.003)	(0.001)
Media topic	0.001	0.018*	0.020***
	(0.004)	(0.009)	(0.004)
Rep score topic	0.029***	0.073***	0.021***
	(0.006)	(0.014)	(0.007)
Dem score topic	0.032***	0.050***	0.025***
	(0.007)	(0.014)	(0.007)
Own performance topic	0.007**		0.016***
	(0.003)		(0.004)
Question FE	No	Yes	Yes
Observations	8785	7902	8785
R^2	0.01	0.03	0.03

Standard errors in parentheses

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

Notes: For each topic, estimated motives winsorized at the 5% level. Columns correspond to different independent and dependent variables. Column 1 measures the mean estimated motive by question in the direction hypothesized in Table 1. Estimated motives are multiplied by 1 for Pro-Motive and -1 for Anti-Motive. Column 2 regresses estimated motives on a dummy for Pro-Rep for each question, multiplying by the direction in Table 1. Column 2 regresses estimated motives on the z score of the initial guess for each question; the guess is winsorized at the 5% level.

A.3 Study Materials: Exact Question Wordings

Crime Under Obama

Some people believe that the Obama administration was too soft on crime and that violent crime increased during his presidency, while others believe that President Obama's pushes towards criminal justice reform and reducing incarceration did not increase violent crime.

This question asks how murder and manslaughter rates changed during the Obama administration. In 2008 (before Obama became president), the murder and manslaughter rate was 54 per million Americans.

In 2016 (at the end of Obama's presidency), what was the per-million murder and manslaughter rate?

Correct answer: 53.

Source linked on results page: `http://bit.ly/us-crime-rate`

Upward Mobility

In 2017, Donald Trump signed into law the largest tax reform bill since Ronald Reagan's 1981 and 1986 bills. Some people believe that Reagan's reforms accelerated economic growth and allowed lower-income Americans to reap the benefits of lower taxes, while other people believe that this decreased the government's spending to help lower-income Americans get ahead.

This question asks whether children who grew up in low-income families during Reagan's tenure were able to benefit from his tax reforms.

Of Americans who were born in the lowest-income (bottom 20%) families from 1980-1985, what percent rose out of the lowest-income group as adults?

(Please guess between 0 and 100.)

Correct answer: 64.9.

Source linked on results page: `http://bit.ly/us-upward-mobility` *(page 47)*

Racial Discrimination

In the United States, white Americans have higher salaries than black Americans on average. Some people attribute these differences in income to differences in education, training, and culture, while others attribute them more to racial discrimination.

In a study, researchers sent fictitious resumes to respond to thousands of help-wanted ads in newspapers. The resumes sent had identical skills and education, but the researchers gave half of the (fake) applicants stereotypically White names such as Emily Walsh and Greg Baker, and gave the other half of the applicants stereotypically Black names such as Lakisha Washington and Jamal Jones.

9.65 percent of the applicants with White-sounding names received a call back. What percent of the applicants with Black-sounding names received a call back?

(Please guess between 0 and 100.)

Correct answer: 6.45.

Source linked on results page: `http://bit.ly/labor-market-discrimination`

Gender and Math GPA

In the United States, men are more likely to enter into mathematics and math-related fields. Some people attribute this to gender differences in interest in or ability in math, while others attribute it to other factors like gender discrimination.

This question asks whether high school boys and girls differ substantially in how well they do in math classes. A major testing service analyzed data on high school seniors and compared the average GPA for male and female students in various subjects.

Male students averaged a 3.04 GPA (out of 4.00) in math classes. What GPA did female students average in math classes?

(Please guess between 0.00 and 4.00.)

Correct answer: 3.15.

Source linked on results page: `http://bit.ly/gender-hs-gpa`

Refugees and Violent Crime

Some people believe that the U.S. has a responsibility to accept refugees into the country, while others believe that an open-doors refugee policy will be taken advantage of by criminals and put Americans at risk.

In 2015, German leader Angela Merkel announced an open-doors policy that allowed all Syrian refugees who had entered Europe to take up residence in Germany. From 2015-17, nearly one million Syrians moved to Germany. This question asks about the effect of Germany's open-doors refugee policy on violent crime rates.

In 2014 (before the influx of refugees), the violent crime rate in Germany was 224.0 per hundred-thousand people.

In 2017 (after the entrance of refugees), what was the violent crime rate in Germany per hundred-thousand people?

Correct answer: 228.2.

Sources linked on results page: Main site: `http://bit.ly/germany-crime-main-site`. 2014 and 2015 data: `http://bit.ly/germany-crime-2014-2015`. 2016 and 2017 data: `http://bit.ly/germany-crime-2016-2017`.

Climate change

Some people believe that there is a scientific consensus that human activity is causing global warming and that we should have stricter environmental regulations, while others believe that scientists are not in agreement about the existence or cause of global warming and think that stricter environmental regulations will sacrifice jobs without much environmental gain.

This question asks about whether most scientists think that global warming is caused by humans. A major nonpartisan polling company surveyed thousands of scientists about the existence and cause of global warming.

What percent of these scientists believed that "Climate change is mostly due to human activity"?

(Please guess between 0 and 100.)

Correct answer: 87.

Source linked on results page: `http://bit.ly/scientists-climate-change`

Gun Reform

The United States has a homicide rate that is much higher than other wealthy countries. Some people attribute this to the prevalence of guns and favor stricter gun laws, while others believe that stricter gun laws will limit Americans' Second Amendment rights without reducing homicides very much.

After a mass shooting in 1996, Australia passed a massive gun control law called the National Firearms Agreement (NFA). The law illegalized, bought back, and destroyed almost one million firearms by 1997, mandated that all non-destroyed firearms be registered, and required a lengthy waiting period for firearm sales.

Democrats and Republicans have each pointed to the NFA as evidence for/against stricter gun laws. This question asks about the effect of the NFA on the homicide rate in Australia.

In the five years before the NFA (1991-1996), there were 319.8 homicides per year in Australia. In the five years after the NFA (1998-2003), how many homicides were there per year in Australia?

Correct answer: 318.6.

Sources linked on results page: `http://bit.ly/australia-homicide-rate` *(Suicides declined substantially, however. For details:* `http://bit.ly/impact-australia-gun-laws`*.)*

Media Bias

Some people believe that the media is unfairly biased towards Democrats, while some believe it is balanced, and others believe it is biased towards Republicans.

This question asks whether journalists are more likely to be Democrats than Republicans.

A representative sample of journalists were asked about their party affiliation. Of those either affiliated with either the Democratic or Republican Party, what percent of journalists are Republicans?

(Please guess between 0 and 100.)

Correct answer: 19.8.

Source linked on results page: http://bit.ly/journalist-political-affiliation

Party Relative Performance

Subjects are randomly assigned to see either the Democrats' score (and asked to predict the Republicans' score) or to see the Republicans' score (and asked to predict the Democrats' score).

Democrats' Relative Performance

This question asks whether you think Democrats or Republicans did better on this study about political and U.S. knowledge. I've compared the average points scored by Democrats and Republicans among 100 participants (not including yourself).

The Republicans scored 70.83 points on average.

How many points do you think the Democrats scored on average?

(Please guess between 0 and 100)

Correct answer: 72.44.

Republicans' Relative Performance

This question asks whether you think Democrats or Republicans did better on this study about political and U.S. knowledge. I've compared the average points scored by Democrats and Republicans among 100 participants (not including yourself).

The Democrats scored 72.44 points on average.

How many points do you think the Republicans scored on average?

(Please guess between 0 and 100)

Correct answer: 70.83.

Own Relative Performance

How well do you think you performed on this study about political and U.S. knowledge? I've compared the average points you scored for all questions (prior to this one) to that of 100 other participants.

How many of the 100 do you think you scored higher than?

(Please guess between 0 and 100.)

Correct answer: Depends on participant's performance.

Random Number

A computer will randomly generate a number between 0 and 100. What number do you think the computer chose?

(As a reminder, it is in your best interest to guess an answer that is close to the computer's choice, even if you don't perfectly guess it.)

Correct answer: Randomly generated for each participant.

Latitude of Center of the United States

The U.S. National Geodetic Survey approximated the geographic center of the continental United States. (This excludes Alaska and Hawaii, and U.S. territories.)

How many degrees North is this geographic center?

(Please guess between 0 and 90. The continental U.S. lies in the Northern Hemisphere, the Equator is 0 degrees North, and the North Pole is 90 degrees North.)

Correct answer: 39.833.

Source linked on results page: `http://bit.ly/center-of-the-us`

Longitude of Center of the United States

The U.S. National Geodetic Survey approximated the geographic center of the continental United States. (This excludes Alaska and Hawaii, and U.S. territories.)

How many degrees West is this geographic center?

(Please guess between 0 and 180. The continental U.S. lies in the Western Hemisphere, which ranges from 0 degrees West to 180 degrees West.)

Correct answer: 98.583.

Source linked on results page: `http://bit.ly/center-of-the-us`

Comprehension Check: Current Year

In 1776 our fathers brought forth, upon this continent, a new nation, conceived in Liberty, and dedicated to the proposition that all men are created equal.

What is the year right now?

This is not a trick question and the first sentence is irrelevant; this is a comprehension check to make sure you are paying attention. For this question, your lower and upper bounds should be equal to your guess if you know what year it currently is.

Correct answer: 2018.

Source linked on results page: `http://bit.ly/what-year-is-it`

A.4 Replication

I preregistered a replication for the findings from this paper. In particular, I ran this in conjunction with a debiasing treatment; the replication tests whether the control group from that sample satisfies the hypotheses from this experiment. This section reports all replication results that were specified in the pre-analysis plan in Thaler (2019).

There are a few differences between the replication sample and the original sample. The replication was conducted approximately one year later, on July 8-9, 2019. The replication questions included additional topics and variants of the original questions.[11] There were no neutral questions.

The sample includes 1,050 subjects recruited from Amazon's Mechanical Turk platform that passed pre-specified comprehension checks that are akin to those in the original experiment. 356 subjects never received a treatment and 694 subjects received a treatment after the end of round 3. As such, the control group includes 1,050 subjects for the first three rounds, and 356 subjects in the remaining rounds. The debiasing treatment group observations are dropped from all analyses. There are 982 subjects who are either Pro-Rep or Pro-Dem, and these subjects give 5,314 news veracity assessments on politicized topics.

A.4.1 Primary Outcomes

The most important primary outcome results are strongly replicated. As seen in the first column of Table A.5, subjects give statistically significantly higher assessments to Pro-Party news than to Anti-Party news ($p < 0.001$).[12] The second column shows that this gap is increasing in partisanship ($p = 0.006$).

[11] In particular, two new politicized topics were added: Wage Growth and Healthcare. On six of the politicized topics, subjects received different versions of the original question as part of a separate experiment on positivity-motivated reasoning.

[12] The coefficient is smaller in the replication, due in large part to the new added questions. On the questions with the exact same wording as the original study, the treatment effect for is 7.1 percentage points (s.e. 1.2 percentage points). On other politicized questions, the treatment effect is 3.5 percentage points (s.e. 1.0 percentage points). Of the original questions, the effects on the following topics were significant at $p < 0.05$ in the predicted direction: Climate Change, Race, Refugees, Gun Laws, Party Performance, Own Performance. The effects on the following topics were not significant at $p < 0.05$: Obama and Crime, Gender, Media.

The next-most important primary outcome results are strongly replicated. Table A.5 shows that subjects give statistically significantly higher assessments to Fake News than to True News. This holds both when Pro-Party / Anti-Party news is not controlled for (column 3) and when Pro-Party / Anti-Party news is controlled for (column 4); both results are significant at $p < 0.001$.

The main alternative measure of motivated reasoning is suggestively replicated. As seen in the first column of Table A.6, results suggest that subjects are more likely to update in the direction of the Pro-Party message compared to the Anti-Party message ($p = 0.055$).[13] The third column shows that, as in Section 1.5.4, this difference vanishes once the news veracity assessment measure is controlled for.

A.4.2 Secondary Outcomes

The underperformance result is strongly replicated. Subjects score 66.3 points (s.e. 0.4 points) on politicized news assessments and 65.5 points (s.e. 1.6 points) on performance news assessments on average. Both of these are statistically significantly lower than 75 points, the score that subjects would receive if they had answered "5/10 chance the message came from True News" ($p < 0.001$).

The result that subjects' confidence intervals are overprecise is strongly replicated. On politicized topics, subjects' 50 percent confidence intervals contain the correct answer 44.1 percent of the time (s.e. 0.8 percent); this is statistically significantly different from 50 percent ($p < 0.001$). As seen in Table A.7, the result that this measure of overprecision is increasing in partisanship is suggestive ($p = 0.066$).

The two polarization results are replicated. On politicized topics, Table A.6 shows that subjects are statistically significantly more likely to follow Polarizing news than anti-Polarizing news ($p = 0.031$).[14] Subjects also state initial medians that are more likely to be in the Pro-

[13] As with the main effect, the coefficient is smaller in the replication, due in large part to the new questions. On the questions with the exact same wording as the original study, the treatment effect for is 5.7 percentage points (s.e. 2.6 percentage points). On other politicized questions, the treatment effect is 2.0 percentage points (s.e. 2.6 percentage points).

[14] As in Section 1.5.4, this difference vanishes once the news assessment measure is controlled for.

Party direction ($p < 0.001$).

A.4.3 Untested Replications

I did not register replication tests for other results. Given the limited sample size, there would be insufficient statistical power for detecting effect sizes similar to the ones in the original experiment. Performance-driven motivated reasoning and overconfidence tests were not pre-specified.[15] Demographic heterogeneity, robustness exercises, and minor results were also not tested. Further work can test whether these results replicate on a larger sample.

[15] Results, however, are broadly similar to those in the main experiment. For instance, subjects assess Pro-Performance news to be 8.0 percentage points higher than Anti-Performance news (s.e. 2.6 percentage points; $p = 0.003$). Subjects expect to place 7.2 percentiles higher than they actually place relative to 100 pilot subjects (s.e. 1.6 percentiles; $p < 0.001$). Both of these are only significant for male subjects.

Table A.5: The Effect of News Direction and Actual Veracity on Perceived Veracity: Replication

	(1)	(2)	(3)	(4)
Pro-Party News	0.053***	0.010		0.046***
	(0.009)	(0.018)		(0.009)
Partisanship x		0.044***		
Pro-Party		(0.016)		
True News			-0.043***	-0.033***
			(0.009)	(0.009)
Question FE	Yes	Yes	Yes	Yes
Round FE	Yes	Yes	Yes	Yes
Subject FE	Yes	Yes	Yes	Yes
Observations	5314	5314	5314	5314
R^2	0.32	0.32	0.32	0.33
Mean	0.578	0.578	0.578	0.578

Standard errors in parentheses

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

Notes: OLS, errors clustered at subject level. Only Pro-Party / Anti-Party news observations. Partisanship is the absolute difference between ratings of the Republican and Democratic parties.

Table A.6: Changing Guess to Follow Message Given News Direction: Replication

	(1)	(2)	(3)	(4)
Pro-Party News	0.038*		-0.020	
	(0.020)		(0.018)	
Polarizing News		0.040**		-0.018
		(0.019)		(0.017)
P(True)			1.108***	1.107***
			(0.055)	(0.055)
Question FE	Yes	Yes	Yes	Yes
Round FE	Yes	Yes	Yes	Yes
Subject FE	Yes	Yes	Yes	Yes
Observations	5314	5314	5314	5314
R^2	0.34	0.34	0.48	0.48
Mean	0.654	0.654	0.654	0.654

Standard errors in parentheses

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

Notes: OLS, errors clustered at subject level. Partisanship is the absolute difference between ratings of the Republican and Democratic parties. Only Pro-Party / Anti-Party news observations. Polarizing News is defined as news that tells subjects that, compared to their initial guess, the answer is in the opposite direction from the population mean. Dependent variable is 1 if subjects change their guess upwards when the message says "Greater Than" or downwards when the message says "Less Than," -1 if they change their guess in the opposite direction, and 0 if they do not change their guess.

Table A.7: Overprecision and Partisanship: Replication

	(1)	(2)
Partisanship	0.055*	0.055*
	(0.030)	(0.030)
Subject controls	No	Yes
Observations	5314	5314
R^2	0.00	0.01

Standard errors in parentheses

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

Notes: OLS, errors clustered at subject level. Only politicized topics included. Partisanship is the absolute difference between ratings of the Republican and Democratic parties. Subject controls are race, gender, age, log(income), education, religion, and whether home state voted for Trump or Clinton in 2016.

A.5 Online Appendix: Additional Robustness for Table 1.2

This section presents additional robustness checks for the main results in Table 1.2. These results are similar for each treatment, if subjects who fail comprehension checks are included instead of excluded, and if the dependent variable is the logit probability of news veracity assessments instead of the assessments themselves.

Main Results by Treatment

First, we look at Table 1.2 by treatment; recall that there are two independent treatment arms. For the first treatment arm, half of subjects are asked to give a second guess to the original question in addition to their news veracity assessment after receiving the message; the other half do not, but have one round in which their willingness-to-pay for a message is elicited.

For the second treatment arm, one-third of subjects are told in the instructions that "Whether you get your message from True News or Fake News is random *and each source is equally likely*." The other two-thirds are told that "Whether you get your message from True News or Fake News is random." That is, one-third of subjects are given a 50-50 prior, and the other two-thirds are not.

The main takeaway is that neither treatment arm substantially affects the main results, and neither arm affects the mean veracity assessment.

Table A.8: The Effect of News Direction and Actual Veracity on Perceived Veracity: Second-Guess Treatment

	(1)	(2)	(3)	(4)	(5)	(6)
Pro-Party News	0.090***	0.092***	0.040**	0.031***		0.081***
	(0.009)	(0.010)	(0.019)	(0.009)		(0.010)
Partisanship x			0.055***			
Pro-Party			(0.018)			
Anti-Party News				-0.057***		
				(0.010)		
True News					-0.061***	-0.035***
					(0.009)	(0.009)
Neutral News	No	No	No	Yes	No	No
Question FE	Yes	Yes	Yes	No	Yes	Yes
Round FE	Yes	Yes	Yes	Yes	Yes	Yes
Subject controls	Yes	No	No	No	No	No
Subject FE	No	Yes	Yes	Yes	Yes	Yes
Observations	4085	4085	4085	5455	4085	4085
R^2	0.05	0.24	0.25	0.20	0.23	0.25
Mean	0.577	0.577	0.577	0.580	0.577	0.577

Standard errors in parentheses

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

Notes: OLS, errors clustered at subject level. Neutral News indicates that Pro-Party / Anti-Party news assessments are compared to assessments on Neutral topics. These classifications are defined in Table 1.1. Controls: race, gender, log(income), years of education, religion, and state. Partisanship is the absolute difference between ratings of the Republican and Democratic parties. Observations only for Second-Guess treatment.

Table A.9: The Effect of News Direction and Actual Veracity on Perceived Veracity: Willingness-to-Pay Treatment

	(1)	(2)	(3)	(4)	(5)	(6)
Pro-Party News	0.095***	0.085***	0.040**	0.043***		0.074***
	(0.009)	(0.009)	(0.018)	(0.010)		(0.009)
Partisanship x			0.045***			
Pro-Party			(0.015)			
Anti-Party News				-0.039***		
				(0.010)		
True News					-0.056***	-0.032***
					(0.009)	(0.009)
Neutral News	No	No	No	Yes	No	No
Question FE	Yes	Yes	Yes	No	Yes	Yes
Round FE	Yes	Yes	Yes	Yes	Yes	Yes
Subject controls	Yes	No	No	No	No	No
Subject FE	No	Yes	Yes	Yes	Yes	Yes
Observations	3817	3817	3817	5097	3817	3817
R^2	0.05	0.25	0.25	0.22	0.24	0.26
Mean	0.569	0.569	0.569	0.568	0.569	0.569

Standard errors in parentheses

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

Notes: OLS, errors clustered at subject level. Neutral News indicates that Pro-Party / Anti-Party news assessments are compared to assessments on Neutral topics. These classifications are defined in Table 1.1. Controls: race, gender, log(income), years of education, religion, and state. Partisanship is the absolute difference between ratings of the Republican and Democratic parties. Observations only for Willingness-to-Pay treatment.

Table A.10: The Effect of News Direction and Actual Veracity on Perceived Veracity: Given 50-50 Prior Treatment

	(1)	(2)	(3)	(4)	(5)	(6)
Pro-Party News	0.090***	0.088***	0.068***	0.046***		0.078***
	(0.011)	(0.011)	(0.021)	(0.011)		(0.012)
Partisanship x			0.021			
Pro-Party			(0.019)			
Anti-Party News				-0.040***		
				(0.013)		
True News					-0.056***	-0.029**
					(0.011)	(0.012)
Neutral News	No	No	No	Yes	No	No
Question FE	Yes	Yes	Yes	No	Yes	Yes
Round FE	Yes	Yes	Yes	Yes	Yes	Yes
Subject controls	Yes	No	No	No	No	No
Subject FE	No	Yes	Yes	Yes	Yes	Yes
Observations	2674	2674	2674	3568	2674	2674
R^2	0.06	0.27	0.27	0.22	0.25	0.27
Mean	0.572	0.572	0.572	0.571	0.572	0.572

Standard errors in parentheses

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

Notes: OLS, errors clustered at subject level. Neutral News indicates that Pro-Party / Anti-Party news assessments are compared to assessments on Neutral topics. These classifications are defined in Table 1.1. Controls: race, gender, log(income), years of education, religion, and state. Partisanship is the absolute difference between ratings of the Republican and Democratic parties. Observations only for Given 50-50 Prior treatment.

Table A.11: The Effect of News Direction and Actual Veracity on Perceived Veracity: Not Given 50-50 Prior Treatment

	(1)	(2)	(3)	(4)	(5)	(6)
Pro-Party News	0.093***	0.088***	0.024	0.033***		0.077***
	(0.008)	(0.008)	(0.016)	(0.008)		(0.008)
Partisanship x			0.065***			
Pro-Party			(0.014)			
Anti-Party News				-0.052***		
				(0.009)		
True News					-0.061***	-0.037***
					(0.008)	(0.008)
Neutral News	No	No	No	Yes	No	No
Question FE	Yes	Yes	Yes	No	Yes	Yes
Round FE	Yes	Yes	Yes	Yes	Yes	Yes
Subject controls	Yes	No	No	No	No	No
Subject FE	No	Yes	Yes	Yes	Yes	Yes
Observations	5228	5228	5228	6984	5228	5228
R^2	0.05	0.24	0.24	0.20	0.22	0.24
Mean	0.574	0.574	0.574	0.576	0.574	0.574

Standard errors in parentheses

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

Notes: OLS, errors clustered at subject level. Neutral News indicates that Pro-Party / Anti-Party news assessments are compared to assessments on Neutral topics. These classifications are defined in Table 1.1. Controls: race, gender, log(income), years of education, religion, and state. Partisanship is the absolute difference between ratings of the Republican and Democratic parties. Observations only for Not Given 50-50 Prior treatment.

Results Without Comprehension Checks

The main results do not include subjects who fail attention and comprehension checks. As such, 313 of 1300 subjects are removed from the analysis. This subsection does the analysis without remove any subjects. Results are directionally identical when all subjects are included.

Table A.12: The Effect of News Direction and Actual Veracity on Perceived Veracity: Including Subjects Who Fail Comprehension Checks

	(1)	(2)	(3)	(4)	(5)	(6)
Pro-Party News	0.075***	0.071***	0.027***	0.031***		0.064***
	(0.005)	(0.006)	(0.010)	(0.006)		(0.006)
Partisanship x			0.048***			
Pro-Party			(0.010)			
Anti-Party News				-0.038***		
				(0.006)		
True News					-0.043***	-0.026***
					(0.006)	(0.006)
Neutral News	No	No	No	Yes	No	No
Question FE	Yes	Yes	Yes	No	Yes	Yes
Round FE	Yes	Yes	Yes	Yes	Yes	Yes
Subject controls	Yes	No	No	No	No	No
Subject FE	No	Yes	Yes	Yes	Yes	Yes
Observations	10478	10478	10478	13991	10478	10478
R^2	0.03	0.30	0.30	0.27	0.29	0.30
Mean	0.560	0.560	0.560	0.561	0.560	0.560

Standard errors in parentheses

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

Notes: OLS, errors clustered at subject level. Neutral News indicates that Pro-Party / Anti-Party news assessments are compared to assessments on Neutral topics. These classifications are defined in Table 1.1. Controls: race, gender, log(income), years of education, religion, and state. Partisanship is the absolute difference between ratings of the Republican and Democratic parties. Observations *include* subjects who failed comprehension checks.

Results Using Logit Veracity Assessments

The model suggests that the relevant dependent variable is logit(P(True)) instead of P(True). Table A.13 is the same as Table 1.2 but with this new dependent variable. Technically, since logit(0) and logit(1) are undefined, they are replaced here with logit(0.025) and logit(0.975).[16]

[16] Subjects choose P(True) = 0 to maximize expected earnings if and only if they believe P(True) $\in [0, 0.05]$. 0.025 is the midpoint of this range. Results are similar if 0.05 is chosen or if these observations are removed.

Table A.13: The Effect of News Direction and Actual Veracity on Perceived Veracity: Logit Veracity Assessments

	(1)	(2)	(3)	(4)	(5)	(6)
Pro-Party News	0.507***	0.486***	0.208***	0.165***		0.423***
	(0.037)	(0.039)	(0.076)	(0.041)		(0.040)
Partisanship x			0.287***			
Pro-Party			(0.068)			
Anti-Party News				-0.302***		
				(0.044)		
True News					-0.330***	-0.193***
					(0.038)	(0.038)
Neutral News	No	No	No	Yes	No	No
Question FE	Yes	Yes	Yes	No	Yes	Yes
Round FE	Yes	Yes	Yes	Yes	Yes	Yes
Subject controls	Yes	No	No	No	No	No
Subject FE	No	Yes	Yes	Yes	Yes	Yes
Observations	7902	7902	7902	10552	7902	7902
R^2	0.04	0.25	0.25	0.21	0.23	0.25
Mean	0.391	0.391	0.391	0.405	0.391	0.391

Standard errors in parentheses

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

Notes: Dependent variable is logit(P(True)). OLS, errors clustered at subject level. Neutral News indicates that Pro-Party / Anti-Party news assessments are compared to assessments on Neutral topics. These classifications are defined in Table 1.1. Controls: race, gender, log(income), years of education, religion, and state. Partisanship is the absolute difference between ratings of the Republican and Democratic parties.

Table A.14: The Effect of News Direction and Actual Veracity on Perceived Veracity: Rounds 1-6

	(1)	(2)	(3)	(4)	(5)	(6)
Pro-Party News	0.080***	0.077***	0.047**	0.030***		0.069***
	(0.009)	(0.011)	(0.021)	(0.010)		(0.011)
Partisanship x			0.031*			
Pro-Party			(0.019)			
Anti-Party News				-0.045***		
				(0.011)		
True News					-0.046***	-0.024**
					(0.011)	(0.011)
Neutral News	No	No	No	Yes	No	No
Question FE	Yes	Yes	Yes	No	Yes	Yes
Round FE	Yes	Yes	Yes	Yes	Yes	Yes
Subject controls	Yes	No	No	No	No	No
Subject FE	No	Yes	Yes	Yes	Yes	Yes
Observations	3521	3521	3521	4919	3521	3521
R^2	0.04	0.36	0.36	0.28	0.35	0.36
Mean	0.562	0.562	0.562	0.563	0.562	0.562

Standard errors in parentheses

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

Notes: OLS, errors clustered at subject level. Neutral News indicates that Pro-Party / Anti-Party news assessments are compared to assessments on Neutral topics. These classifications are defined in Table 1.1. Controls: race, gender, log(income), years of education, religion, and state. Partisanship is the absolute difference between ratings of the Republican and Democratic parties. Observations only for rounds 1-6.

Table A.15: The Effect of News Direction and Actual Veracity on Perceived Veracity: Rounds 7-12

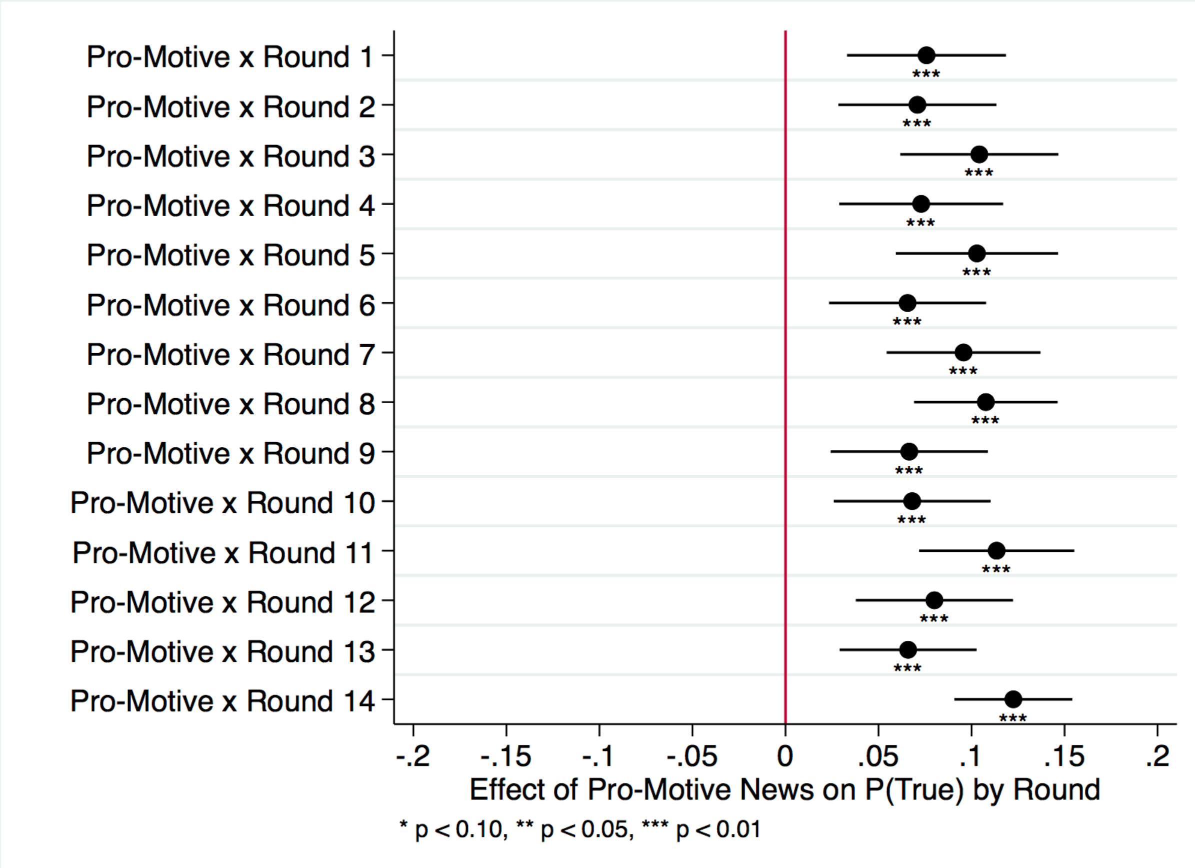
	(1)	(2)	(3)	(4)	(5)	(6)
Pro-Party News	0.095***	0.090***	0.042*	0.040***		0.078***
	(0.009)	(0.011)	(0.022)	(0.010)		(0.011)
Partisanship x			0.050***			
Pro-Party			(0.019)			
Anti-Party News				-0.046***		
				(0.011)		
True News					-0.067***	-0.044***
					(0.011)	(0.011)
Neutral News	No	No	No	Yes	No	No
Question FE	Yes	Yes	Yes	No	Yes	Yes
Round FE	Yes	Yes	Yes	Yes	Yes	Yes
Subject controls	Yes	No	No	No	No	No
Subject FE	No	Yes	Yes	Yes	Yes	Yes
Observations	3485	3485	3485	4737	3485	3485
R^2	0.04	0.37	0.37	0.30	0.36	0.37
Mean	0.586	0.586	0.586	0.587	0.586	0.586

Standard errors in parentheses

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

Notes: OLS, errors clustered at subject level. Neutral News indicates that Pro-Party / Anti-Party news assessments are compared to assessments on Neutral topics. These classifications are defined in Table 1.1. Controls: race, gender, log(income), years of education, religion, and state. Partisanship is the absolute difference between ratings of the Republican and Democratic parties. Observations only for rounds 7-12.

Figure A.5: Round-by-Round Effects of News Direction on Perceived Veracity



Notes: OLS regression coefficients, errors clustered at subject level. FE included for round number and topic. Pro-Motive (vs. Anti-Motive) news is defined in Table 1.1. Error bars correspond to 95 percent confidence intervals.

A.6 Online Appendix: Study Materials

A.6.1 Flow of Experiment

Subjects see a series of pages in the following order:

- Introduction and Consent
- Demographics and Current Events Quiz
- Opinions
- Instructions for Question Pages
- Question 1
- Instructions for News Assessment Pages
- News Assessment 1
- Question 2, News Assessment 2, ..., Question 14, News Assessment 14
- Feedback
- Results and Payment

Screenshots for each of the pages are in the following subsection. Exact question wordings are in the following subsection. Red boxes are not shown to subjects and are included for illustration purposes only. Results pages here are cut off after three questions, but all results are shown to subjects. Choices on the Demographics page and statements on the Opinions page are randomly ordered.

Subjects in the Willingness-To-Pay treatment see the News Valuation page between Question 12 and News Assessment 12. They see the black bar page if their elicited valuation is lower than the random number.

Subjects in the Second Guess treatment see the version of the News Assessment page with the message "After seeing this message and assessing its truthfulness, what is your guess of the answer to the original question?"

A.6.2 Study Materials

Introduction

You are invited to participate in this online study on political attitudes. This is a research project being conducted by Michael Thaler, a PhD student in economics at Harvard University.

Your participation in this survey is entirely voluntary. You may refuse to take part in the research or exit the survey at any time without penalty.

If you choose to be in the study, you will complete a series of questions related to issues affecting the United States today. The study should take approximately 20 minutes to complete, but you may take up to 45 minutes. You will have a chance to earn a bonus of $10.00 in addition to your participation earnings.

Your specific answers will not be shared with anyone, and for the purpose of privacy please do not include your name or other personally identifiable information in your responses. Please make sure to mark your Amazon Profile as private if you do not want it to be accessible via your Mechanical Turk Worker ID.

If you have any questions or concerns, please contact Michael Thaler at michaelthaler@g.harvard.edu.

You may print or save a copy of this information sheet for your own records. **Please do not press the back button, refresh, or leave the page at any time or else you might have a server error; if this happens, you will not be able to reenter the study or earn your payment.**

If you wish to participate in the study, please indicate below that you have read the instructions and enter your Mechanical Turk Worker ID for payment.

What is your MTurk Worker ID number? This is required for payment.

○ I have read the above information and would like to participate in the study.

Next

Demographic Information and Current Events Quiz

It is important for this study that you answer these questions honestly.

Your earnings and bonus are not affected by your answers to these questions.

What is your age?

What is your gender?

- Male
- Female
- Other / Prefer not to answer

What is your race/ethnicity?

- Hispanic or Latino
- Asian
- White
- American Indian
- Black or African American
- Two or more of these
- Other / Prefer not to answer

What is the highest level of education you have completed?

- Did not graduate high school
- High school graduate or GED
- Began college, no degree
- Associate's degree
- Bachelor's degree
- Postgraduate or professional degree

What religious group do you consider yourself affiliated with?

- Mainline Protestant
- Historically black Protestant
- Evangelical Protestant
- Catholic
- Other Christian
- Jewish
- Muslim
- Other religion or faith
- Atheist
- Agnostic
- Unaffiliated

Which US state or territory do you currently live in?

What was your total household income before taxes during the past 12 months?

- Less than $20,000
- $20,000 to $29,999
- $30,000 to $39,999
- $40,000 to $49,999
- $50,000 to $69,999
- $70,000 to $99,999
- $100,000 to $149,999
- $150,000 or more

In politics today, do you consider yourself a Republican, a Democrat, or an Independent?

- Democrat
- Republican
- Independent

Where do you see yourself on the liberal/conservative spectrum?

- Extremely liberal
- Liberal
- Slightly liberal
- Moderate
- Slightly conservative
- Conservative
- Extremely conservative

Please rate how you feel about the Republican Party using a scale of 0 to 100. The higher the number, the more favorable you feel toward the Republican Party.

50

Please rate how you feel about the Democratic Party using a scale of 0 to 100. The higher the number, the more favorable you feel toward the Democratic Party.

50

Who is the current President of France?

- Theresa May
- Charles de Gaulle
- Emmanuel Macron
- Marine Le Pen
- Justin Trudeau

Who won the recent special election in Alabama for the U.S. Senate?

- Doug Jones
- Roy Moore
- Richard Shelby
- Luther Strange
- Thad Cochran

Who was Hillary Clinton's running mate in the 2016 presidential election?

- Martin O'Malley
- Jim Webb
- Joe Biden
- Bernie Sanders
- Tim Kaine

Who is the most recently-appointed Supreme Court Justice?

- Merrick Garland
- Anthony Kennedy
- John Roberts
- Stephen Breyer
- Neil Gorsuch

Next

Opinions

For each of the following statements, please indicate whether you agree or disagree.

Your earnings and bonus are not affected by your answers to these questions.

Statement					
Donald Trump is doing a good job as president.	○ Strongly agree	○ Somewhat agree	○ Neither agree nor disagree	○ Somewhat disagree	○ Strongly disagree
The United States media tends to be biased in favor of Democrats over Republicans.	○ Strongly agree	○ Somewhat agree	○ Neither agree nor disagree	○ Somewhat disagree	○ Strongly disagree
One reason why there are more men than women working in science, technology, engineering, and math is that men are inherently more interested in these fields.	○ Strongly agree	○ Somewhat agree	○ Neither agree nor disagree	○ Somewhat disagree	○ Strongly disagree
The Obama administration did a good job at dealing with violent crime.	○ Strongly agree	○ Somewhat agree	○ Neither agree nor disagree	○ Somewhat disagree	○ Strongly disagree
The United States has a responsibility to accept refugees into the country.	○ Strongly agree	○ Somewhat agree	○ Neither agree nor disagree	○ Somewhat disagree	○ Strongly disagree
Racial discrimination is a major reason why many black people can't get ahead these days.	○ Strongly agree	○ Somewhat agree	○ Neither agree nor disagree	○ Somewhat disagree	○ Strongly disagree
Gun laws should be more strict than they are today.	○ Strongly agree	○ Somewhat agree	○ Neither agree nor disagree	○ Somewhat disagree	○ Strongly disagree
There is solid evidence that the Earth is getting warmer and that this is due to human activity.	○ Strongly agree	○ Somewhat agree	○ Neither agree nor disagree	○ Somewhat disagree	○ Strongly disagree
The recent tax reform bill will help lower-income Americans get ahead.	○ Strongly agree	○ Somewhat agree	○ Neither agree nor disagree	○ Somewhat disagree	○ Strongly disagree

Next

Instructions for Question Pages

Throughout this study, you will see several types of pages, including 14 Question pages.

On each of the Question pages, you will be asked to guess the answer to a factual question; each question has a correct numerical answer. In addition to your guaranteed HIT payment, you will have a chance to win an additional bonus of $10.00 based on your guesses to these questions and questions on other pages. At least one question is an "attention check" for which the correct answer will be obvious.

You will also be asked to provide an upper bound and lower bound for your guess. You should choose these bounds in a way such that you think the answer has a 50% chance of falling between your bounds. The more confident you are, the smaller the difference should be between your upper and lower bound.

The details of the point system used to determine your chance of winning the prize are a bit complicated, but explained below if you are interested. **What is important to know is that the way your earnings are determined ensures that your chances of winning the bonus are maximized by carefully and honestly answering these questions.**

At the end of the study, the points you receive on all choices you make will be averaged, and this will determine the chance (out of 1000) that you win the bonus. For example, if you earn 90 points on average, you will have a 90 out of 1000 chance of winning the bonus.

Your final score, whether you won the prize, and a list of correct answers and sources will be provided at the end of the study.

You will see a Question page on the next screen.

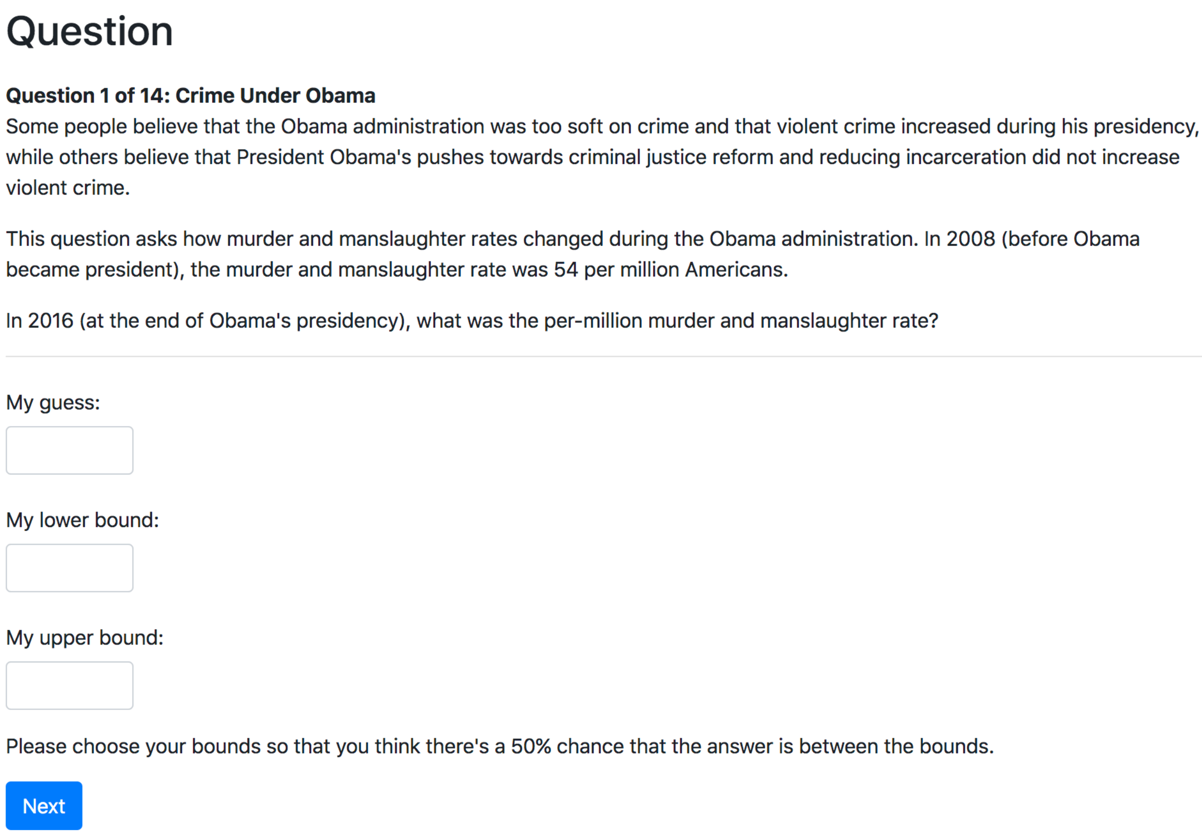
Point system for your guess:
You will receive between 0 and 100 points for each guess you give. The closer your guess is to the correct answer, the more likely it is that you'll win the prize.

If you guess the answer correctly, you will receive 100 points (the maximum) for that question.
If your guess is more than 100 away from the answer, you will receive 0 points for that question.
If your guess is less that 100 away from the answer, you will receive points equal to 100 minus the distance from your guess to the correct answer.

It is in your best interest to guess an answer that is in the "middle" of what you believe is likely. For example, if you think the answer is equally likely to be 10, 40, and 60, you should guess 40.

Point system for your bounds:
*If the answer is **above** your **upper** bound, you will receive points equal to 100 minus 3 times the distance from your guess to the correct answer.*
*If the answer is **below** your **upper** bound, you will receive points equal to 100 minus the distance from your guess to the correct answer.*
*If the answer is **above** your **lower** bound, you will receive points equal to 100 minus the distance from your guess to the correct answer.*
*If the answer is **below** your **lower** bound, you will receive points equal to 100 minus 3 times the distance from your guess to the correct answer.*
You cannot earn negative points. All negative point values will be rounded up to zero.

It is in your best interest to choose a lower bound such that you think it's 3 times more likely to be above the bound than below it, and an upper bound such that it's 3 times more likely to be below the bound than above it. For example, if you think the answer is equally likely to be any number from 100 to 200, you should set a lower bound of 125 and an upper bound of 175.



Question

Question 1 of 14: Crime Under Obama
Some people believe that the Obama administration was too soft on crime and that violent crime increased during his presidency, while others believe that President Obama's pushes towards criminal justice reform and reducing incarceration did not increase violent crime.

This question asks how murder and manslaughter rates changed during the Obama administration. In 2008 (before Obama became president), the murder and manslaughter rate was 54 per million Americans.

In 2016 (at the end of Obama's presidency), what was the per-million murder and manslaughter rate?

My guess:

My lower bound:

My upper bound:

Please choose your bounds so that you think there's a 50% chance that the answer is between the bounds.

Next

Figure A.6: Crime Under Obama question page.

Instructions for News Assessment Pages

After most Question pages, you will see a News Assessment page.

There has been a growing debate about the accuracy of news sources, with both the left and the right accusing each other's primary media of spreading "Fake News." News sources like CNN and Fox News have reported extensively on topics such as crime, global warming, and gun laws; some give factual information, while others may distort the truth or lie outright. This part of the study is testing whether people can recognize Fake News and True News.

On each News Assessment page, you will see the previous Question page and be given a message related to your previous guess from either a True News source or Fake News source. In addition to your guaranteed HIT payment, you will have a chance to win an additional bonus of $10.00 based on your answers to these questions and questions on other pages. The message will say either "The answer is *greater than* your previous guess" or "The answer is **less than** your previous guess."

The True News source will *always* tell you the truth, while the Fake News source will *never* tell the truth.

If the answer truly is greater than your previous guess, True News will tell you "The answer is *greater than* your previous guess" and Fake News will tell you "The answer is *less than* your previous guess."

If the answer truly is less than your previous guess, True News will tell you "The answer is *less than* your previous guess" and Fake News will tell you "The answer is *greater than* your previous guess."

Whether you get your message from True News or Fake News is random; different messages may come from different sources. Seeing Fake News on one page does not affect the chances of seeing Fake News on any other page.

After each question, you will assess whether you think it is more likely that the source is True News or Fake News on a scale of 0/10 to 10/10, and your assessment will determine how many points you will earn for that page.

The details of the point system to determine your chance of winning the prize are a bit complicated, but explained below if you are interested. **What is important to know is that the way your earnings are determined ensures that your chances of winning the bonus are maximized by carefully and honestly answering these questions.**

Your final score, whether you won the prize, and a list of correct answers and sources will be provided at the end of the study.

You will see a News Assessment page on the next screen.

Point system:

Your estimate	***Points earned if the source is True News***	***Points earned if the source is Fake News***
0/10 chance it's True News; 10/10 chance it's Fake News	*0 points*	*100 points*
1/10 chance it's True News; 9/10 chance it's Fake News	*19 points*	*99 points*
2/10 chance it's True News; 8/10 chance it's Fake News	*36 points*	*96 points*
3/10 chance it's True News; 7/10 chance it's Fake News	*51 points*	*91 points*
4/10 chance it's True News; 6/10 chance it's Fake News	*64 points*	*84 points*
5/10 chance it's True News; 5/10 chance it's Fake News	*75 points*	*75 points*
6/10 chance it's True News; 4/10 chance it's Fake News	*84 points*	*64 points*
7/10 chance it's True News; 3/10 chance it's Fake News	*91 points*	*51 points*
8/10 chance it's True News; 2/10 chance it's Fake News	*96 points*	*36 points*
9/10 chance it's True News; 1/10 chance it's Fake News	*99 points*	*19 points*
10/10 chance it's True News; 0/10 chance it's Fake News	*100 points*	*0 points*

For instance, if you estimate a 7/10 chance of True News, then for that round you will earn 91 points if the source is True News and 51 points if the source is Fake News.

At the end of the study, the points you receive on all choices you make will be averaged, and this will determine the chance (out of 1000) that you win the bonus. For example, if you earn 90 points on average, you will have a 90 out of 1000 chance of winning the bonus.

News Assessment

Original question 1 of 14: Crime Under Obama
Some people believe that the Obama administration was too soft on crime and that violent crime increased during his presidency, while others believe that President Obama's pushes towards criminal justice reform and reducing incarceration did not increase violent crime.

This question asks how murder and manslaughter rates changed during the Obama administration. In 2008 (before Obama became president), the murder and manslaughter rate was 54 per million Americans.

In 2016 (at the end of Obama's presidency), what was the per-million murder and manslaughter rate?

Message:
The answer is **less than** your previous guess of **57.0**.

Do you think this information is from True News or Fake News?

- ◯ 0/10 chance it's True News; 10/10 chance it's Fake News
- ◯ 1/10 chance it's True News; 9/10 chance it's Fake News
- ◯ 2/10 chance it's True News; 8/10 chance it's Fake News
- ◯ 3/10 chance it's True News; 7/10 chance it's Fake News
- ◯ 4/10 chance it's True News; 6/10 chance it's Fake News
- ◯ 5/10 chance it's True News; 5/10 chance it's Fake News
- ◯ 6/10 chance it's True News; 4/10 chance it's Fake News
- ◯ 7/10 chance it's True News; 3/10 chance it's Fake News
- ◯ 8/10 chance it's True News; 2/10 chance it's Fake News
- ◯ 9/10 chance it's True News; 1/10 chance it's Fake News
- ◯ 10/10 chance it's True News; 0/10 chance it's Fake News

Figure A.7: Crime Under Obama news assessment page.

News Assessment

Original question 1 of 14: Crime Under Obama
Some people believe that the Obama administration was too soft on crime and that violent crime increased during his presidency, while others believe that President Obama's pushes towards criminal justice reform and reducing incarceration did not increase violent crime.

This question asks how murder and manslaughter rates changed during the Obama administration. In 2008 (before Obama became president), the murder and manslaughter rate was 54 per million Americans.

In 2016 (at the end of Obama's presidency), what was the per-million murder and manslaughter rate?

Message:
The answer is **less than** your previous guess of **57.0**.

Do you think this information is from True News or Fake News?

- 0/10 chance it's True News; 10/10 chance it's Fake News
- 1/10 chance it's True News; 9/10 chance it's Fake News
- 2/10 chance it's True News; 8/10 chance it's Fake News
- 3/10 chance it's True News; 7/10 chance it's Fake News
- 4/10 chance it's True News; 6/10 chance it's Fake News
- 5/10 chance it's True News; 5/10 chance it's Fake News
- 6/10 chance it's True News; 4/10 chance it's Fake News
- 7/10 chance it's True News; 3/10 chance it's Fake News
- 8/10 chance it's True News; 2/10 chance it's Fake News
- 9/10 chance it's True News; 1/10 chance it's Fake News
- 10/10 chance it's True News; 0/10 chance it's Fake News

After seeing this message and assessing its truthfulness, what is your guess of the answer to the original question?

Figure A.8: Crime Under Obama news assessment page: Second Guess question.

News Valuation

On the previous News Assessment pages you were given messages that said that the correct answer was either "greater than" or "less than" your guess, and you were then asked to guess how likely it was that this message came from a True News versus Fake News source.

This section is designed to assess how useful you think those messages are. On this page you will decide whether to see the message or whether to receive additional points and see a screen with a black bar as in the following example:

Original question 12: Gender and Math Grades

In the United States, men are more likely to enter into mathematics and math-related fields. Some people attribute this to gender differences in interest in or ability in math, while others attribute it to other factors like gender discrimination.

This question asks whether high school boys and girls differ substantially in how well they do in math classes. A major testing service analyzed data on high school seniors and compared the average GPA for male and female students in various subjects.

Male students averaged a 3.04 GPA (out of 4.00) in math classes. What GPA did female students average in math classes?

(Please guess between 0.00 and 4.00.)

Message:

The answer is [redacted] your previous guess of **3.1**.

Do you think this information is from True News or Fake News?

- 0/10 chance it's True News; 10/10 chance it's Fake News
- 1/10 chance it's True News; 9/10 chance it's Fake News
- 2/10 chance it's True News; 8/10 chance it's Fake News
- 3/10 chance it's True News; 7/10 chance it's Fake News
- 4/10 chance it's True News; 6/10 chance it's Fake News
- 5/10 chance it's True News; 5/10 chance it's Fake News
- 6/10 chance it's True News; 4/10 chance it's Fake News
- 7/10 chance it's True News; 3/10 chance it's Fake News
- 8/10 chance it's True News; 2/10 chance it's Fake News
- 9/10 chance it's True News; 1/10 chance it's Fake News
- 10/10 chance it's True News; 0/10 chance it's Fake News

To determine whether you receive the message or the black bar, you will write down a "valuation" at the bottom of this page. The more that you think the message helps you, the higher your valuation should be.

(If you would prefer to see the message instead of the black bar, you should submit a valuation between 0 and 25 points, where a larger valuation indicates a stronger preference for the message.)

(If you would prefer to see the black bar instead of the message, you should submit a valuation between -25 and 0 points, where a more negative valuation indicates a stronger preference for the black bar.)

The details of the procedure to determine whether you receive the message or the black bar is a bit complicated, but explained below. **What is important to know is that the way your earnings are determined ensures that your chances of winning the bonus are maximized by honestly answering this question.**

Valuation of message (in points):

Point and message procedure given your valuation:

A computer will randomly select a number between -25 and 25 with all numbers being equally likely.

If this number is greater than your valuation in points, this number will be added to the points you earn on the next News Assessment page, but you will receive the black bar instead of the message (as above).

If this number is less than your valuation in points, you will earn the standard amount of points on the next News Assessment page, and you will receive either the "greater than" or the "less than" message (as in previous pages).

News Assessment

Original question 12 of 14: Gender and Math Grades
In the United States, men are more likely to enter into mathematics and math-related fields. Some people attribute this to gender differences in interest in or ability in math, while others attribute it to other factors like gender discrimination.

This question asks whether high school boys and girls differ substantially in how well they do in math classes. A major testing service analyzed data on high school seniors and compared the average GPA for male and female students in various subjects.

Male students averaged a 3.04 GPA (out of 4.00) in math classes. What GPA did female students average in math classes?

(Please guess between 0.00 and 4.00.)

Message:
The answer is [redacted] your previous guess of **3.1**.

Do you think this information is from True News or Fake News?

- 0/10 chance it's True News; 10/10 chance it's Fake News
- 1/10 chance it's True News; 9/10 chance it's Fake News
- 2/10 chance it's True News; 8/10 chance it's Fake News
- 3/10 chance it's True News; 7/10 chance it's Fake News
- 4/10 chance it's True News; 6/10 chance it's Fake News
- 5/10 chance it's True News; 5/10 chance it's Fake News
- 6/10 chance it's True News; 4/10 chance it's Fake News
- 7/10 chance it's True News; 3/10 chance it's Fake News
- 8/10 chance it's True News; 2/10 chance it's Fake News
- 9/10 chance it's True News; 1/10 chance it's Fake News
- 10/10 chance it's True News; 0/10 chance it's Fake News

Next

Feedback

This is the last page before you see your results. Your feedback is not required, but it is very helpful for designing future studies.

Your earnings and bonus are not affected by your answers to these questions.

What did you think of the study? What did you like or dislike about it?

What was your thought process when you received a message and were deciding whether it was True News or Fake News?

Next

Results: Click the Finish button at the bottom of this page to complete the HIT

Sorry, you did not win the bonus. Your additional bonus was $0.00.

You earned 80.32 points on average across all questions in this study. For questions, solutions, points and whether information was from True News or Fake News, see the tables below.

Question	Correct answer	Your initial guess	Message said	Was the News Real or Fake	Your likelihood of this being True News	Points you earned for your likelihood
In a study, researchers sent fictitious resumes to respond to thousands of help-wanted ads in newspapers. The resumes sent had identical skills and education, but the researchers gave half of the (fake) applicants stereotypically White names such as Emily Walsh and Greg Baker, and gave the other half of the applicants stereotypically Black names such as Lakisha Washington and Jamal Jones. 9.65 percent of the applicants with White-sounding names received a call back. What percent of the applicants with Black-sounding names received a call back? (Please guess between 0 and 100.)	6.45	3.0	The answer is less than your previous guess.	Fake News	9/10	19.0
What is the year right now? This is not a trick question and the first sentence is irrelevant; this is a comprehension check to make sure you are paying attention. For this question, your lower and upper bounds should be equal to your guess if you know what year it currently is.	2018.0	2018.0	The answer is equal to your previous guess.	True News	10/10	100.0
How many degrees West is this geographic center? (Please guess between 0 and 180. The continental U.S. lies in the Western Hemisphere, which ranges from 0 degrees West to 180 degrees West.)	98.583	90.0	The answer is less than your previous guess.	Fake News	5/10	75.0

Question	Correct answer	Your guess	Your lower bound	Your upper bound	Points you earned for your guess and bounds	Source
In a study, researchers sent fictitious resumes to respond to thousands of help-wanted ads in newspapers. The resumes sent had identical skills and education, but the researchers gave half of the (fake) applicants stereotypically White names such as Emily Walsh and Greg Baker, and gave the other half of the applicants stereotypically Black names such as Lakisha Washington and Jamal Jones. 9.65 percent of the applicants with White-sounding names received a call back. What percent of the applicants with Black-sounding names received a call back? (Please guess between 0 and 100.)	6.45	3.0	3.0	3.0	94.25	http://bit.ly/labor-market-discrimination
What is the year right now? This is not a trick question and the first sentence is irrelevant; this is a comprehension check to make sure you are paying attention. For this question, your lower and upper bounds should be equal to your guess if you know what year it currently is.	2018.0	2018.0	2018.0	2018.0	100.0	http://bit.ly/what-year-is-it
How many degrees West is this geographic center? (Please guess between 0 and 180. The continental U.S. lies in the Western Hemisphere, which ranges from 0 degrees West to 180 degrees West.)	98.583	90.0	80.0	100.0	90.47	http://bit.ly/center-of-the-us

B

Appendix for Chapter 2

B.1 Additional Tables

B.1.1 Balance Table

	Negative News	Positive News	Neg vs. Pos	p-value
Democrat	0.479	0.468	0.011	0.567
	(0.014)	(0.014)	(0.020)	
Republican	0.195	0.193	0.002	0.911
	(0.011)	(0.011)	(0.016)	
Male	0.550	0.532	0.018	0.353
	(0.014)	(0.014)	(0.020)	
Female	0.442	0.458	-0.016	0.406
	(0.014)	(0.014)	(0.020)	
Age	35.353	35.732	-0.379	0.357
	(0.291)	(0.290)	(0.411)	
Education	14.919	14.873	0.046	0.538
	(0.052)	(0.054)	(0.075)	
Log(Income)	10.833	10.853	-0.021	0.520
	(0.023)	(0.022)	(0.032)	
White	0.745	0.730	0.015	0.374
	(0.012)	(0.012)	(0.017)	
Black	0.090	0.088	0.002	0.881
	(0.008)	(0.008)	(0.011)	
Hispanic	0.046	0.057	-0.012	0.180
	(0.006)	(0.006)	(0.009)	
Asian	0.087	0.090	-0.003	0.789
	(0.008)	(0.008)	(0.011)	
Religious	0.470	0.479	-0.009	0.649
	(0.014)	(0.014)	(0.020)	
N	1248	1306	2554	

Notes: Standard errors in parentheses. Positive and Negative News are defined in text. Education in years. Religious is 1 if subject affiliates with any religious group.

B.2 Study Materials: Exact Question Wordings

Infant Mortality

The CDC provides statistics for mortality rates for infants. In 1997, there were 28.0 thousand infant deaths in the United States.

How many thousands of infant deaths in the United States were there in 2017 (the most recent year available)?

(If you answer X, it means you think that there were X thousand deaths.)

Correct answer: 22.3.

Source linked on results page: https://www.cdc.gov/nchs/data/nvsr/nvsr68/nvsr68_10-508.pdf

Others' Happiness

Many surveys ask the following question about subjective happiness:

"Please imagine a ladder with steps numbered from zero at the bottom to ten at the top. Suppose we say that the top of the ladder represents the best possible life for you and the bottom of the ladder represents the worst possible life for you. If the top step is 10 and the bottom step is 0, on which step of the ladder do you feel you personally stand at the present time?"

In 2006, the average subjective happiness level in the United States was 7.18 out of 10.

What was the average subjective happiness level in the US in 2018?

Correct answer: 6.88.

Source linked on results page: https://ourworldindata.org/happiness-and-life-satisfaction

Cancer in Children

Acute Myeloid Leukemia (AML) is a devastating illness in which cancerous cells emerge in the bone marrow, invade the blood stream, and may spread to the rest of the body. Tragically, hundreds to thousands of children under the age of 15 are diagnosed with AML each year; it is one of the most common cancers among children.

Of children under the age of 15 who are diagnosed with AML, what percent survive for at least 5 years?

(Please guess between 0 and 100.)

Correct answer: 68.8.

Source linked on results page: `https://www.lls.org/facts-and-statistics/overview/childhood-blood-cancer-facts-and-statistics`

Global Poverty

Around the world, many people do not have enough money for basic necessities. The World Bank defines extreme poverty as having less than the equivalent of $1.90 per day.

In 1990, the World Bank estimated that 1897 million people around the world were living in extreme poverty.

As of 2015 (the most recent year available), how many milllions of people around the world live in extreme poverty?

(If you answer X, it means you think that X million people live in extreme poverty.)

Correct answer: 731.

Source linked on results page: `http://povertydata.worldbank.org/poverty/home/`

Armed Conflict

The Department of Peace and Conflict Research estimates that 45.8 thousand people were killed per year in battles in the fifteen years from 1989-2003.

How many thousands of people were killed per year in battles in the fifteen years from 2004-2018?

(If you answer X, it means you think that X thousand people were killed per year.)

Correct answer: 48.12

Source linked on results page: https://www.pcr.uu.se/digitalAssets/667/c_667494-l_1-k_battle-related-deaths-by-region--1989-2018.pdf

Latitude of Center of the United States

The U.S. National Geodetic Survey approximated the geographic center of the continental United States. (This excludes Alaska and Hawaii, and U.S. territories.)

How many degrees North is this geographic center?

(Please guess between 0 and 90. The continental U.S. lies in the Northern Hemisphere, the Equator is 0 degrees North, and the North Pole is 90 degrees North.)

Correct answer: 39.833.

Source linked on results page: http://bit.ly/center-of-the-us

Comprehension Check: Current Year

In 1776 our fathers brought forth, upon this continent, a new nation, conceived in Liberty, and dedicated to the proposition that all men are created equal.

What is the year right now?

This is not a trick question and the first sentence is irrelevant; this is a comprehension check to make sure you are paying attention. For this question, your lower and upper bounds should be equal to your guess if you know what year it currently is.

Correct answer: 2019.

Source linked on results page: http://bit.ly/what-year-is-it

B.3 Study Materials: Pages in Experiment

Introduction

You are invited to participate in this online study on US and world attitudes. This is a research project being conducted by Michael Thaler, a PhD student in economics at Harvard University.

Your participation in this survey is entirely voluntary. You may refuse to take part in the research or exit the survey at any time without penalty.

If you choose to be in the study, you will complete a series of questions related to issues affecting the world today. The study should take approximately 15 minutes to complete, but you may take up to 60 minutes. You will have a chance to earn a bonus of up to $10.00 in addition to your participation earnings.

Your specific answers will not be shared with anyone, and for the purpose of privacy please do not include your name or other personally identifiable information in your responses. Please make sure to mark your Amazon Profile as private if you do not want it to be accessible via your Mechanical Turk Worker ID.

If you have any questions or concerns, please contact Michael Thaler at michaelthaler@g.harvard.edu.

You may print or save a copy of this information sheet for your own records. **Please do not press the back button, refresh, or leave the page at any time or else you might have a server error; if this happens, you will not be able to reenter the study or earn your payment.**

If you wish to participate in the study, please indicate below that you have read the instructions and enter your Mechanical Turk Worker ID for payment.

What is your MTurk Worker ID number? This is required for payment.

○ I have read the above information and would like to participate in the study.

Next

Demographic Information

It is important for this study that you answer these questions honestly.

Your earnings and bonus are not affected by your answers to these questions.

What is your age?

What is your gender?

- Male
- Female
- Other / Prefer not to answer

What is your race/ethnicity?

- Black or African American
- White
- Asian
- American Indian
- Hispanic or Latino
- Two or more of these
- Other / Prefer not to answer

What is the highest level of education you have completed?

- Did not graduate high school
- High school graduate or GED
- Began college, no degree
- Associate's degree
- Bachelor's degree
- Postgraduate or professional degree

What religious group do you consider yourself affiliated with?

- Mainline Protestant
- Historically black Protestant
- Evangelical Protestant
- Catholic
- Other Christian
- Jewish
- Muslim
- Other religion or faith
- Atheist
- Agnostic
- Unaffiliated

Which US state or territory do you currently live in?

What was your total household income before taxes during the past 12 months?

- Less than $20,000
- $20,000 to $29,999
- $30,000 to $39,999
- $40,000 to $49,999
- $50,000 to $69,999
- $70,000 to $99,999
- $100,000 to $149,999
- $150,000 or more

In politics today, do you consider yourself a Republican, a Democrat, or an Independent?

- Republican
- Democrat
- Independent

Where do you see yourself on the liberal/conservative spectrum?

- Extremely liberal
- Liberal
- Slightly liberal
- Moderate
- Slightly conservative
- Conservative
- Extremely conservative

Please rate how you feel about the Republican Party using a scale of 0 to 100. The higher the number, the more favorable you feel toward the Republican Party.

50

Please rate how you feel about the Democratic Party using a scale of 0 to 100. The higher the number, the more favorable you feel toward the Democratic Party.

50

Next

Instructions for Question Pages

Throughout this study, you will see several types of pages, including 7 Question pages.

On each of the Question pages, you will be asked to guess the answer to a factual question; each question has a correct numerical answer. In addition to your guaranteed HIT payment, you will have a chance to win an additional bonus of $10.00 based on your guesses to these questions and questions on other pages. At least one question is an "attention check" for which the correct answer will be obvious.

You will also be asked to provide an upper bound and lower bound for your guess. You should choose these bounds in a way such that you think the answer has a 50% chance of falling between your bounds. The more confident you are, the smaller the difference should be between your upper and lower bound.

The details of the point system used to determine your chance of winning the prize are a bit complicated, but explained below if you are interested. **What is important to know is that the way your earnings are determined ensures that your chances of winning the bonus are maximized by carefully and honestly answering these questions.**

At the end of the study, the points you receive on all choices you make will be averaged, and this will determine the chance (out of 1000) that you win the bonus. For example, if you earn 90 points on average, you will have a 90 out of 1000 chance of winning the bonus.

Your final score, whether you won the prize, and a list of correct answers and sources will be provided at the end of the study.

You will see a Question page on the next screen.

Next

Point system for your guess:
You will receive between 0 and 100 points for each guess you give. The closer your guess is to the correct answer, the more likely it is that you'll win the prize.

If you guess the answer correctly, you will receive 100 points (the maximum) for that question.
If your guess is more than 100 away from the answer, you will receive 0 points for that question.
If your guess is less that 100 away from the answer, you will receive points equal to 100 minus the distance from your guess to the correct answer.

It is in your best interest to guess an answer that is in the "middle" of what you believe is likely. For example, if you think the answer is equally likely to be 10, 40, and 60, you should guess 40.

Point system for your bounds:
*If the answer is **above** your **upper** bound, you will receive points equal to 100 minus 3 times the distance from your guess to the correct answer.*
*If the answer is **below** your **upper** bound, you will receive points equal to 100 minus the distance from your guess to the correct answer.*
*If the answer is **above** your **lower** bound, you will receive points equal to 100 minus the distance from your guess to the correct answer.*
*If the answer is **below** your **lower** bound, you will receive points equal to 100 minus 3 times the distance from your guess to the correct answer.*
You cannot earn negative points. All negative point values will be rounded up to zero.

It is in your best interest to choose a lower bound such that you think it's 3 times more likely to be above the bound than below it, and an upper bound such that it's 3 times more likely to be below the bound than above it. For example, if you think the answer is equally likely to be any number from 100 to 200, you should set a lower bound of 125 and an upper bound of 175.

Question

Question 1 of 7: Infant Mortality
The CDC provides statistics for mortality rates for infants.

In 1997, there were 28.0 thousand infant deaths in the United States.

How many thousands of infant deaths in the United States were there in 2017 (the most recent year available?

(If you answer X, it means you think that there were X thousand deaths.)

My guess:

My lower bound:

My upper bound:

Please choose your bounds so that you think there's a 50% chance that the answer is between the bounds.

Next

Instructions for News Assessment Pages

After most Question pages, you will see a News Assessment page.

There has been a growing debate about the accuracy of news sources, with many people accusing various media of spreading "Fake News." News sources have reported extensively on topics such as health, conflict, and poverty; some give factual information, while others may distort the truth or lie outright. This part of the study is testing whether people can recognize Fake News and True News.

On each News Assessment page, you will see the previous Question page and be given a message related to your previous guess from either a True News source or Fake News source. In addition to your guaranteed HIT payment, you will have a chance to win an additional bonus of $10.00 based on your answers to these questions and questions on other pages. The message will say either "The answer is *greater than* your previous guess" or "The answer is **less than** your previous guess."

The True News source will *always* tell you the truth, while the Fake News source will *never* tell the truth.

If the answer truly is greater than your previous guess, True News will tell you "The answer is *greater than* your previous guess" and Fake News will tell you "The answer is *less than* your previous guess."

If the answer truly is less than your previous guess, True News will tell you "The answer is *less than* your previous guess" and Fake News will tell you "The answer is *greater than* your previous guess."

Whether you get your message from True News or Fake News is random *and each source is equally likely*; different messages may come from different sources. Seeing Fake News on one page does not affect the chances of seeing Fake News on any other page.

After each question, you will assess whether you think it is more likely that the source is True News or Fake News on a scale of 0/10 to 10/10, and your assessment will determine how many points you will earn for that page.

The details of the point system to determine your chance of winning the prize are a bit complicated, but explained below if you are interested. **What is important to know is that the way your earnings are determined ensures that your chances of winning the bonus are maximized by carefully and honestly answering these questions.**

You will see a News Assessment page on the next screen.

Next

Point system:

Your estimate	***Points earned if the source is True News***	***Points earned if the source is Fake News***
0/10 chance it's True News; 10/10 chance it's Fake News	*0 points*	*100 points*
1/10 chance it's True News; 9/10 chance it's Fake News	*19 points*	*99 points*
2/10 chance it's True News; 8/10 chance it's Fake News	*36 points*	*96 points*
3/10 chance it's True News; 7/10 chance it's Fake News	*51 points*	*91 points*
4/10 chance it's True News; 6/10 chance it's Fake News	*64 points*	*84 points*
5/10 chance it's True News; 5/10 chance it's Fake News	*75 points*	*75 points*
6/10 chance it's True News; 4/10 chance it's Fake News	*84 points*	*64 points*
7/10 chance it's True News; 3/10 chance it's Fake News	*91 points*	*51 points*
8/10 chance it's True News; 2/10 chance it's Fake News	*96 points*	*36 points*
9/10 chance it's True News; 1/10 chance it's Fake News	*99 points*	*19 points*
10/10 chance it's True News; 0/10 chance it's Fake News	*100 points*	*0 points*

For instance, if you estimate a 7/10 chance of True News, then for that round you will earn 91 points if the source is True News and 51 points if the source is Fake News.

At the end of the study, the points you receive on all choices you make will be averaged, and this will determine the chance (out of 1000) that you win the bonus. For example, if you earn 90 points on average, you will have a 90 out of 1000 chance of winning the bonus.

News Assessment

Original question 1 of 7: Infant Mortality
The CDC provides statistics for mortality rates for infants.

In 1997, there were 28.0 thousand infant deaths in the United States.

How many thousands of infant deaths in the United States were there in 2017 (the most recent year available?

(If you answer X, it means you think that there were X thousand deaths.)

Message:
The answer is **less than** your previous guess of **25.0**.

Do you think this information is from True News or Fake News?

- 0/10 chance it's True News; 10/10 chance it's Fake News
- 1/10 chance it's True News; 9/10 chance it's Fake News
- 2/10 chance it's True News; 8/10 chance it's Fake News
- 3/10 chance it's True News; 7/10 chance it's Fake News
- 4/10 chance it's True News; 6/10 chance it's Fake News
- 5/10 chance it's True News; 5/10 chance it's Fake News
- 6/10 chance it's True News; 4/10 chance it's Fake News
- 7/10 chance it's True News; 3/10 chance it's Fake News
- 8/10 chance it's True News; 2/10 chance it's Fake News
- 9/10 chance it's True News; 1/10 chance it's Fake News
- 10/10 chance it's True News; 0/10 chance it's Fake News

After seeing this message and assessing its truthfulness, what is your guess of the answer to the original question?

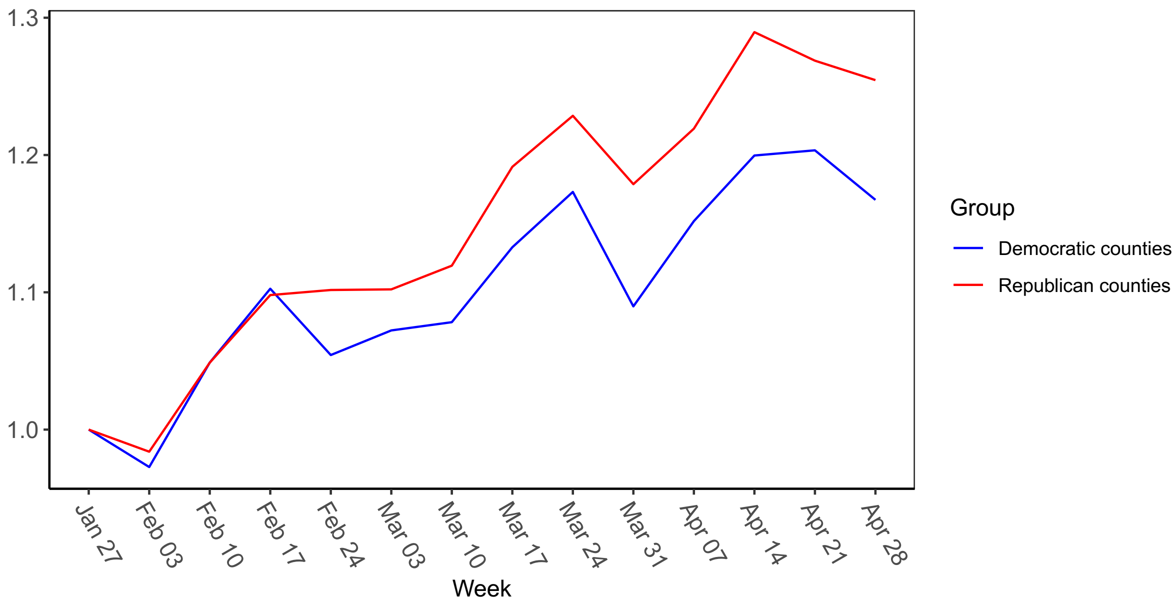
Next

C

Appendix for Chapter 3

C.1 Additional Figures

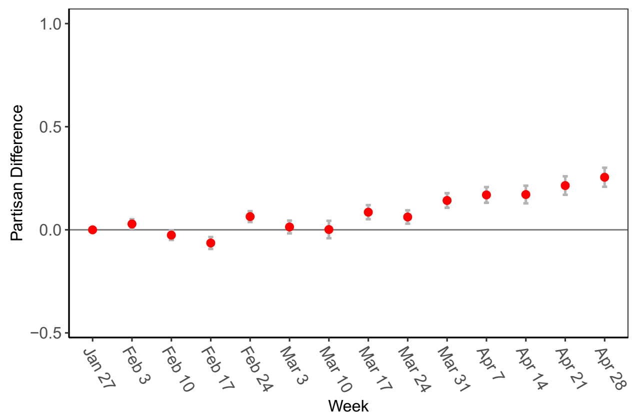
Figure C.1: POI Visits in 2019



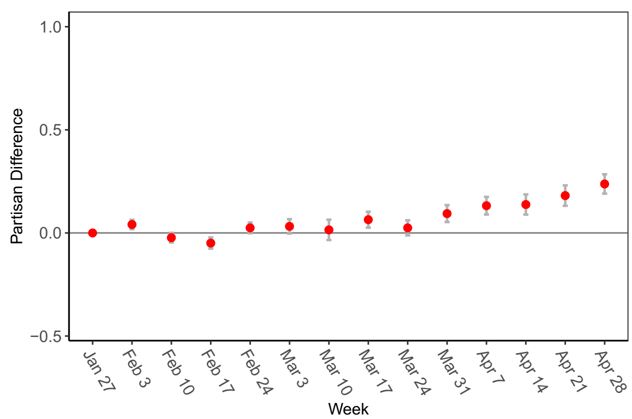
Note: Figure shows the aggregate number of POI visits (normalized to one) for ten weeks starting on January 27, 2019 for Republican counties and Democratic counties. Republican counties are defined to be those whose 2016 Republican vote share is greater than the median vote share across the counties in our sample.

Figure C.2: Partisan Differences in Social Distancing, 2019

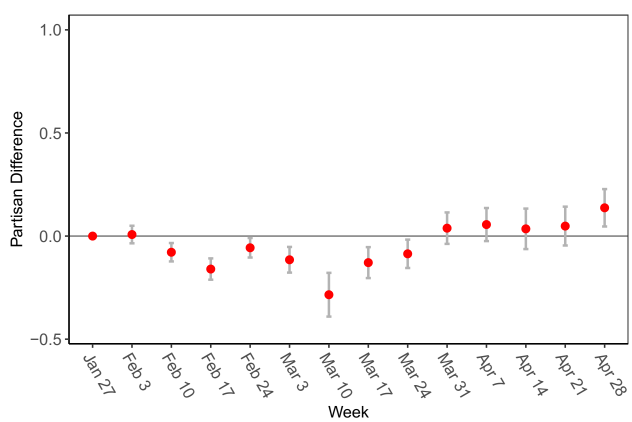
(a) Panel A: Only County & Time FE



(b) Panel B: Adds State-Time FE



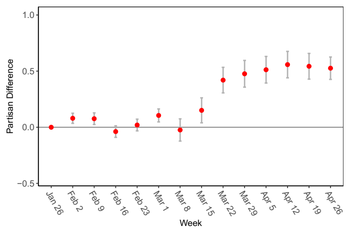
(c) Panel C: Adds Health + Econ Controls



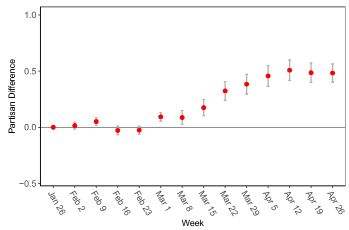
Note: Figure shows the estimated coefficients for county partisanship ρ_i on the log number of POI visits in the county as in Figure 3.4, except that ten weeks of data from January 27, 2019 are used instead of January 26, 2020. For Panel A, only county and time fixed effects are included as controls. Panel B is the same as Panel A except state-time fixed effects replace the time fixed effects. Panel C is the same as Panel B except the health and economic covariates are included. The grey error bars indicate 95 percent confidence intervals constructed using standard errors clustered at the county-level.

Figure C.3: Partisan Differences in Social Distancing, Robustness

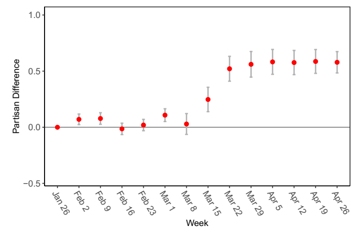
Panel A: Dropping Controls



(a) Drops COVID-19 Controls

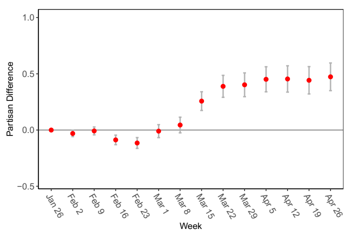


(b) Drops Economic Controls

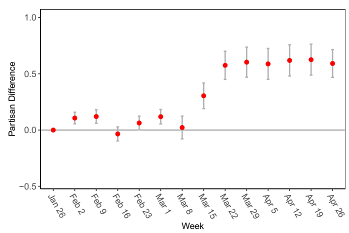


(c) Drops All Health Controls

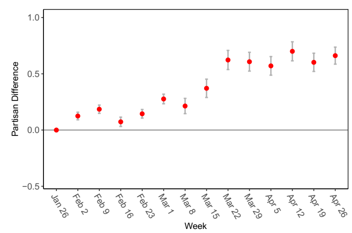
Panel B: Additional Specifications



(d) Linear Controls

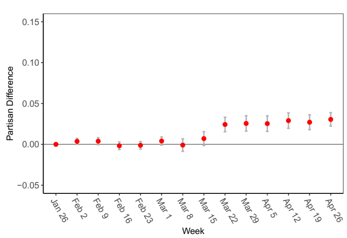


(e) Adds Hispanic and Income

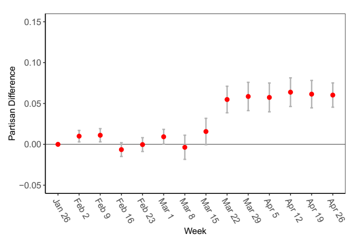


(f) Drops State-Time FE

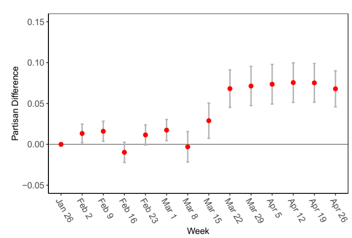
Panel C: Partisanship Indicators



(g) Above or Below Median



(h) Top or Bottom Quartile



(i) Top or Bottom Decile

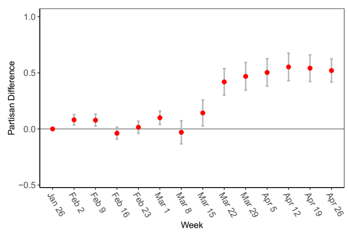
Note: Figure shows the estimated coefficients for county partisanship ρ_i on the log number of POI visits in the county. The specifications are analogous to our baseline in Panel C of Figure 3.4 except with the following deviations.

- Panel A: The first plot drops the COVID-19 cases and deaths controls; the second plot drops the economic controls; and the third plot drops all of the health controls, including the COVID-19 ones.
- Panel B: The first plot does not allow the coefficients on the controls to vary over time and interacts time-invariant controls with a linear time trend; the second plot adds the share Hispanic and the share with income less than 60k with time-varying coefficients; and the third plot drops the state-time fixed effects.
- Panel C: The first plot defines partisanship ρ_i to be 1 if Trump's vote share is greater than the median and -1 otherwise; the second plot defines partisanship ρ_i to be 1 if Trump's vote share is in the top quartile, -1 if in the bottom quartile, and 0 otherwise; and the third plot defines partisanship ρ_i to be 1 if Trump's vote share is in the top decile, -1 if in the bottom decile, and 0 otherwise.

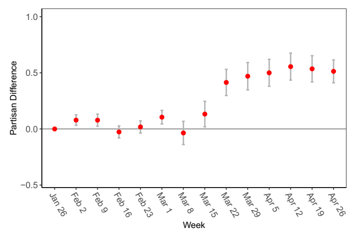
Figure C.3: (continued) Partisan Differences in Social Distancing, Robustness.

Panel D: Sample Restrictions and First Differences

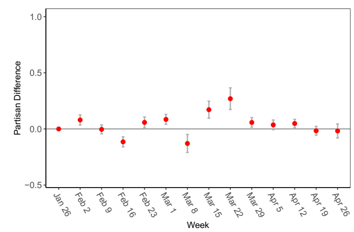
(a) Population below 500,000



(b) Drop CA, WA, and NY

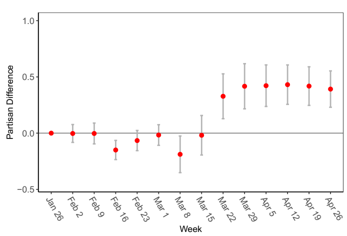


(c) First Differences

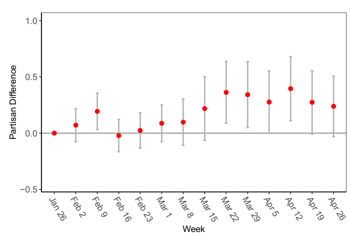


Panel E: Sample Restrictions by Vote Shares

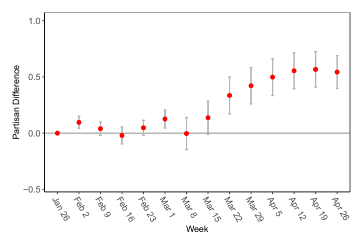
(d) Drop Bottom and Top Deciles



(e) Above Median Vote Share

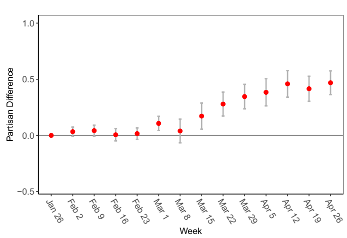


(f) Below Median Vote Share

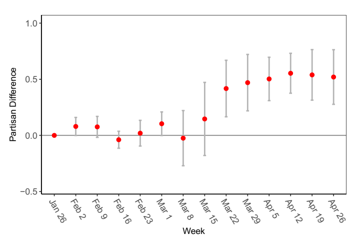


Panel F: Weighting, State Clustering, and Alternative Start Date

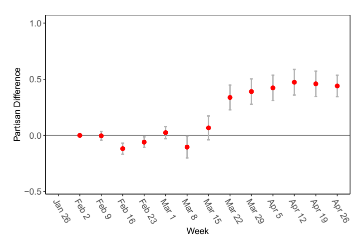
(g) Weight by Population



(h) Cluster SEs by State



(i) Drop Week of January 26



Note: Figure shows the estimated coefficients for county partisanship ρ_i on the log number of POI visits in the county. The specifications are analogous to our baseline in Panel C of Figure 3.4 except with the following deviations.

- Panel D: The first plot only keeps counties with a population below 500,000; the second plot drops California, Washington, and New York; and the third plot shows the estimated coefficients for county partisanship ρ_i on the *change* in the log number of POI visits in the county while dropping county fixed effects.

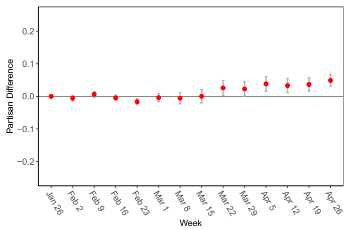
- Panel E: The first plot drops counties for which Trump's vote share was in the bottom or top decile; the second plot keeps counties for which Trump's vote share is greater than the median; and the third plot keeps counties for which Trump's vote share is less than or equal to the median.

- Panel F: The first plot weights observations by the county's population. The second plot clusters standard errors at the state-level. The third plot drops the week of January 26 and normalizes the estimates relative to the week of February 2.

Figure C.3: (continued) Partisan Differences in Social Distancing, Robustness.

Panel G: Alternative Measures

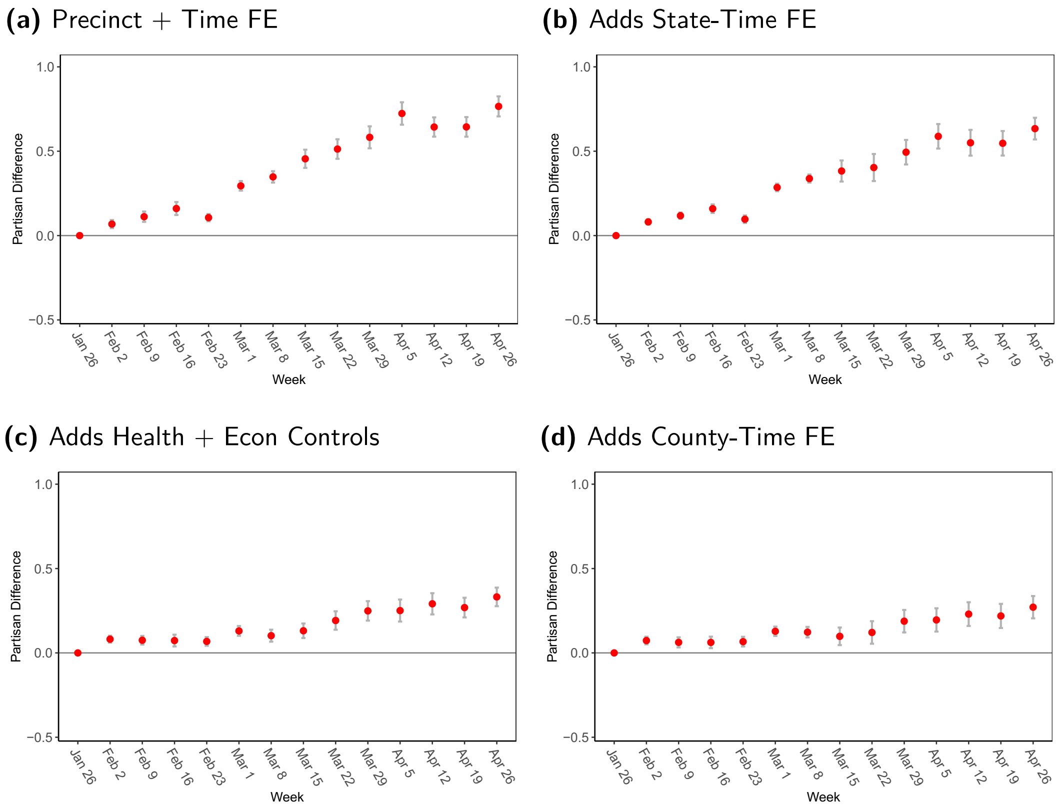
(a) Share Devices Leaving Home, No Correction



Note: Figure shows the estimated coefficients for county partisanship ρ_i on the log number of POI visits in the county. The specifications are analogous to our baseline in Panel C of Figure 3.4 except with the following deviations.

- Panel G: The first plot is analogous to 'Share Devices Leaving Home' in Figure 3.7, except that it does not account for differential sample attrition. Specifically, the outcome is defined to be number of devices observed leaving home divided by the share of devices in the panel for the same period.

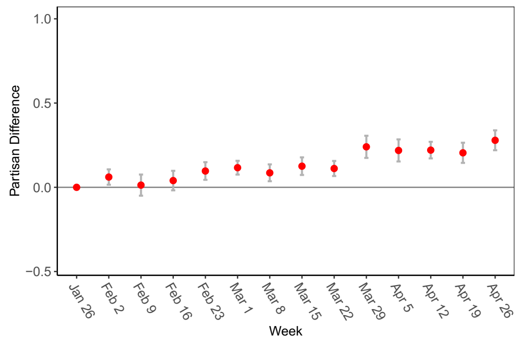
Figure C.4: Partisan Differences in Social Distancing, Precinct



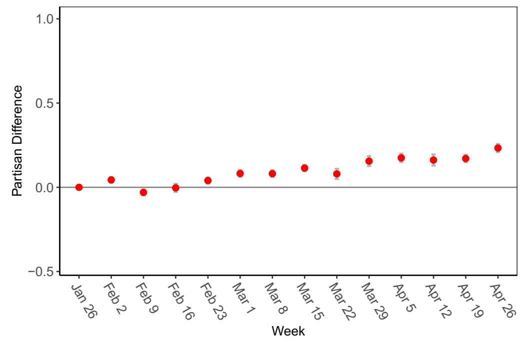
Note: Figure shows the estimated coefficients for precinct partisanship ρ_i on the log number of POI visits in the precinct using the specification outlined in the main text. For Panel A, only precinct and time fixed effects are included as controls. Panel B is the same as Panel A except state-time fixed effects replace the time fixed effects. Panel C is the same as Panel B except the health and economic covariates are included. Panel D is the same as panel C excpet that county-time fixed effects replace the state-time fixed effects, the county-level COVID-19 controls are also dropped in this specification. The grey error bars indicate 95 percent confidence intervals constructed using standard errors clustered at the county-level.

Figure C.5: Partisan Differences in Social Distancing, Precinct 2019

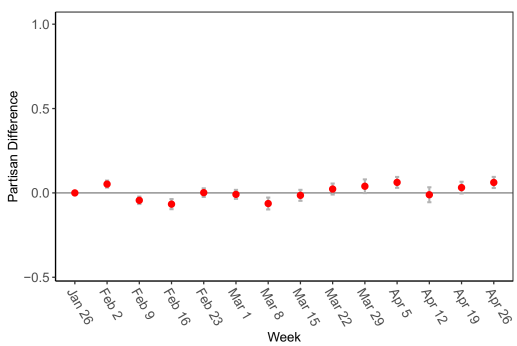
(a) Precinct + Time FE



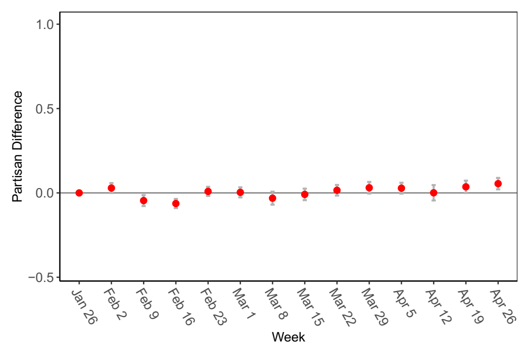
(b) Adds State-Time FE



(c) Adds Health + Econ Controls

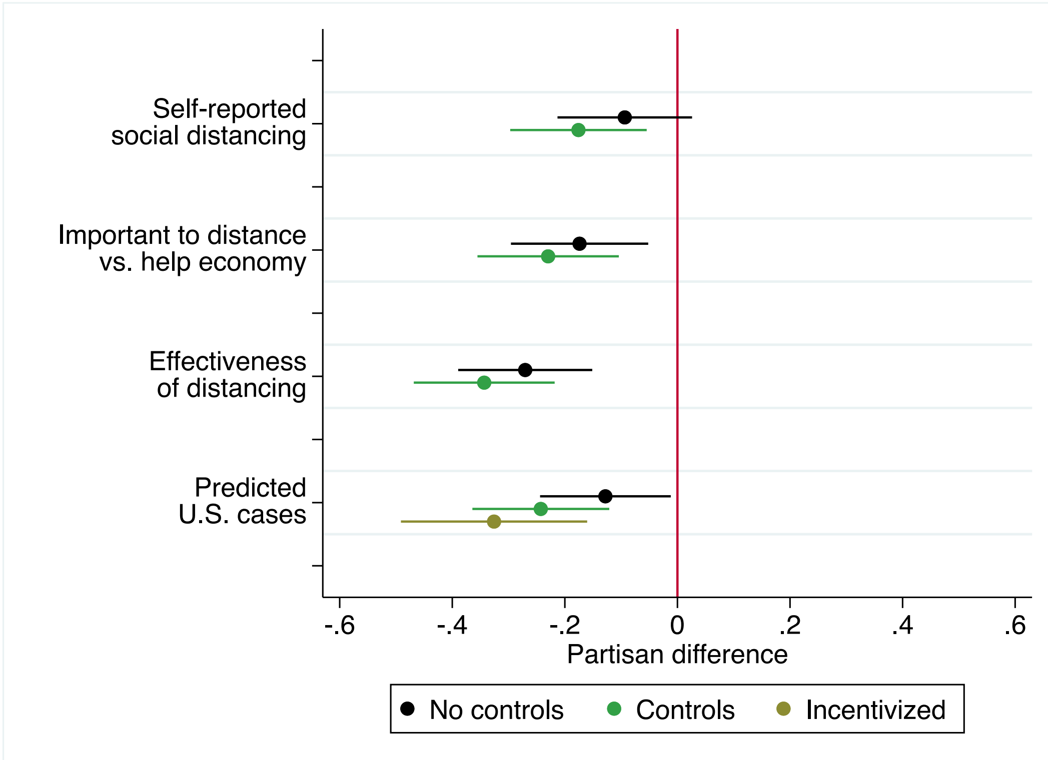


(d) Adds County-Time FE



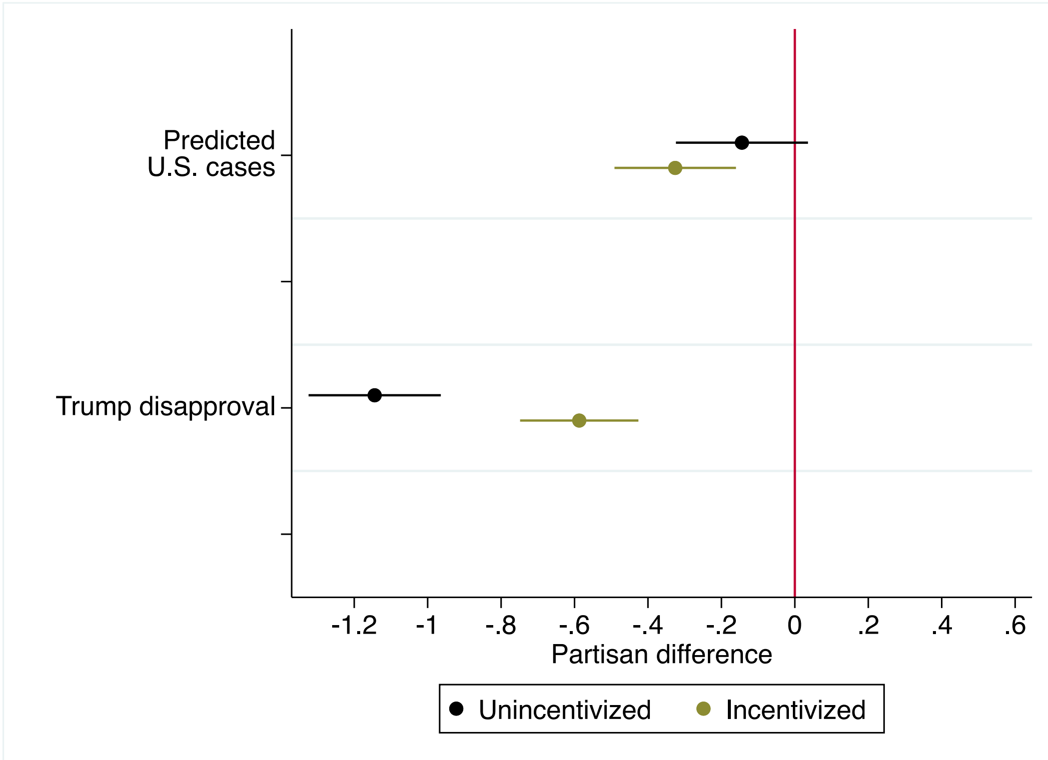
Note: Figure shows the estimated coefficients for precinct partisanship ρ_i on the log number of POI visits in the precinct. The figure mirrors Appendix Figure C.4, except that ten weeks of data from January 27, 2019 are used instead of January 26, 2020. For Panel A, only precinct and time fixed effects are included as controls. Panel B is the same as Panel A except state-time fixed effects replace the time fixed effects. Panel C is the same as Panel B except the health and economic covariates are included. Panel D is the same as panel C excpet that county-time fixed effects replace the state-time fixed effects, the county-level COVID-19 controls are also dropped in this specification. The grey error bars indicate 95 percent confidence intervals constructed using standard errors clustered at the county-level.

Figure C.6: Partisan Differences in Beliefs and Actions: Unweighted



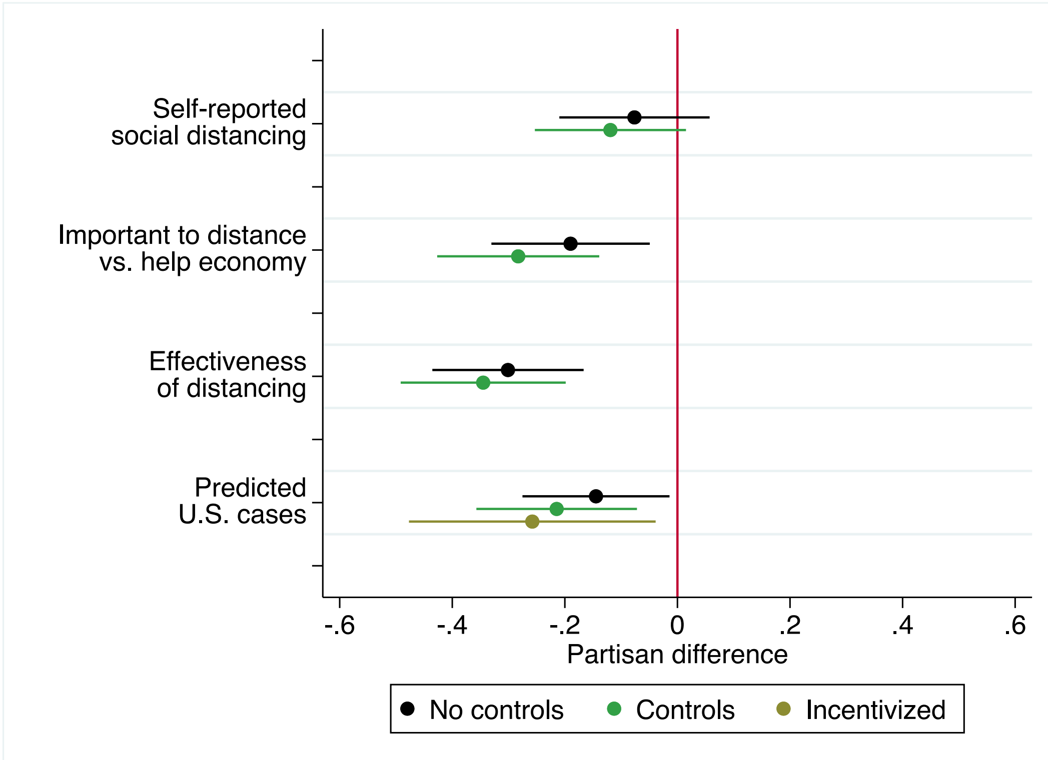
Note: This figure shows coefficient plots of regressing normalized measures of beliefs and actions on party, without weighting observations. Positive values indicate less concern about COVID-19 or social distancing. Demographic controls are age, race, income, education, number of children, ZIP code logged population density, state, county-level deaths and cases. 2 percent of observations are set to the mean due to an invalid ZIP code. Incentivized includes controls and restricts sample to subjects given accuracy incentives. Predicted U.S. cases are predictions about the number of new COVID-19 cases in the U.S. in April; self-reported social distancing is the percent reduction in contact with others over one month; effectiveness of distancing is the estimated likelihood of catching COVID-19 in one month without social distancing; importance of distancing vs. economy is subjects' perception of whether it is more important to go out and stimulate the economy versus staying in and preventing the spread of COVID-19. Error bars represent 95 percent confidence intervals.

Figure C.7: Effect of Incentives on Beliefs



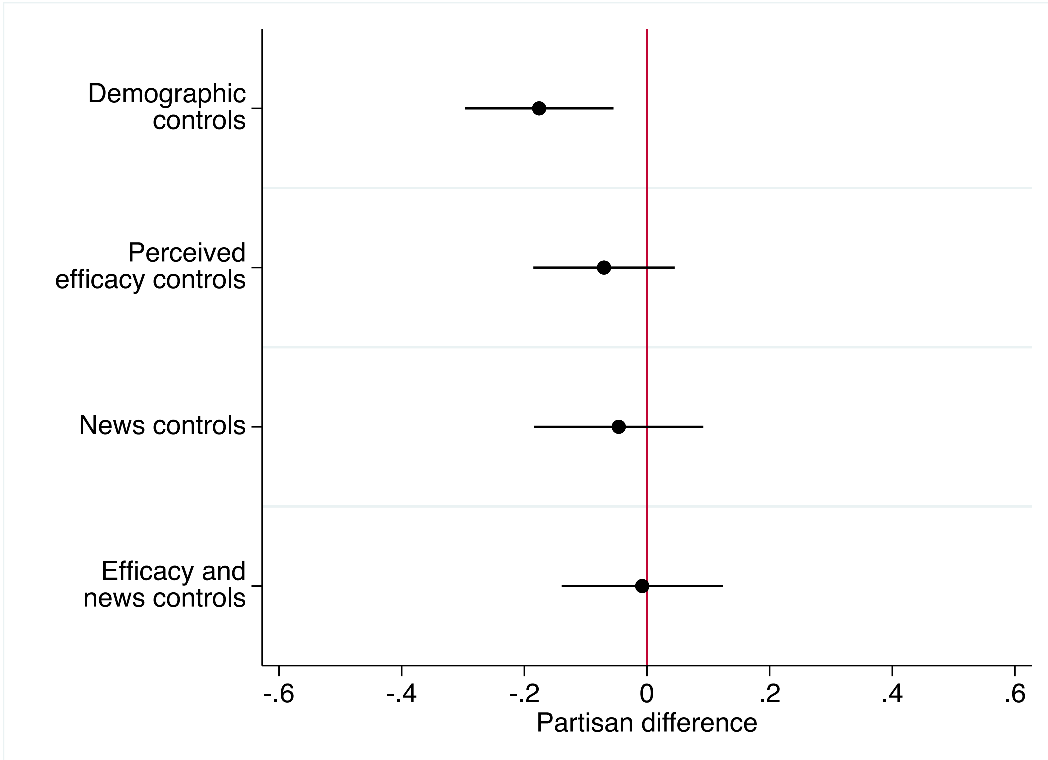
Note: This plot shows coefficient plots of regressing beliefs on party, with and without incentives for getting close to the correct answer. Trump disapproval is a low-stakes question that is susceptible to partisan cheerleading (Bullock et al. 2015; Prior et al. 2015). These results show that predicting COVID-19 cases does not appear susceptible to the same behavior. Observations weighted to mimic a representative sample as described in the text. Error bars represent 95 percent confidence intervals.

Figure C.8: Partisan Differences in Beliefs and Actions: County Fixed Effects



Note: This figure shows coefficient plots of regressing normalized measures of beliefs and actions on party. Positive values indicate less concern about COVID-19 or social distancing. Demographic controls are age, race, income, education, number of children, ZIP code logged population density, county. 21.5 percent of observations are dropped due to an invalid ZIP code or unique county. Observations weighted to mimic a representative sample as described in the text. Incentivized includes controls and restricts sample to subjects given accuracy incentives. Predicted U.S. cases are predictions about the number of new COVID-19 cases in the U.S. in April; self-reported social distancing is the percent reduction in contact with others over one month; effectiveness of distancing is the estimated likelihood of catching COVID-19 in one month without social distancing; importance of distancing vs. economy is subjects' perception of whether it is more important to go out and stimulate the economy versus staying in and preventing the spread of COVID-19. Error bars represent 95 percent confidence intervals.

Figure C.9: Partisan Differences in Social Distancing with Controls for Beliefs and News



Note: This plot shows coefficient plots of regressing self-reported social distancing on party, with and without controls for beliefs and news. Self-reported social distancing is the percent reduction in contact with others over one month. The first row includes only demographic control variables; the second additionally controls for subjects' normalized beliefs about the estimated likelihood of catching COVID-19 in one month without social distancing; the third controls for the partisanship in subjects' news habits; and the fourth controls for both. News habits average the partisanship of news consumption about politics and current events, news consumption about the coronavirus, trust in news about politics and current events, and trust in news about the coronavirus. 2 percent of observations are set to the mean due to an invalid ZIP code. Observations weighted to mimic a representative sample as described in the text. Error bars represent 95 percent confidence intervals.

C.2 Additional Details

C.2.1 Data Details

County-Level Data Build (POI and Visits)

To construct the county-level POI dataset used in the analysis, we proceeded as follows:

1. We use county data on 2016 Presidential votes shares (MIT Election Data and Science Lab 2018). We define the Republican vote share to be the share of votes received by the Republican candidate over the sum of votes across all candidates. We exclude Alaska, and merge with the 2010 TIGER county shapefile.[1] Two counties in the shapefile do not have valid vote data (FIPS: 15005, 51515).

2. We then use the latitude and longitude in the the April 2020 Core POI dataset from SafeGraph to match POIs to counties. We successfully assign more than 99.9 percent of the POIs to a county.

3. We merge the output from (2) with the Patterns dataset from SafeGraph using the safegraph-place-id variable. We sum visits by county for a given day or week, aggregating across POI. We drop all county observations with invalid vote shares at this stage.

4. We use the Open Census data from SafeGraph to construct a county-level dataset of demographic information. We do this by aggregating up the data given at the census block group level to the county level. We then merge the county demographic information with the output from (3).

5. We then merge The New York Times COVID-19 tracking data onto our output from (4). We assume zero cases and deaths for the observations not observed in The New York Times data. We drop the five counties associated with New York City and the four counties which overlap with Kansas City (MO), because The New York Times lists

[1] Downloaded from https://www.census.gov/geo/maps-data/data/cbf/cbf_counties.html on July 24, 2018.

these as geographic exceptions where it either does not assign cases to these counties or excludes cases occurring within the city.

County-Level Data Build (Social Distancing)

To construct the county-level social distancing dataset used in this analysis, we proceeded as follows:

1. We use the Daily Social Distancing SafeGraph data with observations at the census block group-day level for January 26 through April 4. We drop duplicate observations and exclude Alaska. We restrict our sample to census block groups with active devices throughout the entire time period. We also drop one census block group with anomalous behavior as notified by SafeGraph (FIPS: 190570010001).

2. We then aggregate to the county level. For the 'device count' and 'completely home device count' variables, we take the sum. For the 'median home dwell time' variable we take the mean weighted by 'the device count' in the census block group.

3. We then follow steps (4) and (5) described in Section C.2.1.

4. Lastly, we merge on 2016 Presidential vote shares, only keeping observations with valid vote shares.

Precinct-Level Data Build (POI and Visits)

1. We use 2016 precinct-level shapefiles and presidential election votes (Voting and Election Science Team 2018). We define the Republican vote share as in Section C.2.1 step (1). This data covers the following 38 states: AK, AR, AZ, CA, CO, DE, FL, GA, HI, IA, IL, KS, LA, MA, MD, ME, MI, MN, MO, MT, NC, ND, NE, NH, NM, NV, OK, OR, RI, SC, TN, TX, UT, VA, VT, WA, WI, WY

2. Using the set of POIs matched to county in Section C.2.1 step (2), we then use the latitude and longitude of these POIs and the precinct shapefiles from (1) to identify the

precinct containing a given POI. We are able to match 100 percent of the POIs from Section C.2.1 step (2). 55 of these POIs (0.001%) are matched to two precincts.

3. As in Section C.2.1 step (3), we merge the output from (2) with the Patterns dataset from SafeGraph using the safegraph-place-id variable. We sum visits by precinct, aggregating across POIs.

4. We use the Open Census data from SafeGraph to construct a precinct-level dataset of demographic information. We do this by first constructing the geographic intersections formed by our precinct shapefiles and 2019 Tiger census block group shapefiles.[2] Let a_p, a_b, and a_{cp} denote the area of precinct p, census block group b, and of their intersection respectively. For a given count variable x_b given at the block group level in SafeGraph's Open Census data, we construct a precinct-level estimate as: $\hat{x}_p := \sum_b \frac{a_{bp}}{a_b} x_b$. This estimate is exactly correct if a given demographic x_b is evenly distributed across a census block group's area. We then form ratios (e.g., population density or share hispanic) using these summed precinct-level estimates. We merge the precinct demographic information with the output from (3).

5. As in Section C.2.1 step (5), we then merge the New York Times COVID-19 tracking data onto our output from (4).

C.2.2 Survey Details

Data

We clean the survey data from Qualtrics as follows:

1. We match participant IDs from Qualtrics with a list of emailled IDs from CloudResearch and drop observations that do not match to remove test subjects. There is one exception, where the ID on Qualtrics did not correctly generate. We find exactly one remaining participant with the same demographics in the CloudResearch, so we keep this participant.

[2] Downloaded from ftp://ftp2.census.gov/geo/tiger/TIGER2019/BG/ on April 1, 2020.

2. We change one miscoded age from .23 to 23 and one miscoded ZIP code from ,43011 to 43011.

3. We merge ZIP code data with 2010 U.S. Census data and match ZIP codes to states and get population density.

4. We match ZIP codes to counties and use the week of March 29-April 4 and get county-level COVID cases and deaths via the New York Times. All ZIP codes in New York City are matched to the city-level cases and deaths since county-level data is unavailable from the New York Times. For analyses, we control for log(county cases + 1) and log(county deaths + 1).

5. We weight observations across age category, gender, race/ethnicity, and party affiliation using the stata ebalance command. Weights are prespecified in the pre-analysis plan.

6. News sources are numbered in the data in the following order: (1) Network news; (2) Breitbart; (3) CNN; (4) Facebook; (5) Fox News; (6) MSNBC; (7) New York Times; (8) Wall Street Journal; (9) Twitter; (10) Wikipedia; (11) CDC; (12) WHO. For analyses, we average consumption of news about politics and current events, trust in news about politics and current events, consumption of news about the coronavirus, and trust in news about the coronavirus.

We have the following demographic groups prior to weighting:

- Age: 45.7% 18-39, 33.8% 40-59, 20.5% 60+

- Gender: 51.9% Female, 47.75% Male, 0.35% Other / Non-binary

- Race: 66.6% White (Not Hispanic or Latinx), 15.25% Hispanic or Latinx, 11.2% Black or African American (Not Hispanic or Latinx), 4.95% Asian or Pacific Islander, 2.0% Other.

- Party: 34.65% Democratic, 31.25% Republican, 32.8% Independent, 1.3% Other

C.2.3 Survey Questions

Screening

- What is your gender? [Male; Female; Other / Non-binary]
- What race/ethnicity best describes you? [American Indian or Alaska Native; Asian or Pacific Islander; Black or African American (Not Hispanic or Latinx); Hispanic or Latinx; White (Not Hispanic or Latinx); Other]
- Do you consider yourself a Republican, a Democrat, or an Independent? [Democrat (Strongly Democratic); Democrat (Weakly Democratic); Independent (Lean toward the Democratic Party); Independent (Do not lean towards either party); Independent (Lean toward the Republican Party); Republican (Weakly Republican); Republican (Strongly Republican); Other / prefer not to say]
- What is your age?
- Do you currently live in the United States? [Yes; No]

Consent

[Page seen if age > 18, United States = Yes, and not screened out due to demographic quotas.]

Congratulations! You are eligible to participate. Please read the consent form below:

DESCRIPTION: You are invited to participate in an online research study on your views about the news and predictions of what will happen in the future. This is a research project being conducted by researchers at Harvard University and New York University.

TIME INVOLVEMENT: Your participation will take approximately 20 minutes, and the entire study will take place online.

RISKS AND BENEFITS: We will ensure that your individual responses are strictly confidential, and research results will only be presented in the aggregate. Your responses will not be shared with government officials or any 3rd party. We hope that the knowledge gained from this study will benefit society in general. **We cannot and do not guarantee or promise that you will receive any direct benefits from this study.**

PAYMENTS: If you are eligible for the study, and once you complete the study, you will receive a participation fee. You may also earn a bonus payment of up to $100 via an Amazon gift card. All payments will be through your research provider.

PARTICIPANT'S RIGHTS: If you have read this form and have decided to participate in this project, please understand your **participation is voluntary** and you have the **right to withdraw your consent or discontinue participation at any time without penalty or loss of benefits to which you are otherwise entitled. The alternative is not to participate.** You have the right to refuse to answer particular questions. The results of this research study may be presented at scientific or professional meetings or published in scientific journals.

CONTACT INFORMATION:

Questions: If you have any questions, concerns or complaints about this research, its procedures, risks and benefits, contact the researchers at rb4337@nyu.edu.

Independent Contact: If you are not satisfied with how this study is being conducted, or if you have any concerns, complaints, or general questions about the research or your rights as a participant, please contact the Harvard University Area Institutional Review Board (IRB) to speak to someone independent of the research team at cuhs@harvard.edu, (617)-496-2847. You can also write to the Committee on the Use of Human Subjects, Harvard University, 44-R Brattle Street, Suite 200, Cambridge, MA 02138.

Please retain a copy of this form for your records.

If you wish to participate in this study, please click "I consent" to proceed. This serves as an electronic signature indicating your consent to participate in the study.

[I consent; I do not consent]

[Only consenting subjects proceed]

Demographics

- How many children under the age of 18 do you have? [0; 1; 2; 3; 4; 5 or more]
- What is the highest degree or level of schooling that you have completed? [Less than a high school diploma; High school diploma or equivalent (for example: GED); Some college but no degree; Associate's degree; Bachelor's degree; Graduate degree (for example: MA, MBA, JD, PhD)]
- What was your total income in 2019? Please include only employment income (wages, salary, bonuses, tips, and any income from your own businesses). [I did not earn income in 2019; $1 to $9,999; ...; $50,000 to $59,999; $60,000 to $74,999; $75,000 to $99,999; $100,000 to $124,999; $125,000 to $149,999; $150,000 or more] [Coded as midpoint of range in thousands of dollars except for top bracket, who is coded at 200. Log(income + 1) is used as the control.]
- In what ZIP Code do you currently live? Please enter your 5-digit ZIP Code.
- In general, how would you rate your OVERALL health? [Excellent / Very good / Good / Fair / Poor]
- Has a doctor ever told you that you had the following conditions? [Yes / No]
  - Diabetes or high blood sugar
  - Lung disease such as chronic bronchitis or emphysema
  - A heart attack, coronary heart disease, angina, congestive heart failure, or other heart problems
- Please answer the following yes/no questions:
  - In the past week, have you had to go to a work environment in which you were within six feet of others?
  - Have you smoked at least 100 cigarettes in your entire life?
  - Have you smoked at least 10 cigarettes in the past week?

Information sources

- All of the following questions were asked about the following 12 news sources: Network news (ABC, CBS, NBC); Breitbart; CNN; Facebook; Fox News; MSNBC; The New York Times; The Wall Street Journal, Twitter, Wikipedia, The Centers for Disease Control (CDC); The World Health Organization (WHO).
  - **Last year**, how much **trust and confidence** did you have in each of the following sources when it comes to **reporting about <u>politics and current events</u> fully, accurately, and fairly**? [A great deal / A fair amount / Not very much / None at all / Not familiar with this outlet]
  - **Last year**, how **frequently did you get news and information** from each of the following sources **<u>about politics and current events</u>** through any medium (including reading online, watching on TV, etc.)? [Often / Sometimes / Rarely / Never / Not familiar with this outlet]
  - How much **trust and confidence** do you have in each of the following sources when it comes to **reporting about <u>the coronavirus</u> fully, accurately, and fairly**? [A great deal / A fair amount / Not very much / None at all / Not familiar with this outlet]
  - How **frequently are you getting news and information** from each of the following sources **<u>about the coronavirus</u>** through any medium (including reading online, watching on TV, etc.)? [Often / Sometimes / Rarely / Never / Not familiar with this outlet]

Changes in behavior and effects of social distancing

- Think about the ways you may have changed your daily routine in the past two weeks specifically because of the coronavirus. For example, you may be washing your hands more, avoiding restaurants and other public places, and/or reducing interactions with friends and family.

- By what percent have you reduced your overall contact with other people as a result of the coronavirus outbreak? Please enter a percentage from 0 to 100.
- Think back to two weeks ago.
- As of two weeks ago, by what percent had you reduced your overall contact with other people as a result of the coronavirus outbreak? Please enter a percentage from 0 to 100.
- Imagine that starting today and for the rest of the month, you went back to your **normal daily routine from before the coronavirus**. What do you think is the probability that you would catch the coronavirus in the next month? Please enter a percentage from 0 to 100. [Subjects who answer 0 for the percent reduction question see "continued with" instead of "went back to."]
- Imagine that starting today and for the next month, **you cut off all in-person contact with people outside your household.** What do you think is the probability that you would catch the coronavirus in the next month? Please enter a percentage from 0 to 100.
- We'd like to quantify the overall costs (in terms of time, money, and inconvenience) that social distancing imposes on you. Consider a hypothetical situation in **a normal month in the future, after the coronavirus outbreak is completely over.**

  Imagine you had a choice between:

(A) following your normal routine for one month,

OR

(B) cutting off all in-person contact with people outside your household for one month, AND receiving $X cash.

Presumably if you were offered a large amount of cash ($X is large), you'd be willing to cut off all social contact. If you weren't offered any cash ($X is 0), you'd prefer to stick with your normal routine. What value of X would make you equally happy with these two options? Please answer in dollars.

Economic trade-offs

- When there was no "stay-at-home" order for your area, what did you think was the best way to help the country in this time of crisis? [7-point scale from "Go out more to help the economy" to "Go out less to avoid spreading the coronavirus"]

Predictions

[If unincentivized:]

- You will now be asked to make a few predictions.

[If incentivized:]

- You will now be asked to make a few predictions. Think carefully! We'll randomly select 10 participants for an accuracy reward. If you're selected, we'll pay you up to $100 depending on how accurate your prediction was. For example:
  - If your answer is exactly right, we'll give you $100
  - If your answer is 1% off, we'll give you $99
  - If your answer is 2% off, we'll give you $98
  - ...
  - If your answer is 50% off, we'll give you $50
  - etc.

All subjects see:

- We want to know how well you think the U.S. will limit the spread of the coronavirus in the next month. There had been 177,226 known cases of coronavirus in the U.S. by March 31. How many <u>additional</u> known cases will there be in the U.S. in the month of April?

- RealClearPolitics reports polling data on public approval of President Trump's handling of the coronavirus outbreak. What percent of people will say they approve of Trump's handling of the coronavirus outbreak on the latest poll that ends before April 30?